



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TRANSPORTES E
GESTÃO TERRITORIAL

Rafael Elizeu Beltrão de Azevedo

**INTEGRAÇÃO DE MÉTODOS TRADICIONAIS E DE DISPOSITIVOS
MÓVEIS NA CONSTRUÇÃO DE MATRIZES OD
REDUÇÃO DE CUSTOS E TEMPO EM ESTUDOS DE TRANSPORTES**

Florianópolis

2023

Rafael Elizeu Beltrão de Azevedo

**INTEGRAÇÃO DE MÉTODOS TRADICIONAIS E DE DISPOSITIVOS
MÓVEIS NA CONSTRUÇÃO DE MATRIZES OD
REDUÇÃO DE CUSTOS E TEMPO EM ESTUDOS DE TRANSPORTES**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Transportes e Gestão Territorial da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Engenharia Civil.

Orientador: Prof. Marcos Aurélio Marques Noronha, Dr.

Florianópolis

2023

Azevedo, Rafael Elizeu Beltrão de
INTEGRAÇÃO DE MÉTODOS TRADICIONAIS E DE DISPOSITIVOS MÓVEIS
NA CONSTRUÇÃO DE MATRIZES OD : REDUÇÃO DE CUSTOS E TEMPO EM
ESTUDOS DE TRANSPORTES / Rafael Elizeu Beltrão de Azevedo ;
orientador, Marcos Aurélio Marques Noronha, 2023.
170 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia Civil, Florianópolis, 2023.

Inclui referências.

1. Engenharia Civil. 2. sistemas de transportes. 3. matrizes
OD. 4. machine learning. 5. telefonia móvel. I. Noronha, Marcos
Aurélio Marques. II. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Engenharia Civil. III. Título.

Rafael Elizeu Beltrão de Azevedo

**INTEGRAÇÃO DE MÉTODOS TRADICIONAIS E DE DISPOSITIVOS
MÓVEIS NA CONSTRUÇÃO DE MATRIZES OD
REDUÇÃO DE CUSTOS E TEMPO EM ESTUDOS DE TRANSPORTES**

O presente trabalho em nível de Mestrado foi avaliado e aprovado pela banca examinadora composta pelos seguintes membros:

Prof. Amir Matar Valente, Dr.

Universidade Federal de Santa Catarina (UFSC)

Prof. Eduardo Lobo, Dr.

Universidade Federal de Santa Catarina (UFSC)

Prof. Danilo Rinaldi Bisconsini, Dr.

Membro externo, Universidade Tecnológica Federal do Paraná (UTFPR)

Coordenação do Programa de Pós-Graduação

Prof. Marcos Aurélio Marques Noronha, Dr.

Orientador

Florianópolis, 2023.

*À memória de meu pai,
que teria tido orgulho em ver seu filho dando passos cada vez maiores.*

AGRADECIMENTOS

Agradeço a Deus.

Agradeço aos meus pais, Rafael Gomes de Azevedo (falecido) e Nair Dias Beltrão, por terem me dado a capacidade e inteligência, permitindo-me apaixonar por tantas áreas distintas, gerando sonhos que a cada dia busco realizar.

À minha querida irmã por ter me dado apoio e proteção quando eu precisei, além de todo o amor desse mundo.

Agradeço ao meu orientador final, professor Dr. Marcos Aurélio Marques Noronha, pela amizade, pela parceria, pelo ser humano maravilhoso que é e, acima de tudo, um verdadeiro pai em muitos momentos.

Agradeço ao meu orientador no início do processo, professor Dr. Alexandre Hering Coelho, por todas as dicas, conselhos, paciência e por iluminar um caminho em minha vida que para mim era algo impensável.

À minha queridíssima e amiga professora Dra. Patrícia de Oliveira Faria, que sempre me aconselhou da melhor forma possível, abrindo minha visão para tantas prioridades em minha vida.

Segue aqui também um agradecimento à banca examinadora deste trabalho, sendo o professor Dr. Amir Matar Valente, o professor Dr. Eduardo Lobo, o professor Dr. Danilo Rinaldi Bisconsini e a querida amiga e engenheira Dra. Thaís Ventura Chibiaqui, por terem trazido tantos elementos que puderam transformar este trabalho em algo notável. Sem as suas correções e sugestões, eu nunca teria conseguido atingir esse nível de excelência.

Preciso destacar aqui um agradecimento muito especial à empresa APPE, que transcendeu qualquer nível de companheirismo e confiança, me dando a oportunidade de realizar esse curso, sempre acreditando na minha capacidade e apostando em meu futuro profissional. Assim como, registro aqui um agradecimento a todos os amigos que essa empresa me deu.

Por fim, mas não menos importante, mais um agradecimento especial a toda população brasileira pelo apoio financeiro, pois sem os impostos suados que todos são obrigados a pagar, a realização deste trabalho de pesquisa nunca teria sido possível.

"One of the great challenges in this world is knowing enough about a subject to think
you're right, but not enough about the subject to know you're wrong."

Neil deGrasse Tyson, 2021

RESUMO

A construção de Matrizes Origem-Destino (OD) é uma etapa fundamental na modelagem de transportes e planejamento urbano. Porém, tradicionalmente, a obtenção de dados para sua elaboração constitui um processo oneroso e demorado. Este estudo aborda a problemática dos métodos convencionais de coleta de dados, incluindo pesquisas domiciliares, contagens volumétricas classificatórias e levantamentos em linhas de contorno, destacando os altos custos envolvidos e os longos períodos necessários para a conclusão dessas pesquisas. Contudo, a crescente disponibilidade de dados móveis, resultado de uma expansão da capacidade de atuação de empresas operadoras de telecomunicações, gerou uma infraestrutura global capaz de produzir e armazenar uma quantidade significativa de informações como identificação do usuário, localização, tipo de dispositivo, entre outros. Este trabalho explora como a infraestrutura de informação fornecida por essas tecnologias possibilita a geração de dados necessários para construir Matrizes OD de maneira mais rápida e econômica. Após uma extensa procura, sem êxito, por artigos e trabalhos que pudessem trazer análises comparativas e métodos para substituição e/ou complementação de matrizes, o presente estudo se propôs a realizar essa análise abrangente entre os métodos tradicionais e o uso de dados móveis na construção de Matrizes OD, detalhando todas as etapas necessárias em cada abordagem. Constatou-se uma diferença de 9 vezes mais rápido o tempo de execução para o método de dispositivos de dados móveis além de um custo chegando a quase 23 vezes mais caro no método tradicional. Por fim, a dissertação avalia a compatibilidade entre os métodos, considerando as consistências obtidas e discutindo se a utilização de dados móveis pode validar a otimização de futuros trabalhos no campo da simulação de tráfego. Além disso, este trabalho também introduz um novo método de classificação de matrizes específico para o uso de dados móveis, oferecendo uma possibilidade de otimizar trabalhos relacionados a estudos de tráfego. Além de trabalhar com um tamanho de amostra muito maior, esse método também reduz significativamente o tempo despendido para a realização de trabalhos no setor. Em resumo, este estudo contribui para uma compreensão mais aprofundada das vantagens e desafios associados aos métodos de construção de Matrizes OD em transporte e destaca o potencial disruptivo das tecnologias emergentes no setor.

Palavras-chave: sistemas de transportes; matrizes OD; *machine learning*; telefonia móvel; comparação; dados móveis; big data; algoritmo; classificação modal.

ABSTRACT

The construction of Origin-Destination (OD) matrices is a fundamental step in transportation modeling and urban planning. However, traditionally, obtaining data for their preparation has been a costly and time-consuming process. This study addresses the challenges associated with conventional data collection methods, including household surveys, classification volume counts, and contour line surveys, highlighting the high costs involved and the lengthy periods required to complete these surveys. However, the growing availability of mobile data, as a result of the expanded capabilities of telecommunications operators, has created a global infrastructure capable of producing and storing a significant amount of information such as user identification, location, device type, and more. This work explores how the information infrastructure provided by these technologies enables the generation of the data needed to construct OD matrices in a faster and more cost-effective manner. After an extensive but unsuccessful search for articles and papers that could provide comparative analyses and methods for replacing and/or complementing matrices, this study set out to conduct a comprehensive analysis between traditional methods and the use of mobile data in constructing OD matrices, detailing all the necessary steps in each approach. The execution time was 9 times faster for the mobile data device method in addition to a cost reaching almost 23 times more expensive in the traditional method. Finally, the dissertation assesses the compatibility between the methods, considering the consistencies obtained and discussing whether the use of mobile data can validate the optimization of future work in the field of traffic simulation. Additionally, this work also introduces a new classification method specific to the use of mobile data, offering a possibility to optimize traffic-related studies. In addition to working with a much larger sample size, this method also significantly reduces the time spent on tasks in the sector. In summary, this study contributes to a deeper understanding of the advantages and challenges associated with OD matrix construction methods in transportation and highlights the disruptive potential of emerging technologies in the sector.

Keywords: transportation systems; OD matrices; machine learning; mobile telephony; comparison; mobile data; big data; algorithm; modal classification.

RÉSUMÉ

La création des matrices Origine-Destination (OD) est une étape fondamentale dans la modélisation des transports et de l'urbanisme. Cependant, traditionnellement, l'obtention de données pour leur élaboration s'avère un processus coûteux et chronophage. Cette étude aborde les problèmes liés aux méthodes conventionnelles de collecte de données, notamment les enquêtes auprès des foyers, les comptages volumétriques classificatoires et les relevés de lignes de contour, mettant ainsi en évidence les coûts élevés impliqués et les périodes prolongées nécessaires pour mener à bien ces enquêtes. Toutefois, la disponibilité croissante de données mobiles, résultant de l'expansion des capacités des opérateurs de télécommunications, a créé une infrastructure mondiale capable de produire et de stocker une quantité significative d'informations telles que l'identification de l'utilisateur, la localisation, le type d'appareil, et bien plus encore. Ce travail explore comment l'infrastructure d'information fournie par ces technologies permet de générer les données nécessaires à la construction des matrices OD de manière plus rapide et plus économique. Après une recherche extensive et infructueuse d'articles et de documents pouvant fournir des analyses comparatives et des méthodes de remplacement et/ou de complément des matrices, cette étude s'est fixé pour objectif de réaliser une analyse complète entre les méthodes traditionnelles et l'utilisation des données mobiles dans la construction des matrices OD, en détaillant toutes les étapes nécessaires dans chaque approche. Le temps d'exécution était 9 fois plus rapide pour la méthode du périphérique de données mobile, en plus d'un coût atteignant près de 23 fois plus cher dans la méthode traditionnelle. Enfin, ce mémoire évalue la compatibilité entre les méthodes, en tenant compte des consistances obtenues et en discutant de la possibilité que l'utilisation de données mobiles puisse valider l'optimisation des travaux futurs dans le domaine de la simulation du trafic. En outre, ce travail introduit également une nouvelle méthode de classification spécifique à l'utilisation des données mobiles, offrant la possibilité d'optimiser les études liées à la circulation routière. En plus de travailler avec une taille d'échantillon beaucoup plus grande, cette méthode réduira également de manière significative le temps passé sur les tâches dans le secteur. En résumé, cette étude contribue à une compréhension plus approfondie des avantages et des défis associés aux méthodes de construction de matrices OD dans le domaine des transports et met en évidence le potentiel disruptif des technologies émergentes dans le secteur.

Mots-clés: systèmes de transport; matrices OD; apprentissage automatique; téléphonie mobile; comparaison; données mobiles; mégadonnées; algorithmes; classification modale.

RESUMEN

La construcción de matrices de Origen-Destino (OD) es un paso fundamental en la modelización del transporte y la planificación urbana. Sin embargo, tradicionalmente, obtener datos para su desarrollo ha sido un proceso costoso y que consume mucho tiempo. Este estudio aborda los desafíos con los métodos convencionales de recopilación de datos, incluyendo encuestas domiciliarias, conteos de volumen por clasificación y levantamientos de líneas de contorno, destacando los altos costos involucrados y los largos períodos necesarios para completar estas encuestas. No obstante, la creciente disponibilidad de datos móviles, como resultado de las capacidades ampliadas de los operadores de telecomunicaciones, ha creado una infraestructura global capaz de producir y almacenar una cantidad significativa de información, como la identificación del usuario, la ubicación, el tipo de dispositivo y más. Este trabajo explora cómo la infraestructura de información proporcionada por estas tecnologías permite la generación de los datos necesarios para construir matrices OD de manera más rápida y rentable. Después de una extensa y infructuosa búsqueda de artículos y documentos que pudieran proporcionar análisis comparativos y métodos para reemplazar y/o complementar matrices, este estudio se propuso realizar un análisis exhaustivo entre los métodos tradicionales y el uso de datos móviles en la construcción de matrices OD, detallando todos los pasos necesarios en cada enfoque. El tiempo de ejecución fue 9 veces más rápido para el método del dispositivo de datos móviles además de un costo que alcanza casi 23 veces más caro en el método tradicional. Finalmente, la tesis evalúa la compatibilidad entre los métodos, teniendo en cuenta las consistencias obtenidas y discutiendo si el uso de datos móviles puede validar la optimización de futuros trabajos en el campo de la simulación de tráfico. Además, este trabajo también introduce un nuevo método de clasificación específico para el uso de datos móviles, ofreciendo la posibilidad de optimizar estudios relacionados con el tráfico. Además de trabajar con un tamaño de muestra mucho más grande, este método también reducirá significativamente el tiempo empleado en las tareas del sector. En resumen, este estudio contribuye a una comprensión más profunda de las ventajas y desafíos asociados con los métodos de construcción de matrices OD en el transporte y resalta el potencial disruptivo de las tecnologías emergentes en la industria.

Palabras clave: sistemas de transporte; matrices OD; aprendizaje automático; telefonía móvil; comparación; datos móviles; big data; algoritmo; clasificación modal.

摘要

原-目的 (OD) 矩阵的构建是交通模型和城市规划的基本步骤。然而，传统上，为其制定获取数据是一个昂贵且耗时的过程。本研究探讨了传统数据收集方法的问题，包括家庭调查、分类体积计数和等高线测量，强调了涉及的高昂成本和完成这些调查所需的长时间。然而，移动数据的不断增加，是电信运营商扩大能力的结果，已经创造了一个全球基础设施，能够产生和存储大量信息，如用户身份识别、位置、设备类型等。本研究探讨了这些技术提供的信息基础设施如何以更快速和经济的方式生成构建OD矩阵所需的数据。在广泛但未成功地搜索文章和论文以提供比较分析和替代或补充矩阵的方法之后，本研究旨在对传统方法与使用移动数据构建OD矩阵进行全面分析，详细描述了每种方法中所需的所有步骤。最后，论文评估了方法之间的兼容性，考虑到所获得的一致性，并讨论了移动数据的使用是否能够验证未来交通仿真领域工作的优化。移动数据设备方法的执行时间比传统方法快了 9 倍，而且成本几乎高出传统方法的 23 倍。此外，本研究还引入了一种用于移动数据的新矩阵分类方法，为优化与交通研究相关的工作提供了可能性。除了使用更大的样本大小，这种方法还将大幅减少在该领域工作所需的时间。总之，本研究有助于更深入地了解交通领域中OD矩阵构建方法的优势和挑战，并强调了新兴技术在该领域的潜在颠覆性。

关键词: 交通系统；OD矩阵；机器学习；移动电话；比较；移动数据；大数据；算法；模态分类。

ZUSAMMENFASSUNG

Die Erstellung von Quelle-Ziel-Matrizen (OD) ist ein grundlegender Schritt in der Verkehrsmodellierung und Stadtplanung. Allerdings ist die Beschaffung von Daten für deren Aufbereitung traditionell ein teurer und zeitaufwändiger Prozess. Diese Studie befasst sich mit den Herausforderungen herkömmlicher Datenerfassungsmethoden, einschließlich Haushaltsbefragungen, klassifizierender volumetrischer Zählungen und Konturlinienerhebungen, und hebt die damit verbundenen hohen Kosten und die langen Zeiträume hervor, die für die Durchführung dieser Erhebungen erforderlich sind. Allerdings hat die zunehmende Verfügbarkeit mobiler Daten, die das Ergebnis einer Ausweitung der Betriebskapazitäten von Telekommunikationsunternehmen ist, eine globale Infrastruktur geschaffen, die in der Lage ist, eine erhebliche Menge an Informationen wie Benutzeridentifikation, Standort und Gerätetyp zu produzieren und zu speichern andere. In dieser Arbeit wird untersucht, wie die durch diese Technologien bereitgestellte Informationsinfrastruktur die Generierung der für die Erstellung von OD-Matrizen erforderlichen Daten auf schnellere und wirtschaftlichere Weise ermöglicht. Nach einer umfangreichen, erfolglosen Suche nach Artikeln und Werken, die vergleichende Analysen und Methoden zum Ersetzen und/oder Ergänzen von Matrizen liefern könnten, wurde in der vorliegenden Studie vorgeschlagen, diese umfassende Analyse zwischen traditionellen Methoden und der Verwendung mobiler Daten bei der Konstruktion von OD-Matrizen durchzuführen. detaillierte Beschreibung aller notwendigen Schritte für jeden Ansatz. Bei der Methode mit mobilen Datengeräten ergab sich ein Unterschied von einer neunmal schnelleren Ausführungszeit, zusätzlich zu den Kosten, die bei der herkömmlichen Methode fast 23-mal höher waren. Abschließend bewertet die Dissertation die Kompatibilität zwischen den Methoden, berücksichtigt die erzielten Konsistenzen und diskutiert, ob die Nutzung mobiler Daten die Optimierung zukünftiger Arbeiten im Bereich der Verkehrssimulation validieren kann. Darüber hinaus stellt diese Arbeit auch eine neue Matrixklassifizierungsmethode vor, die speziell auf die Nutzung mobiler Daten zugeschnitten ist und eine Möglichkeit bietet, die Arbeit im Zusammenhang mit Verkehrsstudien zu optimieren. Diese Methode erlaubt nicht nur das Arbeiten mit einer erheblich größeren Stichprobengröße, sondern reduziert auch den Zeitaufwand für die Durchführung der Arbeiten in der Branche erheblich. Zusammenfassend trägt diese Studie zu einem tieferen Verständnis der Vorteile und Herausforderungen bei, die mit OD-Array-Konstruktionsmethoden im Transportwesen verbunden sind, und betont das disruptive Potenzial neuer Technologien in diesem Sektor.

Schlüsselwörter: Transportsysteme; OD-Matrizen; maschinelles Lernen; Mobiltelefonie; Vergleich; mobile Daten; Große Daten; Algorithmus; Modale Klassifizierung.

LISTA DE ABREVIATURAS E SIGLAS

AGERBA	Agência Estadual de Serviços Públicos de Energia, Transportes e Comunicações da Bahia
ANTT	Agência Nacional de Transportes Terrestres
BIRD	Banco Internacional para Reconstrução e Desenvolvimento
CNN	Convolutional Neural Networks
DTNN	Decision Trees Neural Networks
GPS	Geographic Positioning System
HBM	Hierarchical Bayesian Model
HIB-IPF	Hellinger Distance-Based Iterative Proportional Fitting
IBGE	Instituto Brasileiro de Geografia e Estatística
LC	Lei Complementar
OD	Origem – Destino
PCA	Principal Components Analysis
PDLT	Plano Diretor de Logística de Transportes
RMFS	Região Metropolitana de Feira de Santana
RMS	Região Metropolitana de Salvador
SEINFRA	Secretaria de Infraestrutura do Estado da Bahia
SIT	Superintendência de Infraestrutura de Transportes
SNV	Sistema Nacional de Viação
SRE	Sistema Rodoviário Estadual
SVD	Singular Value Decomposition

LISTA DE FIGURAS

Figura 1 - Níveis de precisão no zoneamento	23
Figura 2 - Fluxograma da metodologia da pesquisa	35
Figura 3 - Área de Estudo: Regiões Metropolitanas	62
Figura 4 - Região Metropolitana de Feira de Santana.....	63
Figura 5 - Região Metropolitana de Salvador	65
Figura 6 – Zoneamento externo às regiões metropolitanas	70
Figura 7 - Zoneamento - Polígonos e centroides	71
Figura 8 - Fluxograma geral do método.....	74
Figura 9 - Fluxograma geral de pesquisas.....	75
Figura 10 - Postos provenientes do estudo das RM e do PDLT	79
Figura 11 - Sobreposição de <i>shapefiles</i> (rede e pontos de contagem)	80
Figura 12 - Zonas consideradas nas pesquisas domiciliares.....	81
Figura 13 - Linhas de desejo das viagens externas em dia útil típico	90
Figura 14 - Linha de Desejo Estimada em Dia Útil Típico na Hora Pico 18h	91
Figura 15 - Linhas de Desejo de Salvador em dia útil típico	92
Figura 16 - Linhas de Desejo – Região Barra, Rio Vermelho e Pituba – em um dia útil típico.....	92
Figura 17 - Gráfico das Viagens por Faixa Horária de um Dia Útil Típico.....	93
Figura 18 - Linha de Desejo Estimada para um Dia Útil Típico na Hora Pico 18h	93
Figura 19 - Relação de porcentagem entre os tipos de veículo por OD.....	100
Figura 20 - Mensagem de erro na exportação dos dados do OSM, devido ao tamanho da área.....	107
Figura 21 - Dados de exportações do Visum acessados no ArcGIS	109
Figura 22 - Sobreposição de dados.....	110
Figura 23 - Rede de Interesse Importada para o Visum.....	111
Figura 24 - Rede finalizada.....	112
Figura 25 - Validação de caminho mínimo aeroporto de RMS e anel de RMFS	114
Figura 26 - Validação de caminho mínimo longitudinal RMS e ligação com RMFS.....	115
Figura 27 - Malha viária reduzida	119
Figura 28 - Detalhamento da malha viária reduzida	119
Figura 29 - Relatório de consistência gerado pelo módulo Survey Report	129

LISTA DE QUADROS

Quadro 1 - Diferenças entre as pesquisas quanto aos objetivos.....	29
Quadro 2 – Documentos relevantes da área pesquisada.....	60
Quadro 3 - Arquivos de estudos anteriores.....	73
Quadro 4 - Postos de Pesquisa Cordão Externo	78

LISTA DE TABELAS

Tabela 1 - Informações sobre os conglomerados do zoneamento	68
Tabela 2 - Número de zonas de tráfego por municípios das regiões metropolitanas	69
Tabela 3 - Plano amostral da pesquisa domiciliar - RMFS	84
Tabela 4 - Plano amostral da pesquisa domiciliar - RMS	84
Tabela 5 - Comparação de registros entre os métodos.....	101
Tabela 6 - Médias de distribuição por <i>cluster</i>	104
Tabela 7 - Fatores de ocupação veicular	104
Tabela 8 - Redução do tempo de execução de uma alocação – Sem Estimação.....	120
Tabela 9 - Comparação entre características e valores de ambos os métodos.....	127
Tabela 10 - Comparação de R ² com matrizes não ajustadas.....	130
Tabela 11 – Comparação de R ² com matrizes ajustadas	130

LISTA DE EQUAÇÕES

Equação 1 - Amostragem Casual Cimples	82
Equação 2 - Expansão de pesquisas domiciliares	85
Equação 3 - Expansão de pesquisas LC e LT	85
Equação 4 - Expansão de contagens de 6h para 24h	86
Equação 5 - Correção semanal através de posto sazonal.....	87

SUMÁRIO

1. INTRODUÇÃO	22
1.1. OBJETIVOS.....	24
1.1.1. Objetivo geral	24
1.1.2. Objetivos específicos	25
1.2. JUSTIFICATIVA	25
1.3. METODOLOGIA DA DISSERTAÇÃO	26
1.3.1. Classificação de pesquisas	26
1.3.2. Fontes de dados.....	32
1.3.3. Coleta e pré-processamento de dados.....	32
1.3.4. Modelo e parâmetros de simulação	33
1.3.5. Pesquisa exploratória	33
1.4. ESTRUTURA DA DISSERTAÇÃO	36
2. FUNDAMENTAÇÃO TEÓRICA.....	37
2.1. MATRIZES OD	37
2.2. TRATAMENTO DE DADOS.....	39
2.2.1. Limpeza de dados	39
2.2.2. Agregação de dados	40
2.2.3. Normalização de dados	41
2.2.4. Análise de dados	42
2.3. CLASSIFICAÇÃO DE MATRIZES.....	43
2.3.1. Classificação de veículo baseada em GPS.....	44
2.3.1.1. Árvores de Decisão	46
2.3.1.2. Florestas aleatórias	48
2.3.1.3. Redes neurais.....	49
2.3.2. Pesquisas e dados do censo.....	51
2.3.3. Contagens de tráfego	52
2.3.4. Detecção de <i>Bluetooth</i>	53
2.4. SIMULAÇÃO DE TRÁFEGO – PTV VISUM	55
2.4.1. Modelagem da rede.....	55
2.4.2. Alocação de matrizes	55
2.4.3. Estimacão de matrizes OD.....	56
2.4.4. Comparação de consistências	58

2.5.	REVISÃO SISTEMÁTICA	59
2.5.1.	Definição.....	59
2.5.2.	Resultados encontrados.....	59
3.	ESTUDO DE CASO	62
3.1.	FEIRA DE SANTANA	63
3.2.	SALVADOR	64
3.3.	ZONEAMENTO	66
3.3.1.	Aspectos gerais do zoneamento em estudos de mobilidade urbana	66
3.3.2.	Caracterização das áreas de estudo por meio de dados secundários.....	67
3.3.3.	Definição das zonas internas.....	67
3.4.	LEVANTAMENTO DE DADOS – MÉTODO TRADICIONAL.....	72
3.4.1.	Pesquisas Domiciliares	74
3.4.2.	Linha de Contorno	76
3.4.3.	Localização dos postos de contagem e de pesquisas de OD.....	78
3.4.4.	<i>Shapefile</i> de “ <i>Count Locations</i> ”.....	80
4.	CONSTRUÇÃO DE MATRIZES E MODELAGEM	81
4.1.	MÉTODO TRADICIONAL.....	81
4.1.1.	Pesquisas realizadas	81
4.1.2.	Expansões para construção das matrizes	85
4.2.	MÉTODO DE DADOS MÓVEIS.....	87
4.2.1.	Apresentação dos polígonos	87
4.2.2.	Construção do banco de dados.....	87
4.2.3.	Coleta de dados juntos a operadoras de telefonia celular escolhida	88
4.2.4.	Construção de banco de dados a partir das análises de Big Data	88
4.2.5.	Construção das linhas de desejo	89
4.2.6.	Classificação da matriz OD	94
4.3.	CONSTRUÇÃO DO MODELO	105
4.3.1.	Construção da MALHA VIÁRIA DA ÁREA DE ESTUDO	106
4.3.2.	Mosaico de importação – OSM	109
4.3.3.	Importação da rede no VISUM.....	111
4.3.4.	Validação de caminho mínimo	113
4.4.	ALOCAÇÃO DE TRÁFEGO	115
4.4.1.	Aspectos Gerais	115
4.4.2.	Processo de simplificação do Modelo	116

4.4.3. Processo de estimação das matrizes.....	120
5. ANÁLISE COMPARATIVA	121
5.1. MATRIZES PROVINDAS DE PESQUISAS OD	121
5.2. MATRIZES PROVINDAS DE DADOS MÓVEIS	123
5.3. RELATÓRIOS DE CONSISTÊNCIA	128
6. CONCLUSÕES E RECOMENDAÇÕES	132
6.1. CONSIDERAÇÕES FINAIS	132
6.2. RECOMENDAÇÕES PARA TRABALHOS POSTERIORES.....	133
6.2.1. PLACOM – Plataforma de Compatibilização de Matrizes	133
6.2.2. Pesquisas OD – Online	136
REFERÊNCIAS	137
APÊNDICE A	146
ANEXO A – PROCEDIMENTOS – ANÁLISE DE MOBILIDADE.....	157
A1 FONTE DE DADOS.....	158
A2 REGISTRO DE COMUNICAÇÃO.....	159
A3 DADOS ADICIONAIS INCORPORADOS PELA OPERADORA	160
A4 ETAPAS DE ANÁLISE	162
A4.1 FILTROS E TESTES DE CONSISTÊNCIA DE DADOS BRUTOS	162
A4.2 RECONSTRUÇÃO DA SEQUÊNCIA TEMPORAL DOS EVENTOS.....	163
A4.3 CLASSIFICAÇÃO DE EVENTOS POR MOBILIDADE	163
A4.4 PROJEÇÃO À INFRAESTRUTURA.....	164
A5 PRIVACIDADE.....	166
A6 GEORREFERENCIAMENTO.....	167
A7 EXPANSÃO AMOSTRAL.....	169
A7.1 ELEVÇÃO DA AMOSTRA PARA A POPULAÇÃO TOTAL.....	169
A7.2 EXTRAPOLAÇÃO DA AMOSTRA DE INTERNACIONAIS (ROAMERS)	170

1. INTRODUÇÃO

Matrizes de Origem e Destino (OD) são comumente utilizadas para representar fluxos diversos, como de pessoas, mercadorias ou veículos que se deslocam entre diferentes pontos pré-definidos. O modelo tradicional para desenvolvimento dessas matrizes baseia-se em pesquisas domiciliares e entrevistas com usuários nas rodovias, além de contagens de tráfego. Esses procedimentos possuem uma série de características negativas, como o alto consumo de recursos, limitações em relação ao tamanho de amostra, tempo necessário para sua conclusão e, conseqüentemente, uma frequência de atualização baixa.

Hajek (1977), Kuwabara e Sullivan (1987) e Groves (2006) trazem elementos que mostram a propensão a erros nesse processo, tanto em virtude de vieses na amostragem quanto a erros nos próprios relatórios. No mais, Hazelton (2001) afirma que essa abordagem é muito sensível em virtude de os fluxos de rotas estarem baseados em informações prévias, ou seja, as abordagens para levantamento de dados são influenciadas por suposições ou dados anteriores que podem não refletir completamente as condições atuais ou futuras de tráfego. A sensibilidade mencionada por Hazelton (2001) significa que, se as condições de tráfego mudarem significativamente desde a coleta dos dados históricos, as estimativas baseadas nesses dados podem não ser precisas.

Recentemente, a adoção de um novo método para o desenvolvimento das matrizes vem sendo analisada em congressos de mobilidade, artigos e trabalhos acadêmicos, em virtude da possibilidade de se atingir maior quantidade de informações e, ainda assim, maior precisão, dado a disponibilidade de dados móveis. Essas vantagens sobre o método tradicional são mencionadas por Zuo, Chen e Yu (2014), quando abordam o rastreamento contínuo do usuário e a quantidade significativa das amostras obtidas, com a alta qualidade de dados e garantia de autenticidade e confiabilidade. Os dados possuem origem nos sinais de GPS de aparelhos eletrônicos, fornecendo informações sobre movimentos e atividades dos usuários, constituindo matrizes OD carregadas de detalhamentos dos fluxos que existem entre os pares.

Além dos benefícios citados, esse método traz consigo uma redução de tempo, além de um menor custo de operação, quando comparado ao modelo atual de coleta de dados, pesquisas domiciliares OD e contagens de tráfego, os quais são demorados e caros.

Os dados móveis também fornecem uma imagem mais contínua e em tempo real dos movimentos das pessoas, o que pode ser particularmente útil para entender como diferentes fatores, como congestionamento de tráfego, clima e eventos afetam os padrões de viagem.

Inclusive, Phithakkitnukoon et al. (2010) trazem uma análise na tentativa de compreender o comportamento humano, identificando o padrão de atividade diária através de dados de celular.

Da mesma forma, o uso de dados móveis permite a criação de matrizes OD em diferentes escalas, desde um bairro, uma cidade e até uma região, ou mesmo um país, conforme ilustrado na Figura 1. A quantidade de dados obtida permite uma compreensão mais abrangente das interconexões entre diferentes áreas de estudo, que podem ser usadas para aprimorar o desenvolvimento de redes de transporte, planejamento urbano e outras iniciativas destinadas a melhorar a mobilidade e a acessibilidade.



Fonte: Autor, 2023.

Portanto, o uso de dados móveis na geração de matrizes OD garantem uma maneira mais precisa, em tempo real e econômica de entender os fluxos entre diferentes locais, contribuindo para a tomada de decisões e planejamento de sistemas de transporte mais sustentáveis e eficientes.

Para o desenvolvimento deste trabalho, foram utilizados dados de um recente estudo encomendado por uma instituição financeira internacional. Afortunadamente, o presente autor faz parte da equipe contratada por essa instituição, a qual está financiando as etapas do desenvolvimento. Dessa forma, respeitando as devidas cláusulas de confidencialidade e resguardando certas informações, foi concedida ao autor a permissão necessária para trazer dados essenciais e informações relevantes para esta dissertação.

A área de estudo é composta por duas regiões metropolitanas do estado da Bahia: Salvador e Feira de Santana. Ambas as cidades são as mais populosas do estado e possuem um papel significativo no desenvolvimento econômico e social da região.

Feira de Santana está em uma localização estratégica, próxima a importantes rodovias e ferrovias e, assim, constituindo um importante centro logístico. Os estudos socioeconômicos mostram que também possui uma forte atividade industrial, principalmente na área têxtil, de alimentos e bebidas, além de calçados e metal-mecânica.

Salvador é a maior cidade do estado e conta também com a maior densidade populacional, possuindo uma localização estratégica na costa atlântica, sendo um polo comercial em virtude dos diversos portos e com uma indústria diversificada, incluindo setores como petroquímica, automobilística, alimentícia, têxtil, eletroeletrônica e de construção civil.

Portanto, a dissertação está focada em uma região bastante complexa e, ao mesmo tempo, importante do ponto de vista econômico do país. Destarte, a disponibilidade de uma vasta gama de dados confiáveis e atualizados, aliado a estrutura de softwares, equipamentos e equipe à disposição do autor, torna as informações e conclusões deste trabalho mais relevantes do ponto de vista prático, profissional e acadêmico.

Em resumo, o questionamento técnico que norteará o andamento deste trabalho é:

Será que a adoção de matrizes derivadas de dados móveis, ao invés de matrizes tradicionais provenientes de pesquisas domiciliares, é uma abordagem viável e eficaz para capturar com precisão os padrões de mobilidade da população, levando em consideração as vantagens das tecnologias contemporâneas e potenciais limitações inerentes a esses conjuntos de dados dinâmicos?

1.1. OBJETIVOS

1.1.1. OBJETIVO GERAL

Analisar a viabilidade da aplicação de matrizes OD construídas a partir de dispositivos de telefonia móvel, substituindo as matrizes construídas tradicionalmente com base em pesquisas domiciliares e contagens de tráfego.

1.1.2. OBJETIVOS ESPECÍFICOS

- a. Construir as matrizes OD de acordo com os métodos a serem analisados;
- b. Modelar a rede da área de estudo, abrangendo a conexão existente entre as duas regiões metropolitanas;
- c. Comparar os elementos dos métodos usuais para matrizes OD com a metodologia a partir dos dados de telefonia móvel;
- d. Comparar os custos e tempo despendido em cada um dos métodos analisados;
- e. Comparar os relatórios de consistência de ambos os métodos.

1.2. JUSTIFICATIVA

A necessidade de se construir um trabalho nesta área específica surgiu em virtude da inexistência de estudos com este tipo de análise. O desenvolvimento de um novo método para a caracterização do padrão de viagens atual (pessoas e veículos), principalmente, na rede viária das duas regiões metropolitanas mais populosas da Bahia, pode auxiliar o Estado na otimização da identificação da demanda futura. Trabalhar com um modelo de previsão de demanda de transporte que possa ter matrizes provindas de um sistema automático, economizaria tempo e pouparia elevados custos aos cofres públicos.

Pelas colocações de Zuo, Chen e Yu (2014), dentro de um contexto de ITS (*intelligent transportation system*), as matrizes OD constituem uma informação inicial muito importante, sendo dados essenciais para a gestão de tráfego. À vista disso, a possibilidade de obter matrizes com dados mais confiáveis, sem nenhum tipo de viés ou propensões e, além disso, de uma forma mais prática e menos custosa, justifica a realização de um trabalho na respectiva direção.

As informações obtidas nestes estudos poderão ser utilizadas para fundamentar futuras avaliações de alternativas de intervenções nas malhas viárias destas regiões. Possibilitarão, ainda, a avaliação da implementação de melhorias nos sistemas de transportes.

A utilização dessa tecnologia envolvendo Big Data, exposto por Cui, Yu e Yan (2016) como popularmente caracterizada por “5Vs” (*volume, variety, velocity, value e veracity*), traz o diferencial que este estudo se propõe a investigar, que é a possibilidade de aumento da periodicidade dos estudos, por exemplo. Com a adoção desse novo método, o poder público pode realizar atualizações em suas modelagens com um menor intervalo de tempo, propiciando

a introdução de novos elementos ao longo do horizonte de análise e, conseqüentemente, trazendo mais dinâmica às avaliações.

1.3. METODOLOGIA DA DISSERTAÇÃO

1.3.1. CLASSIFICAÇÃO DE PESQUISAS

Cesário (2020) expõe a existência de diversas formas de pesquisas, as quais foram classificadas segundo a abordagem (quantitativa ou qualitativas), segundo à natureza (básica e aplicada), segundo os objetivos (exploratória, descritiva, explicativa) e segundo os procedimentos técnicos (experimental, bibliográfica, documental, de campo, ex-post-facto, de levantamento, com *survey*, estudo de caso, participante, pesquisa-ação, etnográfica e etnometodológica), evidenciando também as suas diferenças.

O processo de pesquisa sistemática pode ser conduzido tanto através de abordagens quantitativas quanto qualitativas, ou até mesmo uma combinação de ambas. Cada abordagem tem suas características distintas e objetivos específicos na revisão sistemática:

Os passos da abordagem qualitativa podem ser mais bem compreendidos de acordo com os seguintes pontos:

- i. **Síntese Estatística (Metanálise):** Na abordagem quantitativa, a ênfase é a síntese estatística dos resultados dos estudos incluídos. Isso é realizado por meio da metanálise, que é uma técnica estatística que combina os resultados numéricos de diferentes estudos para produzir um efeito geral ou média ponderada.
- ii. **Resultados Numericamente Mensuráveis:** Os estudos quantitativos geralmente produzem resultados numéricos mensuráveis, como médias, desvios padrão, *odds ratio*, risco relativo, entre outros. Esses dados são adequados para a análise estatística na metanálise.
- iii. **Padrões de Evidência:** A abordagem quantitativa busca identificar padrões de evidência consistentes entre os estudos. Ao combinar os resultados numéricos, é possível obter uma estimativa mais precisa do efeito de uma intervenção ou tratamento.

Quanto aos passos da abordagem quantitativa, necessário verificar os seguintes pontos:

- i. **Síntese Interpretativa:** Na abordagem qualitativa, a ênfase é a síntese interpretativa dos resultados dos estudos incluídos. Isso envolve a identificação e análise de temas, conceitos e padrões emergentes nos dados qualitativos coletados nos estudos.
- ii. **Resultados Descritivos e Interpretativos:** Os estudos qualitativos frequentemente fornecem resultados descritivos e interpretativos, como citações de participantes, temas emergentes, experiências, opiniões e significados.
- iii. **Compreensão Profunda do Fenômeno:** A abordagem qualitativa busca uma compreensão profunda do fenômeno estudado, explorando a perspectiva dos participantes e a complexidade do contexto.

O segundo ponto citado, a natureza do processo, pode ser dividido em duas formas específicas de análise. Uma seria pela divisão em pesquisa básica, também conhecida como pesquisa fundamental ou pesquisa pura, tendo como objetivo a geração de conhecimento teórico e a compreensão dos princípios fundamentais de um fenômeno ou área de estudo. Em outras palavras, ela busca responder a perguntas relacionadas ao "como" e "por que" dos fenômenos, sem necessariamente ter uma aplicação prática direta. Seu foco consiste na construção de teorias, no desenvolvimento de modelos conceituais e na exploração de relações entre variáveis, contribuindo para o avanço do conhecimento em uma área específica, fornecendo uma base sólida para pesquisas futuras e para o desenvolvimento de aplicações práticas.

Por outro lado, na pesquisa aplicada, também conhecida como pesquisa prática, o objetivo é a aplicação do conhecimento teórico em contextos práticos e a resolução de problemas do mundo real, buscando responder a perguntas relacionadas ao "como resolver" ou "como melhorar" questões específicas. Seu foco é a solução de problemas e na criação de intervenções, políticas ou estratégias que possam ser aplicadas em situações reais para obter resultados concreto. O resultado é a contribuição para o desenvolvimento de soluções para questões práticas e para a melhoria da sociedade, da saúde, da educação e de outros campos, fornecendo evidências que embasam decisões e intervenções

Ainda em relação à natureza do processo, a pesquisa também precisa estar calcada em uma busca sistemática, na qual o processo envolve uma busca minuciosa e abrangente da literatura disponível em diversas fontes, como bases de dados científicas, revistas, conferências

e outros materiais relevantes. Isso garante que todas as evidências relevantes sejam identificadas. Os artigos e estudos encontrados na busca são submetidos a uma avaliação criteriosa para verificar se atendem aos critérios de inclusão estabelecidos no protocolo. Essa seleção é conduzida de forma independente por dois ou mais revisores, reduzindo o risco de enviesamento na seleção dos estudos.

Os estudos selecionados são avaliados quanto à sua qualidade metodológica, considerando aspectos como o desenho do estudo, tamanho da amostra, métodos de coleta e análise de dados, dentre outros. Essa avaliação crítica ajuda a identificar potenciais vieses que possam afetar a confiabilidade dos resultados. É fundamental que os pesquisadores discutam as limitações do estudo, reconhecendo possíveis fontes de viés ou incertezas que possam influenciar os resultados e interpretações.

Por fim, as conclusões da pesquisa sistemática são baseadas nas evidências obtidas dos estudos incluídos e na análise imparcial e transparente dessas evidências. A natureza da pesquisa sistemática reconhece a importância de atualizar a revisão à medida que novas evidências se tornam disponíveis. Dessa forma, é possível manter a relevância e precisão das conclusões ao longo do tempo.

O terceiro ponto mencionado está associado ao processo de pesquisa sistemática, podendo ser orientado por diferentes objetivos, sendo três dos principais: pesquisa exploratória, pesquisa descritiva e pesquisa explicativa. O Quadro 1 apresenta as diferenças existentes entre esses três tipos de pesquisas.

PESQUISA EXPLORATÓRIA	PESQUISA DESCRITIVA	PESQUISA EXPLICATIVA
<p data-bbox="389 331 582 363" style="text-align: center;">Características</p> <p data-bbox="185 384 788 616">A pesquisa exploratória tem como objetivo principal explorar um tópico de pesquisa pouco conhecido ou pouco explorado. Ela é conduzida quando há pouca ou nenhuma literatura sistemática disponível sobre o assunto, e busca mapear e identificar os principais conceitos, ideias e tendências.</p> <p data-bbox="421 679 551 711" style="text-align: center;">Perguntas</p> <p data-bbox="194 738 779 871">A pesquisa exploratória pode ser iniciada com perguntas amplas, como "Quais são os principais temas abordados em relação a X?", ou "Quais são as perspectivas atuais sobre o tema Y?"</p> <p data-bbox="450 943 521 975" style="text-align: center;">Busca</p> <p data-bbox="194 986 779 1150">A busca de literatura na pesquisa exploratória é mais aberta e não se limita a estudos específicos, pois o objetivo é identificar uma variedade de fontes para construir um panorama mais amplo do tópico.</p> <p data-bbox="416 1206 555 1238" style="text-align: center;">Resultados</p> <p data-bbox="194 1249 779 1374">Os resultados da pesquisa exploratória geralmente fornecem uma visão geral do campo de estudo, identificando as principais áreas de interesse e as lacunas existentes na literatura.</p>	<p data-bbox="1032 331 1225 363" style="text-align: center;">Características</p> <p data-bbox="824 400 1431 600">A pesquisa descritiva tem como objetivo principal descrever um fenômeno ou evento específico de forma detalhada e sistemática. Ela é conduzida quando já existe uma quantidade razoável de literatura sobre o tópico, e busca reunir e resumir os resultados de estudos existentes.</p> <p data-bbox="1061 679 1191 711" style="text-align: center;">Perguntas</p> <p data-bbox="824 722 1431 887">A pesquisa descritiva pode ser iniciada com perguntas mais específicas, como "Quais são os resultados dos estudos sobre a eficácia do tratamento X?", ou "Quais são os fatores associados ao fenômeno Y?".</p> <p data-bbox="1090 943 1167 975" style="text-align: center;">Busca</p> <p data-bbox="824 1002 1431 1134">A busca na pesquisa descritiva é mais focada e direcionada para identificar estudos relevantes que possam responder às perguntas específicas da revisão.</p> <p data-bbox="1057 1206 1196 1238" style="text-align: center;">Resultados</p> <p data-bbox="824 1249 1431 1374">Os resultados da pesquisa descritiva fornecem uma síntese detalhada dos principais achados dos estudos incluídos, permitindo uma compreensão aprofundada do tópico em questão.</p>	<p data-bbox="1671 331 1863 363" style="text-align: center;">Características</p> <p data-bbox="1462 419 2069 584">A pesquisa explicativa tem como objetivo principal identificar e explicar as relações causais entre variáveis e fenômenos. Ela é conduzida quando se busca entender os mecanismos e as razões por trás das associações encontradas nos estudos.</p> <p data-bbox="1702 679 1832 711" style="text-align: center;">Perguntas</p> <p data-bbox="1462 722 2069 887">A pesquisa explicativa pode ser iniciada com perguntas do tipo "Quais são os fatores que influenciam a relação entre X e Y?", ou "Por que uma intervenção específica tem um efeito positivo em um grupo de pessoas e não em outro?".</p> <p data-bbox="1731 943 1807 975" style="text-align: center;">Busca</p> <p data-bbox="1462 1002 2069 1134">A busca na pesquisa explicativa é conduzida para identificar estudos que investiguem relações de causa e efeito, e que apresentem dados relevantes para análise causal.</p> <p data-bbox="1697 1206 1836 1238" style="text-align: center;">Resultados</p> <p data-bbox="1462 1249 2069 1374">Os resultados da pesquisa explicativa geralmente fornecem uma análise mais profunda das relações de causa e efeito entre as variáveis estudadas, buscando elucidar os mecanismos subjacentes.</p>

Fonte: Elaborado pelo Autor

A quarta forma diz que o processo de pesquisa sistemática pode ser conduzido utilizando diferentes procedimentos técnicos, dependendo da natureza da pergunta de pesquisa e dos objetivos da revisão. A seguir, há uma relação que apresenta os principais procedimentos técnicos, expondo a diferença entre suas características e suas aplicações:

CARACTERÍSTICAS

- i. **Pesquisa Experimental:** Envolve a manipulação deliberada de uma ou mais variáveis independentes para observar os efeitos causais nas variáveis dependentes. É considerada uma forma de pesquisa controlada e altamente rigorosa.
- ii. **Pesquisa Bibliográfica:** Envolve a busca, coleta e análise de informações provenientes de fontes bibliográficas, como artigos científicos, livros, teses, monografias, entre outros.
- iii. **Pesquisa Documental:** Envolve a análise de documentos oficiais, relatórios, registros, legislação e outras fontes documentais para obter informações relevantes para a pesquisa.
- iv. **Pesquisa Campo:** Envolve a coleta direta de dados em locais específicos, por meio de observações, entrevistas ou questionários.
- v. **Pesquisa Ex-Post-Facto:** Examina relações de causa e efeito em eventos que já ocorreram, sem a manipulação direta de variáveis independentes.
- vi. **Pesquisa Levantamento:** Envolve a coleta de dados de uma amostra representativa de uma população-alvo por meio de questionários estruturados ou entrevistas.
- vii. **Estudo de Caso:** Envolve uma análise detalhada e aprofundada de um caso ou situação específica.
- viii. **Pesquisa Participante:** Envolve a participação ativa de pesquisadores e participantes do estudo em todas as fases da pesquisa, colaborando para gerar conhecimento de forma conjunta.
- ix. **Pesquisa Ação:** Envolve a colaboração entre pesquisadores e participantes para identificar e implementar soluções práticas para problemas específicos.
- x. **Pesquisa Etnográfica:** Envolve a imersão do pesquisador em um contexto social para observar e entender os comportamentos e práticas culturais dos participantes.

- xii. **Pesquisa Etnometodológica:** Se concentra na análise da maneira como as pessoas dão sentido e interpretam suas ações no contexto social.

APLICAÇÃO NA PESQUISA

- i. **Pesquisa Experimental:** Menos comum em revisões sistemáticas, mas pode ser utilizada quando há ensaios clínicos randomizados disponíveis, pois esses estudos são projetados para avaliar causalidade.
- ii. **Pesquisa Bibliográfica:** Parte essencial do processo de pesquisa sistemática, pois envolve a busca abrangente de estudos relevantes para a revisão.
- iii. **Pesquisa Documental:** Pode ser útil em revisões sistemáticas que visam coletar informações de documentos oficiais ou registros que não foram publicados em periódicos.
- iv. **Pesquisa Campo:** Pode ser usada para complementar a análise de estudos disponíveis e obter informações adicionais de participantes ou contextos específicos.
- v. **Pesquisa Ex-Post-Facto:** Pode ser relevante em revisões sistemáticas que buscam examinar associações causais em estudos observacionais.
- vi. **Pesquisa Levantamento:** Pode ser usada em revisões sistemáticas para agregar dados de estudos primários que utilizaram questionários como instrumento de coleta de dados.
- vii. **Estudo de Caso:** Pode ser relevante em revisões sistemáticas que buscam entender fenômenos complexos em contextos específicos.
- viii. **Pesquisa Participante:** Pode ser usada em revisões sistemáticas que buscam envolver os participantes no processo de pesquisa e na análise dos resultados.
- ix. **Pesquisa Ação:** Pode ser usada em revisões sistemáticas que buscam identificar e avaliar a eficácia de intervenções em situações práticas.
- x. **Pesquisa Etnográfica:** Pode ser utilizada em revisões sistemáticas que buscam entender a cultura e os contextos em que os estudos primários foram conduzidos.
- xi. **Pesquisa Etnometodológica:** Pode ser relevante em revisões sistemáticas que buscam entender a interpretação e o significado dos resultados dos estudos primários.

O método empregado nesta dissertação se concentra na simulação de tráfego rodoviário utilizando duas fontes distintas de dados para analisar as técnicas de aquisição, no intuito de se obter uma melhor compreensão dos levantamentos e suas implicações. A abordagem aplicada envolve a coleta e análise de dados reais para alimentar o processo de simulação.

Em relação à classificação da pesquisa deste trabalho é de natureza aplicada, buscando resolver problemas concretos no campo do tráfego rodoviário e da mobilidade urbana. Para fundamentar esse estudo, recorreu-se a autores como Ribeiro et al. (2019), que enfatizam a importância da aplicação de simulações para melhorar o planejamento e a gestão do tráfego urbano.

Da mesma forma, como é um trabalho que envolveu a coleta e análise de dados numéricos para quantificar fenômenos e padrões, pode-se considerar também como pesquisa quantitativa. No contexto de um estudo de tráfego, pesquisas quantitativas incluem a contagem de veículos em pontos de tráfego, a análise de dados de GPS de veículos em tempo real, a medição de velocidades médias, tempos de viagem e outras métricas quantificáveis. Esses dados ajudam na identificação de congestionamentos, fluxos de tráfego em diferentes horários e tendências de mobilidade.

1.3.2. FONTES DE DADOS

A aquisição de dados é um passo fundamental para alimentar a simulação de tráfego com informações realistas. Foram utilizados dados de fontes variadas, incluindo pesquisas domiciliares, sensores de tráfego (microondas), sistemas de posicionamento global (GPS) e dados de câmeras de vigilância. Autores como Silva et al. (2020) exploram a integração de diferentes fontes de dados para melhorar a precisão das simulações de tráfego.

1.3.3. COLETA E PRÉ-PROCESSAMENTO DE DADOS

Os dados coletados foram pré-processados para remover *outliers* e inconsistências, visando garantir a qualidade dos dados de entrada para a simulação. Dessa forma, foram seguidos os procedimentos recomendados por Zheng et al. (2010) para a limpeza e preparação de dados de mobilidade urbana.

Um dos softwares utilizados em quase todas as etapas de processamento foi o Microsoft Excel, possibilitando as análises através de macros e planilhas dinâmicas, além de fórmulas de busca matricial ao longo de uma série de planilhas.

1.3.4. MODELO E PARÂMETROS DE SIMULAÇÃO

Optou-se pelo emprego do modelo de simulação macroscópica VISUM, da empresa alemã PTV Group, pois este software, além de permitir uma representação adequada para análises de planejamento urbano e regional, com a adoção de elementos da infraestrutura viária, possui uma gama de funcionalidades que permitem um maior aprofundamento nos estudos, tem uma interface amigável e integração com outros softwares e plataformas, tornando-o versátil e capaz de atender às necessidades específicas de projetos complexos de transporte.

Os parâmetros de simulação foram definidos com base em estudos anteriores, como o trabalho de Treiber e Kesting (2013), que discute a importância da calibração adequada dos parâmetros, tais como capacidade de vias, velocidade dos tipos de veículos, curvaturas horizontais e verticais, entre outros, para obter resultados realistas e significativos.

Para validar a simulação, os resultados obtidos foram comparados com os dados reais de tráfego coletados. Esse processo é crucial para garantir a precisão do modelo. Adotou-se a abordagem proposta por van Lint e van Zuylen (2003) para avaliar a validade dos resultados da simulação.

1.3.5. PESQUISA EXPLORATÓRIA

O foco deste trabalho consistiu na análise do comportamento dos fluxos de tráfego entre as duas regiões metropolitanas mais populosas do estado da Bahia: Feira de Santana e Salvador. Os dados obtidos vieram de uma compilação de estudos realizados na região, do ano de 2017 até o presente momento, gerando uma base sólida para o desenvolvimento deste modelo.

Foram também realizadas entrevistas que envolveram um volume da ordem de 25 mil pessoas, resultando em mais de 65 mil viagens, divididas em tipos de veículos de uso privado e de transporte coletivo. Essas entrevistas foram essenciais para montar um banco de dados com informações sobre domicílios, indivíduos, segurança viária, dados socioeconômicos e, principalmente, dados de deslocamento.

A coleta de dados envolveu a consideração de questões éticas, como privacidade e consentimento dos usuários. Com base nessa política, foram seguidas as diretrizes de Sánchez-Medina et al. (2018) para garantir que a pesquisa fosse conduzida de maneira ética e responsável.

Além de entrevistas, outro banco de dados foi montado de forma simultânea com a aquisição de dados móveis provindos de empresas de telefonia. Esse outro banco de dados é o foco deste trabalho, o qual serviu para o desenvolvimento da análise comparativa entre as matrizes utilizadas para a simulação realizada.

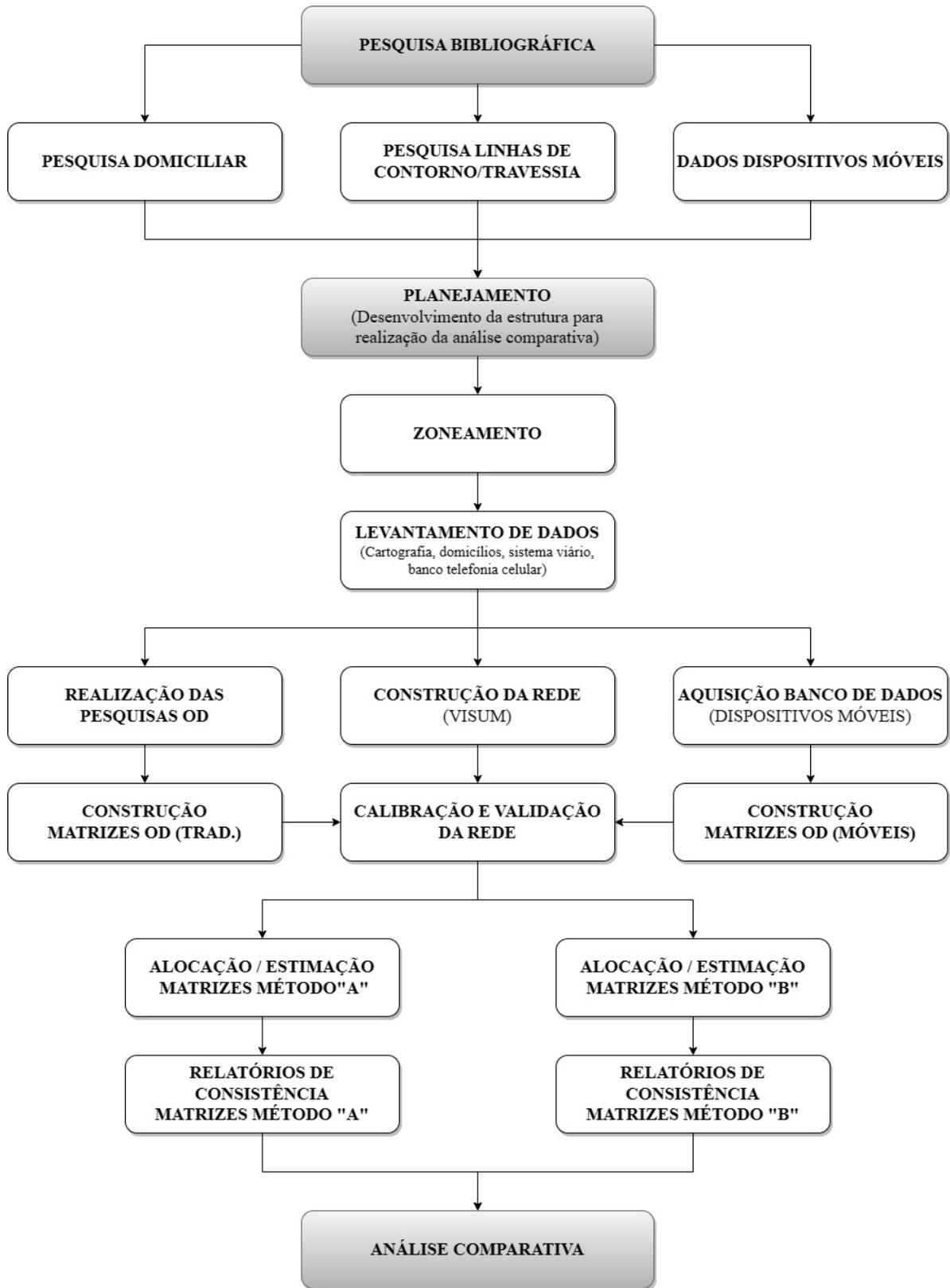
Todo o processo de construção das duas regiões metropolitanas foi realizado com a utilização simultânea dos softwares ArcGIS 10.4 e VISUM Expert 2022 (SP 1-10). Essa combinação permitiu utilizar as capacidades de modelagem espacial e de tráfego para criar uma representação mais precisa da rede de transporte e a criação de modelos de transporte que consideram tanto os aspectos geográficos quanto os fluxos de tráfego, proporcionando melhor análise de mobilidade urbana e planejamento de transporte.

A construção de uma rede com a integração desses programas permitiu maior detalhamento das regiões metropolitanas, como:

- Precisão geográfica: Vias, cruzamentos, zonas urbanas e outros elementos geográficos foram mapeados de forma precisa.
- Conectividade: A modelagem garantiu a conectividade entre todos os nós da malha, com vias interligadas e topologia da rede consistente.
- Atributos da rede: Correta atribuição da capacidade das vias, velocidades máximas, restrições de direção e outros fatores que influenciam a modelagem.
- Simulação de tráfego: Simulação com cenários considerando fluxos de tráfego, tempos de viagem, congestionamentos e demais aspectos do transporte.
- Planejamento de transporte: Avaliação de impactos de projetos de infraestrutura, planejamento de transporte público, gerenciamento de tráfego e otimização de rotas.

O fluxograma apresentado na

Figura 2 ilustra as etapas de desenvolvimento deste trabalho.



Fonte: Autor, 2023.

1.4. ESTRUTURA DA DISSERTAÇÃO

A organização deste trabalho foi desenvolvida com o propósito de elucidar as questões referentes a este novo método de levantamento de dados para, posteriormente, apresentar a área de estudo seguida de sua modelagem. Da mesma forma, será abordado o tratamento necessário dos dados, a alocação e, ao final, uma análise comparativa entre o método tradicional e este que está sendo investigado pelo autor.

A seguir, apresenta-se a estrutura desta dissertação:

Capítulo 1: Introdução – Onde são abordadas as questões pertinentes ao presente trabalho, apresentando um contexto geral, objetivos e justificativa para o desenvolvimento dentro deste tema.

Capítulo 2: Fundamentação Teórica – Trazendo elementos necessários para a aplicação do método a ser analisado.

Capítulo 3: Estudo de Caso – Detalhamento sobre as regiões que compõem a área de estudo, com informações socioeconômicas e de infraestrutura existente, para a definição do zoneamento. O capítulo também traz o processo de levantamento de dado, tanto pelo método tradicional (e posterior tratamento para sua importação e utilização no modelo) quanto pelo método de dados móveis.

Capítulo 4: Construção de Matrizes e Modelagem – São listadas aqui as duas formas de construção de matrizes, através do método tradicional e o de dados móveis. Em cada uma, foram abordados os passos necessários para o respectivo desenvolvimento. Neste capítulo também se encontram os processos de modelagem, validação, alocação e estimação, com as devidas observações sobre os processos.

Capítulo 5: Análise Comparativa – Neste momento ambos os resultados são confrontados, juntamente com seus respectivos relatórios de consistência, tempo despendido para levantamento e relatório de custos, para que seja possível verificar a viabilidade (ou não) de migrar do modelo tradicional para o novo método.

Capítulo 6: Conclusões e Recomendações – Por fim, o último capítulo traz as conclusões do trabalho e recomendações relevantes para a continuidade desse processo de análise.

Referências: Lista de todas as fontes utilizadas ao longo do desenvolvimento do trabalho, que tornaram possível a sua elaboração.

Apêndice e Anexo: Respectivamente, arquivos construídos e extraídos de fontes.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. MATRIZES OD

As matrizes Origem-Destino (OD) são uma ferramenta essencial para análise e planejamento de transportes (Ben-Akiva & Lerman, 1985), tanto para entornos rurais quanto no âmbito de mobilidade urbana. Segundo Chiu et al. (2011), essas matrizes fornecem informações sobre as viagens realizadas pelos usuários do sistema de transporte, dentro de uma determinada área geográfica, incluindo a origem e o destino, o modo de transporte utilizado e o tempo gasto em cada viagem.

As matrizes OD podem ser obtidas por meio de diversas fontes de dados, como pesquisas de origem-destino, dados de bilhetagem eletrônica e dados de GPS em dispositivos móveis (Zheng, 2015). De acordo com Nunes et al. (2021), uma das principais vantagens da utilização de dados de GPS é a possibilidade de obter informações em tempo real sobre o comportamento dos usuários do transporte.

No entanto, para a obtenção de matrizes OD a partir de dados de GPS, é necessário realizar etapas de pré-processamento e tratamento dos dados (Ma et al., 2020). Além disso, é necessário considerar a classificação dos modos de transporte utilizados pelos usuários, a fim de se obter matrizes OD específicas para cada modo (Wu et al., 2017).

Diversos algoritmos têm sido propostos para a construção de matrizes OD a partir de dados de GPS, como árvores de decisão (Chen et al., 2012), redes neurais (Liu et al., 2014) e florestas aleatórias (Xie et al., 2016). Além disso, técnicas de *clustering* têm sido aplicadas para a classificação dos modos de transporte (Sun et al., 2018).

Uma das abordagens mais recentes para a obtenção de dados para matrizes OD é o uso de dados móveis, como registros de torres de celular e dispositivos de GPS. Esses dados podem ser usados para identificar padrões de deslocamento e inferir matrizes OD a partir de algoritmos de aprendizado de máquina e técnicas de análise espacial.

Um estudo recente de Jiang et al. (2021) mostrou que dados de localização de smartphones podem ser utilizados para a criação de matrizes OD com alta precisão. O estudo foi realizado na cidade de Pequim, China, e utilizou dados de localização de mais de 10 milhões de smartphones. Os resultados mostraram que as matrizes OD obtidas a partir desses dados eram altamente precisas e poderiam ser utilizadas para planejamento urbano e tomada de decisão.

Outro estudo, de Liu et al. (2020), utilizou dados de GPS de veículos de transporte público para criar matrizes OD em uma cidade chinesa. Os resultados mostraram que os dados de GPS podem ser uma fonte viável para a obtenção de matrizes OD precisas, especialmente em cidades em desenvolvimento, onde a coleta de dados tradicional pode ser mais difícil.

Esses estudos mostram que o uso de dados móveis para a obtenção de matrizes OD pode ser uma alternativa viável e precisa para a coleta de dados tradicional. No entanto, ainda é necessário trabalhar as lacunas deixadas por outras pesquisas para avaliar a aplicabilidade dessas técnicas em diferentes contextos urbanos e para diferentes tipos de viagens.

Em adição, um aspecto fundamental no campo de estudos de matrizes origem e destino é a distinção entre sistemas e fluxos de mobilidade. Um **sistema de mobilidade** refere-se à infraestrutura, redes de transporte, padrões de deslocamento e elementos que compõem a estrutura física e organizacional que permite a movimentação de pessoas ou bens em uma determinada área. Conforme apontado por Ortúzar e Willumsen (2011), sistemas de mobilidade abrangem desde ruas e rodovias até sistemas de transporte público, ciclovias e estratégias de gestão de tráfego.

Por outro lado, **fluxos de mobilidade** são as viagens ou deslocamentos individuais que ocorrem dentro desse sistema. De acordo com Hansen (1959), um dos pioneiros nesse campo, os fluxos representam os movimentos reais das pessoas entre diferentes zonas ou locais ao longo do tempo. Eles podem ser medidos em termos de números de viagens, distâncias percorridas ou duração dos deslocamentos, e fornecem *insights* cruciais para entender como as pessoas se movem dentro de um sistema de mobilidade.

Essa distinção entre sistemas e fluxos é essencial para a compreensão abrangente dos padrões de mobilidade em uma região ou área urbana. Enquanto os sistemas de mobilidade estabelecem a estrutura que permite a movimentação, os fluxos de mobilidade revelam como essa movimentação realmente ocorre no dia a dia das pessoas. Integrar esses conceitos é crucial para o planejamento eficaz de transportes e políticas de mobilidade urbana, conforme destacado por Hansen (1967).

2.2. TRATAMENTO DE DADOS

O tratamento de dados é uma etapa crucial na análise de matrizes OD de dados móveis. Devido à grande quantidade de dados gerados por dispositivos móveis, é importante que sejam coletados, processados e analisados de forma eficiente para obter informações úteis.

2.2.1. LIMPEZA DE DADOS

Uma das principais etapas de tratamento de dados em matrizes OD de dados móveis é a limpeza de dados. A limpeza de dados envolve a identificação e correção de dados incompletos, inconsistentes ou incorretos. Esses problemas de qualidade de dados podem surgir devido a vários fatores, como erros de sensor, falhas de rede ou imprecisões de localização. A limpeza de dados pode envolver a remoção de valores nulos, a correção de erros de sensor ou a detecção de *outliers*. A limpeza de dados pode ser feita manualmente ou com o uso de algoritmos de detecção de anomalias e técnicas de aprendizado de máquina.

Através de fontes como McKinney (2017) e Smith e Jones (2019), foi possível compreender o detalhamento dos processos envolvidos na limpeza dos dados:

- i. **Identificação de dados inválidos:** A primeira etapa no processo de limpeza de dados é identificar dados inválidos, que incluem valores nulos, valores fora do intervalo válido ou dados inconsistentes.
- ii. **Eliminação de dados duplicados:** A eliminação de dados duplicados é uma etapa importante na limpeza de dados, pois dados duplicados podem levar a resultados enganosos. Para eliminar dados duplicados, é necessário identificar registros que possuem valores idênticos em todas as colunas e manter apenas um deles.
- iii. **Correção de erros de sensor:** Em alguns casos, os sensores podem produzir valores incorretos ou inconsistentes. Para corrigir erros de sensor, pode ser necessário substituir valores inválidos por valores estimados ou interpolados.
- iv. **Detecção de outliers:** *Outliers* são valores extremos que não seguem o padrão dos outros dados. A detecção de *outliers* pode ser feita utilizando várias técnicas, como análise de *boxplot*, análise de desvio padrão ou técnicas de aprendizado de máquina. *Outliers* podem ser removidos ou substituídos por valores estimados.
- v. **Validação de dados:** A validação de dados envolve a verificação dos dados para garantir que eles estejam corretos e precisos. Para validar dados, é necessário compará-los com fontes de dados confiáveis ou estabelecer regras para verificar a integridade dos dados.

Uma das técnicas mais comuns é a imputação de dados faltantes ou inconsistentes. Por exemplo, Jiang et al. (2018) propuseram uma abordagem baseada em Krigagem para imputação de dados faltantes em matrizes OD de transporte público. A técnica de Krigagem é uma técnica estatística para previsão espacial, que pode ser usada para preencher valores ausentes em matrizes OD.

Outra abordagem é a utilização de técnicas de detecção de *outliers* para identificar e remover valores anômalos nas matrizes OD (Zhang et al., 2018). Essa abordagem é útil para remover valores inconsistentes que podem afetar a precisão das análises realizadas com as matrizes OD.

Além disso, também pode ser útil utilizar técnicas de normalização para ajustar os valores nas matrizes OD (Jin et al., 2019). A normalização pode ser usada para lidar com diferentes escalas de valores nas matrizes OD, o que pode afetar a precisão das análises realizadas.

2.2.2. AGREGAÇÃO DE DADOS

Outra etapa importante no tratamento de dados em matrizes OD de dados móveis é a agregação de dados. A agregação de dados envolve a combinação de dados individuais em intervalos de tempo específicos, como horas ou dias. A agregação de dados pode ser útil para reduzir o tamanho da matriz OD e remover o ruído de dados, o que pode tornar a análise mais eficiente. A escolha do intervalo de agregação depende do objetivo da análise. Intervalos menores, como minutos ou segundos, podem ser mais adequados para análises de curto prazo, enquanto intervalos maiores, como horas ou dias, podem ser mais apropriados para análises de longo prazo.

Um maior detalhamento dos processos envolvidos na agregação dos dados é abordado por Longley et al. (2015), descrito a seguir:

- i. **Definição da unidade de agregação:** A primeira etapa no processo de agregação de dados é definir a unidade de agregação. A unidade de agregação é a escala em que os dados serão agregados. Por exemplo, a unidade de agregação pode ser definida como município, bairro ou quadra.
- ii. **Seleção de variáveis:** A segunda etapa no processo de agregação de dados é selecionar as variáveis que serão agregadas. As variáveis selecionadas devem ser relevantes para a análise de fluxo de tráfego, como o número de viagens, a duração das viagens, a distância percorrida ou o tempo gasto em congestionamento.

- iii. **Agregação de dados:** A terceira etapa no processo de agregação de dados é a agregação propriamente dita. Os dados individuais são combinados em grupos maiores, utilizando técnicas de agregação, como soma, média, contagem ou mediana. Por exemplo, o número de viagens pode ser somado para cada unidade de agregação.
- iv. **Análise dos resultados:** A última etapa no processo de agregação de dados é a análise dos resultados. Os resultados agregados podem ser visualizados em um mapa ou em gráficos, para ajudar a entender os padrões de fluxo de tráfego. A análise dos resultados pode ajudar a identificar áreas com altos volumes de tráfego, áreas de congestionamento ou padrões de deslocamento da população.

Uma abordagem comum é a agregação temporal, que consiste em agrupar os dados coletados em intervalos de tempo pré-determinados (Hosseini-Motlagh et al., 2014). Por exemplo, pode-se agrupar as informações coletadas em intervalos de 15 minutos ou uma hora para obter matrizes OD com menor granularidade temporal.

Outra abordagem é a agregação espacial, que consiste em agrupar as informações coletadas em áreas geográficas pré-determinadas (Xie et al., 2015). Por exemplo, pode-se agrupar as informações coletadas em bairros ou zonas para obter matrizes OD com menor granularidade espacial.

Além disso, também é possível combinar as abordagens de agregação temporal e espacial para obter matrizes OD com menor granularidade tanto temporal quanto espacial (Wang et al., 2019).

A agregação de dados em matrizes OD de dados móveis é uma área de estudo em constante evolução, com diversas técnicas e metodologias sendo propostas na literatura para melhorar a qualidade das matrizes OD.

2.2.3. NORMALIZAÇÃO DE DADOS

A normalização de dados é outra etapa importante no tratamento de matrizes OD de dados móveis. A normalização envolve a transformação de dados brutos em uma escala comum, o que permite a comparação direta entre diferentes conjuntos de dados. A normalização de dados pode ser feita por meio de várias técnicas, como normalização *min-max* e normalização *z-score*. A normalização de dados é importante para garantir que diferentes conjuntos de dados sejam comparáveis.

García et al. (2020), bem como Zheng, A. e Casari, A. (2018) abordam passos detalhados dos processos envolvidos na normalização dos dados, como descritos a seguir:

- i. **Seleção das variáveis:** A primeira etapa do processo de normalização de dados é a seleção das variáveis que serão normalizadas. Essas variáveis devem ser escolhidas com base na relevância para a análise de fluxos de tráfego, como o número de viagens, a duração das viagens ou a distância percorrida.
- ii. **Definição da escala de normalização:** A segunda etapa do processo é a definição da escala de normalização. A escala de normalização pode ser baseada na média e no desvio padrão das variáveis selecionadas ou em uma escala predefinida.
- iii. **Cálculo dos valores normalizados:** A terceira etapa é o cálculo dos valores normalizados. O objetivo é transformar os valores das variáveis em uma escala comum, para que possam ser comparados. Existem várias técnicas de normalização, incluindo a normalização min-max, z-score e a normalização por escalonamento decimal.
- iv. **Análise dos resultados normalizados:** A última etapa é a análise dos resultados normalizados. Os dados normalizados podem ser usados para identificar tendências e padrões, e para comparar diferentes regiões, períodos ou grupos populacionais.

Uma das técnicas mais comuns é a normalização min-max, que ajusta os valores para um intervalo entre 0 e 1 (Ghosh et al., 2016). Essa técnica é amplamente utilizada em matrizes OD para garantir que as diferentes variáveis tenham a mesma escala e possam ser comparadas de forma equilibrada.

Outra técnica comum é a normalização z-score, que ajusta os valores para uma escala com média zero e desvio padrão um (Kim et al., 2019). Essa técnica é útil quando os valores dos dados coletados possuem uma distribuição normal.

Além disso, também é possível utilizar técnicas de normalização personalizadas para lidar com problemas específicos, como a normalização logarítmica para lidar com valores extremamente grandes (Liu et al., 2018).

2.2.4. ANÁLISE DE DADOS

A análise de dados é a etapa final no tratamento de dados em matrizes OD de dados móveis. A análise pode envolver a utilização de técnicas de visualização de dados, análise de *cluster*, análise de rede ou outras técnicas de mineração de dados para extrair informações úteis da matriz OD. O objetivo final da análise é fornecer *insights* úteis sobre o fluxo de tráfego de dados móveis, permitindo que sejam tomadas decisões informadas para melhorar a qualidade de serviços e a mobilidade urbana.

Uma das técnicas mais utilizadas é a análise de *cluster*, que permite identificar grupos de dados que possuem características similares (Zheng et al., 2016). Essa técnica é útil para identificar padrões de deslocamento e comportamento dos usuários de transporte.

Outra técnica comum é a análise de rede, que permite modelar a interação entre as diferentes regiões ou locais e identificar padrões de fluxo (Cervero & Kockelman, 1997). Essa técnica é particularmente útil para a análise de fluxos de transporte e mobilidade urbana.

Além disso, também é possível utilizar técnicas de análise espacial para identificar padrões geográficos nos dados coletados (Silva et al., 2018). Essa técnica é útil para identificar regiões com maior demanda de transporte e planejar rotas mais eficientes.

2.3. CLASSIFICAÇÃO DE MATRIZES

A classificação de matrizes Origem-Destino (OD) é a etapa de agrupamento de viagens entre zonas em diferentes categorias ou modos de viagens. É uma importante etapa na análise de mobilidade urbana, pois permite identificar padrões de fluxo e comportamento dos usuários em uma determinada área geográfica. Com o aumento do uso de dados móveis para a obtenção de matrizes OD, torna-se importante também aprimorar as técnicas de classificação dessas matrizes.

Uma das abordagens mais recentes para a classificação de matrizes OD é o uso de algoritmos de aprendizado de máquina. Por exemplo, Zhou et al. (2021) utilizaram algoritmos de clusterização para identificar diferentes padrões de deslocamento em uma cidade chinesa. O estudo mostrou que a classificação das matrizes OD de acordo com esses padrões pode ser útil para aprimorar as políticas de transporte urbano.

Wei et al. (2019) utilizaram redes neurais artificiais (RNA) para classificar matrizes OD de diferentes cidades chinesas. Os resultados mostraram que essa abordagem pode ser utilizada para identificar padrões de fluxo em diferentes contextos urbanos.

Além das técnicas de aprendizado de máquina, também é possível utilizar métodos mais tradicionais, como a análise de componentes principais (PCA). Por exemplo, um estudo de Qian et al. (2019) utilizou PCA para identificar padrões de fluxo de transporte público em uma cidade chinesa. Os resultados mostraram que essa abordagem pode ser útil para a identificação de padrões de demanda e planejamento de serviços de transporte.

Outra abordagem comum é a utilização de técnicas de *clustering* para agrupar as origens e destinos em *clusters*, de modo que as matrizes OD possam ser classificadas por similaridade entre os *clusters*. Por exemplo, Castro et al. (2014) utilizaram o algoritmo k-means para agrupar as estações de bicicleta compartilhadas de Paris em 13 *clusters* e classificar as matrizes OD de bicicleta compartilhada geradas a partir dos dados dessas estações. Outra abordagem é a utilização de técnicas de decomposição de matrizes, como a decomposição em valores singulares (SVD), para extrair padrões e estruturas das matrizes OD (Ji et al., 2018).

Uma técnica mais recente é a utilização de redes neurais convolucionais (CNN) para classificação de matrizes OD (Ma et al., 2021). Essa abordagem utiliza uma arquitetura de CNN para aprender características das matrizes OD e classificá-las em diferentes categorias. A vantagem dessa técnica é que ela é capaz de aprender padrões complexos e não-lineares nas matrizes OD, o que pode levar a uma melhor precisão na classificação.

Esses estudos mostram que a classificação de matrizes OD provinda de dados móveis pode ser realizada com diferentes técnicas, desde métodos tradicionais até algoritmos de aprendizado de máquina. No entanto, ainda são necessárias mais pesquisas para avaliar a aplicabilidade dessas técnicas em diferentes contextos urbanos e para diferentes tipos de viagens.

A seguir serão apresentados quatro diferentes métodos para realizar tal classificação.

2.3.1. CLASSIFICAÇÃO DE VEÍCULO BASEADA EM GPS

Esta forma de classificação independe de qualquer operação manual ou presencial, nos quais se baseiam os modelos tradicionais. Para esta classificação são utilizados algoritmos de aprendizado de máquina, que analisam os veículos com base em seus sinais de GPS e, através da leitura de vários recursos dos sinais de GPS, como velocidade, aceleração e localização, conseguem determinar o tipo de veículo, classificando-o como um veículo de uso coletivo ou privado, por exemplo. Este método é particularmente útil para conjuntos de dados de grande escala, pois pode classificar automaticamente um grande número de viagens em diferentes tipos de veículos.

Este modelo de classificação oferece benefícios, como a classificação de veículos em tempo real, possibilitando identificar o tipo de veículo mesmo em movimento. Essas informações podem ser usadas para uma variedade de aplicações, como monitoramento de tráfego, planejamento de transporte e análise de segurança rodoviária.

O método baseado em GPS pode ser dividido em duas etapas principais: a primeira é a identificação das trajetórias dos usuários, através da coleta de dados de GPS de um número expressivo de veículos, com a posterior divisão desses dados em diferentes categorias com base no tipo de veículo, e a segunda é a inferência dos pares OD correspondentes a essas trajetórias (Han et al., 2016), podendo envolver o uso de técnicas como árvores de decisão, florestas aleatórias ou redes neurais. A identificação de trajetórias pode ser feita por meio de técnicas de *clustering*, como o algoritmo de DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), ou por meio de técnicas de correspondência de pontos de trajetória.

A inferência dos pares OD pode ser feita por meio de técnicas de correspondência de trajetórias, como o algoritmo de DynaMatch, que compara as trajetórias dos usuários com as rotas possíveis e seleciona o par OD mais provável (Zheng et al., 2011). Outra abordagem é a inferência de OD com base em modelos de escolha modal, que levam em consideração as características dos usuários, como a idade e a renda, para estimar os pares OD correspondentes (Zheng et al., 2014).

Apesar de ser uma abordagem popular, o método baseado em GPS apresenta algumas limitações, como a necessidade de uma boa cobertura de GPS e a falta de precisão nas estimativas de trajetórias em áreas com alta densidade de edifícios (Hosseini et al., 2019). Portanto, é importante considerar outras abordagens, como o método baseado em dados de bilhetagem, para complementar a análise de dados de mobilidade.

"A construção de um algoritmo de classificação segue uma sequência lógica" (Müller & Guido, 2017, p. 123). Os seguintes passos são descritos por Müller & Guido, 2017:

- i.** Construção da matriz OD de dados móveis;
- ii.** Determinação dos critérios de classificação através de dados de texto, numéricos ou imagens a serem reconhecida, ou com base no tamanho da matriz;
- iii.** Pré-processamento dos dados móveis, ajustando os dados necessário para sua classificação, através da remoção de valores ausentes, dimensionamento dos dados ou transformação dos dados em um formato mais apropriado;
- iv.** Extração de recursos que sejam relevantes para os critérios de classificação;
- v.** Treinamento do modelo de classificação para o aprendizado de máquina, baseado nos recursos extraídos para classificar a matriz de dados móveis.
- vi.** Avaliação de desempenho do modelo de classificação treinado, através de um conjunto de validação. Nessa etapa, pode vir a ser necessário ajustar os parâmetros do modelo ou tentar diferentes algoritmos para melhorar o desempenho.
- vii.** Aplicação do modelo após aprovação do desempenho do modelo de classificação.

A seguir são apresentados três diferentes métodos de algoritmos comumente utilizados.

2.3.1.1. Árvores de Decisão

As árvores de decisão são algoritmos de aprendizado de máquina simples, com alto poder de análise, desenvolvido para problemas de regressão e classificação. É um tipo de aprendizado supervisionado, onde o algoritmo aprende a partir de dados rotulados para fazer previsões. O primeiro passo do processo é a divisão dos dados em subconjuntos cada vez menores para, em seguida, começar uma série de questionamentos sobre os recursos dos dados, como "Qual a velocidade média?" ou "Essa região possui muitas motos?" Com base nas respostas obtidas, os dados são subdivididos em grupos cada vez menores até que o algoritmo chegue a uma decisão.

O algoritmo de árvores de decisão tem sido amplamente estudado na literatura e existem diversas variações e aprimoramentos disponíveis. Uma abordagem comum é o algoritmo ID3 (Iterative Dichotomiser 3), proposto por Quinlan em 1986 (Quinlan, 1986), que utiliza a medida de ganho de informação para selecionar o atributo mais importante para dividir o conjunto de dados em subconjuntos.

Outro algoritmo popular é o C4.5, uma evolução do ID3 também proposto por Quinlan (Quinlan, 1993), que utiliza a medida de ganho de informação normalizada para lidar com atributos com diferentes números de valores possíveis. O C4.5 também inclui uma técnica de poda de árvore para evitar o *overfitting* e melhorar a generalização do modelo.

Mais recentemente, o algoritmo de árvores de decisão foi aprimorado com a utilização de técnicas de aprendizado profundo, como as redes neurais convolucionais, para melhorar a precisão das previsões e a robustez do modelo (Zhang et al., 2018). Essa abordagem, conhecida como árvores de decisão baseadas em redes neurais (*Decision Trees with Neural Network – DTNN*), tem mostrado resultados promissores em diversas aplicações, como o diagnóstico médico e a previsão de preços de ações.

O resultado é uma estrutura em forma de árvore, onde os nós internos representam as perguntas feitas e as folhas representam as decisões ou previsões finais. As decisões em cada nó são baseadas na característica que proporciona maior ganho de informação, ou seja, que separa melhor os dados em diferentes classes ou grupos.

Aqui está um exemplo de código para o algoritmo de árvore de decisão em Python, retirado do livro "*Python Machine Learning*" de Raschka e Mirjalili (2019):

```

from sklearn.tree import DecisionTreeClassifier
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

# Carrega o conjunto de dados de exemplo Iris
iris = datasets.load_iris()
X = iris.data[:, [2, 3]]
y = iris.target

# Divide o conjunto de dados em um conjunto de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1,
stratify=y)

# Cria um modelo de árvore de decisão com profundidade máxima de 3
tree = DecisionTreeClassifier(criterion='gini', max_depth=3, random_state=1)

# Treina o modelo com o conjunto de treinamento
tree.fit(X_train, y_train)

# Faz previsões no conjunto de teste
y_pred = tree.predict(X_test)

# Avalia a acurácia do modelo
accuracy = accuracy_score(y_test, y_pred)
print("Acurácia:", accuracy)

```

Neste exemplo, foi usado o conjunto de dados “Iris”, que é um conjunto de dados de exemplo que faz parte da biblioteca Scikit-Learn, para criar um modelo de árvore de decisão usando essa mesma biblioteca. Em seguida, dividiu-se o conjunto de dados em um conjunto de treinamento e teste e, por fim, treinou-se o modelo com o conjunto de treinamento. Seguindo essa sequência, foi possível realizar previsões no conjunto de teste e avaliar a acurácia do modelo.

Müller & Guio (2016) afirmam que as árvores de decisão são amplamente utilizadas por sua capacidade de interpretação e de lidar com dados contínuos e categóricos, podendo solucionar problemas complexos. Esse modelo também é robusto para *outliers* e não requer muita preparação de dados. No entanto, tal algoritmo pode vir a realizar ajustes a mais do que o necessário, levando a um desempenho de generalização ruim, especialmente quando a árvore é profunda e possui muitos ramos. Para evitar esse erro de análise, técnicas como poda, ensacamento e florestas aleatórias podem ser usadas.


```
# Cria o classificador da floresta aleatória
clf = RandomForestClassifier(n_estimators=1000, max_depth=2,
                             random_state=0)

# Treina o classificador com os dados sintéticos
clf.fit(X, y)

# Usa o classificador para prever novos dados
print(clf.predict([[0, 0, 0, 0]]))
```

Este código usa a biblioteca scikit-learn para criar um conjunto de dados sintético e treinar um classificador de floresta aleatória com 1000 árvores de decisão e uma profundidade máxima de 2.

Müller & Guio (2016) reiteram que o algoritmo de florestas aleatórias é amplamente utilizado em diferentes áreas, como finanças, saúde e biologia, para tarefas de classificação, regressão e detecção de anomalias. O modelo de florestas aleatórias é amplamente utilizado em instituições bancárias, como avaliação de risco de crédito, detecção de fraude e segmentação de clientes. O motivo são as vantagens que o modelo oferece, como boa precisão, capacidade de lidar com grandes conjuntos de dados e capacidade de trabalhar com variáveis contínuas e categóricas. Outra grande vantagem é a velocidade do treinamento, pois cada árvore pode ser cultivada em paralelo.

No entanto, de acordo com Bishop (2006), trabalhar com um volume tão expressivo de dados em virtude do agrupamento de árvores de decisão, faz com que o método de florestas aleatórias apresente algumas desvantagens, como alto uso de memória, tempo de previsão mais longo em comparação com árvores de decisão únicas e dificuldade na interpretação do modelo. Além disso, assim como ocorre com o modelo de árvores de decisão, esse algoritmo também pode gerar análises equivocadas, dado o alto volume de ajustes, ou ainda, no caso de os subconjuntos de dados e recursos não forem selecionados aleatoriamente.

2.3.1.3. Redes neurais

As redes neurais são um tipo de algoritmo de aprendizado de máquina modelado de acordo com a estrutura e a função do sistema nervoso humano, que tem como objetivo simular a capacidade do cérebro em aprender, reconhecer padrões e realizar tarefas complexas. Eles consistem em um grande número de nós de processamento interconectados, conhecidos como neurônios artificiais, que são organizados em camadas.

Cada neurônio recebe entrada de outros neurônios, processa a entrada usando uma função de ativação e então passa a saída para outros neurônios na próxima camada. As conexões entre os neurônios são associadas a pesos, que podem ser ajustados durante o treinamento para minimizar o erro de predição.

Uma das primeiras redes neurais propostas foi a Perceptron, proposta por Rosenblatt (1958), que foi utilizada para problemas de classificação binária. Desde então, muitos avanços foram feitos na área, incluindo a proposta de redes neurais profundas, com muitas camadas ocultas.

No site/portal da “keras” (keras.io), foi possível encontrar um exemplo de código em Python para esse tipo de algoritmo, utilizando a sua própria biblioteca, conforme exposto a seguir:

```
from keras.models import Sequential
from keras.layers import Dense
import numpy as np

# Define os dados de treinamento
X = np.array([[0, 0], [0, 1], [1, 0], [1, 1]])
y = np.array([0, 1, 1, 0])

# Cria a rede neural
model = Sequential()
model.add(Dense(4, input_dim=2, activation='relu'))
model.add(Dense(1, activation='sigmoid'))

# Compila a rede neural
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])

# Treina a rede neural
model.fit(X, y, epochs=1000, verbose=0)

# Usa a rede neural para prever novos dados
print(model.predict(np.array([[0, 0], [0, 1], [1, 0], [1, 1]])))
```

Este código cria uma rede neural com uma camada oculta de 4 neurônios e uma camada de saída com um neurônio. A rede neural é treinada com os dados de entrada X e saídas esperadas y usando a função de perda de entropia cruzada binária e o otimizador Adam.

Existem muitos outros diferentes tipos de redes neurais, incluindo redes *feedforward*, redes neurais recorrentes e redes neurais convolucionais, cada um com seus próprios pontos fortes e fracos. As redes neurais são amplamente utilizadas para tarefas como classificação de imagens, reconhecimento de fala e processamento de linguagem natural.

Um dos principais pontos fortes das redes neurais é sua capacidade de aprender relacionamentos não lineares complexos entre entradas e saídas. Essas redes também são capazes de lidar com grandes quantidades de dados e podem extrair recursos dos dados automaticamente, reduzindo a necessidade de engenharia manual de recursos.

No entanto, as redes neurais também apresentam alguns desafios. Elas podem ser computacionalmente caras para treinar, especialmente para grandes conjuntos de dados e requerem muitos recursos computacionais. Elas também podem ser propensas a *overfitting*, onde o modelo se torna muito complexo e funciona bem nos dados de treinamento, mas mal em dados novos e não vistos. Técnicas de regularização como abandono, parada antecipada e decaimento de peso podem ser usadas para superar esse problema.

2.3.2. PESQUISAS E DADOS DO CENSO

A utilização de pesquisas e dados do censo é uma importante fonte de informações para a construção de matrizes OD (FLORIO et al., 2019). Essas pesquisas podem fornecer informações detalhadas sobre a origem, destino e características dos viajantes (MARTINS et al., 2019). A partir desses dados é possível identificar padrões de viagens, como o deslocamento de trabalhadores para o local de trabalho (SILVA et al., 2019).

Neste método, o processo de pesquisas tradicionais é necessário para o processo classificativo. Pesquisas e dados do censo são usados para coletar informações sobre os tipos de veículos que as pessoas possuem e usam. Esta informação pode então ser usada para estimar a proporção de viagens feitas por diferentes tipos de veículos e alocá-las à matriz OD correspondente.

Além disso, as pesquisas de origem e destino são utilizadas para aprimorar as matrizes OD, fornecendo informações sobre os motivos das viagens, como trabalho, lazer, entre outros (Ozturk et al., 2019). Isso permite uma melhor segmentação dos viajantes e aprimoramento dos modelos de transporte.

No entanto, é importante considerar que a utilização de pesquisas e dados do censo possui limitações, como o custo elevado e a dificuldade em obter informações precisas sobre os deslocamentos de pessoas (Hadayeghi et al., 2019). Muitas vezes as informações fornecidas pelos usuários são incompletas ou equivocadas, causando grande distorção nos dados. Por isso, é necessário utilizar outras fontes de dados complementares, como os dados de GPS e de redes sociais (LUO et al., 2019).

A construção de matrizes móveis, também conhecidas como matrizes origem-destino, são usadas para mostrar o fluxo de pessoas ou mercadorias entre diferentes locais, como o número de pessoas que viajam de uma cidade para outra ou a quantidade de mercadorias enviadas de um armazém para outro. Por esse motivo, a obtenção dos dados de pesquisas tradicionais é importante para obter as porcentagens dos tipos de veículos em cada par OD para, posteriormente, serem importadas na matriz geral OD de dados móveis para classificá-la.

Outra possibilidade é a criação de matrizes OD entre dois períodos de censo. Isso possibilita a compreensão dos padrões de migração, como quais cidades ou regiões estão ganhando ou perdendo população e as características das pessoas que estão se mudando.

2.3.3. CONTAGENS DE TRÁFEGO

As contagens de tráfego em pontos pré-determinados podem ser usadas para observar e contar o número de diferentes tipos de veículos que passam por um local específico. Segundo Abdel-Aty et al. (1996), essa técnica é eficaz para separar os fluxos de tráfego em diferentes tipos de veículos, como automóveis, ônibus e caminhões. Esta informação, além de ser usada para realizar a calibração das matrizes durante o processo de alocação, assim como o processo descrito no item anterior, em pesquisas e dados do censo, também pode ser usada para estimar a proporção de viagens feitas por diferentes tipos de veículos, fundamental para a construção de matrizes OD precisas (Xie et al., 2007).

No entanto, é importante destacar que a classificação dos veículos por meio de contagens de tráfego pode apresentar algumas limitações. De acordo com Wang et al. (2005), a precisão da classificação pode ser afetada por fatores como a velocidade dos veículos e o tamanho da janela de contagem. Além disso, as contagens de tráfego podem não ser representativas o suficiente para alguns tipos de veículos, como motocicletas e bicicletas (Luan et al., 2017).

Khisty e Lall (2002) abrangem uma série de questões relacionadas à construção de algoritmos para planejamento de transportes e análise de mobilidade. Segue abaixo um exemplo genérico de construção de algoritmo que pode ser utilizado para dividir uma matriz OD em diferentes tipos de veículos a partir de contagens volumétricas classificatórias.

Entrada:

- i. Matriz OD: com as informações de origem e destino das viagens;
- ii. Contagens volumétricas classificatórias: informações sobre o número de veículos de cada tipo que passaram por determinado trecho da via;
- iii. Fatores de distribuição de viagens: informações sobre a proporção de viagens realizadas por cada tipo de veículo a partir das contagens volumétricas classificatórias.

Saída:

- i. Matriz OD dividida por tipo de veículo.

Passos:

- i. Obter as contagens volumétricas classificatórias para cada trecho da via.
- ii. Calcular os fatores de distribuição de viagens para cada tipo de veículo a partir das contagens volumétricas classificatórias.
- iii. Aplicar os fatores de distribuição de viagens na matriz OD para estimar a distribuição de viagens entre os diferentes tipos de veículo.
- iv. Verificar se os resultados obtidos estão dentro dos padrões esperados e realizar ajustes, se necessário.
- v. Dividir a matriz OD por tipo de veículo, utilizando as proporções estimadas a partir das contagens volumétricas classificatórias.

É importante ressaltar que a aplicação desse algoritmo deve ser realizada com cuidado, considerando as particularidades da área analisada e as fontes de informação disponíveis. Além disso, é necessário que as contagens de tráfego sejam realizadas em diferentes horários e dias da semana, de forma a capturar a variação temporal da demanda por transporte.

2.3.4. DETECÇÃO DE *BLUETOOTH*

A detecção de *Bluetooth* envolve o uso de sensores colocados em locais estratégicos para detectar os IDs *Bluetooth* exclusivos dos veículos que passam, provindos de dispositivos móveis, como smartphones e sistemas de áudio automotivos, para rastrear seus movimentos e inferir padrões de viagem. Esta informação pode ser usada para estimar a proporção de viagens feitas por diferentes tipos de veículos e alocá-las à matriz OD correspondente. De acordo com

a literatura, a detecção por *Bluetooth* é uma técnica que tem se mostrado eficiente na classificação de veículos e na determinação de trajetórias de deslocamento. (LI et al., 2015; JIAO et al., 2015; ZHENG et al., 2015)

A adoção desse método exige a implantação de sensores em vários locais ao longo da rede rodoviária, como semáforos e sinais de trânsito, para detectar os dispositivos habilitados (e ativos) dos usuários. Quando um dispositivo habilitado para *Bluetooth* chega ao alcance do sensor, o sensor detecta o endereço MAC (*Media Access Control*) exclusivo do dispositivo e registra sua localização e registro de data e hora. Esse é um endereço físico e único, que é associado a interfaces de comunicação utilizadas em dispositivos de rede. Ao coletar e analisar esses dados, os planejadores e engenheiros de transporte podem inferir a origem e o destino do dispositivo, bem como o caminho percorrido. Além disso, essa técnica tem se mostrado útil na detecção de padrões de tráfego em tempo real, auxiliando na gestão e no controle do tráfego urbano (ZHENG et al., 2015).

Uma das vantagens da detecção *Bluetooth* é sua capacidade de coletar grandes quantidades de dados de forma rápida e econômica. É menos trabalhoso do que os métodos manuais tradicionais de coleta de dados e é menos intrusivo do que algumas outras tecnologias de coleta de dados móveis, como rastreamento por GPS. A detecção *Bluetooth* também pode fornecer uma visão mais abrangente dos padrões de viagem, incluindo meios de transporte não motorizados, como pedestres e bicicletas, que podem ser difíceis de rastrear usando outros métodos.

Segundo Lopes et al. (2019), a detecção de dispositivos *Bluetooth* em rodovias permite a geração de matrizes OD precisas e com baixo custo, sendo uma alternativa interessante para regiões com pouca infraestrutura de monitoramento de tráfego.

No entanto, é importante destacar que a detecção por *Bluetooth* apresenta limitações, como a necessidade de instalação de dispositivos de detecção em pontos estratégicos, a dependência da presença de dispositivos com *Bluetooth* ativo nos veículos e a limitação na identificação de veículos sem dispositivos *Bluetooth* (JIAO et al., 2015; ZHENG et al., 2015). É necessário que o dispositivo esteja ativado para que seja reconhecido pelos sensores da via.

Há também limitações em áreas com baixa penetração de dispositivos *Bluetooth* ou de grande interferência. Além disso, os dados coletados da detecção de *Bluetooth* podem não ser representativos de toda a população, pois podem capturar apenas dados de um determinado segmento da população.

2.4. SIMULAÇÃO DE TRÁFEGO – PTV VISUM

2.4.1. MODELAGEM DA REDE

A modelagem de redes de transporte é uma tarefa complexa que envolve a análise e o planejamento de sistemas viários, permitindo a representação de diferentes modos de transporte, como automóveis, ônibus, metrô, entre outros. O software VISUM é uma ferramenta amplamente utilizada para este fim (MUNIZ et al., 2016), com sua modelagem envolvendo a definição de vias, nós, zonas e outras características relevantes da rede viária (MIRANDA e DIAS, 2018).

O processo de modelagem depende da definição das características da rede, tais como geometria das vias, fluxo de tráfego, capacidade das vias, locais de interseção, pontos de embarque e desembarque, entre outros. Para isso, é possível utilizar dados de campo, como levantamentos de tráfego e de contagem de passageiros, ou dados secundários, como informações de mapas digitais e imagens de satélite. As vias são representadas por segmentos lineares que conectam dois nós da rede viária (MIRANDA e DIAS, 2018), com cada via devendo ter uma capacidade de tráfego associada, que representa o número máximo de veículos que podem passar pela via em uma hora (MUNIZ et al., 2016).

Os nós da rede viária no VISUM representam os pontos onde as vias se conectam e onde as decisões de direção podem ser tomadas pelos usuários da rede (MIRANDA e DIAS, 2018). A definição correta dos nós é importante para que a modelagem da rede de transporte no VISUM seja eficiente e precisa (MUNIZ et al., 2016).

Diversos estudos têm utilizado o VISUM para modelagem de redes de transporte, como a modelagem de redes de transporte urbano em cidades brasileiras (SILVA, 2016), a análise de impacto de obras viárias em congestionamentos (ZHU, 2019) e a avaliação da capacidade de fluxo de rotatórias em estradas de alta velocidade (ZHANG, 2018).

2.4.2. ALOCAÇÃO DE MATRIZES

O processo de alocação de matrizes no software VISUM envolve a distribuição das viagens entre as origens e destinos conhecidos, com base em dados de contagem de tráfego e outros tipos de informações disponíveis. O algoritmo de alocação utilizado pelo VISUM é o método iterativo de ponderação da probabilidade de Hellinger (*Hellinger Distance-Based*

Iterative Proportional Fitting, ou HIB-IPF), que é um método de estimação de parâmetros baseado em máxima verossimilhança (WANG et al., 2011), análise bayesiana e ajuste proporcional iterativo (IPF) (Li et al., 2019).

Segundo Li et al. (2019), o HIB-IPF incorpora uma abordagem hierárquica que permite a modelagem da incerteza em várias escalas e o ajuste dos dados em níveis diferentes, o que pode melhorar a precisão da alocação da matriz OD. O algoritmo também inclui um procedimento de reamostragem de dados para estimar a distribuição posterior dos parâmetros e a incerteza dos resultados.

O método HIB-IPF tem sido aplicado em diversas áreas, como no estudo de demanda de tráfego em grandes cidades (Chen et al., 2017) e na alocação de matrizes OD em áreas rurais (Shan et al., 2019). Em ambos os estudos, os resultados mostraram que o HIB-IPF pode melhorar significativamente a precisão da alocação da matriz OD em comparação com outros métodos de alocação de matrizes.

Além disso, o HIB-IPF também tem sido comparado com outros algoritmos de alocação de matrizes, como o IPF tradicional e o *Hierarchical Bayesian Model* (HBM). Os resultados mostraram que o HIB-IPF pode alcançar melhor desempenho em termos de precisão e estabilidade da alocação da matriz OD (Wang et al., 2019).

Para obter uma matriz de demanda consistente, é necessário utilizar técnicas de validação e ajuste dos dados de entrada, bem como de verificação da qualidade da matriz de saída. O VISUM inclui diversas ferramentas de diagnóstico para avaliar a qualidade da matriz alocada, incluindo a comparação de matrizes observadas e alocadas, a análise de resíduos e a validação cruzada (ZHANG et al., 2018).

Além disso, é importante considerar que o processo de alocação de matrizes pode ser afetado por diversos fatores, como a precisão das contagens de tráfego, definição das zonas de origem e destino, modelagem dos comportamentos dos usuários, entre outros. Por isso, é fundamental utilizar técnicas de validação e ajuste adequadas para obter resultados confiáveis e consistentes (KURAUCHI et al., 2015).

2.4.3. ESTIMAÇÃO DE MATRIZES OD

A estimação de matrizes OD no VISUM é um processo importante para garantir a precisão dos resultados de um modelo de transporte. Segundo Monzon et al. (2005), a estimação

é um processo iterativo que ajusta os parâmetros do modelo para tornar os resultados mais próximos da realidade observada.

Para realizar a estimação no VISUM, é necessário ter uma matriz OD inicial que será ajustada durante o processo. De acordo com Bliemer e Raadsen (2010), é importante que essa matriz inicial seja representativa do tráfego real, para garantir que os ajustes serão feitos a partir de uma base sólida.

O processo de estimação no VISUM pode ser feito utilizando diversas técnicas, como a comparação de dados de contagem de tráfego com as matrizes geradas pelo modelo. Conforme Santos et al. (2018), é possível também utilizar informações de tempo de viagem e velocidade média para calibrar as matrizes OD.

Para realizar a estimação de forma eficiente, é importante utilizar ferramentas e técnicas adequadas. De acordo com Ben-Akiva et al. (1996), o processo de estimação pode ser automatizado utilizando algoritmos como o SIMPLEX, que ajusta os parâmetros do modelo de forma mais rápida e precisa.

Além disso, é importante que o processo de estimação seja realizado de forma iterativa, realizando ajustes e testes sucessivos até que os resultados sejam satisfatórios. Segundo Ortúzar e Willumsen (2011), a estimação de matrizes OD é um processo complexo que requer experiência e conhecimento técnico para ser realizado de forma eficiente.

De acordo com o manual do software VISUM, da empresa PTV, a ferramenta TFlowFuzzy, presente no grupo de módulos de matrizes da sequência de procedimentos, é um módulo que utiliza a lógica *fuzzy* para a atribuição de tráfego em modelos de transporte. A lógica *fuzzy* é uma extensão da lógica booleana convencional que lida com conceitos vagos e incertos, permitindo a modelagem de incertezas e imprecisões. A função principal do TFlowFuzzy é a atribuição de volumes de tráfego às redes de transporte, levando em consideração a interação entre diferentes modos de transporte, como carros, ônibus, metrô, bicicletas, etc., e considera fatores como demanda, capacidade da infraestrutura e características do meio de transporte.

Ao utilizar a lógica *fuzzy*, o TFlowFuzzy é capaz de lidar com informações imprecisas e incertas, como a falta de dados detalhados sobre padrões de tráfego. Utilizando regras e funções de pertinência, aliado às informações disponíveis como contagens de tráfego, pesquisas de viagem ou dados de sensores, ou seja, os inputs fornecidos pelo usuário, o TFlowFuzzy é capaz de estimar matrizes de tráfego que sejam coerentes com os dados existentes e as características da rede de transporte.

2.4.4. COMPARAÇÃO DE CONSISTÊNCIAS

A comparação de consistências de matrizes alocadas no *software* VISUM é um processo importante para verificar a qualidade das previsões de tráfego. Diversos estudos já utilizaram essa técnica para avaliar a precisão dos modelos de transporte e alocar as matrizes OD no VISUM.

Gharizadeh e Haghshenas (2018) utilizaram a comparação de consistências para avaliar a qualidade de diferentes modelos de demanda de transporte em uma cidade iraniana. Eles utilizaram o método de comparação de consistências proposto por Gentile e Noekel (2008), que é implementado no *software* VISUM. Os resultados mostraram que o modelo baseado em redes neurais artificiais apresentou a melhor consistência em relação aos dados observados.

Singh e Jain (2017) utilizaram a comparação de consistências para avaliar a precisão de um modelo de demanda de transporte urbano na cidade de Delhi, na Índia. Os autores utilizaram uma técnica de calibração iterativa para ajustar as matrizes OD no VISUM e compararam as consistências do modelo calibrado com os dados observados. Os resultados indicaram que o modelo calibrado apresentou uma melhor consistência em relação ao modelo não calibrado.

Além disso, Li et al. (2015) utilizaram a comparação de consistências para avaliar a precisão dos modelos de demanda de transporte em uma cidade chinesa. Os autores utilizaram uma técnica de calibração baseada em diferenças mínimas quadradas para ajustar as matrizes OD no VISUM e compararam as consistências do modelo calibrado com os dados observados. Os resultados mostraram que o modelo calibrado apresentou uma melhor consistência em relação ao modelo não calibrado.

A comparação de consistências é uma técnica útil para avaliar a precisão dos modelos de transporte e ajustar as matrizes OD no *software* VISUM. Através dessa técnica, é possível identificar e corrigir inconsistências nos modelos, melhorando a qualidade das previsões de tráfego.

2.5. REVISÃO SISTEMÁTICA

2.5.1. DEFINIÇÃO

As revisões sistemáticas são baseadas em métodos de pesquisa que visam sintetizar e analisar de forma sistemática todas as evidências relevantes disponíveis sobre um determinado tema. Elas são consideradas um dos níveis mais altos de evidência científica, pois fornecem uma abordagem rigorosa e transparente para avaliar a literatura existente, evitando vieses e permitindo que os resultados sejam mais confiáveis e generalizáveis.

De acordo com Moher et al. (2009), a revisão sistemática é um método rigoroso e transparente que busca identificar, avaliar criticamente e sintetizar todas as evidências relevantes sobre um tópico específico, seguindo um protocolo pré-estabelecido. São seguidas etapas bem definidas, como formulação da pergunta de pesquisa, busca abrangente de literatura, seleção criteriosa de estudos relevantes, avaliação crítica da qualidade dos estudos, extração e síntese de dados, e análise dos resultados em conjunto.

Moher et al. (2009) também afirmam que o objetivo da revisão sistemática é obter uma resposta imparcial e confiável para a pergunta de pesquisa, minimizando vieses e maximizando a validade das conclusões.

Tanto Cordeiro et al. (2007) quanto Medeiros et al. (2015) concordam que uma das principais características desse processo é a definição de métodos explícitos e sistematizados para o sistema de busca, permitindo que outras pessoas possam reproduzir o mesmo modelo, já que seus protocolos são replicáveis. Esse modelo de estrutura de revisão permite a diminuição da subjetividade na seleção de publicações.

2.5.2. RESULTADOS ENCONTRADOS

Para o desenvolvimento da base bibliográfica deste trabalho foram utilizados diversos meios, partindo de publicações em congressos, websites de bancos como o Banco Interamericano de Desenvolvimento (BID), estudos realizados por empresas privadas para contratos com instituições governamentais, ou buscas em plataformas como “PubMed”, “Scopus”, “Web of Science” ou “Google Scholar”.

Foram diversas as palavras-chaves e termos utilizados na realização das buscas, sempre focando na pergunta norteadora apresentada no capítulo introdutório, alternando entre português e inglês. A seguir estão os termos utilizados nas buscas em ambos os idiomas:

PORTUGUÊS: Comparação entre matrizes OD.
 Matrizes de dados móveis e matrizes de pesquisas domiciliares.
 Análise de matrizes para simulações de tráfego.
 Classificação de matrizes OD.
 Dados móveis em estudos de tráfego.
 Processo de pesquisa domiciliar em transporte.
 Limitações de pesquisas domiciliares e dados móveis

INGLÊS: Comparison between OD matrices.
 Mobile data matrices and household survey matrices.
 Matrix analysis for traffic simulations.
 Classification of OD matrices.
 Mobile data in traffic studies.
 Household survey process in transport.
 Limitations of household survey and mobile data

Foram encontrados números expressivos de resultados, como a PubMed entregando 933 trabalhos, a Scopus mais de 1.500 e a Google Scholar em torno de 995.000 resultados entre diversos portais exibidos.

No entanto, em virtude dessa análise comparativa ser um assunto muito específico e em atual desenvolvimento e aperfeiçoamento, o número de estudos/artigos que tem uma relação direta é relativamente baixo. A comparação entre os métodos ainda está sendo explorada em diferentes instituições, entre elas o próprio BIRD, no intuito de otimizar o referido processo. Portanto, o processo de busca desse assunto específico (análise comparativa de matrizes) revelou um universo de dados ainda escasso.

Através dos portais mencionados, utilizando todos os termos expostos, foram encontrados 8 resultados que atenderam de forma plena, ou parcial, a busca realizada. Esses títulos podem ser verificados no Quadro 2.

ARQUIVOS ANALISADOS	ANO	AUTORES	PORTAL
Feel Old Yet? Updating Mode of Transportation Distributions from Travel Surveys using Data Fusion with Mobile Phone Data	2023	Eduardo Graells-Garrido; Daniela Opitz; Francisco Rowe; Jacqueline Arriagada	arXiv

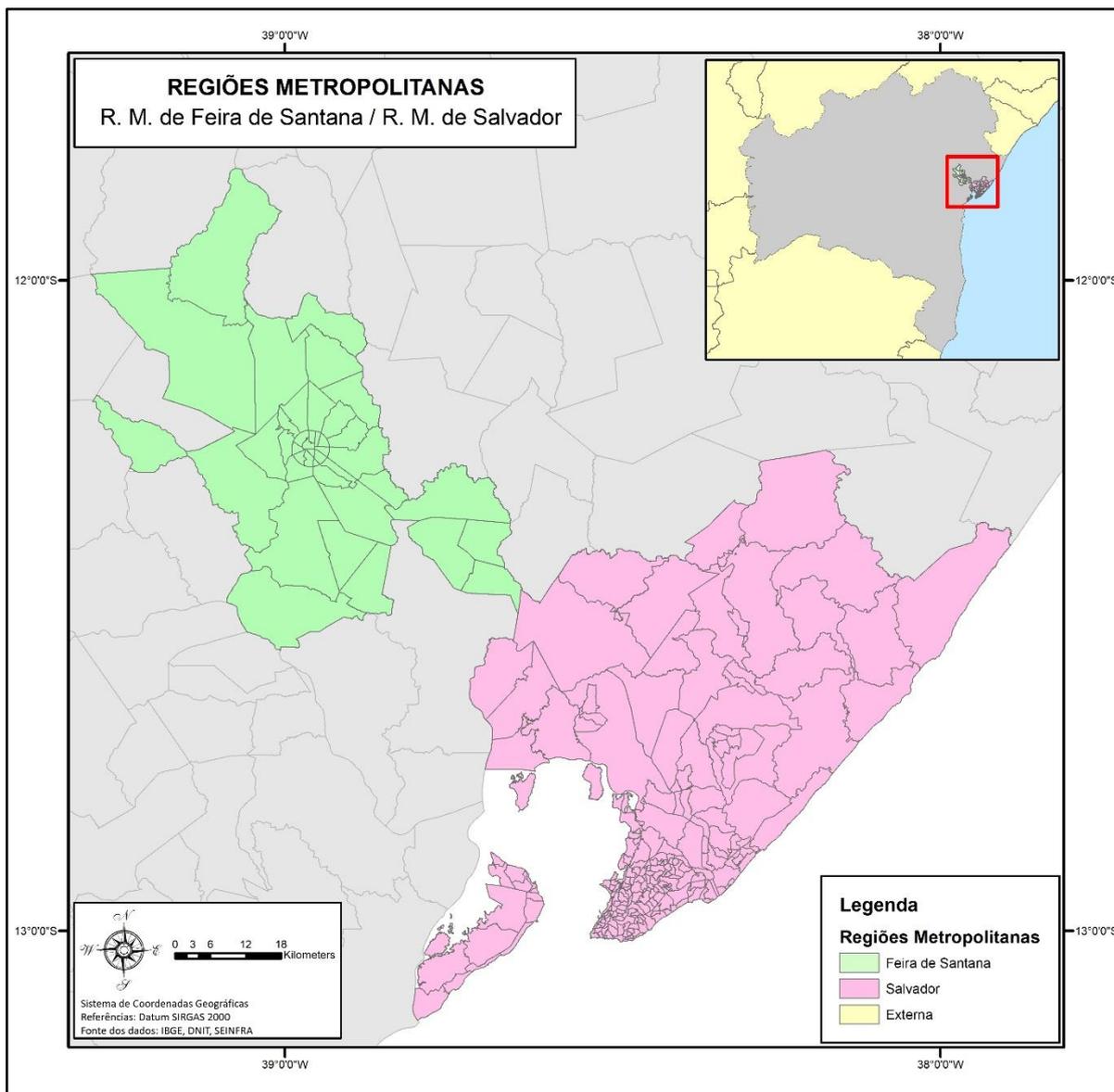
ARQUIVOS ANALISADOS	ANO	AUTORES	PORTAL
Exploring strengths and weaknesses of mobility inference from mobile phone data vs. travel surveys	2020	Noelia Cáceres; L. M. Romero; Francisco G. Benítez	Taylor & Francis Online
Getting the best of both worlds: a framework for combining disaggregate travel survey data and aggregate mobile phone data for trip generation modelling	2020	Andrew Bwambale; Charisma F. Choudhury; Stephane Hess; Md. Shahadat Iqbal	SpringerLink
Fusing Mobile Phone and Travel Survey Data to Model Urban Activity Dynamics	2020	Chao Yang; Yuliang Zhang; Xianyuan Zhan; Satish V. Ukkusuri; Yifan Chen	Wiley Hindawi
Land use inference from mobility mobile phone data and household travel surveys	2019	Noelia Cáceres; Francisco G. Benítez; Luis M. Romero	ScienceDirect
Fusing mobile phone data with other data sources to generate input OD matrices for transport models	2018	L. Montero; X.Ros-Roca; R. Herranz; J.Barceló	ScienceDirect
Review of GPS Travel Survey and GPS Data-Processing Methods	2014	Li Shen; Peter Stopher	Taylor & Francis Online
A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys	2012	Philippe Nitschea; Peter Widhalma; Simon Breussa; Peter Maurera	ScienceDirect

Fonte: Elaborado pelo autor, 2023

Tais artigos foram analisados e, em determinadas questões, puderam contribuir com o presente trabalho. Tanto na elucidação de imprecisões quanto na solução de indefinição, bem como, na adição de novos pontos de vista, os trabalhos citados puderam auxiliar no desenvolvimento deste estudo.

3. ESTUDO DE CASO

A área de estudo deste trabalho compreende as regiões metropolitanas do estado da Bahia, Salvador e Feira de Santana que, juntas, totalizam mais de 31,40% da população estadual, conforme Figura 3.

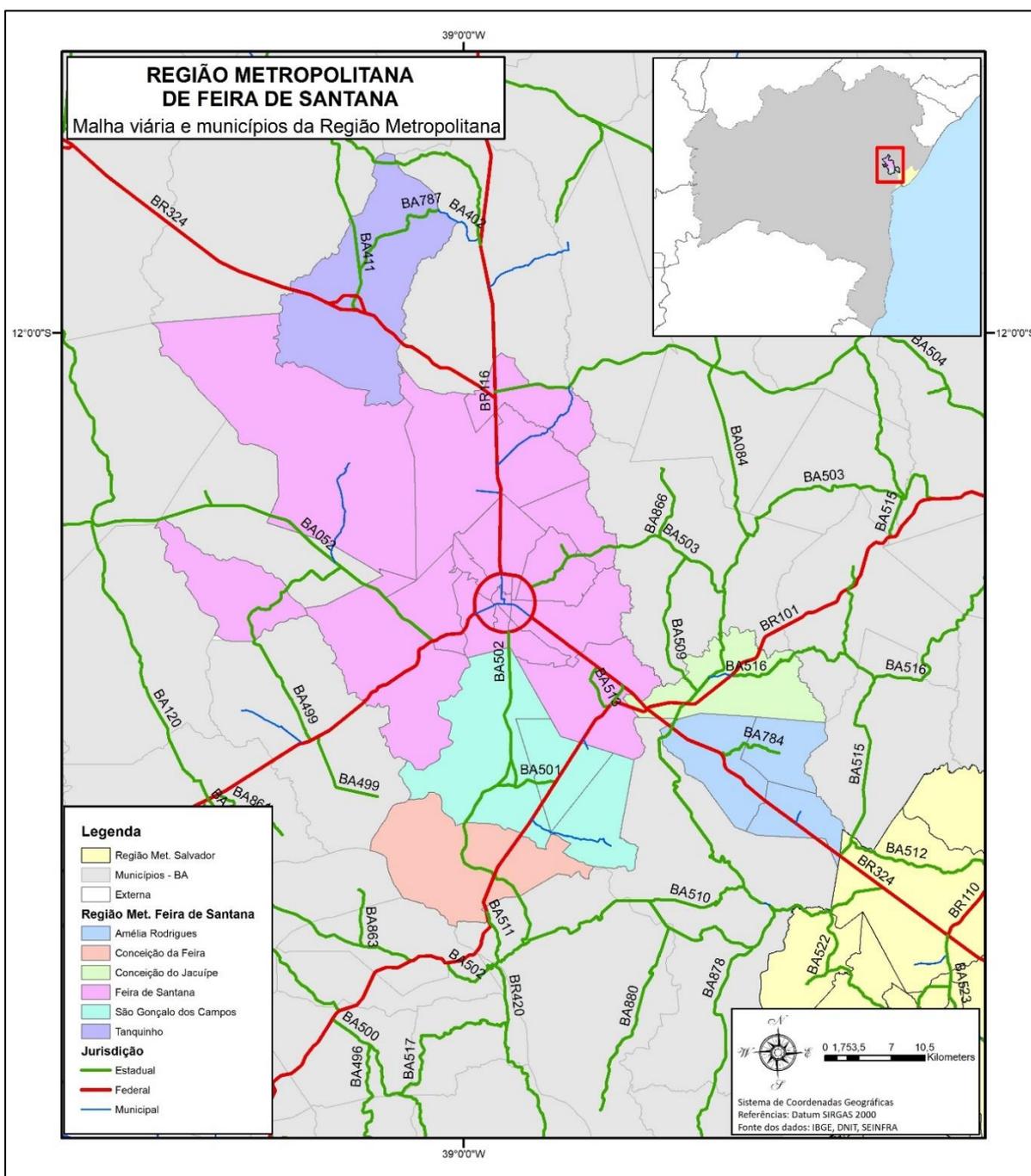


Fonte: APPE, 2023 – Adaptado pelo autor

De acordo com o IBGE (2019), as duas regiões metropolitanas apresentam uma população de 4.670.385 habitantes, correspondendo a uma densidade demográfica média de 10.583,94 hab/ km² (IBGE 2010).

3.1. FEIRA DE SANTANA

A Região Metropolitana de Feira de Santana - RMFS, que foi instituída por meio da Lei Complementar Estadual nº 35/2011 (Lei Orgânica do Estado), é composta por seis municípios: Amélia Rodrigues, Conceição da Feira, Conceição do Jacuípe, Feira de Santana, São Gonçalo dos Campos e Tanquinho, como apresentado na Figura 4.



Fonte: elaborado pelo autor

Há também uma área de expansão da região metropolitana, composta pelos municípios de Anguera, Antônio Cardoso, Candeal, Coração de Maria, Ipecaetá, Irará, Santa Bárbara, Santanópolis, Serra Preta e Riachão do Jacuípe, totalizando dezesseis municípios. No entanto, esses dez municípios da área de expansão não serão considerados neste estudo.

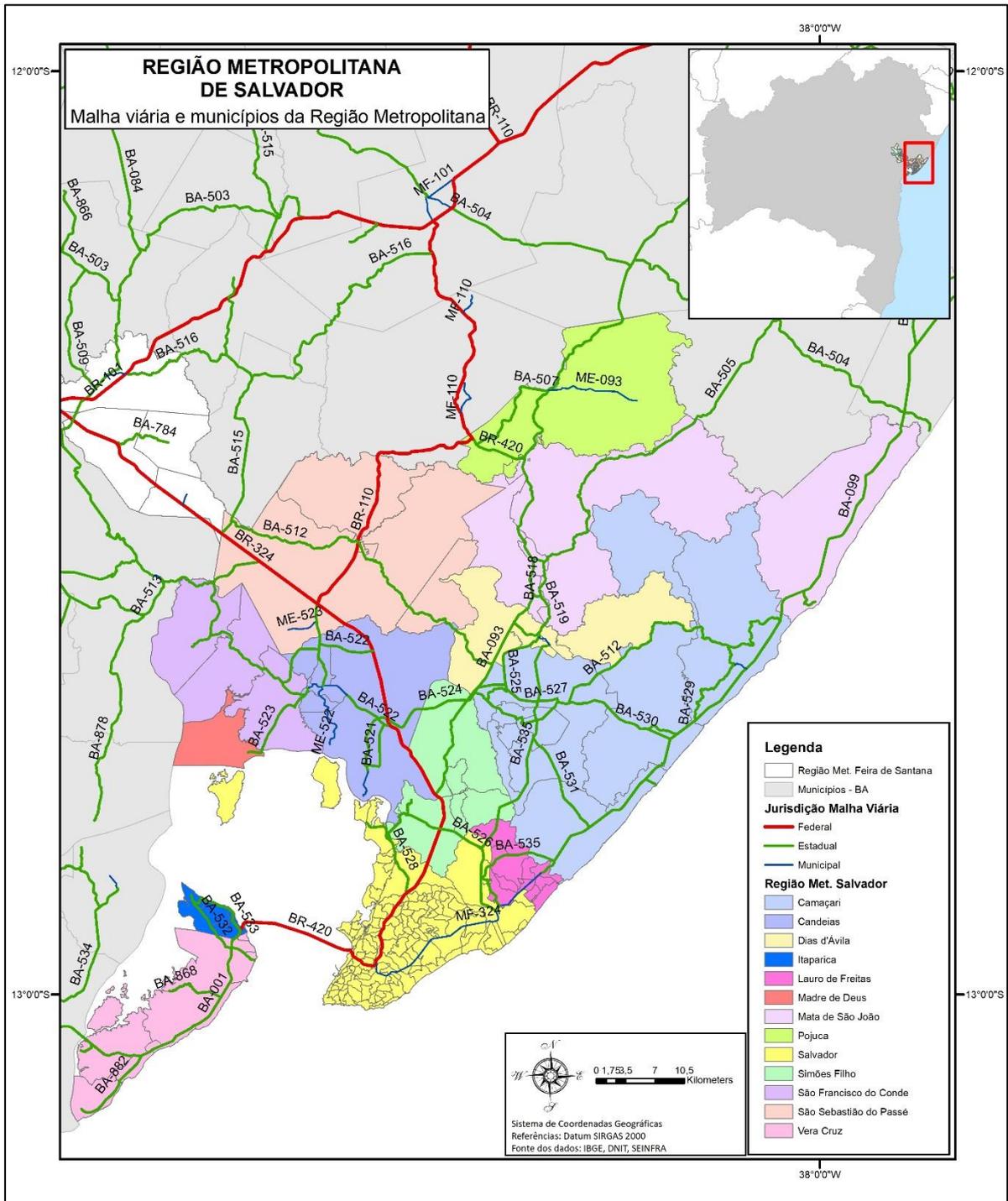
Feira de Santana é um dos sete municípios que podem ser considerados importantes centros econômicos da Bahia (Salvador, Feira de Santana, Vitória da Conquista, Camaçari, Ilhéus, Itabuna e Juazeiro), localizado em uma região estratégica, contando com um anel rodoviário que conecta a BR-116 com a BR-324, sendo esta, a principal ligação com a Região Metropolitana de Salvador. Dessa forma, o atual processo de validação de caminho mínimo precisa simular fluxos partindo de uma gama de pontos externos a este anel para compreender o real comportamento dos motoristas na região.

Devido ao elevado número de rodovias que passa pela cidade, o fluxo de veículos em Feira de Santana se intensifica em certos momentos do dia. De forma a reduzir esse fluxo de veículos, a cidade conta com um sistema de trânsito e transporte público, incluindo ônibus urbanos e uma estação de metrô. Além disso, a cidade também possui uma série de viadutos, pontes e passarelas que auxiliam no aumento da capacidade e na manutenção do nível de serviço das vias em áreas congestionadas.

3.2. SALVADOR

A Região Metropolitana de Salvador - RMS, que foi instituída por meio da Lei Complementar Federal nº 14/1973 (Lei Orgânica do Ministério Público da União), compreende atualmente treze municípios: Itaparica; Camaçari; Madre de Deus; Pojuca; Mata de São João; Vera Cruz; Salvador; Candeias; Lauro de Freitas; Simões Filho; São Francisco do Conde; São Sebastião do Passé; e Dias d'Ávila. Esta Região está representada na Figura 5.

Salvador possui um sistema de vias expressas como a Avenida Paralela, que é uma das principais vias expressas da cidade e conecta várias partes da cidade, incluindo a região central e o aeroporto. Há também um sistema de transporte público bem desenvolvido, que inclui ônibus urbanos, metrô e trem, sendo o metrô um adendo relativamente novo, já que foi inaugurado em 2014 e, atualmente, conta com duas linhas que se estendem por cerca de 33 km de extensão.



Fonte: Elaborado pelo autor, 2023

Assim como Feira de Santana, a cidade também conta com uma série de viadutos, mas em uma densidade muito maior, motivo pelo qual o Consórcio tem levado um tempo maior em seu processo de validação. Para uma simulação que possa atingir uma consistência relevante (maior que 90%), todos os detalhes precisam estar muito bem representados no modelo, para que o comportamento do fluxo alocado tenha a representatividade necessária.

3.3. ZONEAMENTO

3.3.1. ASPECTOS GERAIS DO ZONEAMENTO EM ESTUDOS DE MOBILIDADE URBANA

O zoneamento é considerado uma das etapas de maior importância em todo o processo de modelagem. As zonas de tráfego são representadas em termos computacionais, como se todos seus atributos e propriedades fossem concentrados em um simples ponto chamado centroide. De acordo com as premissas de agregação adotadas, define-se a dimensão das zonas de tráfego, e, conseqüentemente, determina-se o nível de agregação das variáveis que serão utilizadas para a calibração e validação dos modelos de demanda e oferta. Desta forma, a definição das zonas de tráfego será realizada de forma que o erro de agregação causado pela suposição de que todas as suas características sejam concentradas em seu centroide está dentro de padrões aceitáveis.

As zonas de tráfego serão as mais homogêneas possível com relação ao uso do solo e/ou composição de sua população. Zonas de tráfego com grandes diferenças neste aspecto, como zonas residenciais com acentuadas diferenças no nível de renda, a princípio, não serão agregadas, mesmo se forem muito pequenas.

As dimensões da zona de tráfego são um fator relevante que também afetam a modelagem da escolha modal, pois, ao adotar um zoneamento excessivamente agregado, as distâncias de caminhada do centroide às paradas de ônibus tornam-se cada vez mais discordantes com os reais, comprometendo a eficácia da modelagem em redes de transporte público, principalmente quando há vários modos envolvidos.

Os limites das zonas de tráfego serão compatíveis com os limites dos setores censitários utilizados no Censo do IBGE, possibilitando que as informações censitárias possam ser usadas no estudo. As principais barreiras físicas e geográficas aos deslocamentos e ao desenvolvimento do solo, como ferrovias, rios, rodovias, entre outras, também serão considerados limites.

3.3.2. CARACTERIZAÇÃO DAS ÁREAS DE ESTUDO POR MEIO DE DADOS SECUNDÁRIOS

Os levantamentos preliminares para caracterização das áreas de estudo visam fornecer subsídios para o presente trabalho, contemplando as características demográficas, socioeconômicas, ambientais, uso e ocupação do solo, circulação e sistema viário.

Neste estudo, serão utilizadas informações socioeconômicas proveniente dos setores censitários (IBGE, 2010) pois, quando este trabalho começou a ser desenvolvido, eram as últimas referências de dados oficiais especializados com informações socioeconômicas da população, como total de pessoas, idade, gênero, renda, domicílio por condição, entre outros.

Informações mais recentes de órgãos oficiais não contêm dados segregados em níveis de agregação similares aos setores censitários, o que dificulta análises espaciais com dados recentes referentes a bairros ou mesmo as zonas de origem e destino que serão estabelecidas neste estudo.

A atividade de levantamento e coleta de bases cartográficas digitais georreferenciadas (com limites municipais, de bairros e de setores censitários) é fundamental para o estabelecimento da abrangência e dos limites da área de estudo, sua segmentação interna e articulação com a área externa e a consolidação quanto ao zoneamento. A base de dados do IBGE possibilitou, para cada unidade do setor censitário, compreender seus limites, quantidade de população, domicílios e renda média, entre outras variáveis do território.

A utilização dos setores censitários como unidade de análise da população possibilitou o agrupamento dos mesmos para formação das zonas de origem e destino (OD), as quais, por sua vez, serão a menor unidade territorial de análise das viagens realizadas pela população da área de estudo. O método utilizado para concepção das zonas internas e externas será abordado nos próximos itens deste relatório.

3.3.3. DEFINIÇÃO DAS ZONAS INTERNAS

Para concepção do zoneamento, após a análise das informações secundárias obtidas, é possível compreender as características socioeconômicas e de uso do solo nas áreas de estudo. Desta forma, o zoneamento interno foi estruturado observando os seguintes aspectos:

- variáveis socioeconômicas da população – o modelo de geração de viagens representa os padrões de viagens urbanas por meio da correlação entre a produção de viagens e as

características socioeconômicas dos grupos de indivíduos em cada zona. Desta forma, as zonas serão agregadas de maneira a representarem populações de características socioeconômicas similares;

- setores censitários – através de dados obtidos do IBGE, essas áreas foram utilizadas como base para o início da definição já que, de acordo com o próprio instituto, essa é a menor porção de área para o planejamento, coleta e disseminação de resultados de censos e pesquisas;

- uso do solo – observou-se o uso do solo predominante corresponde a edificações residenciais, comerciais, industriais e mistos. Além disso, será analisado o padrão das construções presentes;

- sistema viário – o sistema viário é utilizado como divisor natural entre regiões. Vias arteriais relevantes nem sempre constituem limites de zonas, visto que geralmente o uso do solo dos dois lados são similares, situando-os, portanto, dentro da mesma zona;

- sistema de transporte público – analisa-se com o intuito de representar a demanda de passageiros no modelo de simulação com informações mais próximas da realidade em termos de passageiros*km e trechos críticos;

- sistema de transporte individual – representa o efeito do transporte individual no sistema viário, com a premissa de que zonas muito grandes podem apresentar um percentual elevado de viagens internas; e

- relevo e hidrografia – geralmente, a conformação do relevo e a presença dos rios, por se tratar de barreiras naturais, influenciam nos limites das zonas.

A consideração de todos os aspectos apresentados resultou em uma divisão com 277 zonas, sendo 41 zonas para a Região Metropolitana de Feira de Santana e 230 zonas para a Região Metropolitana de Salvador, além da divisão externa à área de estudo em 6 zonas. O detalhamento dessa informação pode ser verificado na Tabela 1.

A Tabela 2 traz o número de zonas definidas em cada um dos 19 municípios de ambas as regiões metropolitanas, além das zonas externas, ou seja, fora dos limites dessas regiões.

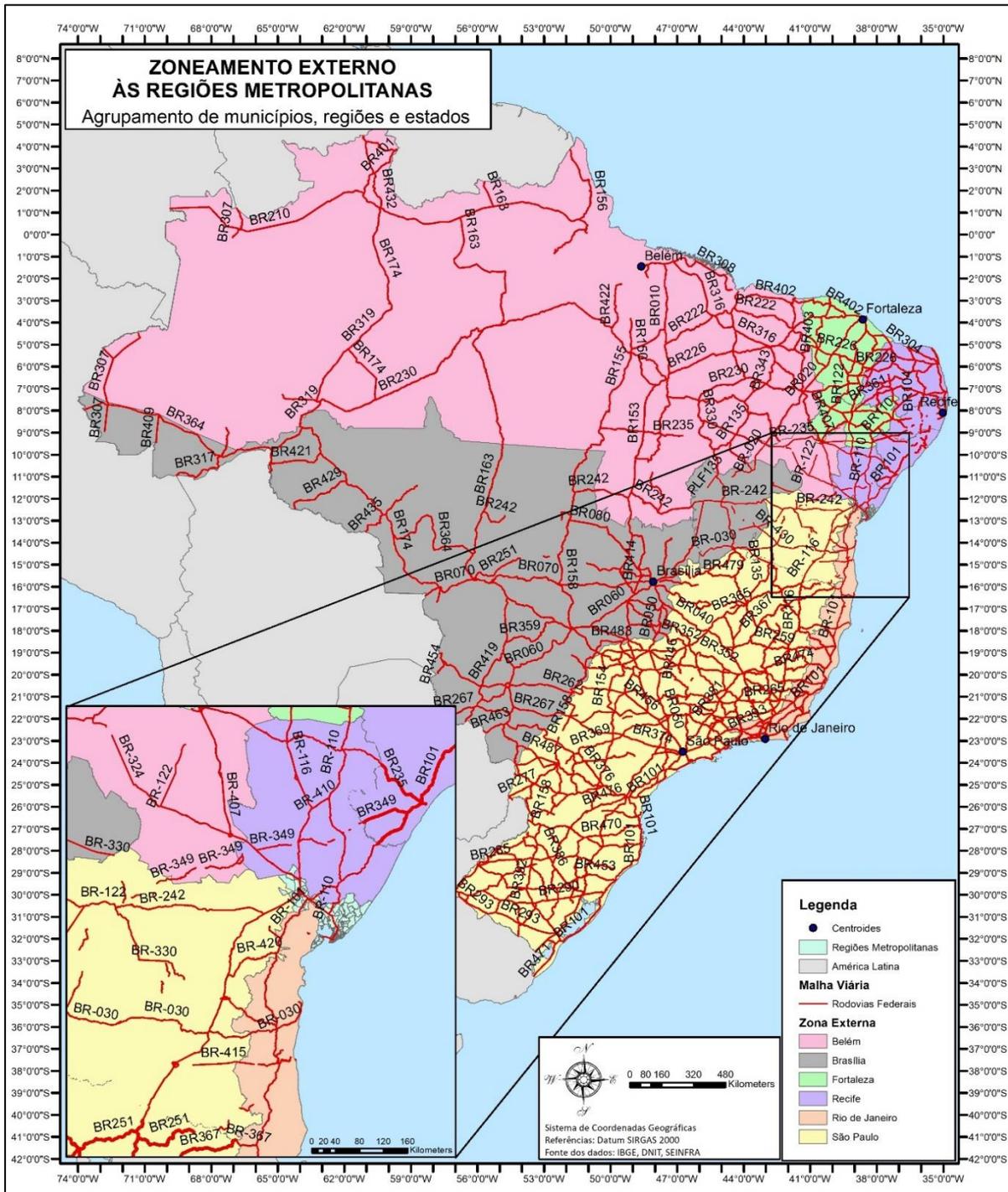
CONGLOMERADOS	MUNICÍPIOS	TOTAL DE BAIROS	SETORES CENSITÁRIOS	ÁREA (KM ²)	ZONAS
Reg. Metrop. Feira de Santana	6	149	1.290	2.242,87	41
Reg. Metrop. Salvador	13	358	6.644	9.455,00	230
Zonas externas	-	-	-	-	6
TOTAL	19	507	7.934	11.697,87	277

Fonte: APPE (2022) – Adaptado pelo autor

REGIÃO METROPOLITANA	MUNICÍPIOS	ZONAS
Feira de Santana	Amélia Rodrigues	3
	Conceição da Feira	1
	Conceição do Jacuípe	1
	Feira de Santana	32
	São Gonçalo dos Campos	3
	Tanquinho	1
	Subtotal de Feira de Santana:	41
Salvador	Camaçari	13
	Candeias	5
	Dias d'Ávila	4
	Itaparica	5
	Lauro de Freitas	15
	Madre de Deus	1
	Mata de São João	5
	Pojuca	3
	Salvador	151
	São Francisco do Conde	3
	São Sebastião do Passé	5
	Simões Filho	9
	Vera Cruz	11
Subtotal de Feira de Santana:	230	
Zonas externas	Belém	1
	Brasília	1
	Fortaleza	1
	Recife	1
	Rio de Janeiro	1
	São Paulo	1
Subtotal de Zonas Externas:	6	
TOTAL		277

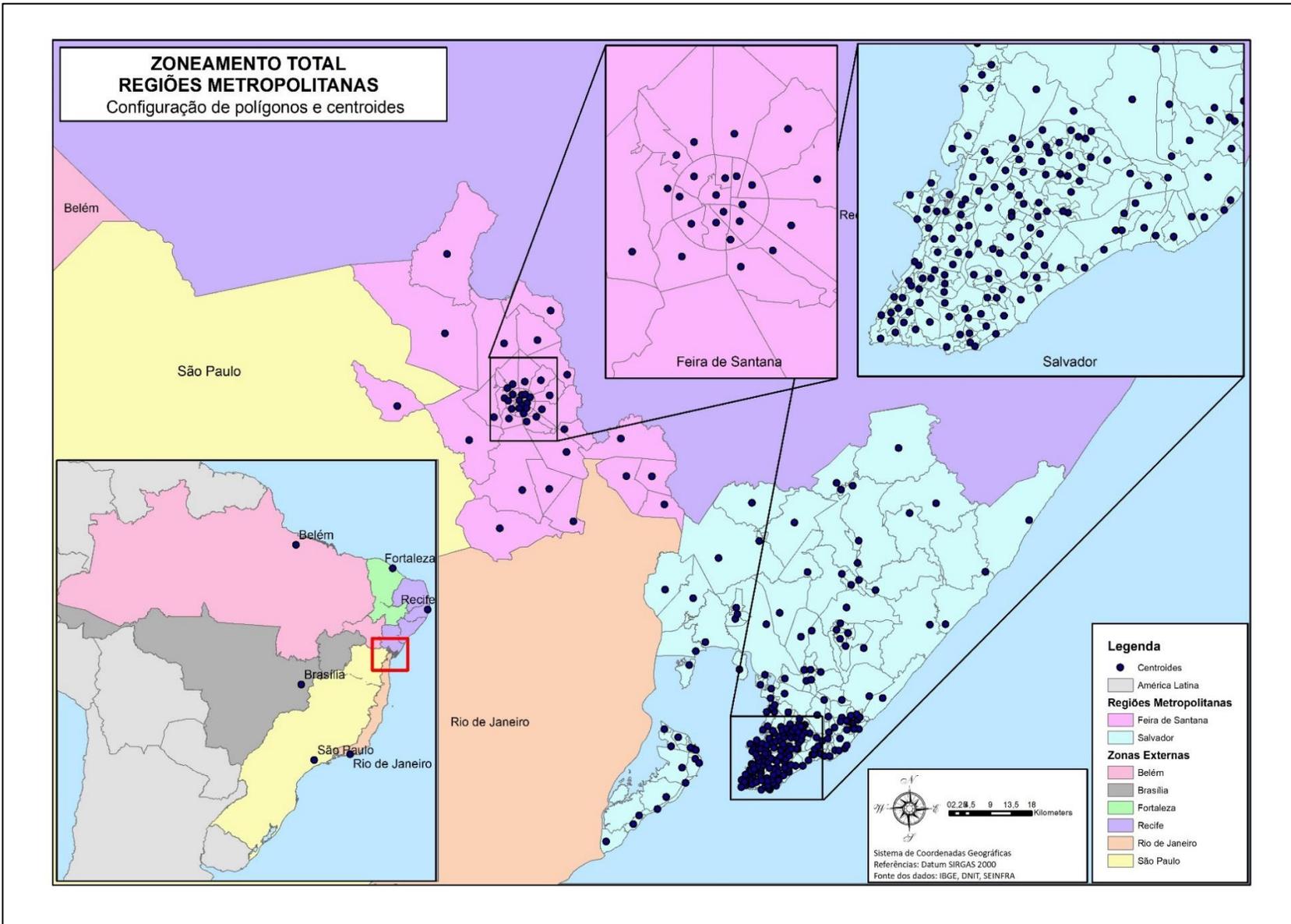
Fonte: APPE (2022) – Adaptado pelo autor

Para a definição do zoneamento externo, foram levados em consideração os pontos principais de acesso às regiões em estudo. A Figura 6 mostra essas conexões ao restante do país, com a junção de municípios, microrregiões, mesorregiões e estados. A conexão ao nordeste das regiões metropolitanas para o centroide Recife é realizada através da rodovia BR-101. Para o sentido norte, no centroide de Fortaleza, há a conexão através da rodovia BR-116. A conexão com o noroeste do país, definido pelo centroide em Belém, é feita pela BR-407 e pela BR-324. Para o centro-oeste, com o centroide definido em Brasília, a principal conexão é feita pela BR-349. O sul e sudeste do país possui sua conexão com as regiões através da BR-116. Por fim, o litoral que abrange os estados de ES e RJ possui sua conexão pela BR-101.



Fonte: Elaborado pelo autor

A Figura 7 apresenta o zoneamento interno das regiões metropolitanas e, ao fim deste trabalho, os Quadros 1 e 2 dispostos no Apêndice A, contêm a configuração do zoneamento externo e do zoneamento total, respectivamente.



Fonte: Elaborado pelo autor

3.4. LEVANTAMENTO DE DADOS – MÉTODO TRADICIONAL

Como qualquer estudo, a aquisição de dados de qualidade é essencial para o sucesso do trabalho. Cox (1984) enfatiza a importância da qualidade dos dados na tomada de decisões e no desenvolvimento de modelos estatísticos precisos. Também defende que a obtenção de dados confiáveis e precisos é essencial para garantir resultados confiáveis e úteis em qualquer área de pesquisa ou aplicação. Mais precisamente, Cox discute a importância do modelo estatístico na análise de dados, mas também ressalta a importância da qualidade dos dados subjacentes ao modelo. Por fim, ainda argumenta que, sem dados confiáveis e precisos, o modelo estatístico não pode produzir resultados confiáveis ou úteis e que, portanto, a qualidade dos dados deve ser uma preocupação primordial em qualquer análise de dados.

Para o desenvolvimento deste trabalho, foram levantados nesse estudo uma série de dados e informações de quatro diferentes fontes:

- i. pesquisas de Origem/Destino Domiciliares;
- ii. análises de Dados de Telefonia Móvel;
- iii. pesquisas de Linha de Contorno; e
- iv. pesquisas de Linha de Travessia.

Para que houvesse a possibilidade de gerenciar um volume tão grande de informações, tanto levantadas quanto produzidas, foi necessário classificar e dividir o conteúdo em:

- a. *shapefiles*: compilação de todos os arquivos usados no processo de análise e criação de mapas;
- b. dados estatísticos: arquivos em excel contendo a relação de setores e zonas, juntamente com as variáveis V001, V002, etc.
- c. análises de Dados de Telefonia Móvel;
- d. pesquisas de Linha de Contorno; e
- e. pesquisas de Linha de Travessia.

Baseado nos dados adquiridos, os resultados a serem desenvolvidos têm por finalidade proporcionar conclusões mais precisas sobre os deslocamentos das populações entre as duas regiões metropolitanas referidas, possibilitando a definição de estratégias que permitam planejar, ao longo dos anos, as melhorias mais adequadas e que possam corresponder às exigências das respectivas populações e municípios.

Primeiramente, verificou-se a existência de estudos anteriores sobre dados relevantes para o desenvolvimento deste trabalho. O Quadro 3 lista os arquivos obtidos.

Região	Arquivos analisados	Observações
Salvador	Arquivos_kmz_do_zoneamento.rar	230 zonas
	Zoneamento_RMS.rar	230 zonas
	Banco de Dados.rar	Arquivos no formato MapInfo e EMME
	Banco de Dados OD Domiciliar rev 2.rar	Arquivos no formato MapInfo e EMME
Feira de Santana	Zoneamento_Feira_de_Santana.rar	30 zonas
	LinhasFeiradeSantana.kml	209 linhas
	Rede de Simulação da RMFS.rar	Arquivos no formato MapInfo e EMME

Fonte: APPE – Adaptado pelo autor

Em relação ao zoneamento, foram constatados problemas em ambas as regiões metropolitanas. Em Feira de Santana, foi necessário realizar uma nova configuração, em virtude de o zoneamento anterior não contemplar a totalidade da região. Esta nova configuração foi construída com base em três condicionantes:

- i. **Setores censitários:** obtidos no sistema do IBGE
- ii. **Eixos de transporte:** analisados de acordo com o SRE-BA
- iii. **Eixos de desenvolvimento:** baseado na evolução industrial e residencial das regiões do município.

O zoneamento da Região Metropolitana de Salvador também apresentou erros, como a não compatibilidade dos limites das zonas com as condicionantes aplicadas para Feira de Santana. A construção do novo zoneamento, no entanto, também usou como referência a configuração anterior.

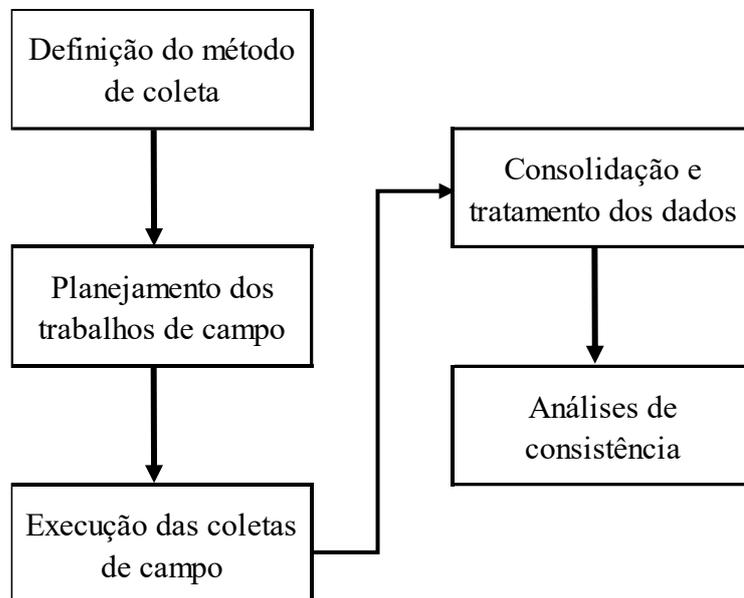
No que tange ao zoneamento externo à área de estudo, conforme apresentado no capítulo anterior, foram definidos agrupamentos ao redor de seis centroides, também levando em consideração os eixos de transporte. Para tal, foram levados em consideração os tipos de unidades geográficas: Município, Microrregião, Mesorregião e Estado. Para o zoneamento interno, considerou-se como menor unidade os setores censitários IBGE (2010), cujas bases foram obtidas diretamente do sítio deste órgão.

3.4.1. PESQUISAS DOMICILIARES

As pesquisas domiciliares envolvem métodos diversos, específicos para cada tipo de levantamento, e são importantes ferramentas utilizadas na coleta de dados para estudos de tráfego, fornecendo informações valiosas sobre os padrões de deslocamento dos indivíduos em uma determinada região.

Para todas as pesquisas existe uma estrutura básica comum, que é explicitada no fluxograma da Figura 8:

Figura 8 - Fluxograma geral do método



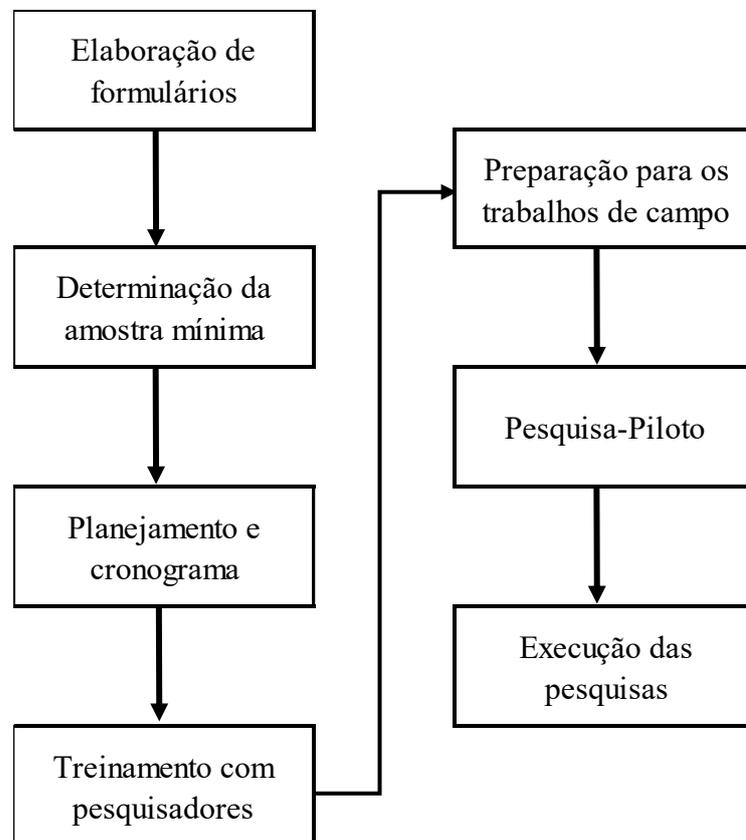
Fonte: APPE (2022) – Adaptado pelo autor

A seleção adequada da amostra é um passo fundamental para garantir a representatividade dos resultados obtidos. Para isso, é importante considerar fatores como tamanho da população, distribuição geográfica, características socioeconômicas e demográficas. Diversas abordagens podem ser utilizadas, tais como amostragem aleatória simples, estratificada ou por conglomerados (HAN, 2014).

A coleta de dados nas pesquisas domiciliares pode ser realizada por meio de diferentes técnicas, como entrevistas presenciais, questionários autoaplicáveis e até mesmo utilizando aplicativos móveis. Tendo em vista a maior importância das pesquisas domiciliares, tanto em termos de recursos despendidos quanto da importância dessas pesquisas no processo de modelagem e planejamento de transportes, o método de coleta dessas pesquisas é mais

complexo. Isso decorre, inclusive, de haver a necessidade de adentrar os domicílios para a realização de entrevistas, o que tende a tornar o processo de coleta mais delicado.

Cada abordagem possui vantagens e desvantagens, sendo necessário avaliar a mais adequada para o contexto específico do estudo. É importante garantir uma abordagem amigável e acessível aos participantes, para obter uma alta taxa de resposta (JOHNSON et al., 2019). Por isso é necessário um treinamento adequado para os pesquisadores, bem como, uma boa distribuição das tarefas realizadas em termos de definição de metodologias e planejamento da coleta de dados. O fluxograma da Figura 9 apresenta de forma resumida cada um dos passos necessários.



Fonte: APPE (2022) – Adaptado pelo autor

Após a coleta dos dados, é necessário realizar o processamento e a análise das informações obtidas. Isso envolve a organização dos dados em formato adequado, a identificação de padrões e tendências, bem como a aplicação de técnicas estatísticas para a

obtenção de resultados significativos. A utilização de software estatístico, como o SPSS ou R, pode facilitar esse processo (WU et al., 2018).

3.4.2. LINHA DE CONTORNO

Como a entrevista domiciliar caracteriza os perfis das viagens internas na área de estudo, torna-se necessária a realização de outros tipos de pesquisas para conhecimento dos deslocamentos externos à área, mas que têm impacto direto no seu sistema de transporte e na sua malha viária. O instrumento adequado para identificação destas viagens externas é a chamada pesquisa de linhas de contorno. Essa técnica consiste em realizar observações e medições ao longo de uma linha traçada perpendicularmente à via, permitindo a obtenção de informações precisas sobre o volume de tráfego, velocidade dos veículos e outras variáveis relevantes.

O planejamento adequado e a seleção cuidadosa do local de estudo são passos cruciais para o sucesso das pesquisas na linha de contorno. É importante considerar fatores como o tipo de via, o volume de tráfego esperado, a disponibilidade de pontos de observação e a representatividade da amostra. A escolha estratégica do local garantirá a coleta de dados relevantes e confiáveis (JIA et al., 2018).

O método para aplicação dessa pesquisa é entrevistar motoristas de veículos de passeio e de carga, ou passageiros de veículos que trafegam nas vias que ligam as regiões metropolitanas ao estado (no entorno), ou seja, nos limites da área de estudo. Para conhecimento do universo (número total de veículos passantes), deverão ser realizadas contagens volumétricas classificatórias de veículos.

Antes de iniciar as observações, é necessário definir os parâmetros de medição que serão registrados ao longo da linha de contorno. Esses parâmetros podem incluir o volume de tráfego, velocidade média dos veículos, composição do tráfego (carros, caminhões, motocicletas, etc.), entre outros. A definição clara dos parâmetros garantirá a consistência e comparabilidade dos dados coletados (ZHOU et al., 2020).

Para realização da pesquisa nas linhas de contorno é necessário o apoio da Polícia Rodoviária Estadual e da Policial Rodoviária Federal, dependendo da jurisdição da rodovia. Este apoio é essencial à realização das entrevistas junto aos motoristas e passageiros, uma vez que os veículos precisam ser parados pelos policiais em locais previamente escolhidos e, somente

após o estacionamento do veículo no acostamento ou em outra área definida para sua parada, o entrevistador poderá abordar o usuário para aplicação do respectivo questionário.

A coleta de dados na linha de contorno envolve a instalação dos equipamentos ao longo da via e o registro contínuo das informações. É necessário garantir que os pontos de observação estejam adequadamente distribuídos e que os equipamentos estejam funcionando corretamente. A duração da coleta de dados pode variar de acordo com a necessidade do estudo e é importante considerar fatores como a sazonalidade e eventos especiais que possam afetar o tráfego (ZHOU et al., 2018).

A escolha dos equipamentos e instrumentação adequados é essencial para a coleta precisa dos dados na linha de contorno. Dentre os equipamentos comumente utilizados, destacam-se os contadores de tráfego automáticos, radares, câmeras de vídeo e sensores de velocidade. A seleção correta dos instrumentos garantirá a obtenção de dados confiáveis e de alta qualidade (YANG et al., 2019).

A seguir, apresenta-se uma sequência de atividades que foram desenvolvidas para realização da pesquisa nas linhas de contorno, a ser realizada nos postos, em dias úteis, entre 6:00 e 20:00 horas, em ambos os sentidos das vias. Foram realizadas as pesquisas de contagem e entrevistas. A seguir, é listada a sequência de atividades:

- (i) Recrutamento, seleção e treinamento de equipes de campo;
- (ii) Preparação dos trabalhos de campo e organização da logística;
- (iii) Execução do plano de comunicação das pesquisas;
- (iv) Aplicação da pesquisa nos locais previamente definidos;
- (v) Consolidação e armazenamento do banco de dados em servidores seguros;
- (vi) Apresentação dos resultados sob forma de Sumário de Dados.

Após a coleta dos dados, é necessário realizar o processamento e a análise das informações obtidas. Isso pode envolver a tabulação dos dados, a identificação de padrões de tráfego, a análise estatística e a geração de relatórios. Técnicas como regressão linear, análise de séries temporais e modelagem estatística podem ser aplicadas para compreender os padrões de tráfego e obter insights relevantes (MADIREDDY et al., 2021).

É importante considerar a validade e as limitações das pesquisas na linha de contorno ao interpretar os resultados obtidos. Essa técnica oferece uma visão detalhada do tráfego em uma via específica, mas pode não ser representativa de todas as condições de tráfego em uma

área. Além disso, fatores externos, como obras de infraestrutura ou eventos especiais, podem afetar temporariamente os padrões de tráfego registrados (ZHOU et al., 2020).

3.4.3. LOCALIZAÇÃO DOS POSTOS DE CONTAGEM E DE PESQUISAS DE OD

Com o objetivo de fornecer subsídios para a calibração das redes de transporte, pesquisas domiciliares e via telefonia móvel, foi definido um conjunto de postos de pesquisa para coleta de informações sobre volume de veículos, sobre origem e destino e demais características das viagens.

Os levantamentos nestes postos, localizados em eixos principais na periferia das Áreas de Estudo, tem o objetivo de identificar viagens que ingressam, saem ou atravessam essas áreas - as Regiões Metropolitanas de Salvador e Feira de Santana.

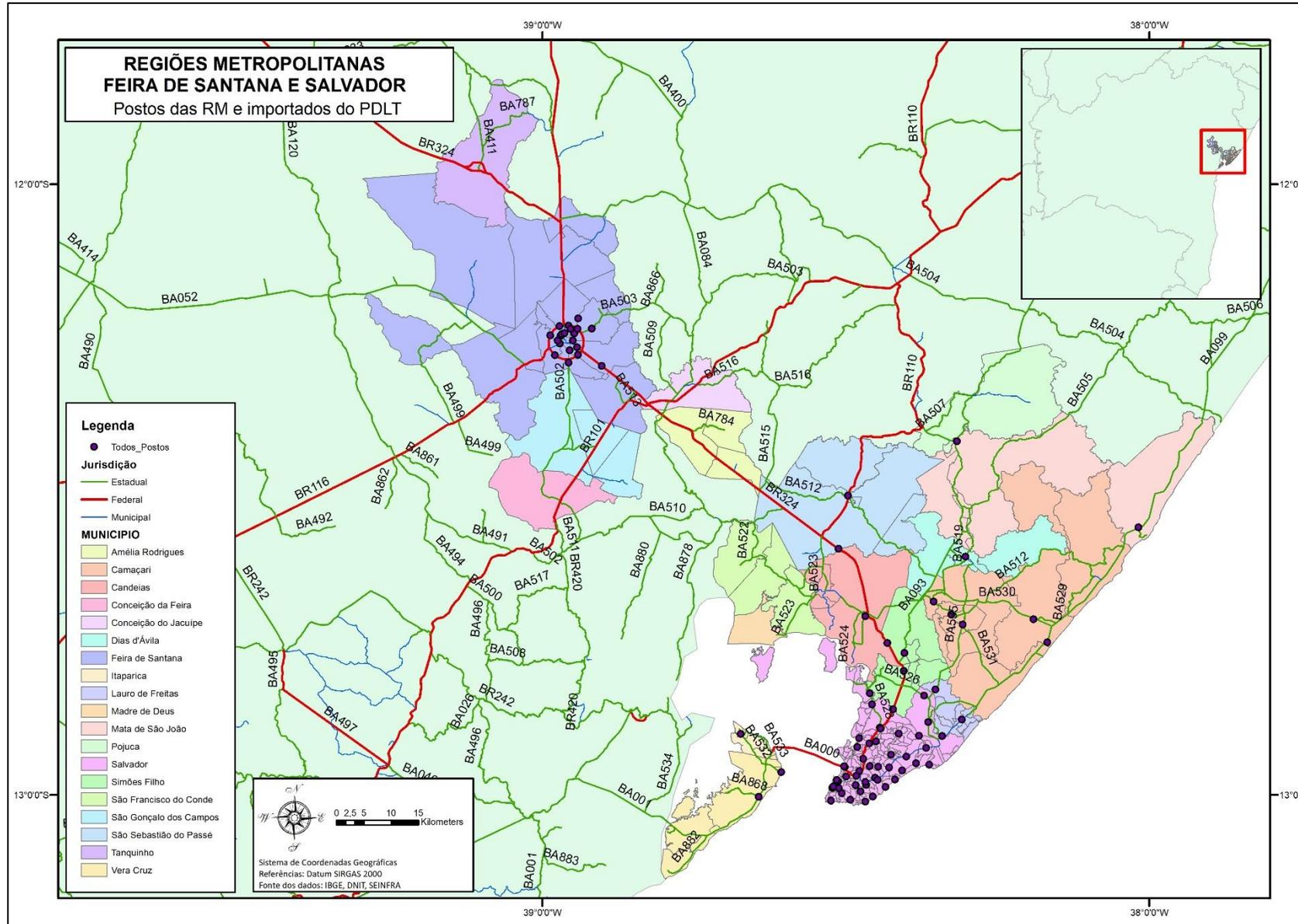
Foram selecionados 8 (oito) postos na Linha de Contorno da Região Metropolitana de Salvador e 6 (seis) postos na Linha de Contorno da Região Metropolitana de Feira de Santana, conforme disposto no Quadro 4:

REGIÃO	POSTO	LOCALIZAÇÃO	REFERÊNCIA	JURISDIÇÃO
Salvador	1	BA-093	Pojuca x Araçás	Estadual
	2	BA-505	Mata de São João x Itanagra	Estadual
	3	BA-099	Mata de São João x Entre Rios	Estadual
	4	BR-110	Pojuca x Catu	Federal
	5	BR-110	S. Sebastião do Passé x Catu	Federal
	6	BA-515	S. Sebastião do Passé x A. Rodrigues	Estadual
	7	BA-030	Ponte do Funil - Itaparica	Estadual
	8	BR-420	Santo Amaro x S. Francisco do Conde	Federal
F. Santana	1	BR-101, km 196	km 196	Federal
	2	BR-324, km 563	Apoio ao Usuário Via Bahia	Federal
	3	BR-101, km 136	Posto Trevo BR	Federal
	4	BR-116, km 408	Posto São Cristóvão	Federal
	5	BR-116, km 429	Posto PRF	Federal
	6	BR-084, km 040	Fazenda/Placa km 40	Estadual

Fonte: APPE (2022) – Adaptado pelo autor

Outros postos foram agregados ao estudo, os quais vieram do Plano Diretor Logístico do Estado da Bahia. Esses postos serviram também como pontos de calibração para as matrizes.

A Figura 10 mostra a disposição desses postos ao longo das regiões metropolitanas.

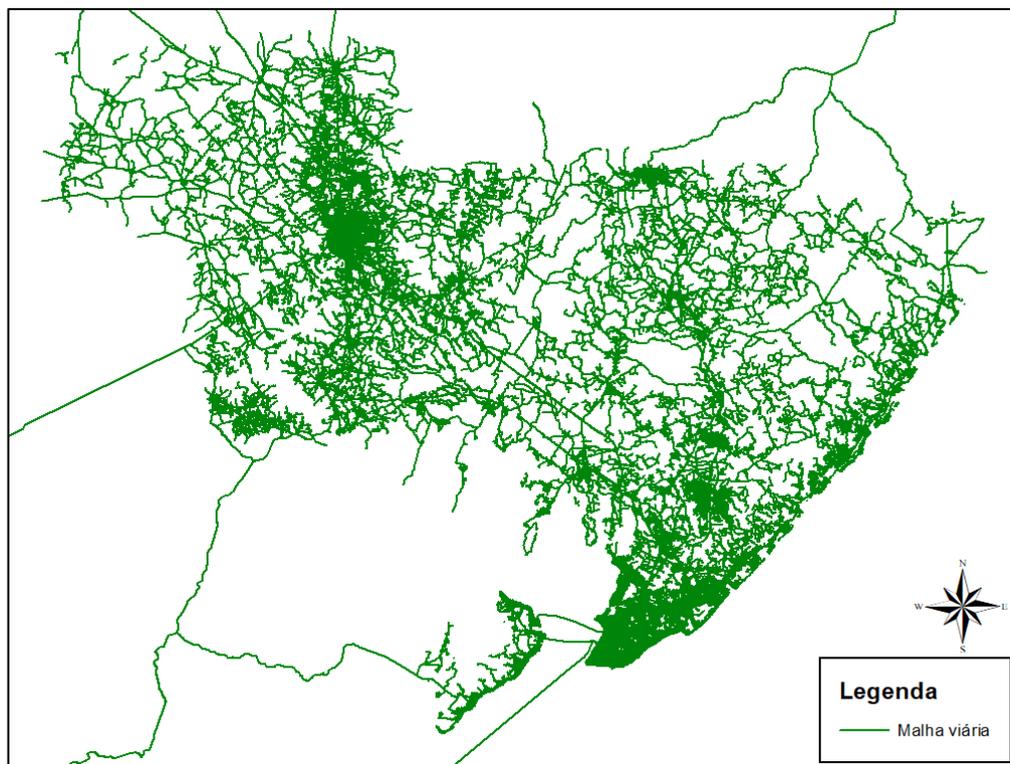


Fonte: APPE (2022) – Adaptado pelo autor

3.4.4. *SHAPEFILE DE “COUNT LOCATIONS”*

A utilização do módulo “*Count Location*”, projetado para importar, visualizar e processar dados de contagem de tráfego em modelos de transporte, auxilia na melhora da precisão e representatividade das matrizes de tráfego usadas em simulações de transporte e análises de redes viárias. Através da importação do *shapefile* com os postos de contagem devidamente configurados, se torna possível a integração de dados de contagem de tráfego no processo de modelagem e planejamento de transporte.

Esse *shapefile* foi montado através da sobreposição da malha rodoviária com os pontos de contagem, conforme apresentado na Figura 11, utilizando a ferramenta ArcGIS. Utilizando como referência as informações dos nós (origem e destino – FROMNODENO/TONODENO), foi feita uma junção espacial (*Spatial Join*) para criar um outro *shapefile* de pontos com tais informações.



Fonte: Elaborado pelo autor

Logo após, esse *shapefile* foi importado no VISUM, utilizando as informações de nós coletadas no processo anterior para definir a posição do *Count Location* no *link* correto e, principalmente, no sentido correto.

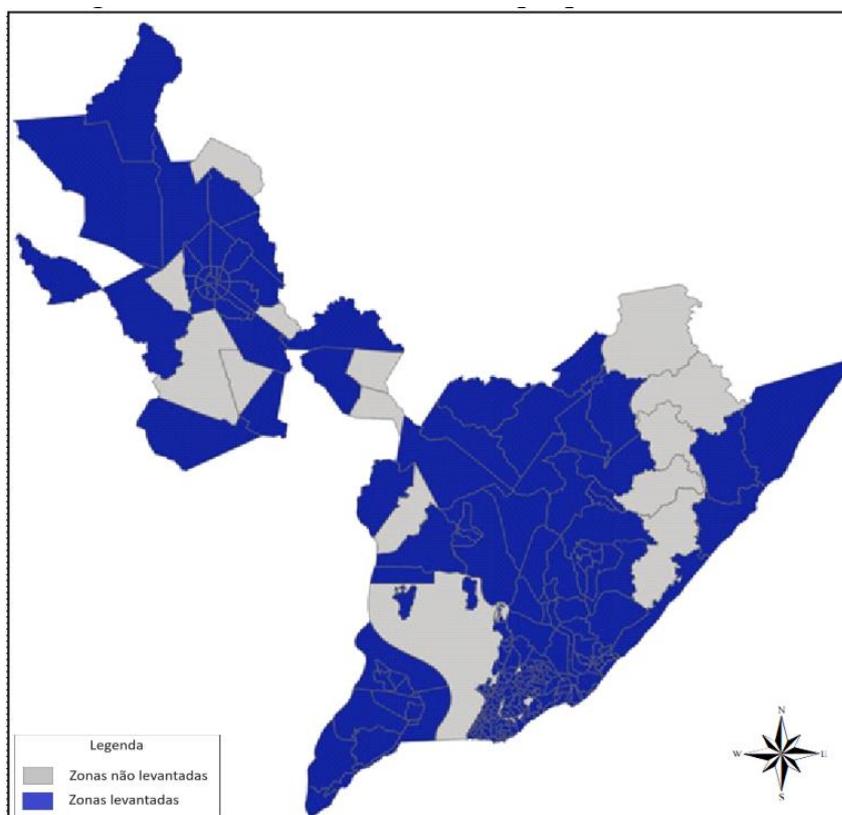
4. CONSTRUÇÃO DE MATRIZES E MODELAGEM

Serão elucidados a seguir os métodos utilizados no desenvolvimento das matrizes. O método tradicional será abordado de forma holística, tendo em vista que já é um método muito consolidado e, além disso, o foco deste trabalho está voltado para o procedimento envolvendo dados móveis e sua posterior análise comparativa.

4.1. MÉTODO TRADICIONAL

4.1.1. PESQUISAS REALIZADAS

Ao total, foram pesquisados 25.003 domicílios nas duas regiões metropolitanas, envolvendo 49.791 cidadãos e totalizando 65.515 viagens registradas no banco de dados, conforme consta a Figura 12.



Fonte: Certare, 2023 – Adaptado pelo autor

Observando a Figura 12, pode-se verificar que algumas zonas (na cor cinza) foram incorporadas a outras em virtude da infraestrutura presente, a qual condiciona o fluxo a compartilhar o mesmo percurso, ou em razão da baixa densidade presente encontrada nas análises dos setores censitários. As zonas incorporadas foram:

- Feira de Santana: Tiquaruçu
- Feira de Santana: UFRB
- Feira de Santana: São Gonçalo dos Campos
- Feira de Santana: Sergi
- Feira de Santana: Bom Viver
- Feira de Santana: Inhatá
- Feira de Santana: Mata da Aliança
- Salvador: Monte Recôncavo
- Salvador: Miranga / Petro Recôncavo
- Salvador: Vargem de Baixo / Fazenda Juruaba
- Salvador: Camassari / Visconde
- Salvador: Comunidade Católica Cidade Santa / Biribeira
- Salvador: Jacuípe / Campos / Garajuba

A técnica de amostragem utilizada para dimensionar o número de domicílios a serem pesquisados em cada uma das zonas OD é a Amostragem Casual Simples. O tamanho da amostra em cada zona de pesquisa obedeceu a Equação (1), baseada no coeficiente de variação da variável **número médio de viagens por domicílio**:

$$n_i = \frac{N_i \cdot z_{\alpha}^2 \cdot C^2(Y)}{(N_i - 1) \cdot \varepsilon^2 + z_{\alpha}^2 \cdot C^2(Y)} \quad (1)$$

Em que:

n_i : tamanho da amostra i ;

N_i : número de domicílios na zona i ;

Y : número de viagens por domicílio na zona i ;

$C(Y)$: coeficiente de variação da variável Y ou [desvio padrão (Y) / média (Y)]

z_{α} : ponto da curva normal associado ao intervalo com nível de confiança de $(1 - \alpha)$

ε : erro absoluto máximo admissível.

Dado o cenário de falta de dados relacionados à variável "número médio de viagens por domicílio" em Feira de Santana, assim como a falta de informações sobre seu coeficiente de variação correspondente, adotou-se um parâmetro de referência para calcular o tamanho da amostra. Esse parâmetro, representado por $C(Y) = 0,53$, corresponde à mediana dos coeficientes de variação do número médio de viagens por domicílio, conforme obtido na Pesquisa Origem e Destino de Salvador, em 2011.

A amostra em questão abrange um total de 2.001 domicílios, distribuídos de acordo com as 34 zonas de pesquisa, conforme detalhado na tabela subsequente. Esses cálculos levaram em consideração um intervalo de confiança de 87% e um erro absoluto máximo admissível de 10% ($\varepsilon = 0,10$).

Com o objetivo de manter o erro máximo admissível para essa estimativa dentro dos limites estabelecidos em 10%, foi necessário agregar algumas zonas que apresentavam características semelhantes ou uma conexão direta com o centro de Feira de Santana ou Salvador.

No contexto da Região Metropolitana de Salvador, as zonas classificadas como "vazias" foram aquelas desprovidas de uso residencial ou que englobavam grandes infraestruturas, como shoppings centers, terminais rodoviários, áreas de mananciais e áreas de proteção ambiental. Essas zonas foram identificadas como 43, 53, 63, 84, 85, 92, 127, 179, 197, 201, 225, 241 e 244.

Para calcular os tamanhos da amostra, foi utilizado um parâmetro referencial para a constante $C(Y)$ com o valor de 0,53, que corresponde à mediana dos coeficientes de variação do número médio de viagens por domicílio, obtidos nas zonas da Pesquisa Origem e Destino de Salvador realizada em 2011.

A amostra total abrange 23.001 domicílios, distribuídos em 217 zonas de pesquisa, conforme apresentado na tabela subsequente. Esses cálculos foram efetuados considerando um intervalo de confiança de 95% (ou $\alpha = 5\%$) e um erro absoluto máximo admissível de 9,8% ($\varepsilon = 0,098$).

A Tabela 3 e a Tabela 4 apresentam os planos amostrais para realização das pesquisas nas respectivas regiões metropolitanas.

Município	Estimativa População 2021 IBGE	Estimativa Domicílios 2021	Tamanho da Amostra (questionários válidos)
Amélia Rodrigues	24.997	7.118	59
Conceição da Feira	22.933	6.400	60
Conceição do Jacuípe	33.631	9.946	60
Feira de Santana	624.107	183.543	1.706
São Gonçalo dos Campos	38.315	10.394	58
Tanquinho	7.936	2.203	58
TOTAL:	751.919	219.604	2.001

Fonte: APPE (2022), adaptado pelo autor

Município	Estimativa População 2021 IBGE	Estimativa Domicílios 2021	Tamanho da Amostra (questionários válidos)
Camaçari	309.206	94.441	1.183
Candeias	87.819	26.370	539
Dias D'Ávila	83.705	25.166	326
Itaparica	22.440	6.895	510
Lauro de Freitas	204.670	62.291	1.553
Madre de Deus	21.754	6.493	109
Mata de São João	47.643	13.998	424
Pojuca	40.401	11.825	216
Salvador	2.940.720	946.400	15.674
São Francisco do Conde	40.664	11.684	219
São Sebastião do Passé	44.555	13.061	507
Simões Filho	137.117	40.893	963
Vera Cruz	44.186	13.908	995
TOTAL:	4.024.880	1.273.425	23.218

Fonte: APPE (2022), adaptado pelo autor

A partir dessa base, iniciaram-se os processos de pesquisas de origem/destino e contagens volumétricas classificatórias da linha de contorno e das linhas de travessia, sendo que essas ocorreram de forma automática e manual. Também foram realizadas as pesquisas de ocupação de forma visual nas linhas de travessia, além das pesquisas domiciliares em ambas as regiões metropolitanas.

4.1.2. EXPANSÕES PARA CONSTRUÇÃO DAS MATRIZES

Para criar as matrizes OD para o referido estudo, foram utilizadas as seguintes premissas:

- i. **Pesquisas Domiciliares:** Para expandir as pesquisas e obter uma representação adequada da população de cada zona, foi aplicada uma expansão proporcional com base na população de cada zona, de acordo com a Equação (2)(2).

$$OD_{\text{expandido}} = OD_{\text{observado}} * \frac{P_{\text{total}}}{P_{\text{amostra}}} \quad (2)$$

Onde:

- $OD_{\text{expandido}}$: matriz OD expandida
- $OD_{\text{observado}}$: matriz OD observada nas pesquisas domiciliares
- P_{total} : população da zona de estudo
- P_{amostra} : população da amostra das pesquisas domiciliares

- ii. **Pesquisas na Linha de Contorno e na Linha de Travessia:** As pesquisas na linha de contorno e na linha de travessia foram realizadas para capturar os fluxos de tráfego que não se originam ou terminam em zonas residenciais. Para expandir essas pesquisas, utilizou-se contagens de volume médio diário (VMD) corrigido para uma média anual através da Equação (3).

$$OD_{\text{expandido}} = OD_{\text{observado}} * \frac{VMDA_{\text{total}}}{VMD_{\text{amostra}}} \quad (3)$$

Onde:

- $OD_{\text{expandido}}$: matriz OD expandida
- $OD_{\text{observado}}$: matriz OD observada nas pesquisas na linha de contorno e na linha de travessia
- $VMDA_{\text{total}}$: VMDA total da zona de estudo
- VMD_{amostra} : VMD da amostra das pesquisas linha de contorno e na linha de travessia

- iii. **Expansão de Contagens de 6h para 24h:** Para representar os fluxos de tráfego ao longo de um dia completo, ou seja, realizar a expansão diária, foi criada uma macro para que as contagens fossem expandidas de 6 horas para 24 horas. Isso

foi feito com base nos postos alocados na linha de contorno e na linha de travessia com que possuía contagens de 24 horas no mesmo dia da semana. Essa expansão é fundamental para criar matrizes OD precisas ao longo do dia e foi feita utilizando a Equação (4).

Na linha de travessia havia 7 postos, sendo 6 na Região Metropolitana de Salvador e apenas 1 na Região Metropolitana de Feira de Santana. Na linha de contorno, o estudo compreendeu 14 postos, estando 8 na RMS e 6 na RMFS.

$$Cont_local_OD_{expandido} = Cont_local_OD_{observado} * \frac{VMDA_{24h,posto_LT_LC}}{VMD_{6h,amostra}} \quad (4)$$

Onde:

- $Cont_local_OD_{expandido}$: matriz OD expandida
- $Cont_local_OD_{observado}$: matriz OD observada nas pesquisas na linha de contorno e na linha de travessia
- $VMDA_{24h,posto_LT_LC}$: VMDA de 24h do posto representativo
- $VMD_{6h,amostra}$: VMD de 6 horas da amostra de contagens

A construção da macro foi baseada na estrutura criada, na qual a planilha apresentava primeiramente o posto de cobertura seguido da data e dos horários iniciais e finais de contagens na parte da manhã e da tarde. Primeiramente, a macro calculava o tráfego total por classe de veículo e, posteriormente, a planilha trazia os fatores de expansão para o VMDA para calcular o VMDA por categoria de veículo no posto de cobertura. Por fim, foram agrupados os volumes para a categorias de veículos usadas no estudo, sendo motos, veículos leves, ônibus e veículos de carga (caminhões).

- iv. **Uso de Posto Sazonal:** Para capturar os efeitos sazonais nas viagens na intenção de realizar a correção semanal, utilizou-se a Equação (5), com dados dos postos sazonais com contagens de tráfego durante uma única semana representativa. Além dos postos específicos do próprio estudo, foram também considerados postos do PNCT. Essas contagens sazonais foram usadas para expandir o volume horário para 24 horas e levar em consideração variações sazonais no tráfego.

$$OD_{expandido} = OD_{observado} * \frac{VMD_{24h,posto_sazonal}}{VMD_{24h,amostra}} \quad (5)$$

Onde:

- $OD_{expandido}$: matriz OD expandida
- $OD_{observado}$: matriz OD observada nas pesquisas na linha de contorno e na linha de travessia
- $VMD_{24h,posto_sazonal}$: VMD de 24h do posto sazonal
- $VMD_{24h,amostra}$: VMD de 24 horas da amostra de contagens

- v. **Volume de Pedágios:** Embora não tenha sido possível obter acesso aos volumes horários diários de pedágios, foram utilizados os dados do acumulado mensal para realizar a correção mensal dos volumes médios diários semanais anteriormente calculados.

4.2. MÉTODO DE DADOS MÓVEIS

4.2.1. APRESENTAÇÃO DOS POLÍGONOS

Para a análise comparativa entre os dois métodos utilizados (tradicional e com dados de telefonia) é essencial que o zoneamento seja o mesmo. Esse zoneamento foi solicitado pela operadora, a qual verificou a possibilidade de realizar a construção do banco de dados de telefonia relativos aos dados de 2022.

Já para determinar as zonas externas, levou-se em conta a rede viária e as características topográficas dos centros populacionais das áreas externas, sendo 6 zonas de tráfego externas. Dessa forma, o zoneamento é formado por 273 zonas internas e 6 zonas externas, conforme apresentados na Figura 6 e Figura 7.

4.2.2. CONSTRUÇÃO DO BANCO DE DADOS

A empresa escolhida para a realização do estudo foi a Kido Dynamics, em virtude da gama de informações disponibilizadas através de sua plataforma e seu processamento mais ágil.

Os dados disponibilizados foram:

- i. Matrizes OD por hora, total diário, por dia útil e final de semana;

- ii. Matrizes OD por motivo de viagem, por faixa de renda, gênero e faixa etária;
- iii. Trajetórias e velocidades, para viagens acima de 5 km.

4.2.3. COLETA DE DADOS JUNTOS A OPERADORAS DE TELEFONIA CELULAR ESCOLHIDA

A operadora coleta os dados através de redes de antenas de telefonia, sendo sua principal fonte de dados. Denominando todo registro captado como um evento, onde podem ser coletados por meio de chamadas telefônicas, SMS, conexões à internet e informações da própria operadora para garantir o controle e qualidade do serviço.

Cada registro possui informações de espaço-tempo do dispositivo relacionando sua posição em uma determinada área de cobertura, que está associada a uma antena.

Contudo para concluir a análise de mobilidade, esses dados complementam-se com arquivos de geolocalização das antenas, conexões 2G, 3G, 4G e 5G e arquivos contendo informações úteis dos clientes, como: faixa etária, gênero, classe social, nacionalidade, nacionais e estrangeiros viajando pelo país. Com a captação desses conjuntos de dados a análise de mobilidade foi viabilizada.

4.2.4. CONSTRUÇÃO DE BANCO DE DADOS A PARTIR DAS ANÁLISES DE BIG DATA

Para as etapas de construção do banco de dados, a operadora telefônica trabalha através de quatro fases de análise. São elas:

- i. **Filtro de erros de transcrição:** uma limpeza de dados incompletos é realizada;
- ii. **Filtro de massa crítico do usuário:** em torno de 25% dos dispositivos com o menor número de eventos registrados são eliminados;
- iii. **Filtro de relevância de amostra:** eventos redundantes ou considerados irrelevantes são eliminados;
- iv. **Filtro de ação:** a dinâmica de uma sequência de eventos registrados de cada usuário é analisada, filtrando os que mostram um movimento incompatível com a física no tempo e no espaço.

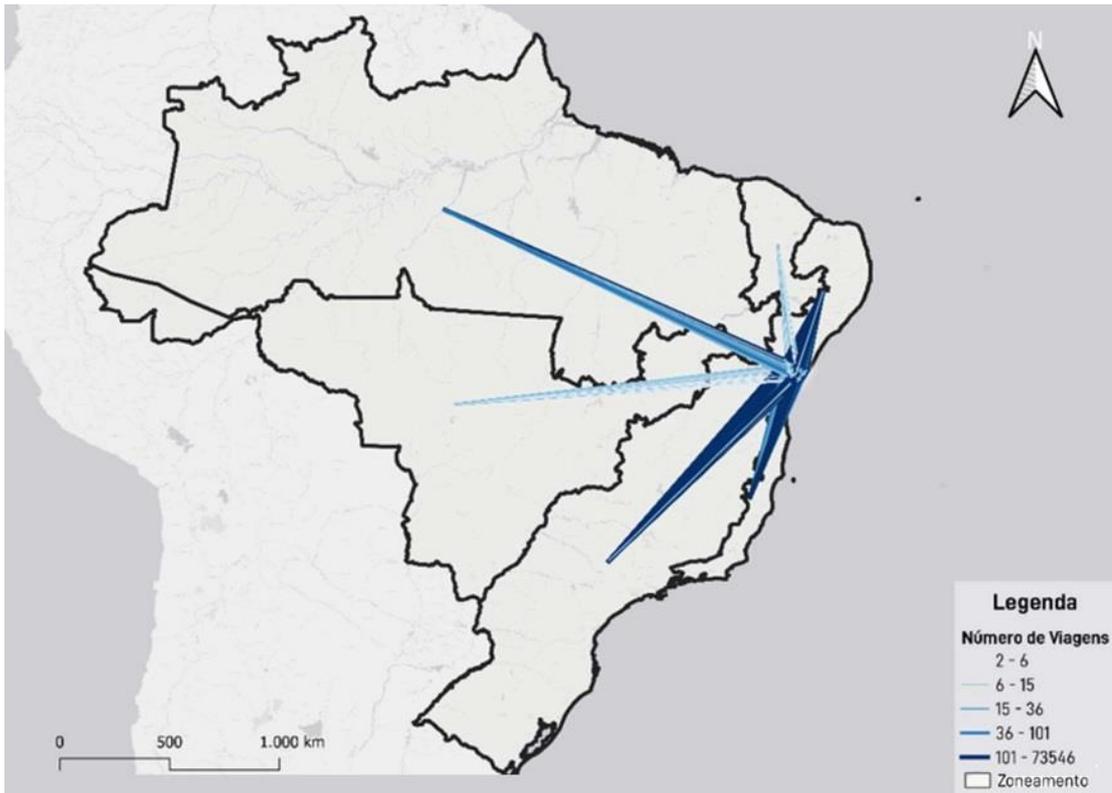
Apesar da aplicação desses filtros, o número de eventos registrados da amostra permanece significativo. Mesmo com o banco de dados sofrendo uma redução de 10% do original, ainda dispõe de uma quantidade de dados relevantes para a mobilidade.

No Anexo A está disposto todo o processo de tratamento dos dados para obtenção das matrizes. O texto apresentado foi desenvolvido pela empresa parceira da telefonia que faz o processo de coleta dos dados e posterior tratamento.

4.2.5. CONSTRUÇÃO DAS LINHAS DE DESEJO

As empresas de telefonia celular operam com rede de antenas que fornecem cobertura contínua para os celulares dos clientes. Com base nas conexões entre os aparelhos móveis e as antenas da operadora, estima-se a quantidade de viagens realizadas dentro de uma determinada região. A partir disso, é possível realizar análises a respeito dos locais, horários e dias com maiores volumes de viagens.

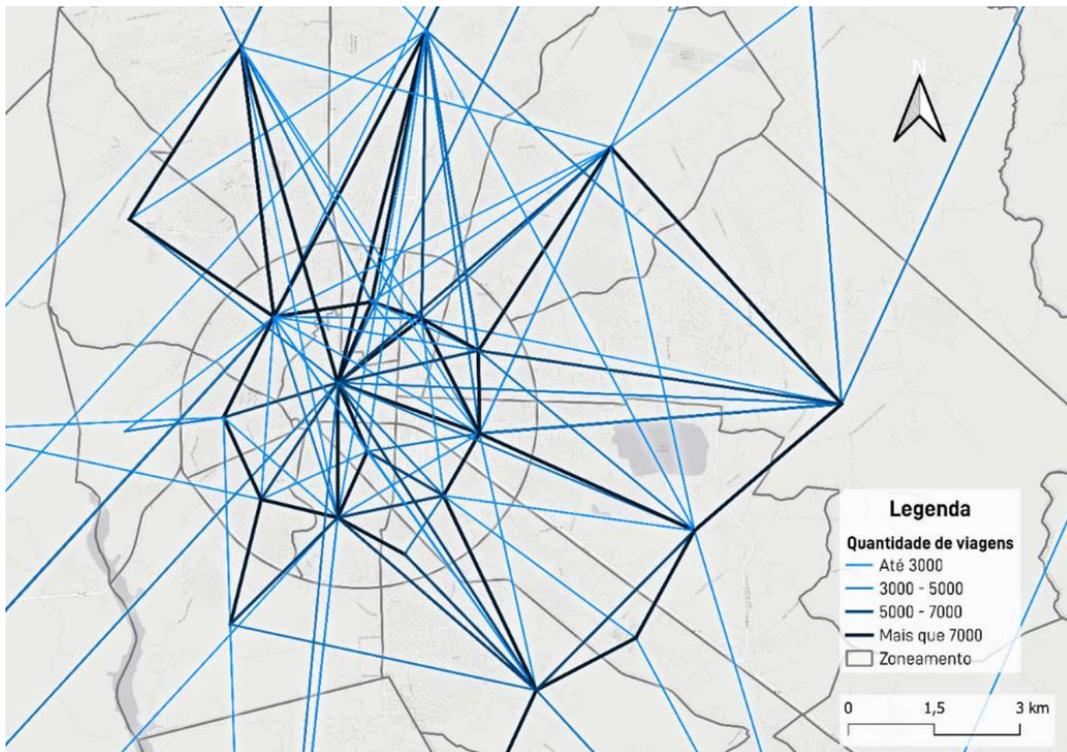
A seguir apresenta-se maiores detalhes quanto às Linhas de Desejo referentes às regiões metropolitanas de Feira de Santana e Salvador. Quanto às viagens que possuem origem e/ou destino em regiões externas à RM, nota-se que há maior intensidade de viagens entre a região sudeste e região nordeste - com foco nos estados do litoral leste. Essa composição é percebida também a partir dos DDD mais observados, conforme mostrado no Relatório 3 – nota-se, ao se excluir os DDDs 75 e 71, que os predominantes são os DDDs de São Paulo, Santa Catarina, Pernambuco, Rio de Janeiro e Ceará. A Figura 13 apresenta as linhas de desejo para essas regiões externas.



Fonte: Kido Dynamics, 2022

De forma mais detalhada para a Região Metropolitana de Feira de Santana, percebe-se um padrão de viagens de forma radial, com algumas centralidades. As maiores concentrações de viagens estão na região central, havendo, porém, demandas entre todas as zonas que delimitam a Av Eduardo Froes da Mota – importante delimitação de zonas dentro da cidade.

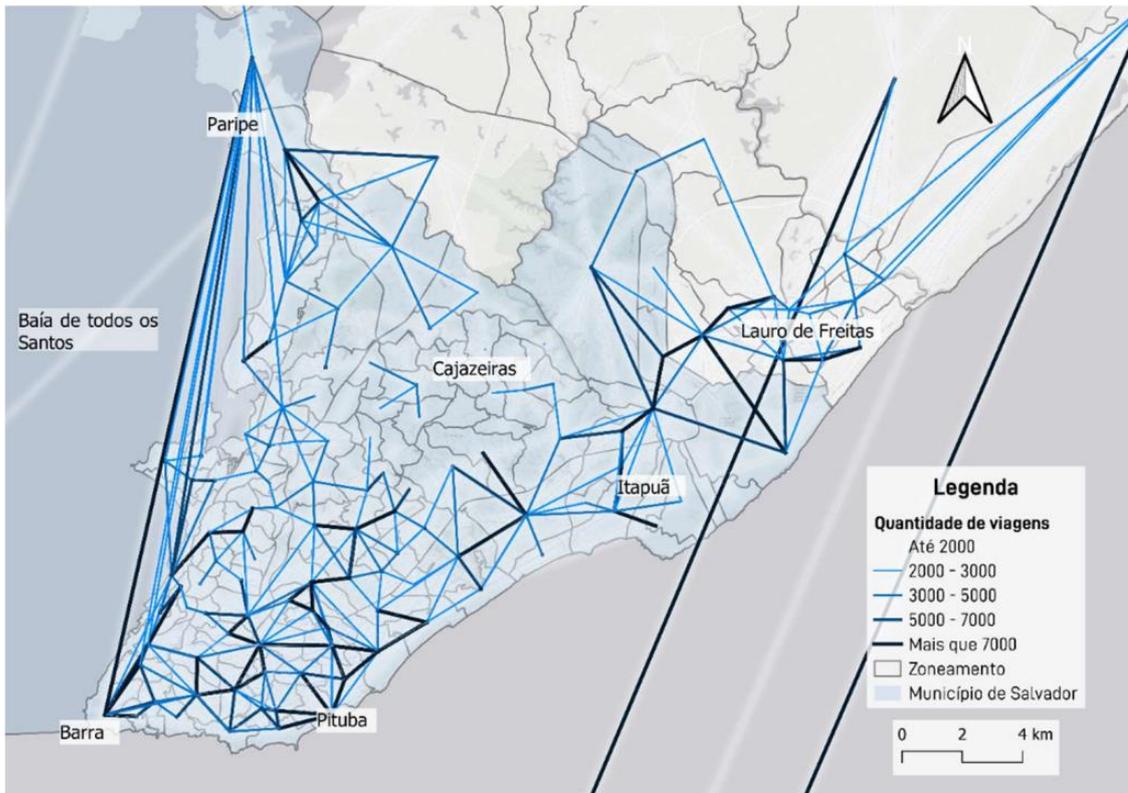
A Figura 14 apresenta esses padrões verificados em Feira de Santana, considerando-se as viagens médias dentro de um dia útil típico.



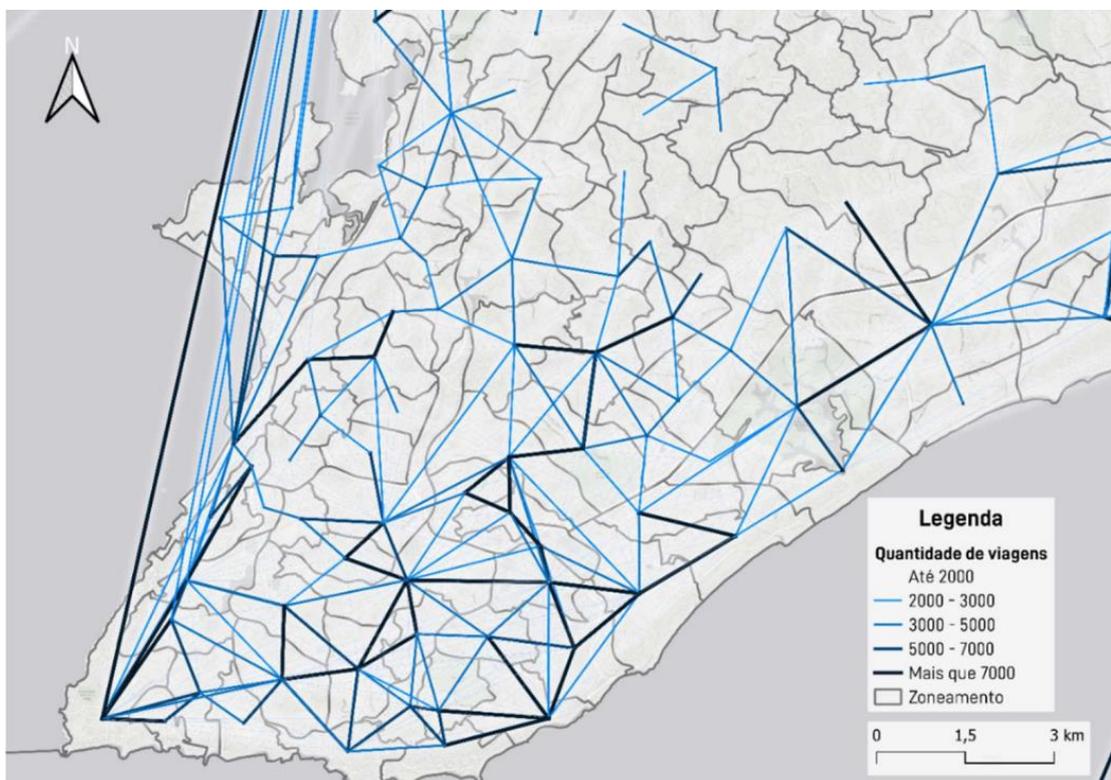
Fonte: Kido Dynamics, 2022

Considerando apenas Salvador, observa-se maior desigualdade entre essas viagens. Alguns postos funcionam como centralidades do município, gerando alta demanda de viagens no entorno, não sendo notadas linhas de desejo entre zonas muito distantes umas das outras, com exceção das viagens realizadas entre Paripe e a região da orla da Baía de Todos os Santos.

O padrão radial não é observado de forma intensa em toda a cidade e região metropolitana. Os bairros Barra, Pituba, Caminho das árvores e Brotas apresentam elevadas demandas. Além disso, o município de Lauro de Freitas apresenta expressiva participação nas viagens para Salvador, principalmente nos bairros mais próximos ao limite municipal entre esses dois municípios. A Figura 15 e a Figura 16 apresentam os padrões observados para Salvador, considerando-se as viagens médias em um dia útil típico.



Fonte: Kido Dynamics, 2022



Fonte: Kido Dynamics, 2022

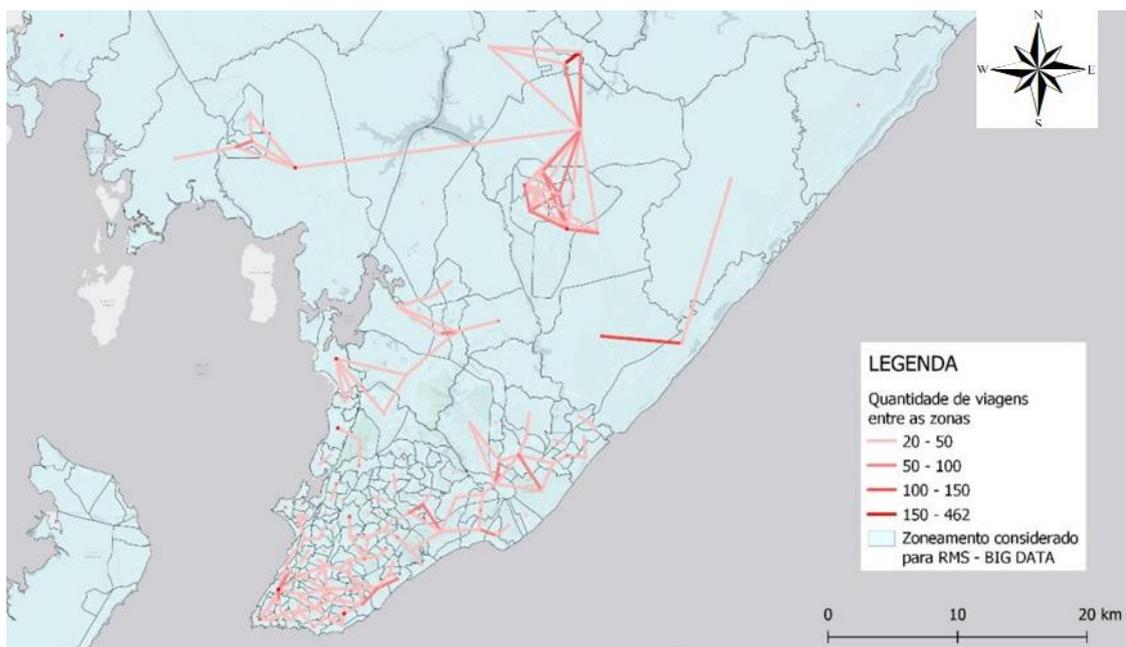
O gráfico correspondente à Figura 17 apresenta o comportamento dos deslocamentos ao longo de um dia útil típico, onde percebe-se picos nos horários de almoço e fim de dia, especialmente entre 18h e 19h.



Fonte: Kido Dynamics, 2022

Ressalta-se que, para estudos futuros envolvendo análises em transportes coletivos, a análise horária deverá ser realizada também para todos os dias da semana separadamente, bem como para o fim de semana, a fim de entender flutuações nos valores.

A Figura 18 mostra uma espacialização da realização dessas viagens, apresentando os locais entre os quais há maiores números de viagens dentro da hora pico das 18h às 19h.



Fonte: Kido Dynamics, 2022

4.2.6. CLASSIFICAÇÃO DA MATRIZ OD

Conforme abordado, as empresas de telefonia celular operam com rede de antenas que fornecem cobertura contínua para os celulares dos clientes. Com base nas conexões entre os aparelhos móveis e as antenas da operadora, estima-se a quantidade de viagens realizadas dentro de uma determinada região.

No entanto, assim como exposto no capítulo de revisão bibliográfica, para a classificação das matrizes pelo método de GPS, ou seja, através de algoritmos que pudessem analisar o comportamento dos fluxos, seria necessário ter acesso a uma gama de informações da *big data*, como código do aparelho, código da antena, *timestamp*, entre outros. Porém, dada às diretrizes da Lei Geral de Proteção de Dados (LGPD), o autor não conseguiu acesso aos dados completos, apenas valores representando a quantidade de celulares entre os pares OD, impossibilitando uma análise nessa profundidade.

A LGPD é uma legislação brasileira que estabelece diretrizes rigorosas para a proteção dos dados pessoais dos indivíduos. De acordo com a LGPD, as empresas privadas não podem simplesmente obter detalhes de dados de telefonia, como *timestamp*, código da antena, código do usuário e outros dados do "*big data*" para montar estudos de tráfego sem uma base legal sólida e o consentimento explícito dos titulares dos dados. Existem várias razões pelas quais isso não é possível de acordo com a LGPD:

- i. **Consentimento:** A LGPD exige que as empresas obtenham o consentimento claro e informado dos titulares dos dados antes de coletar e processar qualquer informação pessoal. Isso se aplica especialmente a dados sensíveis, como informações de telefonia que podem revelar a localização e os movimentos das pessoas. Sem o consentimento explícito dos usuários, a coleta desses dados seria ilegal.
- ii. **Princípio da Necessidade:** A LGPD estabelece o princípio da necessidade, o que significa que as empresas só podem coletar e processar os dados necessários para a finalidade específica para a qual foram obtidos. Para um estudo de tráfego, as informações pessoais detalhadas podem não ser estritamente necessárias e, portanto, não podem ser coletadas sem justificativa adequada.
- iii. **Privacidade e Proteção de Dados:** A LGPD visa proteger a privacidade dos indivíduos e estabelece diretrizes rigorosas para o tratamento de dados pessoais.

A coleta de informações de telefonia, especialmente em detalhes, pode ser considerada invasiva e, portanto, está sujeita a restrições significativas.

- iv. **Responsabilidade e Segurança:** A LGPD também impõe a responsabilidade das empresas na proteção dos dados que coletam. Isso inclui a implementação de medidas de segurança para proteger os dados contra acesso não autorizado ou vazamento. A obtenção de dados de telefonia sem os devidos cuidados de segurança violaria os princípios da LGPD.

Em resumo, a LGPD estabelece uma estrutura rigorosa para a proteção de dados pessoais e requer que as empresas obtenham o consentimento adequado e sigam os princípios de necessidade, privacidade, responsabilidade e segurança ao coletar e processar informações, o que torna necessário ter uma base legal e garantir o respeito aos direitos de privacidade dos indivíduos ao conduzir estudos de tráfego ou qualquer outra atividade de coleta de dados.

Nesse cenário, a classificação da matriz se deu através das pesquisas domiciliares, com a utilização das porcentagens de cada tipo de veículo obtidas nas pesquisas domiciliares.

Outro fator também ligado às diretrizes da LGPD, criou uma dificuldade na interpretação dos resultados, que foi a presença de valores "<10" no lugar da quantidade de dispositivos móveis em deslocamento. Essa informação indica que o valor real é menor do que 10 e, portanto, para preservar a privacidade e evitar a identificação individual, os valores exatos foram anonimizados ou truncados.

Ao realizar análises com essas matrizes, é comum substituir o valor "<10" por um valor adequado que não comprometa a privacidade e que seja estatisticamente razoável. Aqui estão algumas abordagens que podem ser consideradas na substituição:

- i. **Substituição Média ou Mediana:** Os valores "<10" podem ser substituídos pela média ou mediana dos valores conhecidos na matriz. Isso ajuda a manter uma estimativa razoável do fluxo sem introduzir informações específicas de identificação.
- ii. **Valor Mínimo:** Esses valores "<10" também podem ser substituídos pelo valor mínimo observado nos dados reais. Isso fornece uma estimativa conservadora e minimiza a possibilidade de superestimar o fluxo.
- iii. **Imputação Estatística:** Podem ser utilizadas técnicas de imputação estatística para preencher os valores faltantes. Métodos como regressão linear ou regressão

por média ponderada podem ser aplicados para prever os valores faltantes com base em padrões existentes.

- iv. **Randomização Controlada:** Se a preocupação com a privacidade ainda for um fator preponderante na análise, pode-se aplicar um processo de randomização controlada para ajustar os valores "<10". Isso envolve a introdução de alguma variação controlada nos valores, mantendo as características gerais da matriz.
- v. **Agrupamento:** Se os valores "<10" ocorrem em várias células próximas umas das outras, estes podem ser agrupados e substituídos pela média ou outro valor apropriado para a categoria.

A técnica escolhida foi a de substituição média para esse trabalho, no entanto, independentemente da abordagem escolhida, é importante documentar claramente as substituições realizadas e as razões por trás delas. Isso garante a transparência na análise e ajuda a interpretar corretamente os resultados. Além disso, ao utilizar essas técnicas, é necessário sempre considerar a natureza dos dados, o contexto da análise e as melhores práticas de privacidade.

O banco de dados recebido da empresa telefônica traz colunas representando o código de origem, de destino e a quantidade de dispositivos móveis do respectivo par OD, além da data de contagem. São sete contagens distintas, cada uma realizada em um dia da semana, em quatro diferentes arquivos, totalizando um mês de levantamento.

Para a utilização da técnica referida, foi necessário criar uma macro para ler a relação de dados de cada par OD e trazer a quantidade de valores numéricos presentes ao longo de todo o período de análise. Após isso, a macro realiza a soma dos valores e calcula sua média. Esse valor é utilizado na substituição dos valores “<10” presentes nas planilhas para o somatório total do fluxo de dispositivos.

A macro desenvolvida encontra-se a seguir:

```
Sub AnaliseZonas_Otimizada()
    Dim wsWeek1a As Worksheet
    Dim wsSoma As Worksheet
    Dim lastRow As Long
    Dim codigo As Variant
    Dim valor As Variant
    Dim dict As Object

    ' Defina as planilhas onde estão os dados
    Set wsWeek1a = ThisWorkbook.Sheets("Week1a")
    On Error Resume Next
```

```

Set wsSoma = ThisWorkbook.Sheets("Soma")
On Error GoTo 0

' Crie a guia "Soma" se não existir
If wsSoma Is Nothing Then
    Set wsSoma = ThisWorkbook.Sheets.Add(After:=wsWeek1a)
    wsSoma.Name = "Soma"
End If

' Encontre a última linha com dados na guia "Week1a"
lastRow = wsWeek1a.Cells(wsWeek1a.Rows.count, "K").End(xlUp).Row

' Carregar os dados da planilha em uma matriz
Dim data As Variant
data = wsWeek1a.Range("K2:L" & lastRow).Value

' Crie um dicionário para armazenar códigos com "<10"
Set dict = CreateObject("Scripting.Dictionary")

' Loop para identificar códigos com "<10"
For i = LBound(data, 1) To UBound(data, 1)
    codigo = data(i, 1)
    valor = data(i, 2)

    If Not dict.Exists(codigo) Then
        dict(codigo) = False
    End If

    If valor = "<10" Then
        dict(codigo) = True
    End If
Next i

' Preencha a guia "Soma" com códigos contendo "<10" e contagens de valores numéricos e médias
wsSoma.Cells(1, "A").Value = "Códigos com <10"
wsSoma.Cells(1, "B").Value = "Contagem de Valores Numéricos"
wsSoma.Cells(1, "C").Value = "Média de Valores Numéricos"
Dim rowSoma As Long
rowSoma = 2

For Each codigo In dict.Keys
    If dict(codigo) Then
        wsSoma.Cells(rowSoma, "A").Value = codigo
        wsSoma.Cells(rowSoma, "B").Value = ContarValoresNumericos(data, codigo)
        wsSoma.Cells(rowSoma, "C").Value = CalcularMediaValoresNumericos(data, codigo)

        ' Substitua valores zero por 1 na coluna C
        If wsSoma.Cells(rowSoma, "C").Value = 0 Then
            wsSoma.Cells(rowSoma, "C").Value = 1
        End If

        rowSoma = rowSoma + 1
    End If
Next codigo

Set dict = Nothing
End Sub

Function ContarValoresNumericos(data As Variant, codigo As Variant) As Long
    Dim count As Long

```

```

For i = LBound(data, 1) To UBound(data, 1)
    If data(i, 1) = codigo Then
        valor = data(i, 2)
        If IsNumeric(valor) And valor <> "<10" And valor <> "<1" Then
            count = count + 1
        End If
    End If
End If
Next i

    ContarValoresNumericos = count
End Function

Function CalcularMediaValoresNumericos(data As Variant, codigo As Variant) As Double
    Dim total As Double
    Dim count As Long

    For i = LBound(data, 1) To UBound(data, 1)
        If data(i, 1) = codigo Then
            valor = data(i, 2)
            If IsNumeric(valor) And valor <> "<10" And valor <> "<1" Then
                total = total + valor
                count = count + 1
            End If
        End If
    Next i

    If count > 0 Then
        CalcularMediaValoresNumericos = total / count
    Else
        CalcularMediaValoresNumericos = 0
    End If
End Function

```

Para pares OD onde não havia registros numéricos, apenas a informação “<10”, a contagem de valores numéricos ficou zerada e, em virtude disso, a média também. Portanto, nestes casos foi atribuído o valor igual a 1.

Para o próximo passo, foi necessário trazer cada uma das sete matrizes em extensão MXT, criadas com a partir das pesquisas domiciliares, e criar uma estrutura contendo todas as 76.729 possibilidades de relações entre as zonas. Para tal, outra macro foi construída realizar essa tarefa de montagem da estrutura, conforme exposto a seguir:

```

Sub GerarRelacoesDeDeslocamento()
    Dim ws As Worksheet
    Set ws = ThisWorkbook.Sheets.Add

    Dim bairros() As String
    Dim numBairros As Integer
    numBairros = 277
    ReDim bairros(1 To numBairros)

    Dim i As Integer, j As Integer
    For i = 1 To numBairros

```

```
    bairros(i) = i
Next i

Dim rowNum As Long
rowNum = 1

ws.Cells(rowNum, 1).Value = "Origem"
ws.Cells(rowNum, 2).Value = "Destino"
rowNum = rowNum + 1

For i = 1 To numBairros
    For j = 1 To numBairros
        ws.Cells(rowNum, 1).Value = bairros(i)
        ws.Cells(rowNum, 2).Value = bairros(j)
        rowNum = rowNum + 1
    Next j
Next i
End Sub
```

Esse passo foi necessário para encontrar as relações de tipos de veículo entre cada OD, ou seja, saber qual a porcentagem de motos, carros ou quaisquer outros tipos de veículos que existem em determinada relação entre duas zonas.

A Figura 19 mostra a interface do programa Microsoft Excel com uma parte dessa análise.

Origem	Destino	MO	VL	VON	VC	METRO	FERRY	BARCO	TOTAL	MO	VL	VON	VC	METRO	FERRY	BARCO	TOTAL
1	1	218,521	43,704	0,000	0,000	0,000	0,000	0,000	262,225	83,33%	16,67%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
1	2	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	3	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	4	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	5	0,000	43,704	0,000	0,000	0,000	0,000	0,000	43,704	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
1	6	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	7	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	8	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	9	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	11	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	12	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	13	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	14	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	15	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	16	0,000	43,704	0,000	0,000	0,000	0,000	0,000	43,704	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
1	17	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	18	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	19	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	20	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	21	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	22	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	23	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	24	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	25	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	26	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	27	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	28	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	29	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	30	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	31	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	32	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	33	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	34	87,408	349,634	34,000	0,000	0,000	0,000	0,000	471,042	18,56%	74,23%	7,22%	0,00%	0,00%	0,00%	0,00%	100,00%
1	35	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	36	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	37	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	38	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	39	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	40	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-	-	-	-	-	-	-	-
1	-	-	-	-	-	-	-	-
SOMA_TOTAL		OCUP_MO	OCUP_VL	OCUP_VON	OCUP_VC	OCUP_METRO	OCUP_FERRYBOAT	OCUP_BARCO_LANCHA	OCUP_TOTAL								

Fonte: APPE (2023) – Adaptado pelo autor.

Ao juntar as matrizes de ambos os métodos, constatou-se que a pesquisa por dados móveis traz uma quantidade muito superior de relações OD do que a pesquisa domiciliar, com mais viagens registradas. A Tabela 5 mostra a diferença ocorrida nos registros de viagens.

	PESQUISAS DOMICILIARES		PESQUISAS DADOS MÓVEIS	
PARES OD SEM REGISTROS	62.846	81,91%	5.596	7,29%
PARES OD COM REGISTRO	13.883	18,09%	71.133	92,71%
TOTAL:	76.729		76.729	

Fonte: Elaborado pelo autor

Analisando a Tabela 5, o banco de dados produzido por dispositivos móveis atinge 92,71% de pares OD com registro, contra apenas 18,09% da pesquisa domiciliar. Isso dificulta o processo de classificação pois, a diferença entre a quantidade de pares com registro entre as duas pesquisas, resulta em 57.520 pares de OD com viagens que não possuem classificação.

Dessa forma, para realização a classificação de pares OD sem dados domiciliares, foi decidido utilizar um **agrupamento não supervisionado**, ou seja, algoritmos de agrupamento, como o K-means, para agrupar os pares OD semelhantes com base em características como distância, tempo de viagem, características das vias, etc. Isso pode permitir que se identifique padrões e classes de viagens mesmo sem ter dados domiciliares para todas as viagens.

Em seguida, a determinação da proporção de viagens de diferentes modos de transporte (como motos, carros, ônibus, etc.), foi realizada por uma análise estatística dos dados. De acordo com o agrupamento realizado, foi calculada a proporção de viagens para cada tipo de veículo em relação ao número total de viagens, dentro do respectivo agrupamento. Isso pode ser feito dividindo o número de viagens de um tipo de veículo pelo número total de viagens.

O primeiro passo foi a criação de um algoritmo utilizando a biblioteca sklearn que, conforme mencionado na revisão bibliográfica deste trabalho, oferece uma gama de algoritmos e ferramentas para tarefas de aprendizado de máquina e mineração de dados. O K-means é um dos algoritmos de agrupamento incluídos na biblioteca scikit-learn.

A seguir é apresentado o algoritmo utilizado para esse agrupamento:

```

import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Carregar os dados da planilha Excel
data = pd.read_excel("D:\\APPE\\RMS_RMFS\\Matrizes - Dados
Móveis\\SOMA_Matrizes_out_values.xlsx")

# Selecionar as colunas relevantes para o agrupamento
features = data[['Distancia', 'Tempo', 'Velocidade']]

print("Length of data before preprocessing:", len(data))

# Reset the index after removing rows
features.reset_index(drop=True, inplace=True)

# Normalizar as features para média zero e desvio padrão um
scaler = StandardScaler()
scaled_features = scaler.fit_transform(features)

# Escolher o número de clusters desejado
num_clusters = 10

# Criar o modelo K-means
kmeans = KMeans(n_clusters=num_clusters, random_state=42)

# Treinar o modelo e obter os rótulos dos clusters
cluster_labels = kmeans.fit_predict(scaled_features)

# Adicionar os rótulos dos clusters a uma nova coluna 'Cluster'
features['Cluster'] = cluster_labels

# Visualizar os resultados
plt.scatter(data['Distancia'], data['Tempo'], c=data['Cluster'], cmap='rainbow')
plt.xlabel('Distância')
plt.ylabel('Tempo de Viagem')
plt.title('Agrupamento K-means')
plt.show()

# Salvar o resultado no arquivo Excel
result_file = " D:\\APPE\\RMS_RMFS\\Matrizes - Dados Móveis\\Classificado.xlsx"
data.to_excel(result_file, index=False)

print(f'Resultado da classificação salvo em {result_file}')

```

Após a divisão da planilha em 10 *clusters* (grupos), foi necessário calcular a média das porcentagens existentes em cada tipo de veículo. Para essa execução, foi criada uma macro que coleta os dados da planilha de entrada, calculando médias das porcentagens por *cluster* e exibindo esses resultados na planilha de saída, conforme exibido a seguir.

```

Sub GenerateSummarySheet()
    Dim inputSheet As Worksheet
    Dim outputSheet As Worksheet
    Dim lastRow As Long
    Dim clusterData As Object
    Dim clusterCode As Variant
    Dim percentages As Variant
    Dim i As Long, j As Long

    ' Set input and output sheets
    Set inputSheet = ThisWorkbook.Sheets("MATRIZ")
    Set outputSheet = ThisWorkbook.Sheets.Add(After:=inputSheet)
    outputSheet.Name = "Classificados"

    ' Initialize dictionary to store aggregated data
    Set clusterData = CreateObject("Scripting.Dictionary")

    ' Find the last row in the input sheet
    lastRow = inputSheet.Cells(inputSheet.Rows.Count, "G").End(xlUp).Row

    ' Iterate through rows to gather data
    For i = 2 To lastRow
        clusterCode = inputSheet.Cells(i, "G").Value
        percentages = Application.Index(inputSheet.Range("K" & i & ":Q" & i).Value, 1, 0)

        If Not clusterData.Exists(clusterCode) Then
            clusterData.Add clusterCode, Array(0, percentages)
        End If

        clusterData(clusterCode)(0) = clusterData(clusterCode)(0) + 1

        ' Handle the case where percentages is not an array
        If IsArray(percentages) Then
            For j = 1 To UBound(percentages, 2)
                clusterData(clusterCode)(1)(j) = clusterData(clusterCode)(1)(j) + percentages(1, j)
            Next j
        End If
    Next i

    ' Write headers
    outputSheet.Cells(1, 1).Value = "Cluster"
    For i = 1 To 7
        outputSheet.Cells(1, i + 1).Value = "Average " & inputSheet.Cells(1, 10 + i).Value
    Next i

    ' Write aggregated data
    i = 2
    For Each clusterCode In clusterData.Keys
        outputSheet.Cells(i, 1).Value = clusterCode

        If IsArray(clusterData(clusterCode)(1)) Then
            For j = 1 To 7
                outputSheet.Cells(i, j + 1).Value = clusterData(clusterCode)(1)(j) /
clusterData(clusterCode)(0)
            Next j
        End If

        i = i + 1
    Next clusterCode

```

```
' Apply cell formatting
outputSheet.Cells.HorizontalAlignment = xlCenter
```

```
MsgBox "Summary sheet 'Classificados' has been generated.", vbInformation
End Sub
```

O resultado dessa macro pode ser visualizado na Tabela 6.

CLUSTER	MO	VL	VON	VC	METRO	FERRY	BARCO
0	8,95%	51,37%	37,84%	0,39%	1,44%	0,00%	0,02%
1	4,55%	43,74%	44,47%	5,52%	0,00%	1,72%	0,00%
2	8,21%	63,84%	15,41%	12,19%	0,30%	0,00%	0,05%
3	9,22%	46,89%	41,20%	1,55%	1,10%	0,00%	0,04%
4	6,07%	64,65%	13,18%	15,77%	0,33%	0,00%	0,00%
5	15,16%	47,92%	33,29%	0,54%	0,44%	0,00%	2,64%
6	6,89%	59,64%	25,51%	7,97%	0,00%	0,00%	0,00%
7	7,70%	50,90%	40,04%	0,17%	1,19%	0,00%	0,00%
8	7,30%	67,14%	9,77%	15,79%	0,00%	0,00%	0,00%
9	9,77%	48609%	39,32%	0,70%	1,54%	0,00%	0,06%

Fonte: Elaborado pelo autor

Essas médias foram distribuídas ao longo dos 76.729 pares OD, de acordo com os respectivos grupos de *cluster*, para serem incorporadas ao total. No entanto, as colunas de fluxos pertencentes ao tipo Metrô, *Ferryboat* e Barcos, não foram incorporadas à planilha. A maior parte da malha não está habilitada para esses tipos de transporte, impossibilitando que certos fluxos fiquem restritos apenas a alguns pares OD.

Em virtude dessa questão, decidiu-se manter estes tipos de veículos restritos apenas a alguns pares e a redistribuição de dados de deslocamento foi realizada levando esse critério em consideração.

Por fim, juntamente com as porcentagens de classificação, foram aplicados os fatores de ocupação veicular, presente na Tabela 7, para transformar a quantidade de dispositivos móveis em quantidade de veículos.

MO	VL	VON	VC	METRO	FERRY	BARCO
1,1	1,7	15,37	1,16	80	50	5

Fonte: Elaborado pelo autor

Assim, foram gerados os sete arquivos em extensão MTX, para as matrizes providas no levantamento com dados móveis.

4.3. CONSTRUÇÃO DO MODELO

Conforme mencionado, o software utilizado para realizar o processo de modelagem foi o VISUM, da empresa alemã PTV. O processo de modelagem é um procedimento multifacetado que compreende várias etapas interdependentes, como criação de tipos de veículo, matrizes, importações de *shapefiles*, análises do modelo, sequência de procedimentos e geração de relatórios de consistência.

A seguir estão algumas etapas essenciais para a construção do modelo:

- i. Criação de Tipos de Veículo: Na seção "*Demand Data*", foram criados os diferentes tipos de veículos com suas características específicas, como capacidade, velocidade máxima, etc. Esses tipos de veículo foram usados para representar diferentes classes de veículos no modelo.
- ii. Criação de Matrizes: Após a construção dos arquivos MTX, as matrizes foram importadas diretamente no modelo pela ferramenta "*Read matrix contents from file*"
- iii. Conexão entre Matrizes e Tipos de Veículo: Depois da criação das matrizes, foi necessário conectá-las aos tipos de veículo definidos anteriormente na guia "*Demand*". Dessa forma, são associadas as demandas de viagens de cada tipo de veículo às matrizes correspondentes.
- iv. Importações de *Shapefiles*: Conforme está na sequência desse capítulo, para criar a representação da rede de transporte, os *shapefiles* da malha viária foram construídos e importados no *software*, bem como as zonas e pontos de contagem (*count locations*). Para tal, os *shapefiles* devem ser compatíveis com o VISUM e conter informações geoespaciais relevantes.
- v. Análise do Modelo através do "*Network Check*": Antes de prosseguir para os procedimentos de alocação, foi necessário realizar uma verificação da rede usando a função "*Network Check*" para garantir que não houvesse problemas na representação da rede, como nós desconectados ou segmentos de estrada faltantes.
- vi. Sequência de Procedimentos ("*Procedure Sequence*"): A sequência de procedimentos é usada para executar várias etapas de processamento de dados e cálculos no VISUM. Para este trabalho, foi criado um procedimento para a

combinação de matrizes e vetores, realizando um backup das matrizes antes dos processos de estimação.

- vii. Alocação: Foi realizada uma alocação estocástica, que é um método utilizado para distribuir a demanda de viagens das matrizes para a rede de transporte, levando em consideração a incerteza e a variabilidade na demanda.
- viii. Módulo TFlowFuzzy para Correções de Matrizes (Estimações): O módulo TFlowFuzzy é usado para realizar as correções nas matrizes, geralmente através de técnicas de estimação, como a estimação da matriz de tráfego a partir de dados observados. Esse algoritmo auxiliou no ajuste das matrizes para que pudessem refletir de forma mais precisa a realidade do tráfego.
- ix. Geração de Relatórios de Consistência pelo Módulo PrT *Survey Report*: Por fim, o módulo PrT *Survey Report* foi utilizado para gerar relatórios que analisassem a consistência dos resultados obtidos no modelo em relação aos dados observados. Essa ferramenta avaliou a qualidade e a precisão do modelo de tráfego, permitindo que melhorias pudessem ser realizadas.

A seguir estão descritas algumas etapas essenciais para a construção do modelo.

4.3.1. CONSTRUÇÃO DA MALHA VIÁRIA DA ÁREA DE ESTUDO

Com a necessidade de se obter uma rede detalhada, definiu-se como ponto de partida a importação das informações disponíveis na plataforma do Open Street Maps –OMS, que além de disponibilizar as feições das rodovias, trilhos e metrô, por exemplo, também dispõe os atributos específicos para cada segmento viário. Essa base de dados é um projeto colaborativo de dados geográfico editável e gratuito. Nessa plataforma, os usuários podem coletar dados usando levantamento manual, dispositivos GPS, fotografia aérea e outras fontes gratuitas.

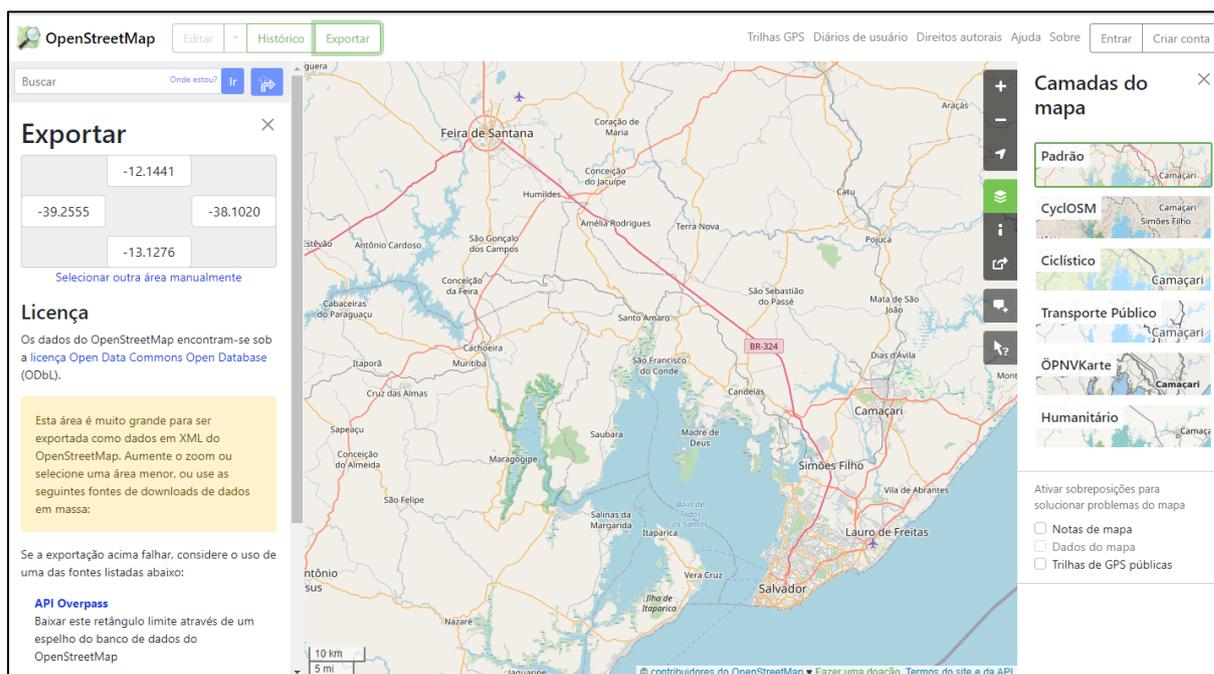
Os atributos citados estão relacionados às informações necessárias que o modelo precisa informar para que a alocação possa ser realizada, como velocidade de fluxo livre (V0), número de faixas, tipos de veículos que podem utilizar a ligação (TSysSet) e até mesmo o tipo característico do segmento (TypeNo). Utilizando-se desta alternativa o tempo despendido para construir, adequar e ajustar a rede é extremamente inferior se comparado a uma alternativa comum, como a importação de *shapefiles* do Sistema Nacional de Viação – SNV, Sistema

Rodoviário Estadual - SRE e levantamento e mapeamento de rodovias municipais de interesse para o projeto.

Os dados estão dispostos no site oficial do projeto (<https://www.openstreetmap.org/>), no qual foram pesquisadas as regiões metropolitanas em estudo.

Os dados do OSM podem ser usados de várias maneiras, incluindo a produção de mapas em papel e mapas eletrônicos, a geocodificação de endereços e nomes de lugares e o planejamento de rotas. Dessa forma, se torna possível atender a todas as demandas do estudo, pois muitos usuários de dispositivos GPS usam dados OSM para substituir os dados de mapa integrados em seus dispositivos.

No entanto, a importação dos dados no respectivo site possui um limite de 5.000 nós, impossibilitando o download de áreas que são demasiadamente grandes, como pode-se verificar na Figura 20.



Fonte: PrtScn de tela do website do Open Street Map

Assim, como a quantidade de nós necessária para o atual estudo excede o limite do sistema, não é possível utilizar a ferramenta do próprio website do Open Street Map. De forma a solucionar esse impasse, foi utilizada a biblioteca OSMnx para lidar com geodados do servidor e fazer o download dos dados no formato de um *shapefile*.

De acordo com a documentação oficial do OSMnx, esse módulo é um pacote Python que permite baixar dados geoespaciais do Open Street Map e modelar, projetar, visualizar e analisar redes de ruas do mundo real e quaisquer outras geometrias geoespaciais. Utilizando esse recurso, se torna possível baixar e trabalhar com outros tipos de infraestrutura, comodidades / pontos de interesse, áreas de construção, dados de elevação, rolamentos / orientações de ruas e velocidade / tempo de viagem.

A utilização desse recurso exige a instalação da respectiva biblioteca OSMnx, sendo também necessário escrever um código onde é construído um objeto instanciado [class GetOSMData() e def __init__(self)] e é inserida a função `ox.graph_from_place()`.

Ao fim do processo, se a variável interna do interpretador do Python for igual a "`__main__`", o programa pode executar o documento.

Dessa forma, basta alterar o nome do município (*Municipality name*) no código para que toda a base de dados da infraestrutura rodoviária do município escolhido seja baixada no formato *shapefile*, conforme mostrado no código a seguir:

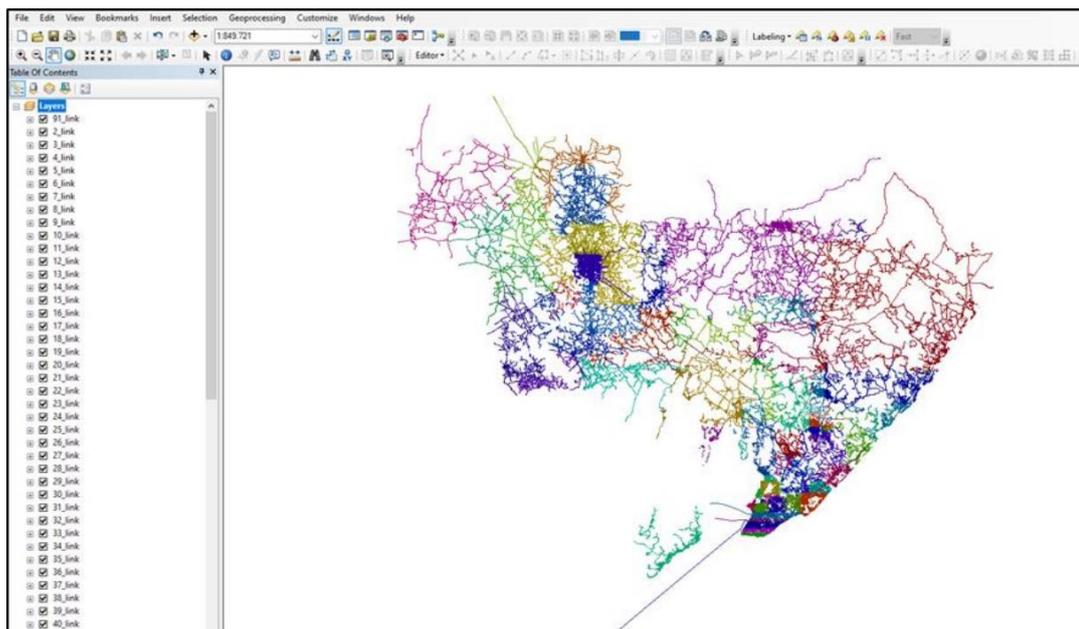
```
place_name = "Municipality name"  
graph = ox.graph_from_place(place_name)
```

Tal alternativa tornou o processo de montagem viável, bastando que fossem realizados os downloads das malhas de cada município de ambas as regiões metropolitanas, com a posterior junção em um único arquivo *shapefile*. Todavia, quando o trabalho estava sendo realizado foi constatado que, por mais que fosse possível importar todas as feições (*links*) da rede de uma única vez, as informações de atributos dessas feições não eram identificadas pelo software PTV Visum, fazendo com que a rede existisse, porém, sem informações necessárias à alocação.

Quando se importa um arquivo *.osm* no Visum, todas as configurações dos tipos de *links* são reconhecidas e montadas automaticamente, mas o download em *.shp* pelo Python faz com que uma gama de informações seja perdida.

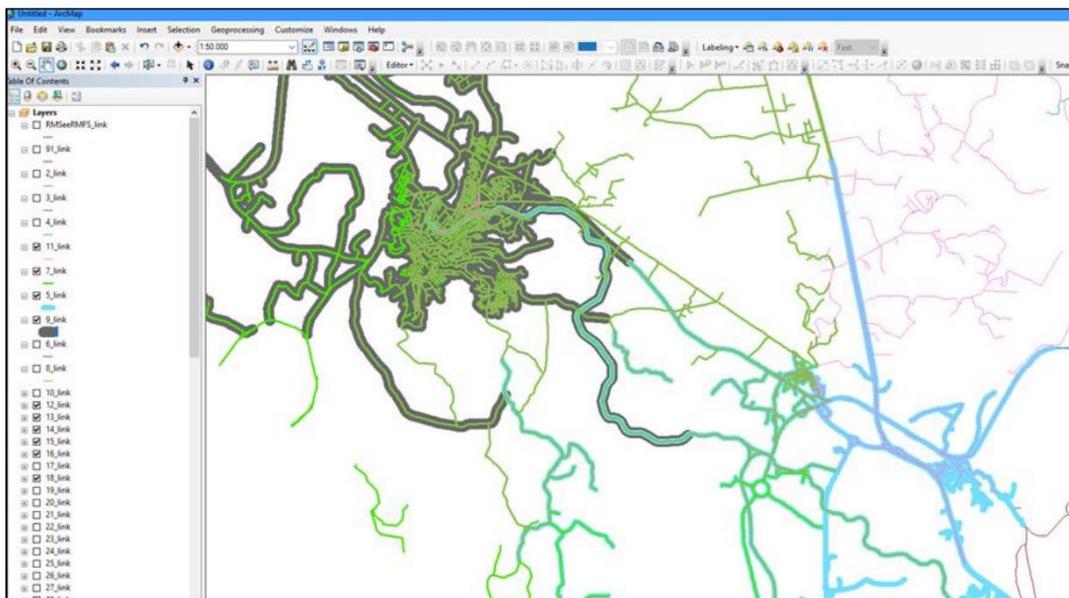
4.3.2. MOSAICO DE IMPORTAÇÃO – OSM

Para que a rede atendesse à área em estudo, foram necessárias importações de 91 áreas no PTV Visum e as devidas exportações em arquivos no formato *shapefile*, posteriormente acessados através do software ArcGIS, conforme apresentado na Figura 21.



Fonte: APPE (2022) - Adaptado pelo autor

Com todo o mosaico da rede no formato *shapefile* no ArcGIS, procedeu-se a união e criação de um *shapefile* único contendo todas as feições exportadas separadamente. Ainda assim, como as importações realizadas não tinham uma limitação bem definida de área, ocorreram uma série de sobreposições entre esses 91 arquivos, conforme se observa na Figura 22. Nota-se que na área analisada, há locais com sobreposição entre 4 arquivos (simbolizados como verde, cinza, azul e rosa)



Fonte: APPE (2022) - Adaptado pelo autor

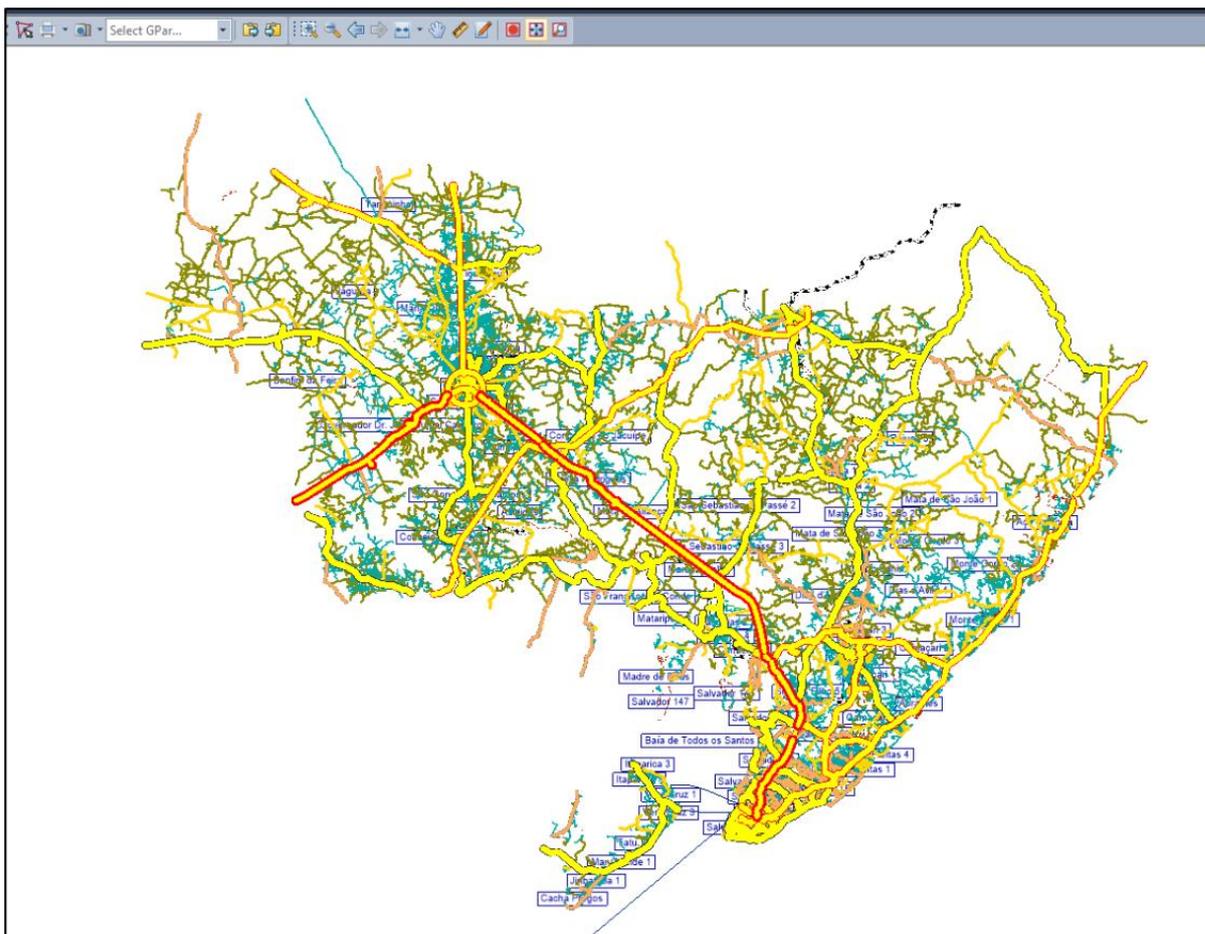
Após a união do mosaico em um único arquivo, foram identificadas muitas feições duplicadas, triplicadas e assim por diante. Antes da resolução desse impasse, criou-se um novo atributo nesse arquivo único denominado “GERAR”, identificando cada objeto por um código único de 0 a 519.313, indicando que o total de feições nesse arquivo único e ainda com sobreposições tinha um total de 519.313 *links*.

Para o problema das sobreposições, através do ArcGIS, especificamente com a ferramenta de geoprocessamento denominada “*Dissolve*”, foram agregadas as feições sobrepostas numa mesma ligação. No entanto, essa ferramenta, quando utilizada, gera um novo arquivo *shapefile* sem os atributos originais. Ou seja, resolvido o problema das sobreposições surgiu um outro, que é um arquivo com as feições corretas, porém sem as informações de atributos originais e necessárias ao modelo.

Para o novo obstáculo, como a única correlação mantida eram as próprias feições georreferenciadas, utilizou-se outra ferramenta do software ArcGIS, denominada “*Spatial Join*” com método de correlação “*Are Identical To*”. Dessa forma foi possível correlacionar o *shapefile* dissolvido (sem sobreposições e sem os atributos necessários) com o *shapefile* original (com sobreposições e com os atributos necessários), copiando então os atributos necessários para o arquivo corrigido.

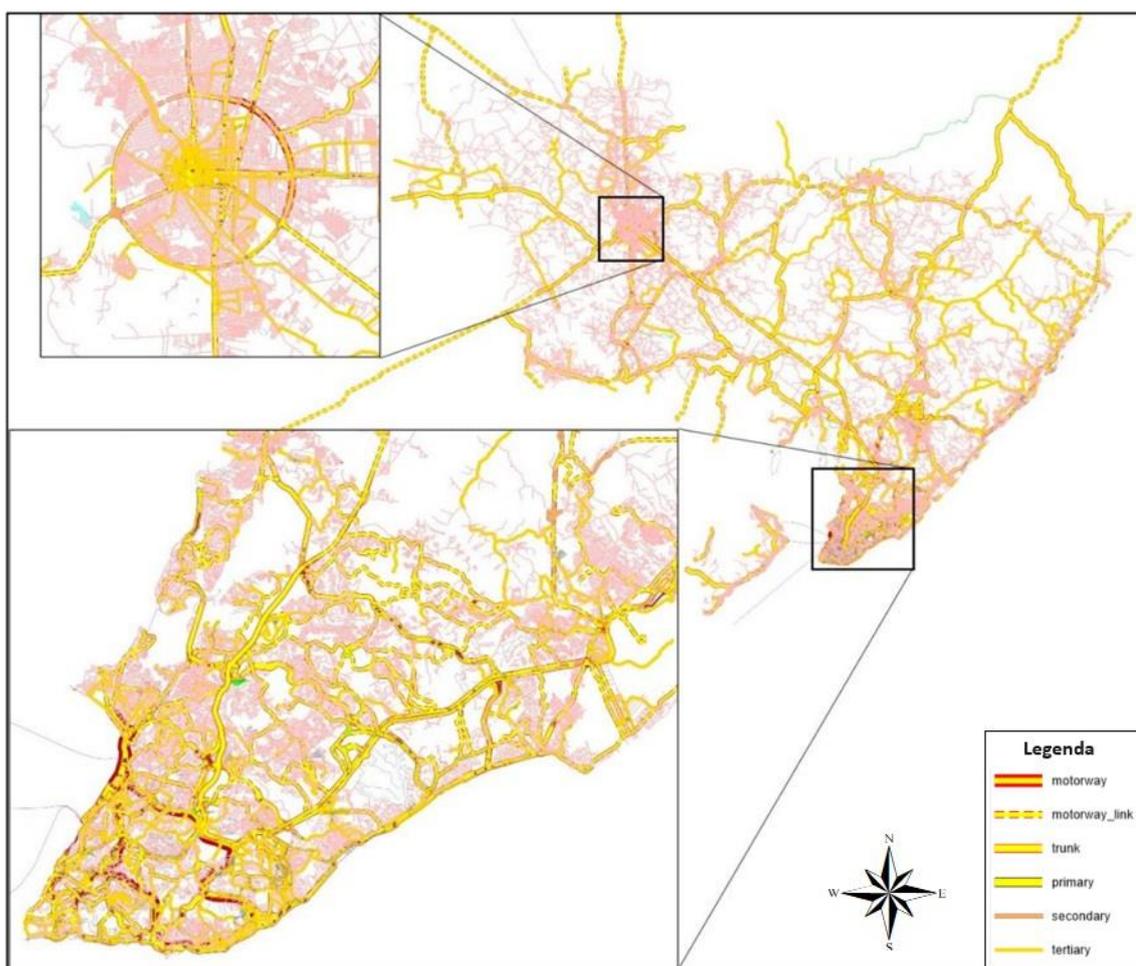
4.3.3. IMPORTAÇÃO DA REDE NO VISUM

Com os trabalhos anteriores finalizados, foi possível realizar a importação da rede para o PTV Visum, conforme mostra a Figura 23.



Fonte: APPE (2022) - Adaptado pelo autor

Com a importação realizada foi possível realizar outras análises internas no modelo. Verificou-se que todos os atributos foram atendidos após as correções explicadas nos itens anteriores, com exceção dos sentidos de fluxo da via. Para a solução desse impasse foi necessária a verificação de todos os *links* e cruzamentos, um procedimento que levou 11 meses, resultando em uma rede que representa fielmente a realidade, conforme exposto na Figura 24.



Fonte: APPE (2022) – Adaptado pelo autor

Após a importação, também foi possível dar início a ligação das Zonas de Tráfego (*centroids*) à rede no PTV Visum, através dos conectores (*Connectors*).

4.3.4. VALIDAÇÃO DE CAMINHO MÍNIMO

Posteriormente, utilizando a ferramenta de rotas do Google Maps, foi possível efetuar processos de validação do modelo através de caminhos mínimos, realizando ajustes finos na malha para que a mesma representasse fielmente o comportamento de variados fluxos, entre diferentes pontos de início e fim. Esse procedimento baseou-se na escolha aleatória de pares OD no VISUM (220 testes efetuados), com a definição desses pontos no Google Maps, para alocação de uma matriz nula com inserção de fluxo apenas entre esse par OD.

A validação de caminho mínimo é um processo essencial em análises de transporte e modelagens de redes. Ela envolve a determinação do caminho mais curto entre dois pontos específicos em uma região, com base em informações de custo, distância, tempo de viagem ou qualquer outra métrica relevante. Quando se utiliza matrizes unitárias em uma modelagem, significa que cada elemento da matriz representa uma única conexão direta entre dois pontos, indicando a disponibilidade ou capacidade de movimento entre esses pontos. Isso é comumente aplicado em sistemas de transporte, redes rodoviárias ou em qualquer contexto em que o conceito de distância ou custo seja relevante.

O processo de validação de caminho mínimo com matrizes unitárias envolveu os seguintes passos:

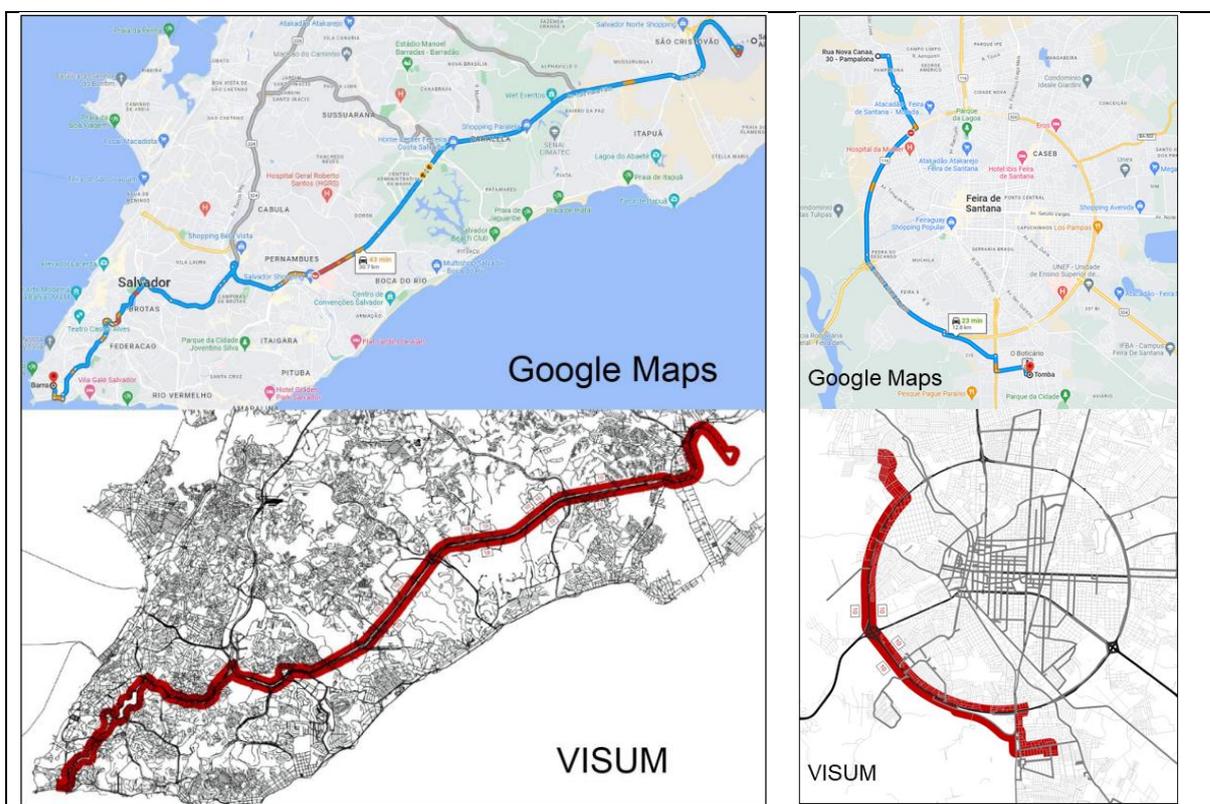
- i. **Preparação dos Dados:** Primeiramente, criou-se uma matriz que representasse as conexões diretas entre os pontos de origem e destino pré-definidos. Cada célula da matriz indicará a presença (ou ausência) de uma conexão direta, muitas vezes usando valores binários (0 para ausência e 1 para presença).
- ii. **Escolha da Métrica:** Definiu-se a métrica pela qual deseja encontrar o caminho mínimo, podendo ser distância, tempo de viagem, custo, ou qualquer outra métrica apropriada para a análise.
- iii. **Algoritmos de Caminho Mínimo:** Foram utilizados algoritmos como *Dijkstra*, *Bellman-Ford* ou A^* para encontrar o caminho mínimo entre os pontos de origem e destino. Esses algoritmos consideraram as conexões disponíveis (representadas pela matriz unitária) e a métrica escolhida para otimizar a rota.
- iv. **Validação e Verificação:** Após calcular o caminho mínimo, validou-se e verificou-se os resultados. Para isso foi necessário se certificar de que os resultados fazem sentido em relação à modelagem e contexto. Além disso, foi

verificado se o caminho mínimo encontrado é realmente o mais curto ou mais eficiente, de acordo com a métrica escolhida.

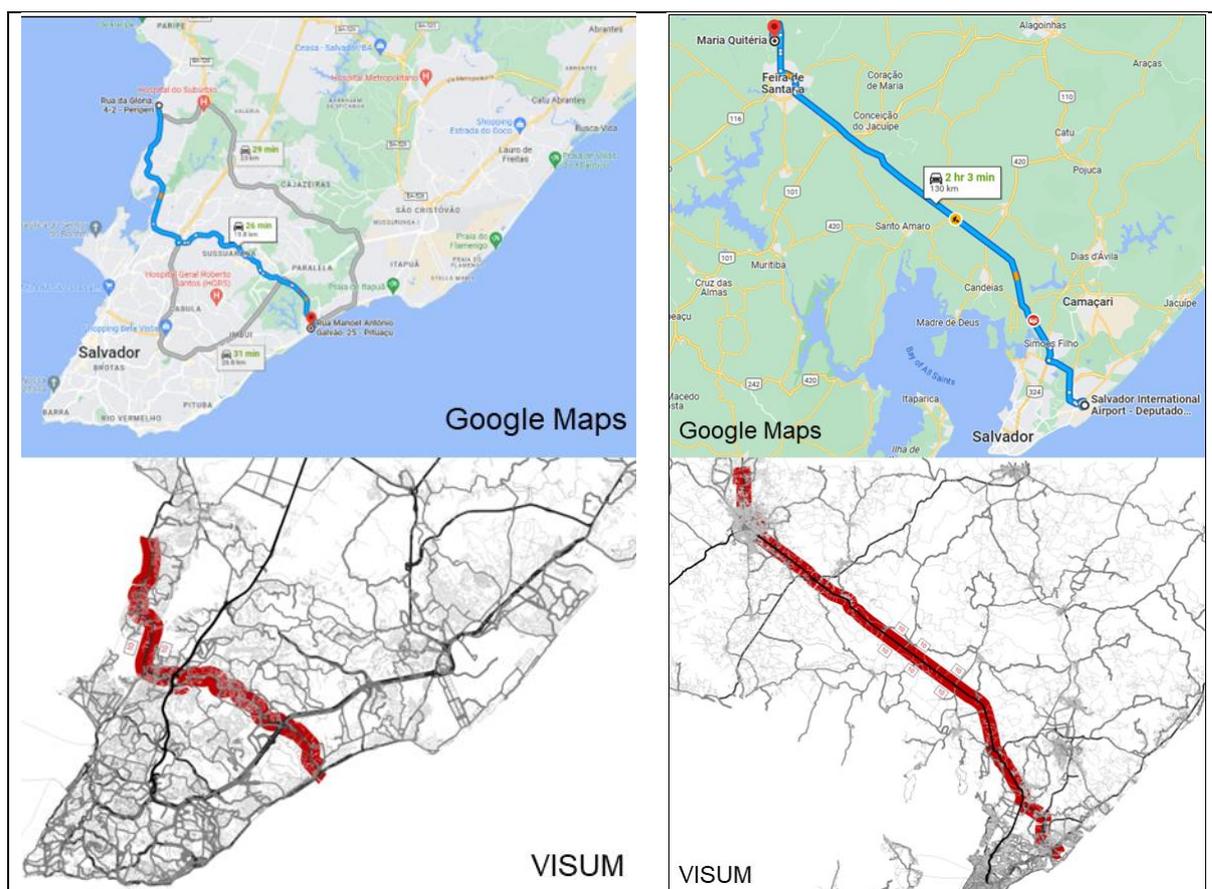
- v. **Análise de Sensibilidade:** Ao fim, foram realizadas análises de sensibilidade, variando os parâmetros ou pesos atribuídos às conexões, para entender como pequenas mudanças afetavam os resultados e a escolha do caminho mínimo.

A utilização de matrizes unitárias simplifica o processo de validação de caminho mínimo, uma vez que as conexões diretas são modeladas de maneira clara e direta. No entanto, é importante lembrar que a representação precisa e atualizada das conexões é crucial para garantir resultados confiáveis. Além disso, a escolha adequada da métrica e a seleção do algoritmo apropriado são fatores chave para obter insights relevantes e úteis para a tomada de decisões em planejamento de transporte e logística.

O processo pode ser adaptado para a utilização do Google Maps, uma ferramenta para encontrar rotas e caminhos mínimos, como apresentado na Figura 25 e Figura 26.



Fonte: APPE (2023) – Adaptado pelo autor



Fonte: APPE (2023) – Adaptado pelo autor

4.4. ALOCAÇÃO DE TRÁFEGO

4.4.1. ASPECTOS GERAIS

Neste capítulo, serão expostos os aspectos gerais da alocação de tráfego realizada com o software VISUM. Estão listados cinco pontos de ajustes que são essenciais na alocação de tráfego e na obtenção de resultados válidos. Estes pontos incluem o limite máximo de iterações, o coeficiente de impedância, o método de média sucessiva para "*smoothed impedances*", o modelo de escolha Kirchhoff e a definição de independência com base no tempo de deslocamento, conforme calculado pela metodologia de Cascetta.

- i. **Limite máximo de iterações:** No contexto da alocação de tráfego com o software VISUM, é importante definir um limite máximo de iterações como condição de término global. Neste estudo, optou-se por utilizar 5 iterações como limite máximo. Esse valor foi escolhido com base na complexidade do sistema

e na convergência dos resultados, garantindo um equilíbrio entre a precisão da alocação e o tempo de processamento.

- ii. **Coefficiente de Impedância:** Neste estudo, foi adotado um coeficiente de 1,00 como parte do processo de alocação. O coeficiente de impedância afeta a sensibilidade da alocação às condições de tráfego e às características da rede. Um valor de 1,00 indica que a alocação considera as impedâncias conforme fornecidas, sem amplificar ou reduzir seu efeito.
- iii. **Método de Média Sucessiva para "Smoothed Impedances":** Para este estudo, decidiu-se adotar o método de média sucessiva para suavizar as impedâncias. Esse método visa reduzir flutuações abruptas nas impedâncias da rede, tornando a alocação mais estável e realista.
- iv. **Modelo de Escolha Kirchhoff:** Para este trabalho, o beta foi ajustado para 4 como parte do modelo Kirchhoff. Esse valor determina a sensibilidade dos usuários à variação nas impedâncias da rede e desempenha um papel fundamental na distribuição do tráfego pelos diferentes caminhos disponíveis.
- v. **Definição de Independência com Base no Tempo de Deslocamento:** A independência na alocação de tráfego é definida pelo tempo de deslocamento, e seu cálculo segue a metodologia de Cascetta. Esta abordagem leva em consideração a precisão na estimativa dos tempos de viagem e, portanto, é essencial para garantir que os resultados refletem de forma precisa as escolhas dos usuários com base em seus tempos de deslocamento esperados.

4.4.2. PROCESSO DE SIMPLIFICAÇÃO DO MODELO

Inicialmente, em virtude da possibilidade de iniciar um trabalho com as linhas de transporte coletivo e, da mesma forma, pensando em futuros trabalhos na região em níveis microscópicos (para utilização em VISSIM), foi decidido manter a malha em uma complexidade elevada, com uma quantidade aproximada de 300.000 *links*. Porém, essa quantidade de *links*, somado ao fato de as matrizes possuírem um volume elevado de dados, fez com que a primeira tentativa de alocação chegasse a 24 horas e 50min de processo, sem nenhum outro tipo de módulo como TFlowFuzzy para realização de estimações. Esse tempo inviabiliza o trabalho, pois um processo envolvendo outros módulos poderia chegar a quase uma semana de processamento.

Além das condições citadas, esse tempo gasto na alocação das matrizes pode envolver diversos outros parâmetros como as características do tráfego, das viagens, dos veículos, das condições das vias, entre outros. Ao longo do processo de construção do modelo, ao invés de utilizar apenas as características dos tipos de *links* cadastrados (*strict*), foram adotadas características específicas em diversos *links*, em virtude da complexidade das vias do município de Salvador. Dessa forma, quanto mais complexo for o modelo, mais tempo levará para processar todas as interações e cálculos necessários.

A simulação de redes grandes requer uma quantidade considerável de recursos de hardware, como CPU, memória RAM e capacidade de armazenamento. O processo de alocação de matrizes envolve algoritmos de otimização que procuram encontrar os melhores valores para os parâmetros do modelo. Em uma simulação iterativa, como é comum em simulações de tráfego, cada iteração envolve recalculer os fluxos de tráfego e as condições da rede. Quanto mais iterações forem necessárias para que o sistema alcance a estabilidade ou a convergência, mais tempo a simulação levará.

Portanto, como o modelo utilizado é muito detalhado, com muitos parâmetros finos, a simulação exige um processamento adicional para calcular todos esses detalhes, resultando em um prolongamento do tempo de simulação.

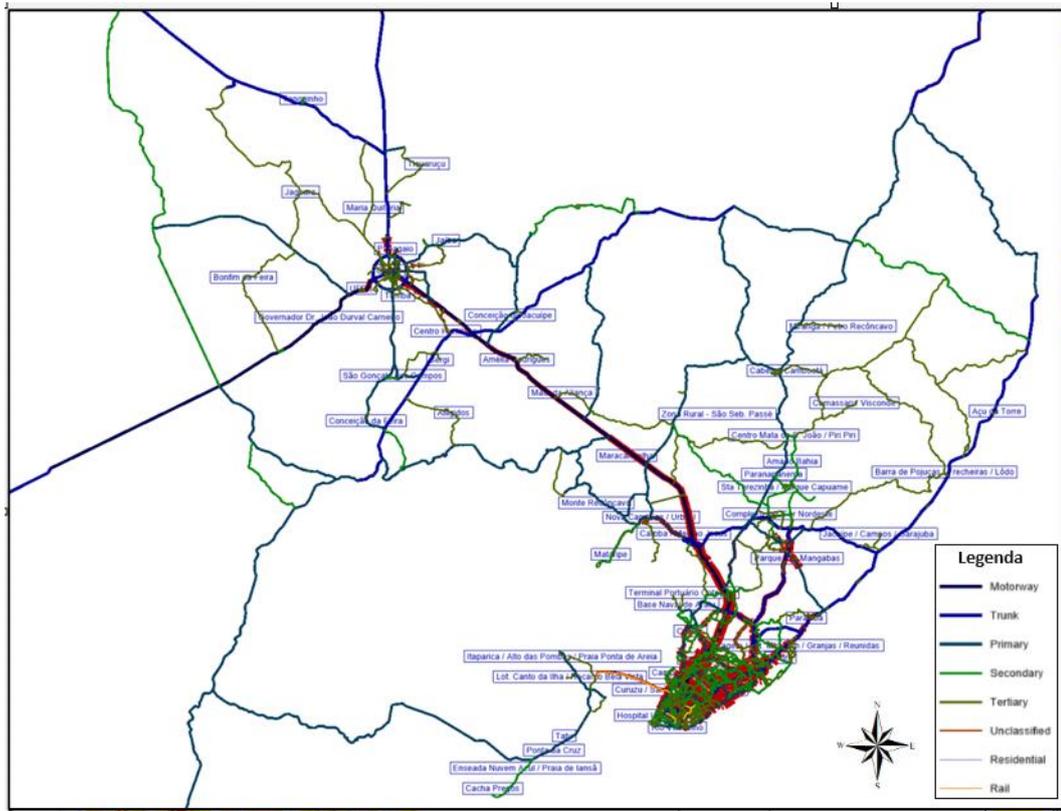
Dessa forma, foram analisadas algumas possibilidades para melhorar esse tempo de processamento:

- i. Simplificação do modelo, removendo detalhes menos relevantes.
 - a. Foram removidos todas os *links* de travessas, ruas residenciais, linhas ferroviárias e linhas marítimas, deixando apenas o *ferryboat* entre Salvador e Ilha de Itaparica. Dessa forma, o modelo ficou reduzido a cerca de 48 mil *links*.
 - b. Foram diminuídas as condições de tráfego, considerando dados agregados através da ferramenta *strict* ao invés de detalhes por veículo, ou seja, utilizando padrões de características pré-definidas.
 - c. Em relação a quantidade de tipos de veículo, optou-se por diminuir a quantidade, retirando primeiramente veículos ferroviários e marítimos, ficando apenas com os veículos rodoviários, os quais possuem pontos de controle para estimação e análise de consistência.
- ii. Haveria uma opção de utilizar um hardware mais potente, porém essa opção se tornou inválida no momento em virtude de restrições financeiras e, ao mesmo

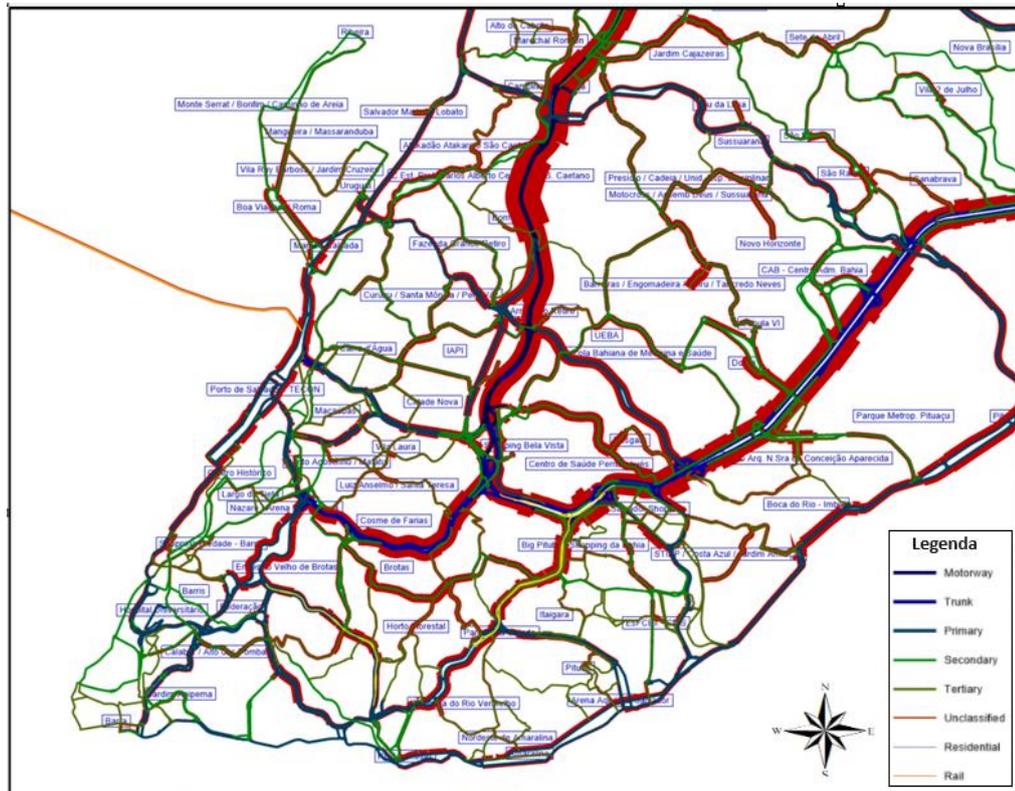
tempo, a máquina utilizada já possui um i7 de oitava geração com 20Gb de RAM.

- iii. Ajuste dos parâmetros da simulação para reduzir a complexidade.
 - a. Optou-se por uma redução da quantidade de iterações, para diminuir o número de cálculos pesados, passando de 10 (*default*) para 5. Foram sendo feitas alocações em cada diminuição e o resultado não apresentou alterações significativas até esse valor.
 - b. Houve um ajuste da resolução temporal através da utilização de intervalos de tempo maiores para a simulação, reduzindo a quantidade de cálculos necessários.
 - c. Haveria também a opção da simplificação dos algoritmos de otimização, pela substituição por algoritmos de otimização mais simples, no entanto, essa opção foi descartada, dado que isso significaria uma busca menos precisa.
- iv. Por fim, poderiam ser utilizadas outras técnicas para aproveitar vários recursos computacionais. Porém, para esse recurso estariam disponíveis duas técnicas: paralelização e distribuição de carga, as quais se mostraram inviáveis no momento.
 - a. Paralelização: Seria necessário dividir a simulação em partes menores e executar cada parte em paralelo usando múltiplos núcleos de CPU. Isso poderia acelerar significativamente o tempo de processamento, mas não havia essa disponibilidade de recursos.
 - b. Distribuição de carga: Tendo acesso a vários computadores, haveria a possibilidade de distribuir partes da simulação em diferentes máquinas e utilize um sistema de coordenação para consolidar os resultados. Porém, em virtude de só haver uma chave do VISUM, essa opção também foi descartada.

Na Figura 27 e na Figura 28 apresenta-se a versão final da redução da rede, a qual foi usada para a elaboração deste trabalho. Posteriormente, a Tabela 8 mostra a diminuição do tempo de acordo com as alterações realizadas no modelo.



Fonte: Elaborado pelo autor



Fonte: Elaborado pelo autor

MÉTODO	TEMPO DE EXECUÇÃO				
	Original	Rede menor v01 Iteração=10	Rede menor v02 Iteração=10	Rede menor v04 Iteração=5	Rede menor v09 Iteração=5
	309.550 links	48.124 links	43.618 links	43.618 links	34.583 links
TRADICIONAL	42min 38s	10min 3s	7min 00s	4min 17s	2min 21s
DADOS MÓVEIS	24h 40min 20s	10h 57min 20s	-	-	50min 07s
Comparação:	Inicial	4,2x menor	6,1x menor	9,9x menor	18,1x menor

Fonte: Elaborado pelo autor

4.4.3. PROCESSO DE ESTIMAÇÃO DAS MATRIZES

A consistência dos dados é crucial para garantir que a simulação de tráfego seja representativa do comportamento real do tráfego em uma área específica. Os ajustes de *links* e nós foram realizados ao longo dos últimos meses para alinhar os dados da simulação com os dados reais coletados. Os ajustes envolveram a calibração de parâmetros, a correção de fluxos de tráfego e a otimização da geometria da rede.

A simulação, os valores eram analisados e os trechos de estrada com maiores discrepâncias entre os dados simulados e reais eram identificados. Repetidamente, eram realizados ajustes nos parâmetros de capacidade, velocidade e tempo de percurso para melhor refletir as condições reais. Também eram realizadas avaliações nos postos de contagem de tráfego e seus respectivos dados para calibrar os volumes nos *links*.

Primeiramente, foram utilizados módulos de combinação de matrizes para criar um *backup* das matrizes originais de ambos os métodos analisados. Depois foi feita uma alocação estocástica de todas as matrizes para, na sequência, aplicar os módulos de correção de demanda em cima de cada matriz, através das referências importadas e inseridas nos *Count Locations*, utilizando o algoritmo TFlowFuzzy. Esse algoritmo faz a estimação das matrizes, realizando alterações nos valores dos pares OD, para que o fluxo presente nos *links* possa se aproximar dos valores referenciais.

Por fim, foi realizada uma nova alocação, trazendo os valores ajustados nas etapas anteriores para a malha viária, com o qual foi possível obter um R^2 maior, indicando uma melhor concordância entre os resultados simulados e os dados reais.

5. ANÁLISE COMPARATIVA

No contexto da pesquisa científica, os métodos de análises comparativas desempenham um papel fundamental na avaliação, compreensão e contextualização de fenômenos complexos e multifacetados. Esses métodos proporcionam uma estrutura sistemática para contrastar e relacionar variáveis, teorias ou abordagens, permitindo a identificação de semelhanças, diferenças e padrões subjacentes. Ao empregar técnicas analíticas como estudos de caso, análises de conteúdo, revisões sistemáticas e sínteses qualitativas ou quantitativas, os pesquisadores podem discernir relações de causa e efeito, delinear tendências evolutivas e destacar implicações significativas. A escolha criteriosa dos métodos de análises comparativas, aliada a uma abordagem metodológica robusta, não apenas fortalece a validade e a confiabilidade das conclusões da pesquisa, mas também enriquece a base de conhecimento em diversas áreas do saber, ao fornecer insights substanciais para a tomada de decisões informadas e o avanço do conhecimento científico.

Com base nessa compreensão sobre métodos comparativos, esse trabalho se propôs a observar os resultados das alocações de ambos os métodos e expor as análises realizadas através de dados majoritariamente quantitativos. Dessa forma, ao longo do desenvolvimento deste estudo, verificou-se que a modelagem de tráfego utilizando matrizes provindas de pesquisas OD (Origem-Destino) e matrizes provindas de dados móveis são abordagens distintas com vantagens e limitações específicas.

Na análise realizada com essas duas abordagens, assim como era esperado, de acordo com a bibliografia analisada, foi possível chegar às seguintes interpretações:

5.1. MATRIZES PROVINDAS DE PESQUISAS OD

➤ **Vantagens:**

- **Detalhamento dos fluxos de tráfego:** Essas pesquisas permitiram capturar informações detalhadas sobre as origens e destinos das viagens, bem como as rotas e modos de transporte utilizados pelos usuários (Ben-Akiva & Lerman, 1985). Ao utilizar o método de construção de matrizes OD, foi possível obter um panorama abrangente e preciso dos padrões de deslocamento, identificando as principais rotas de viagem, os pontos de congestionamento e as áreas com alta demanda (Cascetta, 2009).

- **Informações precisas:** A coleta direta de dados sobre as viagens realizadas pelos usuários de transporte trouxe características individuais como origem, destino, modo de transporte utilizado e propósito da viagem (Vovsha & Petersen, 2011). Essa abordagem detalhada e específica resultou em matrizes OD mais precisas, que refletem de forma mais acurada os fluxos de tráfego e as demandas de mobilidade em uma região específica (Ortúzar & Willumsen, 2011).
- **Representatividade da demanda:** Mesmo com uma amostra menos abrangente (quando comparada com a coleta de dados de dispositivos móveis), foi possível obter uma coleta mais detalhada de dados, a qual se constitui como uma amostra representativa da população e suas viagens diárias (Stopher & Greaves, 2007). Ao utilizar o método de construção de matrizes OD, é possível obter uma visão mais precisa e detalhada da demanda de transporte, levando em consideração as diferentes origens, destinos e padrões de viagem dos usuários (Bhat & Koppelman, 1999). Isso resulta em matrizes OD que melhor representam a diversidade da demanda, permitindo uma análise mais completa e uma base sólida para o planejamento e a tomada de decisões no campo do tráfego.

➤ **Limitações:**

- **Custo e tempo:** Essas pesquisas requerem recursos consideráveis, como equipes de campo, equipamentos de coleta de dados e análise especializada, o que pode resultar em altos custos operacionais (Vovsha & Petersen, 2011). Além disso, o tempo necessário para realizar as pesquisas OD, desde o planejamento até a coleta e análise dos dados, pode ser extenso, demandando meses ou até mesmo anos para obter resultados abrangentes (Stopher & Greaves, 2007). Essas limitações financeiras e temporais podem ser um desafio significativo, especialmente em projetos com recursos limitados ou prazos apertados.
- **Atualidade dos dados:** As pesquisas OD requerem um esforço considerável em termos de tempo e recursos, o que resulta em um intervalo de tempo entre a coleta e a disponibilidade dos dados para análise (Knoop & Axhausen, 2019). Esse atraso pode levar a uma representação

desatualizada dos padrões de demanda de viagens, uma vez que o comportamento dos usuários de transporte pode evoluir ao longo do tempo, influenciado por fatores como mudanças socioeconômicas, desenvolvimento urbano e alterações nas políticas de transporte (Badoe & Miller, 2000). Portanto, é importante considerar essa limitação ao interpretar e utilizar as matrizes OD para tomar decisões relacionadas ao planejamento e gerenciamento do tráfego.

- **Amostragem:** A construção de matrizes de tráfego a partir de pesquisas origem-destino (OD) apresenta limitações em relação à amostragem utilizada. As pesquisas OD exigem uma seleção representativa das zonas de origem e destino, bem como dos indivíduos ou veículos a serem entrevistados, a fim de obter resultados confiáveis e generalizáveis. No entanto, a amostragem pode ser um desafio complexo devido à variabilidade espacial e temporal da demanda de viagens, bem como à diversidade dos padrões de deslocamento da população (Bhat & Koppelman, 1999). A escolha adequada dos pontos de coleta de dados e a determinação do tamanho da amostra são questões críticas para garantir a representatividade dos resultados obtidos por meio das pesquisas OD.

5.2. MATRIZES PROVINDAS DE DADOS MÓVEIS

➤ **Vantagens:**

- **Cobertura abrangente:** Diferentemente das pesquisas origem-destino (OD) tradicionais, que dependem de amostragens pontuais, os dados móveis são gerados a partir de fontes como dispositivos GPS, sensores de veículos e redes de telecomunicações, proporcionando uma cobertura mais ampla e contínua (Calabrese et al., 2015). Esses dados capturam automaticamente informações sobre os deslocamentos dos usuários em tempo real, abrangendo uma variedade de modos de transporte e rotas (Zheng et al., 2014). Com base nisso, as matrizes construídas a partir de dados móveis oferecem uma visão mais abrangente dos padrões de tráfego, permitindo uma compreensão mais precisa e detalhada do comportamento

dos usuários e do desempenho do sistema de transporte (Zheng et al., 2010).

- **Atualidade dos dados:** Como citado, ao contrário das pesquisas tradicionais de origem-destino (OD), que podem ter intervalos de coleta espaçados no tempo, os dados móveis fornecem informações em tempo real e contínuas sobre os deslocamentos dos usuários (Jiang et al., 2016). Essa natureza dinâmica dos dados móveis permite uma atualização frequente das matrizes de tráfego, refletindo os padrões de demanda em constante evolução (Jiang et al., 2016; Liu et al., 2018). Com essa atualização em tempo real, as matrizes construídas a partir de dados móveis oferecem uma visão mais precisa e atualizada do fluxo de tráfego, permitindo uma tomada de decisão mais informada e a implementação de estratégias de gerenciamento de tráfego mais eficazes (Liu et al., 2018).
- **Custo e eficiência:** Ao utilizar dados móveis, como registros de GPS de veículos ou dados de redes de telecomunicações, é possível obter informações de deslocamento de forma automática e em grande escala, eliminando a necessidade de realizar pesquisas de campo tradicionais, que são dispendiosas e demoradas (Zheng et al., 2014; Calabrese et al., 2015). Além disso, a coleta contínua e em tempo real dos dados móveis permite uma atualização frequente das matrizes de tráfego, proporcionando informações atualizadas sobre os padrões de tráfego e a demanda de transporte (Liu et al., 2018). Essa abordagem mais eficiente em termos de custo e tempo torna as matrizes de tráfego construídas a partir de dados móveis uma alternativa viável e econômica para o planejamento e gerenciamento do tráfego.

➤ **Limitações:**

- **Privacidade e confidencialidade:** A utilização de dados móveis, como registros de GPS e informações de localização, pode levantar preocupações éticas e legais devido à natureza sensível das informações pessoais envolvidas (Elmahmudi et al., 2019). O acesso e o uso desses dados requerem estritas medidas de segurança e proteção da privacidade,

a fim de garantir que as informações pessoais dos usuários sejam devidamente anonimizadas e protegidas contra uso indevido (Figueiras et al., 2021). A privacidade e a confidencialidade são questões críticas que devem ser abordadas ao utilizar dados móveis para a construção de matrizes de tráfego, garantindo o respeito aos direitos individuais dos usuários e a conformidade com as regulamentações de privacidade vigentes.

- **Representatividade:** Embora os dados móveis forneçam informações sobre os deslocamentos dos usuários em tempo real, é importante considerar que esses dados podem não ser totalmente representativos de toda a população (Calabrese et al., 2015). Diferentes grupos demográficos e segmentos da população podem ter diferentes padrões de mobilidade e acesso a dispositivos móveis, o que pode resultar em uma sub-representação de certas áreas geográficas ou grupos específicos de usuários (Bieker-Walz et al., 2018). Portanto, ao utilizar matrizes de tráfego construídas a partir de dados móveis, é necessário ter cuidado na interpretação dos resultados, considerando suas limitações em termos de representatividade da amostra.
- **Limitações de precisão:** Embora os dados móveis forneçam uma cobertura abrangente e em tempo real dos deslocamentos dos usuários, a precisão das informações pode ser influenciada por fatores como erros de localização, inconsistências nos registros de GPS e sub-representação de certos grupos demográficos (Zheng et al., 2014; Calabrese et al., 2015), sem falar na dificuldade da classificação das matrizes produzidas. Além disso, a natureza voluntária dos dados móveis pode introduzir vieses, pois nem todos os usuários participam ativamente da coleta de dados. Essas limitações podem afetar a representatividade dos padrões de tráfego e introduzir distorções nos resultados obtidos a partir das matrizes de tráfego construídas com dados móveis (Bieker-Walz et al., 2018). Portanto, é importante considerar essas limitações ao interpretar e utilizar as matrizes de tráfego derivadas de dados móveis para garantir uma análise precisa e robusta.

A combinação de ambas as abordagens pode oferecer benefícios complementares, com a modelagem de tráfego podendo se beneficiar da gama de informações das pesquisas OD, especialmente para entender os padrões de viagem e as características socioeconômicas dos usuários. Ao mesmo tempo, os dados móveis podem fornecer uma visão em tempo real e abrangente do tráfego, permitindo uma análise mais atualizada e eficiente.

De acordo com as análises realizadas, verificou-se uma diferença na ordem de nove vezes em relação ao tempo necessário para os procedimentos de definições, coleta e preparação dos dados. Enquanto as pesquisas domiciliares precisaram de quase dezoito meses para serem plenamente concluídas, todas as etapas do processo por dados móveis levaram menos de dois meses.

Essa diferença também ficou evidente no que tange os custos, com as pesquisas domiciliares chegando a ser 22,9 vezes mais caras do que o método de dados móveis.

A Tabela 3 traz essa análise comparativa de valores e custos entre os dois métodos apresentados nesse trabalho. Em virtude da necessidade de preservar informações confidenciais, não foram trazidos valores monetários. Dessa forma, foram trazidas informações comparativas e custos em termos percentuais do trabalho desenvolvido.

ITEM	MÉTODO TRADICIONAL	MÉTODO DADOS MÓVEIS	COMPARAÇÃO
TEMPO	18 meses	2 meses	9x menor
CVC (Linha de Contorno)	5,34% do valor total	Não há em dados móveis	-
PESQUISA OD (Linha de Contorno)	3,60% do valor total	Não há em dados móveis	-
CVC (Linha de Travessia) AUTOMÁTICA	2,78% do valor total	Não há em dados móveis	-
CVC (Linha de Travessia) MANUAL	14,52% do valor total	Não há em dados móveis	-
PESQUISA OD (Linha de Travessia)	1,80% do valor total	Não há em dados móveis	-
PESQUISA OV (Linha de Travessia) AUTOMÁTICA	1,35% do valor total	Não há em dados móveis	-
PESQUISA OV (Linha de Travessia) MANUAL	15,96% do valor total	Não há em dados móveis	-
PESQUISA OD DOMICILIAR	54,65% do valor total	Não há em dados móveis	-
BIG DATA - DISPOSITIVOS MÓVEIS	Não há em método tradicional	100% do valor total	-
MÃO DE OBRA NÃO ESPECIALIZADA	36 pessoas	Não há em dados móveis	-
PROFISSIONAIS ESPECIALIZADOS	5 profissionais	3 profissionais	1,7x menor
ZONAS		277	-
PARES OD		76.729	-
PARES OD COM REGISTROS	13.883	71.133	5,1x maior
DIAS LEVANTADOS	1	28	28x maior
QTDE DE DESLOCAMENTOS	4.032.625	569.732.807	141x maior
TOTAL DO CUSTO:			23x menor
CUSTO POR DESLOCAMENTO: (R\$/desloc)			3.231x menor

Fonte: elaborado pelo autor

*CVC: Contagem Volumétrica Classificatória

* Pesquisa OD: Pesquisa Origem/Destino

* Pesquisa OV: Pesquisa de Ocupação Visual

5.3. RELATÓRIOS DE CONSISTÊNCIA

Através do módulo do *Survey Report*, foi possível gerar os relatórios de consistência de ambos os cenários, com as matrizes provindas das pesquisas tradicionais e de dados móveis.

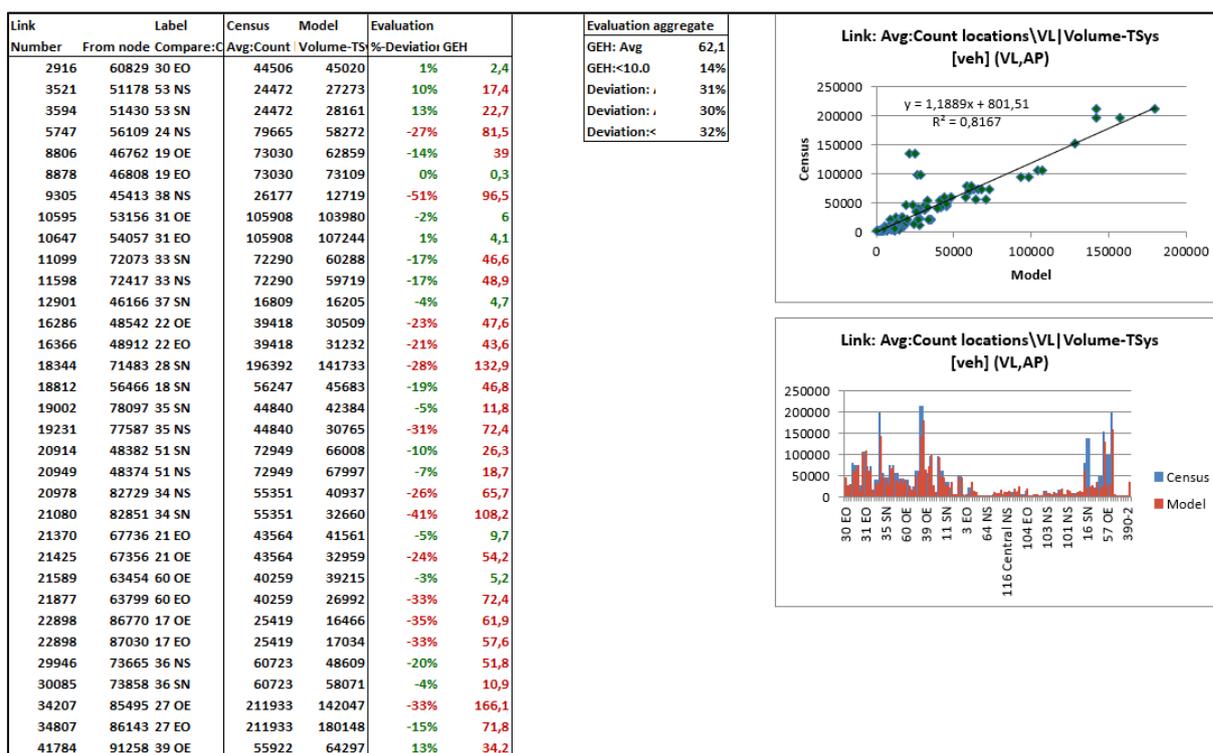
Basicamente, o diagnóstico realizado nos relatórios de consistência segue uma sequência lógica, de acordo com os três itens aqui apresentados:

- i. **R²:** Fornece insights sobre o grau de ajuste ou consistência entre os valores alocados e os valores observados em diferentes cenários ou alocações. O R² é uma medida estatística que varia de 0 a 1 e indica o quão bem os valores observados são explicados pelo modelo ou alocação.
- ii. **Dispersão:** Análise do quanto os valores alocados em relação aos valores observados variam entre diferentes cenários ou alocações. Um baixo percentual de desvio indica que os valores alocados estão próximos dos valores observados, enquanto um alto percentual de desvio indica que há uma diferença significativa entre eles.
- iii. **GEH:** Avaliação da consistência entre valores alocados e observados, levando em consideração o erro geométrico harmônico (GEH). O *GEH Statistic* é uma medida estatística que ajuda a determinar o grau de desvio entre os valores alocados e observados, levando em consideração o erro relativo e a dispersão.

Conforme apresentado na Figura 29, primeiramente confere-se o gráfico da equação da reta para identificar os outliers que distorcem o valor de R², com os valores do modelo no eixo x e, no eixo y, os valores obtidos nos pontos de controle. Nos resultados finais deste trabalho, ao invés de trabalhar com os valores de 0 a 1, o padrão utilizado está em porcentagem.

Logo após, através da inserção de um filtro na coluna de desvio (*Evaluation % Deviation*), verificam-se os postos de controle que estão com valores discrepantes. No referido estudo, definiu-se um valor limite máximo de 20%.

Por fim, analisam-se os valores de GEH para compreender se são necessários refinamentos dos dados de entrada, ajustes de modelos de comportamento, validação de parâmetros de simulação, comparação com dados de campo ou, inclusive, iterações e reajustes com o ajuste gradual de alguns parâmetros e modelos.



Fonte: elaborado pelo autor

Os relatórios foram gerados apenas para os tipos de veículos rodoviários, meio no qual há a disponibilidade de pontos de controle, o que possibilita a análise. Portanto, cada relatório é composto por cinco guias de análises, sendo motos, veículos leves, ônibus, caminhões e uma análise do volume total da via.

São apresentadas aqui duas análises comparativas diferentes. Na primeira estão contidos os resultados de alocações sem o processo de estimação, no intuito de verificar o comportamento das matrizes de demanda (ou matrizes de tráfego de entrada). A segunda análise se refere a consistência das matrizes que já foram ajustadas pelo processo de estimação.

A Tabela 10 e a Tabela 11 apresentam comparações em relação ao tempo de execução e a consistência expressa em R^2 obtida ao final do processo de alocação. É interessante frisar que o intuito deste estudo é realizar comparações entre os métodos, tanto em relação às questões envolvendo o processo construtivo das matrizes quanto no que se refere às consistências obtidas. Dessa forma, devido às restrições temporais e a natureza comparativa do processo, não foi investido um tempo maior para realizar análises mais rigorosas (em *links* e pontos de controle) no intuito de obter uma consistência perto de 90%. Como pode ser verificado a seguir, com a utilização de apenas uma rodada de estimação foi possível chegar perto de 80% e, com este valor, as comparações já apresentaram os dados necessários para a conclusão desse estudo.

A Tabela 10 apresenta a comparação dos valores entre as matrizes originais, sem terem passado pelo processo de estimação com o algoritmo TFlowFuzzy. Nota-se que o tempo de processamento é muito maior, chegando a 22,69 vezes o tempo da tradicional. Isso ocorre em virtude de o método tradicional possuir apenas 13.883 pares OD, enquanto o método por dados móveis gerou matrizes que atingiram 51.495 pares OD. No rodapé da tabela, a comparação é sempre expressa por um resultado do valor presente no método de dados móveis dividido pelo valor do método tradicional.

MÉTODO	TEMPO EXECUÇÃO	MO	VL	VON	VC	VMDA
TRADICIONAL	2 min 21s	12,74%	39,30%	39,55%	28,35%	34,24%
DADOS MÓVEIS	53 min 20s	24,35%	32,97%	29,21%	20,38%	32,97%
Comparação:	22,69 x maior	1,91x maior	0,84x menor	0,74x menor	0,72x menor	0,96x menor

Fonte: elaborado pelo autor

No entanto, conforme pôde ser visualizado na Tabela 10, em quase todos os tipos de veículos os valores de consistência se mantiveram parecidos, com o valor total resultando em uma diferença de apenas 1,27% menor para os dados móveis.

A Tabela 11 apresenta a comparação entre os resultados obtidos com a alocação realizada juntamente com o processo de ajuste das matrizes. É possível verificar que os tempos de execução são maiores em virtude da adição de 4 etapas de algoritmos de estimação, além de uma nova alocação necessária para o processo.

MÉTODO	TEMPO EXECUÇÃO	MO	VL	VON	VC	VMDA
TRADICIONAL	6 min 59s	46,52%	77,95%	75,69%	68,87%	75,83%
DADOS MÓVEIS	1h 37min 53s	61,33%	80,82%	82,01%	73,96%	79,42%
Comparação:	14,02x maior	1,32x maior	1,04x maior	1,08x maior	1,07x maior	1,05x maior

Fonte: elaborado pelo autor

A análise comparativa entre as matrizes de tráfego derivadas de dados móveis e aquelas construídas a partir de pesquisas domiciliares demonstrou que ambas as abordagens

resultaram em matrizes com consistências similares. Observou-se que a variação de melhora das consistências foi de 1,05x a 1,32x quando utilizando matrizes de dados móveis, ou seja, invertendo a situação inicial. Essa análise aponta que os resultados obtidos são comparáveis em termos de consistência.

Uma das investigações realizadas neste estudo envolveu a redução do número de iterações de 10 para 5 vezes durante a estimação das matrizes de tráfego. Notavelmente, os resultados indicaram que essa redução não gerou impacto negativo na consistência das matrizes, contrariando expectativas iniciais. A constatação de que a redução de iterações não resultou em uma queda de consistência é relevante para a eficiência dos processos de estimação, sugerindo que é possível alcançar resultados confiáveis com uma carga computacional menor.

Por fim, verificou-se que, embora o modelo baseado em dados móveis apresente um valor de consistência ligeiramente menor inicialmente, após apenas uma rodada ele atinge uma consistência maior que o modelo tradicional. Isso ocorre porque a base de pares OD é cinco vezes maior. Dessa forma, se houvesse um número maior de pontos de controle, o modelo em dados móveis provavelmente apresentaria uma consistência melhor. No mais, verificou-se também que o número de rodadas para a estimação das matrizes também é menor, pois o fluxo alocado na malha viária é mais bem distribuído, facilitando o trabalho do algoritmo.

As conclusões deste estudo têm implicações diretas para o campo do planejamento de transporte e tráfego. Ao reconhecer que as matrizes de tráfego provenientes de dados móveis e pesquisas domiciliares oferecem consistências comparáveis, os profissionais do setor podem escolher entre essas fontes com maior confiança, baseando sua escolha em fatores como disponibilidade, custo e escopo do projeto.

Além disso, a compreensão de que uma redução nas iterações não prejudica a consistência das matrizes de tráfego permite uma economia de recursos computacionais e tempo durante os processos de estimação, o que pode ser particularmente útil em cenários de simulações de grande escala ou de prazos apertados.

Dessa forma, apesar das conclusões abrangentes obtidas neste estudo, algumas áreas permanecem abertas para futuras pesquisas. Seria valioso investigar como diferentes cenários urbanos ou tipos de rede podem afetar as relações entre matrizes de dados móveis e matrizes de pesquisas domiciliares. Além disso, a análise de outros indicadores além da consistência, como precisão em diferentes horários ou origens/destinos específicos, pode fornecer insights adicionais sobre a aplicabilidade prática dessas abordagens.

6. CONCLUSÕES E RECOMENDAÇÕES

6.1. CONSIDERAÇÕES FINAIS

Neste trabalho, verificou-se que, a análise comparativa entre os métodos de construção de matrizes de Origem-Destino (OD), por meio de pesquisas domiciliares e de dados móveis revelou uma série de considerações para aprimorar a eficiência e a abrangência desses processos fundamentais para o planejamento urbano e de transportes. Os resultados evidenciaram claramente que a pesquisa domiciliar demanda um tempo e um custo significativamente maior e envolve uma quantidade considerável de mão de obra, tanto especializada quanto não especializada, em comparação com a coleta de dados móveis.

Embora as pesquisas domiciliares ofereçam uma riqueza de informações detalhadas sobre a população, permitindo uma compreensão mais profunda das características socioeconômicas e de comportamento de viagem, os dados móveis apresentam uma vantagem em termos de escala e representatividade. A capacidade de capturar milhões de viagens em comparação com as dezenas de milhares obtidas pelas pesquisas domiciliares fornece um panorama mais amplo e preciso do padrão de deslocamento da população.

No entanto, é inegável que a pesquisa domiciliar desempenha um papel crucial ao fornecer informações detalhadas que enriquecem a compreensão das viagens individuais, suas motivações e características demográficas. Portanto, a combinação de ambos os métodos surge como a abordagem mais promissora, permitindo a criação de matrizes OD que não apenas representam a escala das viagens com precisão, mas também incorporam as nuances e particularidades capturadas pelas pesquisas domiciliares.

Além disso, a estratégia de atualização contínua dos valores das matrizes com base nas pesquisas de dados móveis oferece uma maneira otimizada de manter as informações sempre relevantes e atualizadas. Essa abordagem híbrida permitirá que os planejadores urbanos e de transportes tomem decisões embasadas em dados robustos e abrangentes, levando em consideração tanto a escala das viagens quanto as características individuais que influenciam os padrões de mobilidade.

Em suma, respondendo ao questionamento norteador apresentado no capítulo introdutório deste trabalho, esta pesquisa ressalta a importância de uma abordagem flexível e adaptativa na construção de matrizes OD, reconhecendo os pontos fortes e limitações de diferentes métodos. A combinação e integração desses métodos, aliada à atualização regular

dos dados, contribuirão para um modelo de planejamento com potencial para moldar um sistema de transportes que atenda de maneira mais precisa e abrangente às necessidades da população.

A combinação dessas abordagens distintas oferece benefícios complementares. A modelagem de tráfego pode se beneficiar da riqueza de informações das pesquisas OD, especialmente para entender os padrões de viagem e as características socioeconômicas dos usuários. Ao mesmo tempo, os dados móveis podem fornecer uma visão em tempo real e abrangente do tráfego, permitindo uma análise mais atualizada e eficiente.

A escolha entre utilizar matrizes provindas de pesquisas OD ou dados móveis dependerá dos objetivos específicos da análise, da disponibilidade dos dados e dos recursos disponíveis. Em muitos casos, a combinação dessas abordagens pode ser a solução mais adequada para obter uma compreensão abrangente e precisa do tráfego em uma determinada área, auxiliando na otimização de recursos e no planejamento de transporte de forma mais eficiente e eficaz.

6.2. RECOMENDAÇÕES PARA TRABALHOS POSTERIORES

6.2.1. PLACOM – PLATAFORMA DE COMPATIBILIZAÇÃO DE MATRIZES

Para que esse trabalho possa se tornar relevante em âmbito profissional, recomenda-se a continuidade dos estudos para a concepção e desenvolvimento de uma plataforma que utilize dados móveis para complementar matrizes OD, originalmente construídas por meio de pesquisas domiciliares. A proposta visa a otimização dos recursos, permitindo a obtenção de informações detalhadas e atualizadas para análises de demanda de transporte.

Para o pleno êxito do projeto, a elaboração dessa plataforma precisaria seguir os seguintes passos:

- i. Fundamentação da Plataforma:** Para desenvolver a plataforma de complementação de matrizes OD, é crucial estabelecer uma base sólida que integre os seguintes aspectos:

- **Modelo Conceitual:** Definir um modelo que estruture a complementação de matrizes OD, onde as matrizes construídas por pesquisas domiciliares sirvam como base, permitindo a integração de dados móveis e pesquisas de contagem volumétrica classificatória. Isso garantirá que a plataforma forneça resultados confiáveis e consistentes.
 - **Coleta e Processamento de Dados:** Implementar processos eficientes para a coleta, processamento e integração de dados móveis. Isso inclui a definição de critérios de qualidade, tratamento de *outliers* e técnicas de agregação para garantir que os dados móveis sejam incorporados de maneira precisa nas matrizes.
- ii. **Estrutura da Plataforma:** A estrutura da plataforma deve ser projetada de forma a aproveitar os pontos-chave das matrizes de pesquisa domiciliar enquanto incorpora os benefícios dos dados móveis. Recomenda-se:
- **Base de Dados Unificada:** Estabelecer uma base de dados unificada que armazene as matrizes de pesquisa domiciliar, os dados móveis e as pesquisas de contagem volumétrica classificatória. Isso permitirá que a plataforma acesse e integre as informações de maneira eficiente.
 - **Mecanismo de Atualização Contínua:** Desenvolver um mecanismo de atualização que permita a incorporação regular dos dados móveis à plataforma. Isso garantirá que as matrizes sejam atualizadas ao longo do tempo, capturando as mudanças nas tendências de viagem.
- iii. **Utilização Inteligente de Dados Móveis:** Para maximizar a utilidade dos dados móveis na complementação das matrizes, recomenda-se:
- **Análise de Padrões de Viagem:** Realizar análises detalhadas dos padrões de viagem capturados pelos dados móveis. Identificar os principais corredores de tráfego, horários de pico e modos de transporte predominantes.
 - **Estimativas de Viagens Faltantes:** Utilizar algoritmos e técnicas de modelagem estatística para estimar viagens faltantes nas matrizes de

pesquisa domiciliar com base nos dados móveis. Isso contribuirá para uma representação mais completa do fluxo de viagens.

iv. Validação e Avaliação: A plataforma deve ser submetida a um rigoroso processo de validação e avaliação. Recomenda-se:

- **Comparação com Dados Reais:** Realizar comparações entre os resultados da plataforma e dados reais de tráfego para avaliar sua precisão e confiabilidade.
- **Avaliação de Desempenho:** Medir o desempenho da plataforma em termos de tempo de processamento, qualidade das estimativas e eficácia na atualização das matrizes.

v. Interface de Usuário Intuitiva: Desenvolver uma interface de usuário intuitiva e amigável é essencial para a adoção e utilização eficaz da plataforma. A interface deve permitir:

- **Visualização de Resultados:** Fornecer meios para visualizar as matrizes resultantes e comparar as variações ao longo do tempo.
- **Configuração Flexível:** Permitir aos usuários configurarem parâmetros de complementação de acordo com as necessidades específicas de cada análise.

A construção da **PLACOM** (Plataforma de Complementação de Matrizes OD) utilizando dados móveis como fonte de complementação é uma iniciativa para otimizar recursos e obter resultados mais atualizados. A estruturação do sistema através de dados mais detalhados com as matrizes de pesquisa domiciliar e a integração inteligente de dados móveis e pesquisas de contagem volumétrica classificatória permitirão a criação de um ambiente analítico mais dinâmico. A implementação dessas recomendações garantirá que a plataforma otimize uma série de atividades para profissionais envolvidos em análises de demanda de transporte.

6.2.2. PESQUISAS OD – ONLINE

Uma outra recomendação para futuras análises seria a adoção de pesquisas OD de forma online. Esse método pioneiro vem sendo realizado em Recife, no estado de Pernambuco, a partir do ano de 2015. O modelo possui uma periodicidade de 2 anos, atingindo 84mil pessoas em sua primeira edição e mais de 150mil pessoas na segunda edição. O seu formato garante uma abrangência espacial, faixas de renda necessárias para validação específica da amostra e ainda possui um menor custo em comparação com o método tradicional de pesquisas domiciliares.

Foi desenvolvido um software de validação que é programado para executar as etapas de correção, verificando respostas incompletas, incoerentes ou em duplicidade. Esse programa também faz a extrapolação/expansão dos dados, de acordo com o atendimento aos estratos de renda da população e todas as zonas de tráfego, garantindo a representatividade do resultado.

A participação do formulário online é obrigatória para empreendimentos geradores de viagens, com penalidades caso haja o descumprimento, como impossibilidade de renovação de alvarás. Ao fim de cada formulário, o funcionário do respectivo estabelecimento recebe um código, o qual deve ser apresentado ao seu empregador.

Da mesma forma, há outros tipos de pesquisas baseadas em *crowdsourcing*, ou seja, pesquisas colaborativas. Diversas empresas, governos ou diferentes tipos de organização, oferecem aos cidadãos participantes benefícios em forma de descontos em impostos, ingressos de cinema, teatro, ou quaisquer outros tipos de incentivo, para participarem das pesquisas.

Portanto, uma análise adicional, poderia levar em consideração a unificação desse tipo de pesquisa para a construção da base estrutural da PLACOM, com as posteriores atualizações com as matrizes provindas de dispositivos móveis.

REFERÊNCIAS

- Abdel-Aty, M., Kitamura, R., & Jovanis, P. P. (1996). Analysis of traffic data to identify vehicle occupancy levels for HOV facilities. *Transportation Research Record*, 1554(1), 98-103.
- Badoe, D. A., & Miller, E. J. (2000). Transportation-land-use interaction: empirical findings in North America, and their implications for modeling. *Transportation Research Part D: Transport and Environment*, 5(4), 235-263.
- Bhat, C. R., & Koppelman, F. S. (1999). A conceptual framework of individual activity program generation. *Transportation Research Part A: Policy and Practice*, 33(5), 357-383.
- Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.
- Ben-Akiva, M., Morikawa, T., & Koutsopoulos, H. (1996). Calibration of traffic simulation models. *Transportation Research Part B: Methodological*, 30(6), 473-487.
- Bieker-Walz, L., Kaltenbrunner, A., & Helbing, D. (2018). The truth is in the data? On the limits and opportunities of using mobile phone data to understand human behavior. *Journal of the Royal Society Interface*, 15(140), 20180086.
- Bishop, Christopher M. (2006). "Pattern Recognition and Machine Learning." Editora Springer.
- Bliemer, M. C., & Raadsen, M. P. (2010). Improving the efficiency of the estimation of origin-destination matrices. *Transportation Research Part C: Emerging Technologies*, 18(3), 382-391.
- BRASIL. Bahia. Lei Complementar Estadual nº 35, de 2011. Lei Orgânica do Estado da Bahia. *Diário Oficial do Estado, Bahia*, 06 de julho de 2011.
- BRASIL. Lei Complementar nº 14, de 1973. Lei Orgânica do Ministério Público da União. *Diário Oficial da União, Brasília*, 8 de junho de 1973.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Bwambale, A., Choudhury, C.F., Hess, S. et al. Getting the best of both worlds: a framework for combining disaggregate travel survey data and aggregate mobile phone data for trip generation modelling. *Transportation* 48, 2287–2314 (2021).
- Cáceres, N., Benítez, F. G., & Romero, L. M. (2019). Land use inference from mobility mobile phone data and household travel surveys. *ScienceDirect*.
- Caceres, N., Romero, L. M., & Benitez, F. G. (2020). Exploring strengths and weaknesses of mobility inference from mobile phone data vs. travel surveys. *Taylor & Francis Online*.

- Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr, J., Ratti, C., & Sevtsuk, A. (2015). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 58, 252-262.
- Cascetta, E. (2009). *Transportation Systems Engineering: Theory and Methods*. Springer.
- Castro, J., Sobrino, J. A., & Llamas, M. (2014). Mining origin-destination matrices from bike-sharing systems. *Expert Systems with Applications*, 41(5), 1971-1980.
- Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3), 199-219.
- Chauhan, K., & Rana, S. (2019). Analysing the dynamics of urban mobility using mobile data: A case study of Delhi. *Journal of Transport Geography*, 77, 78-89.
- Chao Yang, Yuliang Zhang, Xianyuan Zhan, Satish V. Ukkusuri, Yifan Chen, "Fusing Mobile Phone and Travel Survey Data to Model Urban Activity Dynamics", *Journal of Advanced Transportation*, vol. 2020, Article ID 5321385, 17 pages, 2020.
- Chen, S., Zhang, J., & Wang, F. Y. (2011). Classification of vehicle types based on uncalibrated traffic cameras. *IEEE Transactions on Intelligent Transportation Systems*, 12(2), 594-605.
- Chen, Y., Li, X., & He, X. (2017). Hierarchical iterative Bayesian IPF for predicting OD demand of large urban areas. *Transportation Research Part C: Emerging Technologies*, 81, 350-370.
- Chen, Z., Pang, J., Liu, Y., Wang, F., & Guo, J. (2012). A hybrid model for travel mode identification using GPS data. *Journal of Geographical Systems*, 14(4), 463-477.
- Chiu, Y. C., Chen, C. H., & Su, T. H. (2011). Trip origin-destination matrix estimation using GPS-enabled mobile phones. *Transportation Research Part C: Emerging Technologies*, 19(4), 766-780.
- Cox, D. R. "Role of Models in Statistical Analysis." *Statistical Science*, vol. 5, no. 2, 1990, pp. 169-74. JSTOR, <http://www.jstor.org/stable/2245676>. Accessed 23 Mar. 2023.
- Du, H., & Wang, D. (2021). Analysis of urban mobility patterns using mobile phone data: A case study in Beijing. *Transportation Research Part C: Emerging Technologies*, 124, 103134.
- Elmahmudi, A., Zhang, J., & Venkatachalam, S. (2019). Big data privacy in intelligent transportation systems: State-of-the-art and future research directions. *IEEE Access*, 7, 8688-8711.
- Figueiras, P., Teixeira, C., & Nunes, F. (2021). Location privacy in mobile networks: A systematic literature review. *Future Generation Computer Systems*, 118, 517-533.
- García, S., Luengo, J., & Herrera, F. (2020). *Data Preprocessing in Data Mining*. Springer.

Garrido, Eduardo Graells-Garrido; Opitz, Daniela; Rowe, Francisco; Arriagada, Jacqueline. Feel Old Yet? Updating Mode of Transportation Distributions from Travel Surveys using Data Fusion with Mobile Phone Data. arXiv, 2023.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42.

Ghosh, S., Chakraborty, D., & Dey, A. (2016). Data normalization for clustering of urban mobility patterns using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 71, 42-57.

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations* (Vol. 3). JHU press.

Graells-Garrido, E., Opitz, D., Rowe, F., & Arriagada, J. (2023). Feel Old Yet? Updating Mode of Transportation Distributions from Travel Surveys using Data Fusion with Mobile Phone Data. arXiv.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646-675.

Hajek, J. J. (1977). Optimal sample size of roadside-interview origin-destination surveys (No. RR 208).

Han, J., Zhang, S., & Li, X. (2016). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(2), 1-15.

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, 25(2), 73-76.

Hansen, W. G. (1967). How accessibility shapes land use. *Journal of the American Institute of Planners*, 33(2), 73-76.

Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. (2009). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction." Editora Springer.

Hazelton, M.L., 2001b. Inference for origin–destination matrices: estimation, reconstruction and prediction. *Transportation Research Part B*, 35 (7), 667–676.

Hosseini, S. A., Ramezani, M., Geroliminis, N., & Abadi, M. N. (2019). A comparative analysis of OD estimation approaches using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 105, 449-463.

Hosseini-Motlagh, S. M., Geroliminis, N., & Khani, A. (2014). Temporal aggregation of large-scale origin–destination demand estimation from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 38, 63-80.

<https://keras.io/getting-started/sequential-model-guide/>. Accessed 18 Abr. 2023

[https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html)

[learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html). Accessed 18 Abr. 2023

Hu, T., Yuan, Y., & Zheng, Y. (2020). Inferring travel modes from mobile phone data using sequence-to-sequence learning. *Transportation Research Part C: Emerging Technologies*, 117, 102699.

Iqbal, Md. Shahadat; Choudhury, Charisma F.; Wang, Pu and González, Marta C. "Development of Origin–destination Matrices Using Mobile Phone Call Data." *Transportation Research Part C: Emerging Technologies* 40 (March 2014): 63–74. © 2014 Elsevier Ltd.

Ji, X., Liu, Y., & Lv, Y. (2018). A framework for origin-destination matrix estimation and prediction using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 95, 152-165.

Jiang, S., Yang, Y., Gupta, S., Veneziano, D., & González, M. C. (2016). The timeGeo modeling framework for urban motility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37), E5370-E5378.

JIANG, X., et al. High-resolution spatiotemporal urban flow inference from massive mobile phone data. *IEEE Transactions on Mobile Computing*, v. 20, n. 1, p. 396-410, 2021.

Jiang, Y., Liu, W., Li, H., Sun, Y., & Li, X. (2018). A spatiotemporal kriging approach for imputing missing values in public transport OD matrix. *Transportation Research Part C: Emerging Technologies*, 86, 120-133.

JIAO, L. et al. Bluetooth traffic monitoring: Count and speed measurement. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE, 2015. p. 632-637.

Jin, X., Li, Z., Jiang, S., & Hu, Y. (2019). A novel normalization method for origin-destination matrices in transportation networks. *Journal of Transport Geography*, 74, 77-87.

Kim, J., Kim, K., & Park, H. (2019). Exploring commuting patterns and characteristics using mobile phone data: A case study of Seoul, Korea. *Journal of Transport Geography*, 77, 97-109.

Knoop, V. L., & Axhausen, K. W. (2019). Temporal dynamics of travel behavior: A review of recent empirical evidence. *Transport Reviews*, 39(4), 440-470.

Kung, K. S., & Greaves, M. (2013). On the use of mobile phone data for estimating dynamic OD matrices. *Transportation Research Part C: Emerging Technologies*, 26, 301-313.

KURAUCHI, F. et al. Data processing and validation in urban travel demand modeling using VISUM. *Transportation Research Procedia*, v. 10, p. 472-481, 2015.

Kuwahara, M., and Sullivan, E. C. (1987). Estimating origin-destination matrices from roadside survey data. *Transportation Research Part B*, 21(3), 233-248.

L. Montero, X. Ros-Roca, R. Herranz, J. Barceló, Fusing mobile phone data with other data sources to generate input OD matrices for transport models, *Transportation Research Procedia*, Volume 37, 2019, Pages 417-424, ISSN 2352-1465.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

LI, X. et al. Analysis of travel time and speed using Bluetooth data in different states of connection. *Transportation Research Record*, v. 2536, n. 1, p. 45-53, 2015.

Li, Y., Zhou, Z., & Zhang, M. (2019). Hierarchical iterative Bayesian iterative proportional fitting model for origin-destination demand estimation. *Transportation Research Record*, 2673(8), 1-11.

Li, Z. (2019). *Transportation Asset Management - Methodology and Applications*. CRC Press - Taylor & Francis Group.

LIU, X., et al. Inferring urban commuting patterns using bus GPS data: A case study in Shenzhen, China. *Transportation Research Part C: Emerging Technologies*, v. 117, p. 102777, 2020.

Liu, X., Gong, M., Jiang, Y., & Chen, M. (2018). A comprehensive framework for spatio-temporal visualization and analysis of big taxi trip data. *ISPRS International Journal of Geo-Information*, 7(9), 343.

Liu, Y., Kang, C., Gao, S., & Wang, F. Y. (2018). Real-time traffic prediction with spatiotemporal correlations using mobile phone data. *IEEE Transactions on Intelligent Transportation Systems*, 19(2), 441-450.

Liu, Y., Wu, J., & Yang, H. (2014). Travel mode detection based on neural network and personal data recorder. *Transportation Research Part C: Emerging Technologies*, 38, 151-164.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic Information Science and Systems*. John Wiley & Sons.

LOPES, T. R. et al. Coleta de dados de tráfego em rodovias utilizando tecnologia Bluetooth. In: *Anais do XVIII Congresso Brasileiro de Engenharia de Tráfego, Transporte e Logística*. 2019.

Luan, X., He, Q., Zhang, Y., & Gu, Y. (2017). Classification of motorized and non-motorized vehicles using a low-cost acoustic sensor. *Transportation Research Part C: Emerging Technologies*, 79, 54-67.

Ma, Y., Huang, X., & Cheng, X. (2020). Constructing travel demand matrix based on GPS data: A review

Ma, Y., Li, Z., Li, M., & Li, Y. (2021). A convolutional neural network based classification approach for urban transportation origin-destination matrix. *Transportation Research Part C: Emerging Technologies*, 124, 103120.

McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media

MIRANDA, F. A.; DIAS, G. G. S. Modelagem de transportes com VISUM: conceitos, aplicação e prática. Blucher, 2018.

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.

Montero, L., Ros-Roca, X., Herranz, R., & Barceló, J. (2018). Fusing mobile phone data with other data sources to generate input OD matrices for transport models. ScienceDirect.

Monzon, A., Ortega, E., & Llorca, C. (2005). Calibration of a traffic assignment model: Iterative procedure for parameter adjustment. Transportation Research Record: Journal of the Transportation Research Board, 1915(1), 68-77.

MUNIZ, L.; CASTRO, J. A.; CARVALHO, M. L. M. de. Modelagem de transportes com ênfase em simulação. Elsevier, 2016.

Müller, Andreas C., e Guido, Sarah. (2016). "Introduction to Machine Learning with Python." Editora O'Reilly Media.

Nitschea, P., Widhalma, P., Breussa, S., & Maurera, P. (2012). A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys. ScienceDirect.

Noelia Caceres, L. M. Romero & Francisco G. Benitez (2020) Exploring strengths and weaknesses of mobility inference from mobile phone data vs. travel surveys, Transportmetrica A: Transport Science, 16:3, 574-601

Noelia Caceres, L. M. Romero & Francisco G. Benitez, Land use inference from mobility mobile phone data and household travel surveys, Transportation Research Procedia, Volume 47, 2020, Pages 417-424, ISSN 2352-1465.

Ortúzar, J. de D., & Willumsen, L. G. (2011). Modelling transport. John Wiley & Sons.

Pang, Z., Chen, X., Lin, T., & Song, X. (2019). A two-stage clustering method for classifying travel modes using GPS trajectory data. Transportation Research Part C: Emerging Technologies, 99, 38-54.

Philippe Nitsche, Peter Widhalm, Simon Breuss, Peter Maurer, A Strategy on How to Utilize Smartphones for Automatically Reconstructing Trips in Travel Surveys, Procedia - Social and Behavioral Sciences, Volume 48, 2012, Pages 1033-1046, ISSN 1877-0428.

Phithakkitnukoon, S., and Ratti, C., (2011), Inferring Asymmetry of Inhabitant Flow using Call Detail Records, Journal of Advances in Information Technology, 2 (4), 239-249.

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., and Ratti, C. (2010). Activity - Aware Map: Identifying human daily activity pattern using mobile phone data, Human Behavior Understanding, 6219(3), 14-25, Springer Berlin / Heidelberg.

PTV Group. (2022). VISUM User Manual. PTV Group.

- QIAN, Z., et al. Principal component analysis of public transit flows derived from smart card data. *Transportation Research Part C: Emerging Technologies*, v. 105, p. 188-202, 2019.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Elsevier.
- Reades, J., Calabrese, F., and Ratti, C. (2009). Eigenplaces: analyzing cities using the space-time structure of the mobile phone network, *Environment and Planning B: Planning and Design*, 36(5), pp. 824-836
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408.
- Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296.
- Santos, A., Almeida, C. M., & Rodrigues, F. (2018). Calibration of an urban transport model: the case of the Porto Metropolitan Area. *Journal of Transport Geography*, 71, 126-136.
- Shan, W., Lu, C., & Xie, X. (2019). Origin-destination matrix estimation for rural transport: A hierarchical iterative Bayesian iterative proportional fitting model. *Journal of Advanced Transportation*, 2019.
- Shen, L., & Stopher, P. (2014). *Review of GPS Travel Survey and GPS Data-Processing Methods*. Taylor & Francis Online.
- SILVA, D. F. *Modelagem de Redes de Transporte Urbano no VISUM*. Universidade de Brasília, 2016.
- Silva, T. L., de Queiroz, R. L., & Alves, H. C. (2018). Mobility patterns and socioeconomic variables: a study based on mobile phone data in Brazil. *Journal of Transport Geography*, 66, 1-9.
- Smith, A., & Jones, B. (2019). Detecting Outliers in Data: A Comprehensive Review. *Journal of Data Science*, 5(2), 123-145.
- Stopher, P. R., & Greaves, S. P. (2007). Household travel surveys: where are we going?. *Transportation Research Part A: Policy and Practice*, 41(5), 367-381.
- Yang, C., Zhang, Y., Zhan, X., Ukkusuri, S. V., & Chen, Y. (2020). Fusing Mobile Phone and Travel Survey Data to Model Urban Activity Dynamics. Wiley Hindawi.
- WANG, D. Z. et al. Comparison of three iterative proportional fitting algorithms for regionalizing travel surveys. *Journal of Transport Geography*, v. 19, n. 5, p. 1001-1010, 2011.
- Wang, D., & Zhang, K. (2016). Urban travel mode identification by mobile phone data with multi-dimensional context. *Transportation Research Part C: Emerging Technologies*, 72, 22-32.

- Wang, D., Liu, Y., Xu, S., & Xu, G. (2005). Classifying vehicles using magnetic signature-based sensors. *Transportation Research Part C: Emerging Technologies*, 13(6), 445-462.
- Wang, D., Pedersen, T. B., & Zhang, K. (2016). Exploring origin-destination mobility data from mobile phone call records. *Transportation Research Part C: Emerging Technologies*, 72, 186-195.
- Wang, J., Wang, W., & Sun, H. (2019). An evaluation of origin-destination matrix estimation methods using traffic data in urban road networks. *Journal of Advanced Transportation*, 2019.
- Wang, J., Wang, W., Liu, W., & Sun, L. (2019). Aggregating origin-destination matrices with spatiotemporal constraints for large-scale taxi trip data. *Transportation Research Part C: Emerging Technologies*, 107, 50-64.
- Wang, Y., Zhang, Z., & Zhang, C. (2018). Quantile-Quantile random forests for high-dimensional data regression. *Pattern Recognition*, 77, 307-316.
- WEI, T., et al. Urban passenger travel pattern classification by artificial neural networks using GPS data. *Transportation Research Part C: Emerging Technologies*, v. 99, p. 128-143, 2019.
- Wu, H., Li, X., Li, Q., & Li, H. (2018). Classifying urban travel modes using mobile phone data: A comparative study of machine learning methods. *Transportation Research Part C: Emerging Technologies*, 95, 1-15.
- Vanderlande, J. et al. A multi-period analysis of dynamic user equilibrium traffic assignment with day-to-day learning and short-term departure time choice. *Transportation Research Part B: Methodological*, v. 122, p. 155-178, 2019
- Vovsha, P., & Petersen, E. L. (2011). Travel survey methods: Quality and future directions. *Transport Reviews*, 31(2), 219-236
- Xie, K., Li, M., Li, Y., & Liu, H. (2007). A method for identifying direction of flow with the use of vehicle count data. *Transportation Research Part C: Emerging Technologies*, 15(6), 331-346.
- Xie, Y., Yan, X., Zhang, C., Liu, H., & Zang, X. (2015). Spatial aggregation and disaggregation of origin-destination matrices using auxiliary data. *Transportation Research Part C: Emerging Technologies*, 52, 64-81.
- Zhang, C., & Lou, Y. (2019). Urban travel mode classification from mobile phone data using deep learning. *Transportation Research Part C: Emerging Technologies*, 104, 29-44.
- Zhang, J., Chen, S., & Wang, F. Y. (2012). Vehicle classification using a single uncal
- Zhang, J., Lu, Y., Liu, Y., Yu, Z., & Ma, W. (2018). Outlier detection and processing of origin-destination data based on minimum covariance determinant. *Transportation Research Part C: Emerging Technologies*, 95, 1-15.
- ZHANG, X. et al. An Overview of PTV Visum 2020. PTV Group, 2018.

ZHANG, X. Evaluation of the Traffic Capacity of Roundabouts on High-Speed Roads Based on the VISSIM and VISUM Models. *Journal of Advanced Transportation*, v. 52, n. 4, p. 622-637, 2018.

Zhang, Z., Cui, P., Zhu, W., & Yang, S. (2018). Deep learning on decision trees for efficient text classification. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (pp. 211-220).

Zheng, A., & Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly Media.

ZHENG, J. et al. Bluetooth traffic sensing: A new way of traffic sensing. *IEEE Transactions on Intelligent Transportation Systems*, v. 16, n. 1, p. 110-123, 2015.

Zheng, Y., Liu, Y., Wang, L., & Miao, L. (2014). Inferring transportation mode from anonymous, sparse and noisy data sources. *PloS one*, 9(9), e107559.

Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W. Y. (2014). Understanding mobility based on GPS data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3), 1-33.

Zheng, Y., Liu, Y., Yuan, J., & Xie, X. (2016). Urban computing with taxicabs. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3), 1-19.

Zheng, Y., Liu, Y., Yuan, J., Xie, X., & Sun, G. (2011). Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing* (pp. 89-98). ACM.

Zheng, Y., Liu, Y., Yuan, J., Xie, X., & Yang, Q. (2010). Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing (UbiComp'10)* (pp. 89-98).

Zheng, Y., Zhang, L., Xie, X., & Ma, W. Y. (2011). Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1046-1054).

ZHOU, J., et al. Travel patterns of bike sharing riders: A data-driven study in Beijing. *Transportation Research Part D: Transport and Environment*, v. 91, p. 102642, 2021.

ZHU, S. *Analysis of the Traffic Impact of Road Construction on Congestion in Urban Areas*. University of California, 2019

APÊNDICE A

O Quadro A 1 relaciona os municípios, regiões e estados componentes de cada uma das seis zonas de tráfego externas.

ZONA DE TRÁFEGO EXTERNA	TIPO DE UNIDADE	ESTADO	ESTADO/REGIÃO/MUNICÍPIO
Belém	Estado	AM	Pintadas
	Estado	AP	Antônio Gonçalves
	Estado	MA	Baixa Grande
	Estado	PA	Caém
	Estado	PI	Mairi
	Estado	RR	Morro Do Chapéu
	Estado	TO	Mundo Novo
	Município	BA	Piritiba
	Município	BA	Tapiramutá
	Município	BA	Andorinha
	Município	BA	Caldeirão Grande
	Município	BA	Campo Formoso
	Município	BA	Capim Grosso
	Município	BA	Filadélfia
	Município	BA	Itiúba
	Município	BA	Jacobina
	Município	BA	Jaguarari
	Município	BA	Miguel Calmon
	Município	BA	Mirangaba
	Município	BA	Ourolândia
	Município	BA	Pindobaçu
	Município	BA	Ponto Novo
	Município	BA	Quixabeira
	Município	BA	São José Do Jacuípe
	Município	BA	Saúde
	Município	BA	Senhor Do Bonfim
	Município	BA	Serrolândia
	Município	BA	Umburanas
	Município	BA	Várzea Da Roça
	Município	BA	Várzea Do Poço
Município	BA	Várzea Nova	
Microrregião	BA	Juazeiro	
Brasília	Estado	AC	Acre
	Estado	DF	Distrito Federal
	Estado	GO	Goiás
	Estado	MS	Mato Grosso do Sul
	Estado	MT	Mato Grosso

ZONA DE TRÁFEGO EXTERNA	TIPO DE UNIDADE	ESTADO	ESTADO/REGIÃO/MUNICÍPIO
	Estado	RO	Rondônia
	Mesorregião	BA	Extremo Oeste Baiano
	Microrregião	BA	Barra
	Microrregião	BA	Bom Jesus Da Lapa
	Município	BA	Uibaí
	Município	BA	São Gabriel
	Município	BA	Presidente Dutra
	Município	BA	Lapão
	Município	BA	Jussara
	Município	BA	João Dourado
	Município	BA	Irecê
	Município	BA	Ibititá
	Município	BA	Ibipeba
	Município	BA	Gentio Do Ouro
	Município	BA	Central
	Município	BA	América Dourada
Fortaleza	Estado	CE	Ceará
	Microrregião	BA	Paulo Afonso
	Mesorregião	PE	São Francisco Pernambucano
	Mesorregião	PE	Sertão Pernambucano
Recife	Estado	AL	Alagoas
	Estado	PB	Paraíba
	Estado	RN	Rio Grande do Norte
	Estado	SE	Sergipe
	Mesorregião	BA	Nordeste Baiano
	Município	BA	Água Fria
	Município	BA	Coração De Maria
	Município	BA	Irá
	Município	BA	Pedrão
	Município	BA	Ouriçangas
	Município	BA	Santanópolis
	Município	BA	Teodoro Sampaio
	Município	BA	Catu
	Município	BA	Itanagra
	Município	BA	Terra Nova
	Município	BA	Santa Bárbara
	Mesorregião	PE	Agreste Pernambucano
	Mesorregião	PE	Mata Pernambucana
	Mesorregião	PE	Metropolitana de Recife
Rio de Janeiro	Estado	ES	Espírito Santo
	Estado	RJ	Rio de Janeiro
	Município	BA	Salinas Da Margarida

ZONA DE TRÁFEGO EXTERNA	TIPO DE UNIDADE	ESTADO	ESTADO/REGIÃO/MUNICÍPIO
	Mesorregião	BA	Sul Baiano
	Município	BA	Aratuípe
	Município	BA	Cachoeira
	Município	BA	Conceição Do Almeida
	Município	BA	Cruz Das Almas
	Município	BA	Dom Macedo Costa
	Município	BA	Governador Mangabeira
	Município	BA	Jaguaripe
	Município	BA	Maragogipe
	Município	BA	Muniz Ferreira
	Município	BA	Muritiba
	Município	BA	Nazaré
	Município	BA	Salinas da Margarida
	Município	BA	Santo Amaro
	Município	BA	Santo Antônio de Jesus
	Município	BA	São Félix
	Município	BA	São Felipe
	Município	BA	Sapeçu
	Município	BA	Saubara
	São Paulo	Estado	MG
Estado		PR	Paraná
Estado		RS	Rio Grande do Sul
Estado		SC	Santa Catarina
Estado		SP	São Paulo
Mesorregião		BA	Centro Sul Baiano
Município		BA	Barra Do Mendes
Município		BA	Barro Alto
Município		BA	Cafarnaum
Município		BA	Canarana
Município		BA	Iaçu
Município		BA	Ibiquera
Município		BA	Iraquara
Município		BA	Itaberaba
Município		BA	Lajedinho
Município		BA	Macajuba
Município		BA	Mulungu Do Morro
Município		BA	Ruy Barbosa
Município		BA	Souto Soares
Município		BA	Antônio Cardoso
Município		BA	Ipecaetá
Município		BA	Ipirá
Município		BA	Itatim

ZONA DE TRÁFEGO EXTERNA	TIPO DE UNIDADE	ESTADO	ESTADO/REGIÃO/ MUNICÍPIO
	Município	BA	Rafael Jambeiro
	Município	BA	Santa Teresinha
	Município	BA	Santo Estêvão
	Município	BA	Serra Preta
	Município	BA	Cabaceiras do Paraguaçu
	Município	BA	Castro Alves
	Município	BA	Varzedo
	Município	BA	Boa Vista do Tupim
	Município	BA	Elísio Medrado
	Município	BA	Anguera

Fonte: elaborado pelo autor

O Quadro A 2 apresenta a relação com todas as 277 zonas, listando o código, nome do local, município a qual pertence e sua respectiva região metropolitana.

NO	NOME	MUNICÍPIO	REGIÃO METROPOLITANA
1	Afligidos	São Gonçalo dos Campos	FEIRA DE SANTANA
2	Amélia Rodrigues	Amélia Rodrigues	FEIRA DE SANTANA
3	Bonfim da Feira	Feira de Santana	FEIRA DE SANTANA
4	Centro	Feira de Santana	FEIRA DE SANTANA
5	Queimadinha	Feira de Santana	FEIRA DE SANTANA
6	São João	Feira de Santana	FEIRA DE SANTANA
7	Conceição da Feira	Conceição da Feira	FEIRA DE SANTANA
8	Conceição do Jacuípe	Conceição do Jacuípe	FEIRA DE SANTANA
9	Sobradinho / Jardim Cruzeiro	Feira de Santana	FEIRA DE SANTANA
10	Gov. Dr. João Durval Carneiro	Feira de Santana	FEIRA DE SANTANA
11	Centro Humildes	Feira de Santana	FEIRA DE SANTANA
12	Bom Viver	Feira de Santana	FEIRA DE SANTANA
13	Inhatá	Amélia Rodrigues	FEIRA DE SANTANA
14	Jaguara	Feira de Santana	FEIRA DE SANTANA
15	Jaíba	Feira de Santana	FEIRA DE SANTANA
16	Manga - Subae	Feira de Santana	FEIRA DE SANTANA
17	Papagaio	Feira de Santana	FEIRA DE SANTANA
18	Aeroporto	Feira de Santana	FEIRA DE SANTANA
19	Maria Quitéria	Feira de Santana	FEIRA DE SANTANA
20	Mata da Aliança	Amélia Rodrigues	FEIRA DE SANTANA
21	Matinha	Feira de Santana	FEIRA DE SANTANA
22	Nova Esperança	Feira de Santana	FEIRA DE SANTANA
23	UFRB	Feira de Santana	FEIRA DE SANTANA
24	Gabriela / Pampalona / Campo Gado Novo	Feira de Santana	FEIRA DE SANTANA

NO	NOME	MUNICÍPIO	REGIÃO METROPOLITANA
25	UEFS	Feira de Santana	FEIRA DE SANTANA
26	Santana	Feira de Santana	FEIRA DE SANTANA
27	Calumbi	Feira de Santana	FEIRA DE SANTANA
28	Muchila / Pedra do Descanso	Feira de Santana	FEIRA DE SANTANA
29	Jardim Acácia	Feira de Santana	FEIRA DE SANTANA
30	Santa Mônica	Feira de Santana	FEIRA DE SANTANA
31	Caseb/Estação Nova/Lagoa Grande	Feira de Santana	FEIRA DE SANTANA
32	Brasília	Feira de Santana	FEIRA DE SANTANA
33	Eucalipto	Feira de Santana	FEIRA DE SANTANA
34	São Gonçalo dos Campos	São Gonçalo dos Campos	FEIRA DE SANTANA
35	Sergi	São Gonçalo dos Campos	FEIRA DE SANTANA
36	Conj. Viveiros	Feira de Santana	FEIRA DE SANTANA
37	Tomba	Feira de Santana	FEIRA DE SANTANA
38	IFBA	Feira de Santana	FEIRA DE SANTANA
39	Subaé	Feira de Santana	FEIRA DE SANTANA
40	Tanquinho	Tanquinho	FEIRA DE SANTANA
41	Tiquaruçu	Feira de Santana	FEIRA DE SANTANA
42	Paraciba	Camaçari	SALVADOR
43	Jacuípe / Campos / Garajuba	Camaçari	SALVADOR
44	Complexo Ind. For Nordeste	Camaçari	SALVADOR
45	Parque das Mangabas	Camaçari	SALVADOR
46	Tancredo N./Gleba E/Ulisses Guim.	Camaçari	SALVADOR
47	Burizeiro	Camaçari	SALVADOR
48	Bela Vista / Piaçaveira / Inocoop	Camaçari	SALVADOR
49	Bairro do Natal/Gleba B/Alto d. Cruz	Camaçari	SALVADOR
50	Centro / Camaçari de Dentro	Camaçari	SALVADOR
51	Monte Gordo / Praia de Itacimirim	Camaçari	SALVADOR
52	Barra de Pojuças / Frecheiras / Lôdo	Camaçari	SALVADOR
53	Camassari / Visconde	Camaçari	SALVADOR
54	Abrantes	Camaçari	SALVADOR
55	Caroba / Menino Jesus	Candeias	SALVADOR
56	Nova Candeias / Urbis I	Candeias	SALVADOR
57	Centro / Malemba / Candeias	Candeias	SALVADOR
58	Sarandi / Areias	Candeias	SALVADOR
59	Estaleiro/Most. de N.Sa. Fé e Rosário	Candeias	SALVADOR
60	Paranapanema	Dias d'Ávila	SALVADOR
61	Centro - Dias D'Ávila	Dias d'Ávila	SALVADOR
62	Sta Terezinha / Parque Capuame	Dias d'Ávila	SALVADOR
63	Comun. Cat. Cid. Santa / Biribeira	Dias d'Ávila	SALVADOR
64	Lot. Canto da Ilha/Recant Bela Vista	Itaparica	SALVADOR
65	Engenho	Itaparica	SALVADOR
66	Bom Despacho	Itaparica	SALVADOR
67	Terminal Marítimo Bom Despacho	Itaparica	SALVADOR
68	Itaparica / Alto das Pombas / Praia Ponta de Areia	Itaparica	SALVADOR

NO	NOME	MUNICÍPIO	REGIÃO METROPOLITANA
69	Centro	Salvador	SALVADOR
70	Loteamento Bosque Kiosque	Lauro de Freitas	SALVADOR
71	Miragem / Lot. Miragem / Granjas / Reunidas	Lauro de Freitas	SALVADOR
72	Parque dos Coqueiros / Portão do Sol / Buraquinho	Lauro de Freitas	SALVADOR
73	Vilas do Atlântico	Lauro de Freitas	SALVADOR
74	Vila Praiana	Lauro de Freitas	SALVADOR
75	Jardim Aeroporto	Lauro de Freitas	SALVADOR
76	Jardim Portão / Pomar do Rio	Lauro de Freitas	SALVADOR
77	Encontro Águas	Lauro de Freitas	SALVADOR
78	Jaíba / Jambeiro	Lauro de Freitas	SALVADOR
79	Jardim Castelão / Ipiranga	Lauro de Freitas	SALVADOR
80	Lot. Villas Bosque	Lauro de Freitas	SALVADOR
81	Águas Finas / Vida Nova	Lauro de Freitas	SALVADOR
82	Caixa D'Água / Ipitinga	Lauro de Freitas	SALVADOR
83	Pouso Alegre / Jardim Tarumã / Cidade Nova	Lauro de Freitas	SALVADOR
84	Base Naval de Aratu	Salvador	SALVADOR
85	Vargem de Baixo / Faz. Juruaba	Mata de São João	SALVADOR
86	Cabelo / Camboatá	Mata de São João	SALVADOR
87	Centro Mata de S. João / Piri Piri	Mata de São João	SALVADOR
88	Amado Bahia	Mata de São João	SALVADOR
89	Açu da Torre	Mata de São João	SALVADOR
90	Santiago	Pojuca	SALVADOR
91	Centro - Pojuca	Pojuca	SALVADOR
92	Miranga / Petro Recôncavo	Pojuca	SALVADOR
93	Hospital Universitário	Salvador	SALVADOR
94	Vitória	Salvador	SALVADOR
95	Jardim Apipema	Salvador	SALVADOR
96	Calabar / Alto das Pombas	Salvador	SALVADOR
97	UFBA / Ondina / Zoológico	Salvador	SALVADOR
98	Engenho Velho da Federação	Salvador	SALVADOR
99	Rio Vermelho	Salvador	SALVADOR
100	Amaralina	Salvador	SALVADOR
101	Nordeste de Amaralina	Salvador	SALVADOR
102	Cosme de Farias	Salvador	SALVADOR
103	Brotas	Salvador	SALVADOR
104	Horto Florestal	Salvador	SALVADOR
105	Candeal	Salvador	SALVADOR
106	Engenho Velho de Brotas	Salvador	SALVADOR
107	Chapada do Rio Vermelho	Salvador	SALVADOR
108	Santa Cruz	Salvador	SALVADOR
109	Arena Aquática Salvador	Salvador	SALVADOR
110	Pituba	Salvador	SALVADOR

NO	NOME	MUNICÍPIO	REGIÃO METROPOLITANA
111	Itaigara	Salvador	SALVADOR
112	EsFCEEx - CMS	Salvador	SALVADOR
113	STIEP/Costa Azul / Jardim Armação	Salvador	SALVADOR
114	Caminho das Árvores / Pituba	Salvador	SALVADOR
115	Resgate	Salvador	SALVADOR
116	Centro de Saúde Pernambués	Salvador	SALVADOR
117	UEBA	Salvador	SALVADOR
118	Luiz Anselmo / Santa Teresa	Salvador	SALVADOR
119	Santo Agostinho / Matatu	Salvador	SALVADOR
120	Vila Laura	Salvador	SALVADOR
121	Cidade Nova	Salvador	SALVADOR
122	Macaúbas	Salvador	SALVADOR
123	Caixa d'Água	Salvador	SALVADOR
124	Curuzu / Santa Mônica / Pero Vaz	Salvador	SALVADOR
125	IAPI	Salvador	SALVADOR
126	Bom juá	Salvador	SALVADOR
127	São Gonçalo do Retiro	Salvador	SALVADOR
128	Porto de Salvador / TECON	Salvador	SALVADOR
129	Liberdade	Salvador	SALVADOR
130	Nazaré / Arena Fonte Nova	Salvador	SALVADOR
131	Federação	Salvador	SALVADOR
132	Barris	Salvador	SALVADOR
133	Shopping Piedade - Barris	Salvador	SALVADOR
134	Centro	Salvador	SALVADOR
135	Centro Histórico	Salvador	SALVADOR
136	Santo Antônio	Salvador	SALVADOR
137	Largo da Tieta	Salvador	SALVADOR
138	Hospital Sta Izabel/Colég. Saleziano	Salvador	SALVADOR
139	Boa Viagem / Roma	Salvador	SALVADOR
140	Mont. Serrat/Bonfim/Caminho Areia	Salvador	SALVADOR
141	Marés / Calçada	Salvador	SALVADOR
142	Salvador Marina / Lobato	Salvador	SALVADOR
143	Mangueira / Massaranduba	Salvador	SALVADOR
144	Ribeira	Salvador	SALVADOR
145	Vila Ruy Barbosa / Jardim Cruzeiro	Salvador	SALVADOR
146	Uruguai	Salvador	SALVADOR
147	USF Fiais / Polipor / C. Est. Dantas Jr / S. Caet.	Salvador	SALVADOR
148	Atakadão Atakarejo São Caetano	Salvador	SALVADOR
149	C.Est. Prof. Carlos Alberto Cerqueira - S. Caetano	Salvador	SALVADOR
150	Fazenda Grande Retiro	Salvador	SALVADOR
151	Campinas de Pirajá	Salvador	SALVADOR
152	Boa Vista de São Caetano	Salvador	SALVADOR
153	Arraial do Retiro	Salvador	SALVADOR

NO	NOME	MUNICÍPIO	REGIÃO METROPOLITANA
154	Barreiras / Engomadeira / Beiru / Tancredo Neves	Salvador	SALVADOR
155	Doron	Salvador	SALVADOR
156	Cabula VI	Salvador	SALVADOR
157	Novo Horizonte	Salvador	SALVADOR
158	Boca do Rio - Imbuí	Salvador	SALVADOR
159	Pituaçu	Salvador	SALVADOR
160	Jaguaribe / Piatã / Patamares	Salvador	SALVADOR
161	Alto do Coqueirinho	Salvador	SALVADOR
162	Escola Bahiana de Medicina e Saúde	Salvador	SALVADOR
163	Mata Escura	Salvador	SALVADOR
164	Jardim Sto Inácio / Calabetão	Salvador	SALVADOR
165	Presídio / Cadeia / Unid. Esp. Discip.	Salvador	SALVADOR
166	Porto Seco Pirajá	Salvador	SALVADOR
167	Alto do Cabrito	Salvador	SALVADOR
168	Marechal Rondon	Salvador	SALVADOR
169	Plataforma	Salvador	SALVADOR
170	Pirajá	Salvador	SALVADOR
171	Parque Florestal da Represa	Salvador	SALVADOR
172	Valeria	Salvador	SALVADOR
173	Itacarânia / Alto da Terezinha	Salvador	SALVADOR
174	Praia Grande / Peri Peri	Salvador	SALVADOR
175	Nova Constituinte	Salvador	SALVADOR
176	Coutos	Salvador	SALVADOR
177	Fazenda Couto	Salvador	SALVADOR
178	São Tomé de Paripe	Salvador	SALVADOR
179	Terminal Portuário Cotegipe	Salvador	SALVADOR
180	Paripe	Salvador	SALVADOR
181	Riacho do Golfo	Salvador	SALVADOR
182	Nova Esperança / Areia Branca	Salvador	SALVADOR
183	Itinga	Salvador	SALVADOR
184	EUCAFI / Ped. Valéria de Salvador	Salvador	SALVADOR
185	Fazenda Grande II	Salvador	SALVADOR
186	São Cristóvão	Salvador	SALVADOR
187	Mussurunga	Salvador	SALVADOR
188	Fazenda Grande IV	Salvador	SALVADOR
189	Cajazeiras XI	Salvador	SALVADOR
190	Fazenda Grande I	Salvador	SALVADOR
191	Cajazeiras X	Salvador	SALVADOR
192	Fazenda Grande III	Salvador	SALVADOR
193	Cajazeiras VII	Salvador	SALVADOR
194	Parque Metrop. Pituaçu	Salvador	SALVADOR
195	Paralela / ADELBA	Salvador	SALVADOR
196	Parque da Cidade	Salvador	SALVADOR
197	CAB - Centro Adm. Bahia	Salvador	SALVADOR

NO	NOME	MUNICÍPIO	REGIÃO METROPOLITANA
198	Motocross / Ass. Deus / Sussuarana	Salvador	SALVADOR
199	Castelo Branco	Salvador	SALVADOR
200	Shopping Bela Vista	Salvador	SALVADOR
201	Salvador Shopping	Salvador	SALVADOR
202	Big Pituba / Shopping da Bahia	Salvador	SALVADOR
203	Santuário Arq. N.Sra da Conceição Aparecida	Salvador	SALVADOR
204	Estação Imbuí	Salvador	SALVADOR
205	Piatã / AABB	Salvador	SALVADOR
206	Pátio Transsalvador/DETRAN Orla - Bairro da Paz	Salvador	SALVADOR
207	Bairro da Paz / Estação Bairro da Paz	Salvador	SALVADOR
208	Praia de Itapuã / Assoc. Brasil / Lagoa do Abaet	Salvador	SALVADOR
209	Alphaville II	Salvador	SALVADOR
210	Vila 2 de Julho	Salvador	SALVADOR
211	Canabrava	Salvador	SALVADOR
212	Vale dos Lagos	Salvador	SALVADOR
213	São Marcos	Salvador	SALVADOR
214	Sussuarana	Salvador	SALVADOR
215	Nova Sussuarana	Salvador	SALVADOR
216	São Rafael	Salvador	SALVADOR
217	Jardim Cajazeiras	Salvador	SALVADOR
218	Pau da Lima	Salvador	SALVADOR
219	Vila Canária	Salvador	SALVADOR
220	Sete de Abril	Salvador	SALVADOR
221	Novo Marotinho	Salvador	SALVADOR
222	Jardim Nova Esperança	Salvador	SALVADOR
223	Nova Brasília	Salvador	SALVADOR
224	Dom Avelar	Salvador	SALVADOR
225	Nova Rodoviária de Salvador / Depot Wilport	Salvador	SALVADOR
226	Águas Claras	Salvador	SALVADOR
227	Cajazeiras VI	Salvador	SALVADOR
228	Cajazeiras IV	Salvador	SALVADOR
229	Cajazeiras V	Salvador	SALVADOR
230	Jaguaripe I	Salvador	SALVADOR
231	Cajazeiras VIII	Salvador	SALVADOR
232	ADELBA / Itapuã	Salvador	SALVADOR
233	Colégio Marízia Maior / Praia de Catussaba	Salvador	SALVADOR
234	Stella Maris	Salvador	SALVADOR
235	Praia do Flamengo	Salvador	SALVADOR
236	Parque das Dunas / Praia de Ipiranga	Salvador	SALVADOR
237	Aeroporto Internacional de Salvador	Salvador	SALVADOR
238	Maracanã (Ilha de Maré)	Salvador	SALVADOR

NO	NOME	MUNICÍPIO	REGIÃO METROPOLITANA
239	Sr. dos Passos (Ilha dos Frades)	Salvador	SALVADOR
240	Jardim das Margaridas	Salvador	SALVADOR
241	6º Batalhão de Polícia do Exército	Salvador	SALVADOR
242	Barra	Salvador	SALVADOR
243	Graça	Salvador	SALVADOR
244	Monte Recôncavo	São Francisco do Conde	SALVADOR
245	Mataripe	São Francisco do Conde	SALVADOR
246	São Francisco do Conde	São Francisco do Conde	SALVADOR
247	Maracangalha	São Sebastião do Passé	SALVADOR
248	Zona Rural - São Seb. Passé	São Sebastião do Passé	SALVADOR
249	Taquipe / Cinco Rios / Est. Coletora Almeida	São Sebastião do Passé	SALVADOR
250	Centro São Seb. Passé / Papuçu	São Sebastião do Passé	SALVADOR
251	Lamarão do Passé	São Sebastião do Passé	SALVADOR
252	Subestação Cotegipe / Belov Eng.	Simões Filho	SALVADOR
253	Núcleo Hab. Rubens Costa Cia	Simões Filho	SALVADOR
254	IFBA - Campus Simões Filho	Simões Filho	SALVADOR
255	Pitanguiha / Jardim Sta Terezinha	Simões Filho	SALVADOR
256	Lot. Cristo Rey / São José	Simões Filho	SALVADOR
257	Centro - Simões Filho	Simões Filho	SALVADOR
258	Seguro	Simões Filho	SALVADOR
259	Zona Rural - Simões Filho	Simões Filho	SALVADOR
260	Via das Torres / Binatural BA / CD-Química Amparo	Simões Filho	SALVADOR
261	Lot. Enseada do Sol	Vera Cruz	SALVADOR
262	Lot. Nova Itaparica	Vera Cruz	SALVADOR
263	Terminal Marítimo Vera Cruz	Vera Cruz	SALVADOR
264	Mar Grande / Praia de Corôa	Vera Cruz	SALVADOR
265	Tatu	Vera Cruz	SALVADOR
266	Ponta da Cruz	Vera Cruz	SALVADOR
267	Baiacu Vera Cruz	Vera Cruz	SALVADOR
268	Enseada Nuvem Azul / Praia de Iansã	Vera Cruz	SALVADOR
269	Matarandiba	Vera Cruz	SALVADOR
270	Cacha Pregos	Vera Cruz	SALVADOR
271	Base Naval de Aratu	Madre de Deus	SALVADOR
272	Belém	Belém	EXTERNO
273	Brasília	Brasília	EXTERNO
274	Fortaleza	Fortaleza	EXTERNO
275	Recife	Recife	EXTERNO
276	Rio de Janeiro	Rio de Janeiro	EXTERNO
277	São Paulo	São Paulo	EXTERNO

ANEXO A – PROCEDIMENTOS – ANÁLISE DE MOBILIDADE

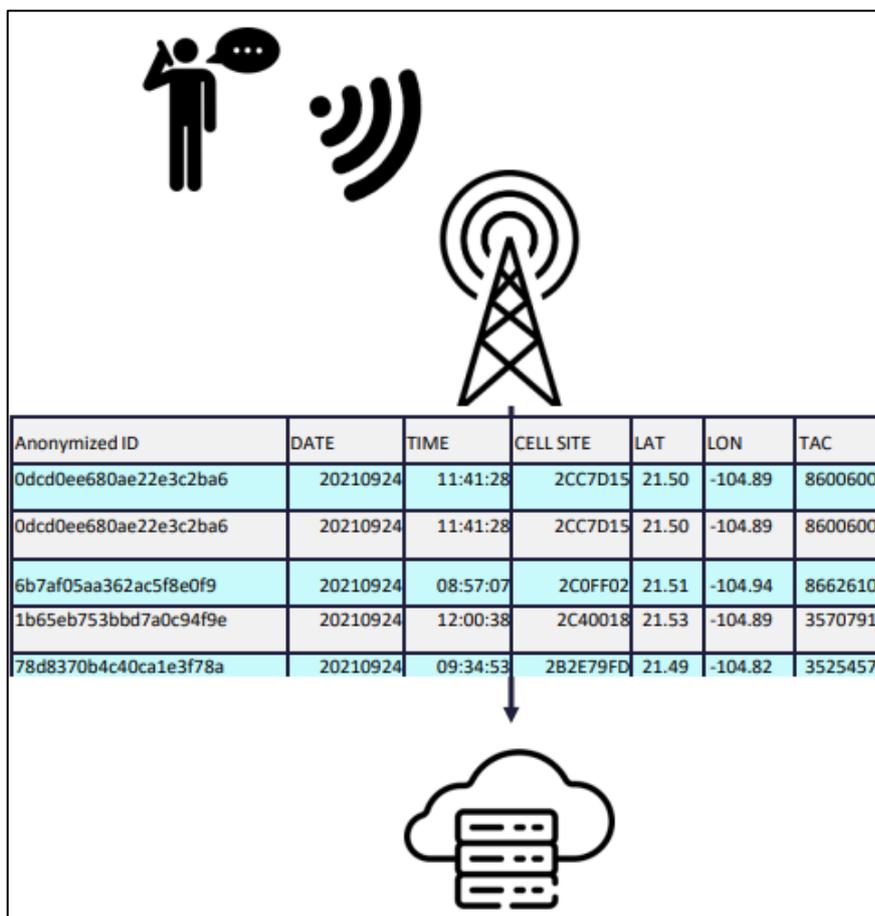
A1 FONTE DE DADOS

A empresa desenvolvedora utiliza os CDRs (*Call Detail Records*) e a rede de antenas de telefonia móvel como a principal fonte de dados.

Toda vez que um usuário de celular faz uso de seu terminal móvel, a operadora móvel registra um evento de comunicação (registro) que inclui um identificador exclusivo do usuário (ID), o horário do evento e a localização (ID da antena), além de diferentes informações técnicas.

Os usuários geram uma média de 100 a 500 eventos por dia, o que representa uma fonte precisa e diversificada de informações disponíveis para analisar de forma confiável a mobilidade das pessoas.

Durante o processo de tratamento e análise dos dados, nenhuma informação individualizada é coletada ou relacionada à própria comunicação do usuário (conteúdo de mensagens, chamadas ou aplicativos).



Fonte: Kido Dyamics

A2 REGISTRO DE COMUNICAÇÃO

CDRs (Call Detail Records) são registros que contêm dados relacionados a chamadas telefônicas, SMS, conexões à Internet e informações geradas pela própria rede para garantir o controle e a qualidade do serviço.

Do ponto de vista dos estudos de mobilidade, cada registro fornece informações espaço-tempo do dispositivo relacionando sua posição a uma área de cobertura Chamada célula e que está associada a uma antena.

Os registros de comunicação móvel gerados pelos diferentes elementos da rede móvel da operadora (MSC, SMSC, MMSC, SGSN/GGSN) são consolidados por seus sistemas de gestão e embalados em arquivos específicos por tipo de comunicação realizada.

Esses arquivos incluem todos os registros gerados para os clientes da operadora e outros usuários que usam a rede móvel como uma rede visitante.

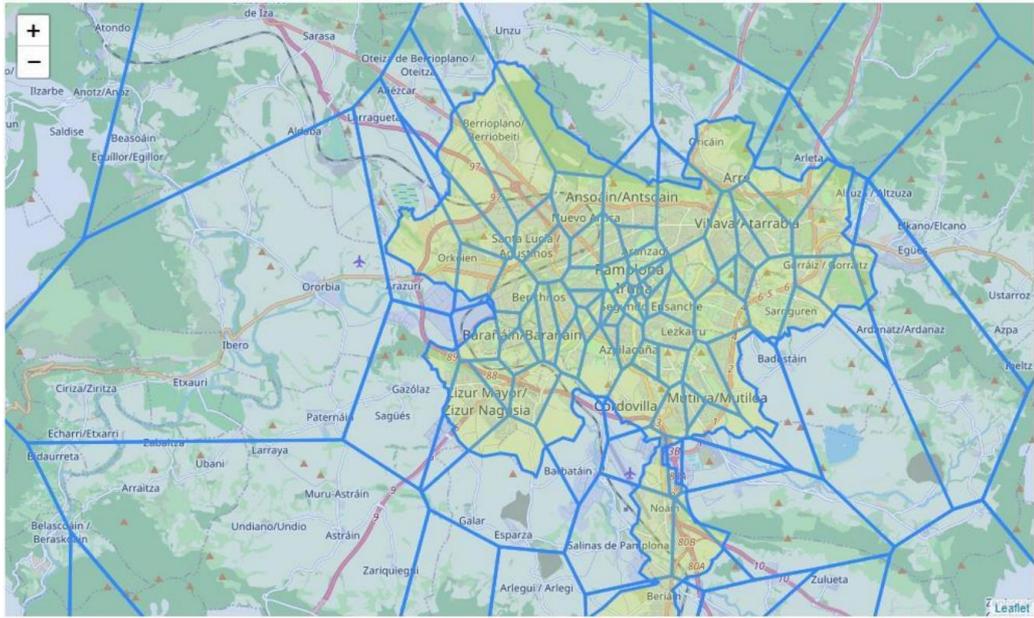
Para o processamento, esses arquivos geralmente são identificados com seu tipo de registro, o sistema que os gerou e o período de tempo a que correspondem. Do ponto de vista geral, os campos contidos em cada registro (sinalização e CDR) de interesse para aplicações de *BigData* aplicadas aos estudos de mobilidade são:

- i. **Identidade do usuário:** identificador anonimizado correspondente a cada usuário.
- ii. **Tipo cliente:** permite que o usuário seja identificado como cliente do operador responsável pela infraestrutura ou como usuário em trânsito (roamer).
- iii. **Data:** data de início da comunicação.
- iv. **Tempo:** tempo de início e fim da comunicação.
- v. **Célula inicial:** célula onde o celular está localizado no início do registro.
- vi. **Célula final:** célula onde o celular está localizado no final do registro.

A3 DADOS ADICIONAIS INCORPORADOS PELA OPERADORA

Além disso, para concluir a análise de mobilidade é necessário ter uma série de arquivos que complementem as informações fornecidas pelas operadoras móveis, especificamente:

- i.** Arquivo com informações sobre a geolocalização de todas as estações base (antenas) da operadora correspondentes às diferentes tecnologias implantadas: 2G, 3G, 4G e 5G.
- ii.** Arquivos com informações do cliente que serão úteis para incorporar a capacidade de segmentação solicitada e como referência para a realização do "censo próprio" no processo de expansão da amostra definido abaixo. As seguintes informações serão incorporadas:
 - Faixa etária;
 - Gênero;
 - Classe social;
 - Nacionalidade;
 - Nacionais (identificando CEP / município de residência);
 - Estrangeiros viajando pelo país (identificando o país de origem de seu operador).



Fonte: Kido Dyamics

A4 ETAPAS DE ANÁLISE

A4.1 FILTROS E TESTES DE CONSISTÊNCIA DE DADOS BRUTOS

Como primeiro passo, é realizada uma análise preliminar para garantir a consistência, a qualidade e o anonimato do fluxo de registros gerados pela operadora móvel. As entradas errôneas são filtradas e a atividade de comunicação e o comportamento geral de mobilidade através dos registros do operador são comparados com os padrões esperados para garantir sua validade.

Os seguintes filtros são incorporados:

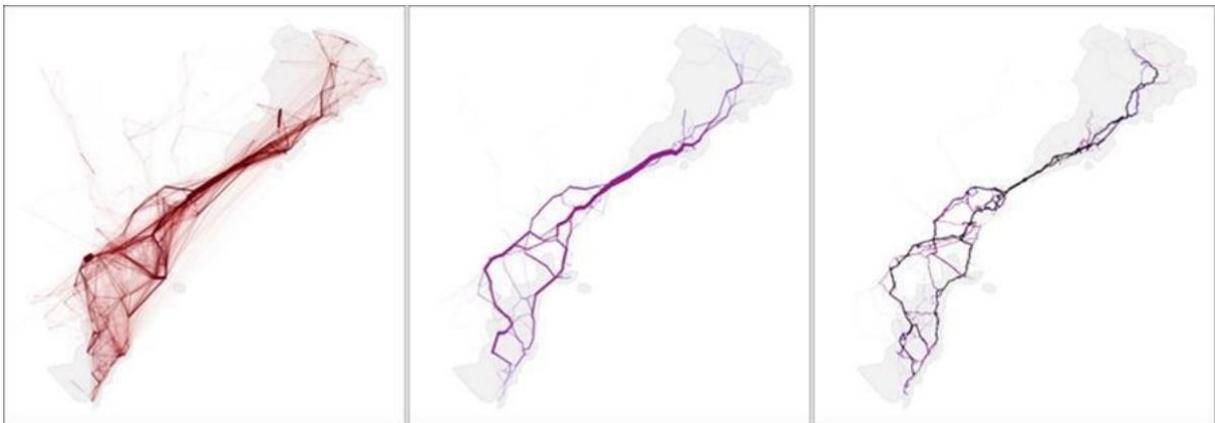
- i. **Filtro de erros de transcrição:** As entradas incompletas ou errôneas são removidas dos dados originais. Em alguns casos existem antenas que não podem ser localizadas, eventos com algum erro de transcrição ou campos vazios. Eles não são comuns, mas precisam ser filtrados.
- ii. **Filtro de massa crítico por usuário:** Os 25% dos dispositivos com o menor número de eventos são eliminados. Estes são tipicamente dispositivos com 1 a 10 eventos por dia, insuficientes para fazer uma análise adequada de mobilidade. Apesar do filtro, o tamanho da amostra permanece significativo.
- iii. **Filtro de relevância da amostra:** Para cada usuário, eventos redundantes ou irrelevantes para mobilidade são filtrados. Isso inclui sequências de eventos na mesma antena ou mudanças de conexão para duas, três e até quatro antenas com sobreposição de cobertura, mesmo que o dispositivo não tenha se movido.
- iv. **Filtro de ação:** A compatibilidade dinâmica da sequência de eventos de cada usuário é analisada, filtrando aqueles que mostram um movimento incompatível com a física no tempo e no espaço. Ocorre com dispositivos que compartilham a mesma identificação ou se há antenas mal localizadas. Dependendo da fonte da incompatibilidade, o dispositivo ou a antena são removidos.

Após a aplicação desses filtros, o tamanho do banco de dados é reduzido a 10% do original, mas contém 100% das informações relevantes para a mobilidade.

A4.2 RECONSTRUÇÃO DA SEQUÊNCIA TEMPORAL DOS EVENTOS

As comunicações registradas pela rede móvel não estão associadas à posição GPS (mas sim às posições das antenas da rede), nem permitem a rastreabilidade contínua dos movimentos ao longo do tempo. A série temporal de eventos por dispositivo mostra saltos entre antenas às vezes de até várias horas e até vários quilômetros.

Para se obter uma visão completa da mobilidade, foi desenvolvido um procedimento de reconstrução a partir dos resultados de pesquisas realizadas no campo da socio-física, presumindo que a atividade demográfica humana responde ao princípio da entropia máxima. Aplicada em estudos de mobilidade, a distribuição mais provável para a trajetória de um usuário é aquela que maximiza a entropia utilizando as observações definidas pelos dados da operadora como condicionantes.



Fonte: Kido Dyamics

A4.3 CLASSIFICAÇÃO DE EVENTOS POR MOBILIDADE

Uma vez concluída a sequência temporal de eventos de mobilidade por usuário, as subsequências são agrupadas em estadias, micro viagens, viagens e macro viagens, com base em suas características cinemáticas.

As estadias e micro viagens são utilizadas na análise de visitas, ocupação e capacidade de público, enquanto as micro viagens, viagens e macro viagens são utilizadas na análise de tráfego. A seguir, estão as definições para cada um dos conceitos envolvidos.

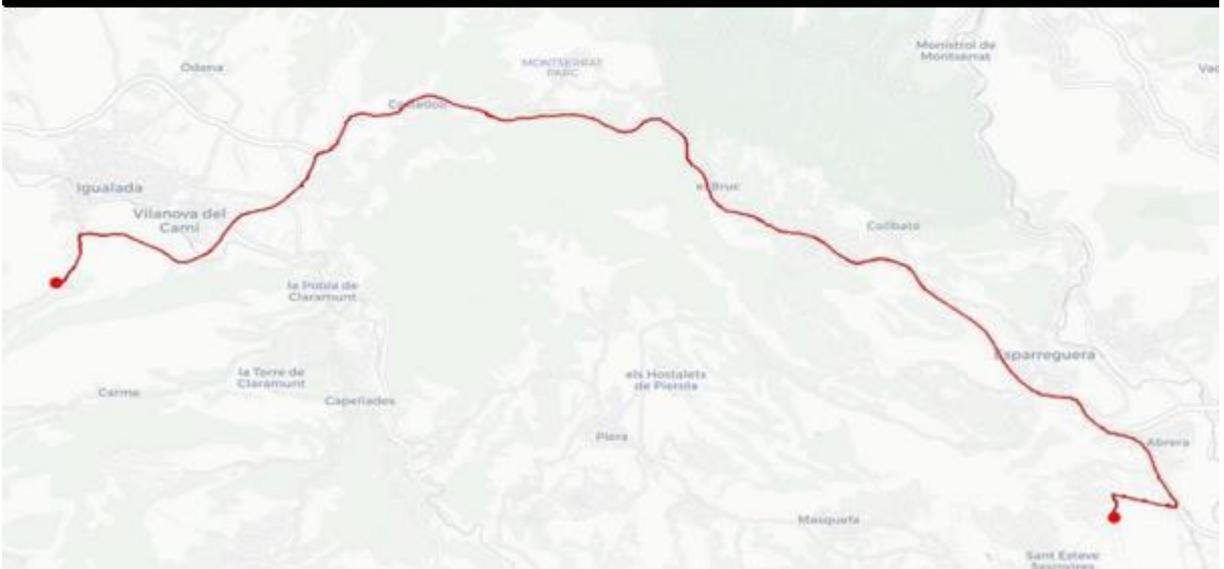
- i. **Viagem:** Uma sequência de eventos é classificada dentro da mesma viagem se o dispositivo tiver viajado pelo menos 5 km no intervalo de última hora. Este filtro indica que muito provavelmente esse referido deslocamento tenha sido realizado em um veículo motorizado.
- ii. **Micro viagem:** Uma sequência de eventos é classificada como uma micro viagem se houve um deslocamento, mas menor do que a condição de viagem do ponto anterior.
- iii. **Macro viagem:** Uma macro viagem é uma sequência de viagens e micro viagens que ocorreram em um intervalo de tempo entre eles menos do que a duração de seus componentes. Essa definição permite abranger viagens de longa distância que tiveram paradas ou viagens multimodais com intercâmbios.
- iv. **Estadia:** Uma estadia é uma sequência de eventos onde nenhum movimento ocorreu de acordo com as definições anteriores.

A4.4 PROJEÇÃO À INFRAESTRUTURA

Finalmente, é implementada uma nova camada tecnológica que possibilita saltar do local espacial com base em antenas para uma projeção baseada na malha viária da área estudada, permitindo detalhes sem precedentes no conhecimento de padrões de mobilidade.

Para isso, são gerados milhões de agentes virtuais que simulam a atividade observada nos eventos registrados pela operadora móvel, para filtrar aqueles que maximizam a probabilidade de suas ações no espaço e no tempo.

Após a reconstrução da série temporal de eventos por usuário, projetamos o movimento da infraestrutura de antenas para a infraestrutura viária de acordo com o caminho mais provável, em tempo e distância.



Fonte: Kido Dyamics

A5 PRIVACIDADE

Tanto durante o processo de análise de dados quanto de fornecimento dos resultados ao destinatário, um princípio básico é assegurado: a incapacidade de identificar uma pessoa física, seja a partir de seus dados de identificação ou de sua localização.

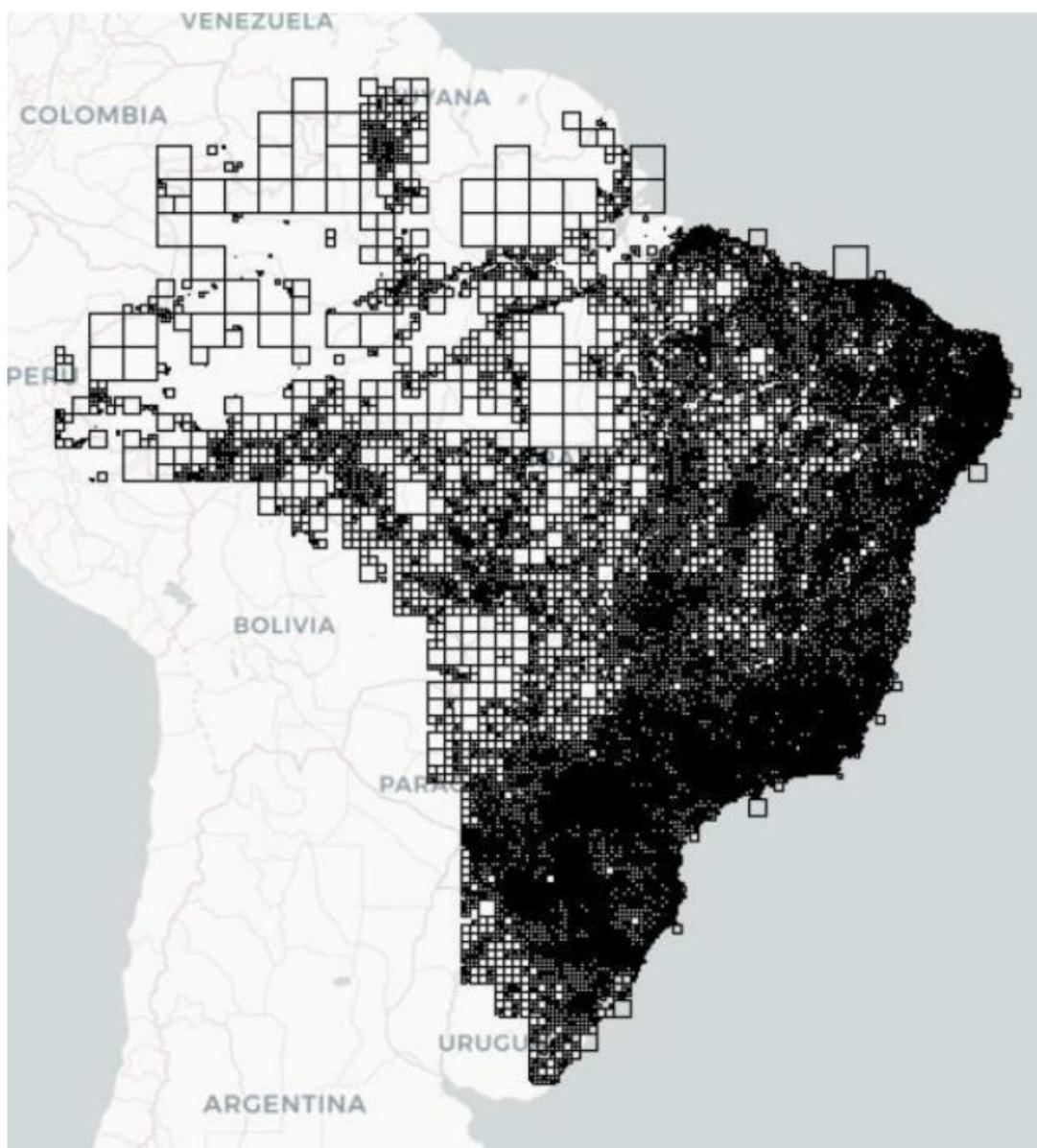
A garantia da privacidade do usuário passa por um processo de “anonimização” e agregação, o qual é constituído por uma sequência de quatro passos, como descritos a seguir:

- i. **Pseudoanonimização:** Os dados fornecidos pelo operador são previamente anonimizados (Hash ID, SHA 128 ou similar), sem qualquer possível acesso a dados pessoais, como número de telefone, endereço ou nome.
- ii. **Privacidade diferencial I:** anonimização individual completa. Ao projetar a sequência de eventos para a infraestrutura viária, os agentes virtuais que representam o movimento dos dispositivos se movem seguindo um componente aleatório em seu movimento, o que impossibilita a localização de pontos exatos como a casa ou o local de trabalho de forma única, impossibilitando o cruzamento de dados no nível individual.
- iii. **Agregação:** A contribuição de cada dispositivo anonimizado é agregada por segmentos em blocos de tamanho mínimo de 10 dispositivos. Em caso de agregados inferiores a 10, os resultados são eliminados.
- iv. **Privacidade diferencial II:** anonimização da agregação. Como medida adicional de segurança, adicionamos uma camada de aleatoriedade (ruído) aos resultados de agregação. Neste processo adicionamos a cada elemento um número aleatório da ordem de 20 dispositivos (uniforme entre -10 e +10).

A6 GEORREFERENCIAMENTO

O arranjo de antenas da rede móvel dentro de uma área geográfica de estudo não é homogêneo, e geralmente responde a questões de conectividade e densidade populacional. Portanto, é necessário desenvolver métodos que permitam que a granularidade espacial seja aumentada e padronizada aos níveis exigidos nos estudos de mobilidade.

A projeção da atividade de antenas para Grid de Resolução Espacial próprio foi alcançada através do desenvolvimento de um modelo de distribuição territorial próprio com base na demografia, na densidade dos nós de comunicação reais e a estrutura da rede móvel.



Fonte: Kido Dyamics

Conforme exposto na figura anterior, para praticamente todo o território nacional, o Grid é composto por milhares de divisões quadriculares, o que possibilita aumentar drasticamente o nível de desagregação das áreas de estudo de mobilidade em uma determinada área geográfica.

Também incorpora uma melhoria em relação à rigidez de ter que se adaptar à estrutura específica das áreas de cobertura da rede móvel da operadora (raio das antenas).

Este modelo, que tem sido utilizado para definir as áreas centrais e periféricas de qualquer país, possibilita aumentar a precisão nos detalhes da análise em áreas geográficas mais específicas.

A7 EXPANSÃO AMOSTRAL

Dado que as operadoras móveis não possuem uma participação de mercado igual a 100%, é necessário realizar uma extrapolação da amostra para representar toda a população analisada da forma mais confiável possível.

A7.1 ELEVAÇÃO DA AMOSTRA PARA A POPULAÇÃO TOTAL

Consiste na extrapolação dos resultados da amostra de dispositivos utilizados para a análise, para a população total de um determinado país. É feito seguindo a seguinte sequência de ações:

- i. Atribuição de residência:** Utilizando apenas os eventos associados às estadias, a cada dispositivo é atribuído o local de residência mais provável usando como referência os polígonos dos blocos censitários disponíveis em um determinado país. Os blocos censitários são projetados de forma a garantir um nível adequado de agregação em um determinado território que garanta privacidade.
- ii. Criação de um censo próprio:** Os resultados anteriores são adicionados em uma única lista para cada bloco censitário. Este censo é cruzado com os dados demográficos da lista de clientes da operadora, atribuindo uma pirâmide de sua própria população a cada bloco censitário.
- iii. Cruzamento do censo:** Para evitar um ajuste severo com o censo e o registro populacional (série anual), perder flutuações naturais (população flutuante) e outras dinâmicas demográficas relevantes, aplica-se um processo estatístico que reproduz a população em diferentes escalas de agregação, aplicando-se a cada bloco censitário vários raios populacionais (5'000, 10'000, 50'000, 100'000 e 500.000 habitantes), cruzados com a participação de mercado do operador, e é calculado para cada bloco censitário, por sexo e faixa etária, o fator de expansão necessário para equiparar a média populacional de cada raio. Isso compensa as flutuações locais na participação de mercado e capta mudanças demográficas em escalas de tempo inferiores a um ano.
- iv. Aplicação de pesos estatísticos:** Uma vez obtidos os fatores de escala para cada segmento demográfico, aplica-se aos agregados nas mesmas proporções para elevar os resultados ao total de habitantes em vez de dispositivos móveis.

A7.2EXTRAPOLAÇÃO DA AMOSTRA DE INTERNACIONAIS (ROAMERS)

Os dispositivos roaming-in (visitantes tipicamente internacionais em um determinado território) não possuem dados demográficos e apenas a nacionalidade de seu operador está disponível.

Neste caso, a expansão é obtida para cada nacionalidade utilizando dados disponíveis publicamente sobre visitantes internacionais como referência e para cada operador de acordo com acordos internacionais de roaming entre operadores.

