

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO TECNOLÓGICO DE JOINVILLE  
CURSO DE ENGENHARIA DE TRANSPORTES E LOGÍSTICA

MURILO COLIN DA SILVEIRA

DESENVOLVIMENTO DE UM MÓDULO EM ALTERYX COM BASE EM DADOS DE  
MODELO DE MACROSIMULAÇÃO EMME PARA ANÁLISE DE ALTERNATIVAS  
EM VIAS AUSTRALIANAS

Joinville

2023

MURILO COLIN DA SILVEIRA

DESENVOLVIMENTO DE UM MÓDULO EM ALTERYX COM BASE EM DADOS DE  
MODELO DE MACROSIMULAÇÃO EMME PARA ANÁLISE DE ALTERNATIVAS  
EM VIAS AUSTRALIANAS

Trabalho apresentado como requisito para obtenção do título de bacharel no Curso de Graduação em Engenharia de Transporte e Logística do Centro Tecnológico de Joinville da Universidade Federal de Santa Catarina.

Orientadora: Dra. Simone Becker Lopes

Joinville

2023

MURILO COLIN DA SILVEIRA

DESENVOLVIMENTO DE UM MÓDULO EM ALTERYX COM BASE EM DADOS DE  
MODELO DE MACROSIMULAÇÃO EMME PARA ANÁLISE DE ALTERNATIVAS  
EM VIAS AUSTRALIANAS

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de bacharel em Engenharia de Transportes e Logística, na Universidade Federal de Santa Catarina, Centro Tecnológico de Joinville.

Joinville (SC), 01 de dezembro de 2023.

**Banca Examinadora:**

---

Profa. Dra. Simone Becker Lopes  
Orientadora/Presidente

---

Prof. Dr. Rodrigo Castelan Carlson  
Membro  
Universidade Federal de Santa Catarina

---

Eng<sup>a</sup>. Raiza Benedecti  
Membro  
Urban Analytics and Complex Systems

Dedico este trabalho a minha família

## **AGRADECIMENTOS**

Agradeço primeiramente aos meus pais que me educaram e me guiaram por toda minha vida para chegar até aqui, amo vocês de uma maneira que não consigo expressar em palavras. Agradeço também a Leidiane por estar ao meu lado durante todo o processo de escrita e mais da metade da faculdade ao meu lado. Agradeço a professora Dra. Simone, a melhor professora da UFSC, e que me inspirou a gostar de absolutamente tudo que eu gosto na área de transportes.

Obrigado ao departamento de Transportes de Queensland Australia por ter fornecido os dados para essa pesquisa e pela UACS por ter me dado essa oportunidade e ideia de estudo.

"No caminho do progresso, cada alternativa é uma escolha, e cada escolha, uma possibilidade para moldar o futuro." - [Autor Desconhecido].

## RESUMO

Com uma previsão de crescimento de 22% para os próximos dez anos a Austrália, a fim de evitar congestionamentos nas vias devido ao aumento do número de veículos e viagens nas estradas, precisa desenvolver infraestruturas e estratégias às tomadas de decisões urbanas. Como alternativa ao problema, o governo australiano criou um modelo integrado de uso de solo e transportes, ao qual este trabalho auxilia em específico no seu módulo de custo e tomada de decisão. Apresenta-se o desenvolvimento de um método para integrar os resultados de macrossimulação de tráfego no software Emme e a análise de dados no software Alteryx Designer. Com o intuito de fazer a escolha de infraestrutura de menor custo e que traga os melhores benefícios para o sistema como um todo, na tomada de decisão em investimentos futuros. Atingiu-se o objetivo com um fluxo de trabalho que treina o modelo de *machine learning Random Forest* e realiza a previsão dos custos totais de infraestrutura dos dados de saída da macrossimulação. Depois é realizada análises de custo benefício para definir os melhores cenários a serem implementados. Este trabalho analisou dois cenários de infraestrutura para os próximos 23 anos do sudeste do estado de Queensland, Austrália, o cenário Referência e o Modificado. Apenas ocorrem diferenças na categoria de estradas e diferindo apenas nos anos de aplicação de algumas infraestruturas nas categorias restantes (vias de ônibus e trilhos). O modelo mostrou-se estável e com relativa precisão. Ambos os cenários tiveram as categorias de ônibus e trilhos muito similares o que indica a confiabilidade do modelo em relação a múltiplas aplicações com baixa variabilidade de predições. O desempenho geral do Coeficiente de Determinação ( $R^2$ ) esteve em uma faixa entre 0,71 e 0,95 para estradas, 0,69 e 0,75 para ônibus, e 0,83 e 0,91 para modos de trilhos. Esses valores de  $R^2$  foram considerados satisfatórios perante os *stakeholders*. O cenário Modificado acabou sendo o preferido nesse estudo com uma diferença de custo economizado por tempo de viagem de A\$0,00772 hora/usuário/km. O valor é baixo, mas deve-se lembrar que o custo é por usuário da rede e por quilômetro de viagem.

**Palavras-chave:** Machine Learning. Alteryx. Custo de Infraestrutura.

## ABSTRACT

With a projected growth of 22% over the next ten years, Australia, in an effort to mitigate traffic congestion arising from the increased number of vehicles and road travels, needs to develop infrastructure and strategies for urban decision-making. As an alternative to this issue, the Australian government has established an integrated model for land use and transportation, to which this work specifically contributes in its cost and decision-making module. This involves the development of a method to integrate the results of traffic microsimulation using the Emme software and data analysis through the Alteryx Designer software. The goal is to select infrastructure with the lowest cost and the greatest benefits for the overall system, aiding in future investment decision-making. The objective is achieved through a workflow that trains the Random Forest machine learning model and predicts the total infrastructure costs from the output data of the microsimulation. Subsequently, cost-benefit analyses are conducted to define the best scenarios for implementation. This study examined two infrastructure scenarios for the next 23 years in the southeast of Queensland, Australia: The Reference scenario and the Modified scenario. Differences are observed solely in the road category, with variations in the years of application for some infrastructures in the remaining categories (bus lanes and rails). The model demonstrated stability and relative accuracy. Both scenarios exhibited very similar bus and rail categories, indicating the model's reliability across multiple applications with low prediction variability. The overall performance of the Coefficient of Determination ( $R^2$ ) ranged between 0.71 and 0.95 for roads, 0.69 and 0.75 for buses, and 0.83 and 0.91 for rail modes. These  $R^2$  values were deemed satisfactory by stakeholders. The Modified scenario emerged as the preferred option in this study, with a cost difference saved per travel time of A\$0.00772 hour/user/km. While the value may seem low, it is essential to note that the cost is per network user and per kilometer traveled.

**Keywords:** Machine Learning. Alteryx. Infrastructure Costing.



## LISTA DE FIGURAS

Figura 1 - Framework do Modelo LUTI do TMR .....	17
Figura 2 - Interface do Software Alteryx Designer .....	22
Figura 3 - Exemplo de Resultado para uma Regressão Linear .....	26
Figura 4 – Metodologia .....	36
Figura 5 – Shape com vias do Sudeste de Queensland.....	38
Figura 6 – Links Utilizados Para Treinamento .....	39
Figura 7 - Parte da Rede de Dados para Realização da Inferência.....	41
Figura 8 - Fluxo de Trabalho para Seleção de Duplicatas .....	42
Figura 9 - Fluxo para o Ano 2031 na Seleção de Duplicatas.....	43
Figura 10 - Ferramenta Filtro do Alteryx Designer.....	44
Figura 11 - Fluxo de Trabalho para Filtragem dos Dados .....	45
Figura 12 - Modelo com Ferramentas de Predição.....	46
Figura 13 - Erro de Acordo com o Número de Árvores.....	46
Figura 14 - Validação da Categoria Trilhos .....	52
Figura 15 - Validação da Categoria Estrada .....	53
Figura 16 - Validação da Categoria Ônibus.....	53

## LISTA DE QUADROS

Quadro 1 - Conjunto de Categorias de Ferramentas do Software Alteryx....23

## LISTA DE TABELAS

Tabela 1 – Resultados dos Coeficientes de Cada Método .....	51
Tabela 2 - Resultados da Previsão de Custos por LGA Cenário Referência	54
Tabela 3 - Custos Totais Previstos por Ano Cenário Referência.....	56
Tabela 4 - Custos Previstos por LGA Cenário Modificado.....	57
Tabela 5 - Custo Total por Ano Cenário Modificado .....	59
Tabela 6 - Custo Economizado por Viagem .....	60

## LISTA DE ABREVIATURAS E SIGLAS

ABS – *Australian Bureau of Statistics*

ACB – Análise Custo Benefício

CREA – Conselho Regional de Engenharia e Agronomia

LGA – Governos Locais

LGIP – *Local Government Infrastructure Plans*

LUTI – *Land use and Transport Integration*

NSW – Nova Gales do Sul

QTRIP – *Queensland Transport and Roads Investment Program*

R<sup>2</sup> – Coeficiente de Determinação

SIG – Sistemas de Informações Geográficas

TMR – *Department of Transport and Main Roads*

VBTT – Valor do Tempo de Viagem Simplificado

VTTS – Valor do Tempo de Viagem

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>13</b>
1.1. OBJETIVOS .....	15
1.1.1. <b>Objetivo Geral</b> .....	<b>15</b>
1.1.2. <b>Objetivos Específicos</b> .....	<b>15</b>
<b>2. FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>16</b>
2.1. MODELOS DE INTEGRAÇÃO DE USO DO SOLO E TRANSPORTES .....	16
2.2. MACROSSIMULAÇÃO DE TRANSPORTES .....	17
2.3. ANÁLISE DE DADOS .....	18
2.3.1 <b>Análise, Coleta e Organização de dados</b> .....	<b>19</b>
2.3.2 <b>Alteryx como Ferramentas de Análise de Dados</b> .....	<b>21</b>
2.4. ALGORÍTMOS DE PREVISÃO .....	24
2.4.1 <b>Regressão Linear</b> .....	<b>25</b>
2.4.2 <b>Floresta Aleatória (Random Forest)</b> .....	<b>27</b>
2.4.3 <b>Método Spline</b> .....	<b>28</b>
2.5. ANÁLISE ECONÔMICA EM INFRAESTRUTURA DE TRANSPORTES .....	29
2.5.1 <b>Análise Custo Benefício</b> .....	<b>30</b>
2.5.1.1. <u>Economia de Tempo de Viagem</u> .....	<u>31</u>
<b>3. METODOLOGIA</b> .....	<b>35</b>
3.1. AQUISIÇÃO DOS DADOS .....	37
3.2. DADOS PARA INFERÊNCIA .....	40
3.3. TRATAMENTO DOS DADOS DE ENTRADA .....	41
3.4. DESENVOLVIMENTO DO MODELO BASE DE TREINAMENTO .....	43
3.5. ANÁLISE DE RESULTADOS .....	47
<b>4. ANÁLISE DE RESULTADOS</b> .....	<b>50</b>
4.1. <b>Métodos Preditivos</b> .....	<b>50</b>
4.2. <b>Inferências nos Cenários</b> .....	<b>54</b>
4.3. <b>Análise de Benefícios</b> .....	<b>59</b>
<b>5. CONCLUSÃO</b> .....	<b>62</b>
<b>REFERÊNCIAS</b> .....	<b>64</b>

## 1. INTRODUÇÃO

A Austrália é o maior país da Oceania e o sexto maior do mundo em área, com 7.741.220 km<sup>2</sup>, e, segundo a *Australian Bureau of Statistics* (ABS, 2021), tem apenas 25.750.198 habitantes, o que o coloca na 55ª posição em população no mundo. Porém, prevê-se para os próximos dez anos um crescimento de cerca de 22% da população para 31.433.000 habitantes (ABS, 2018), fato que ocorre em função da migração que o país recebe historicamente, apesar dos baixos índices de fecundidade (ABS, 2021).

Com maior número de pessoas, mais veículos e viagens acontecerão, resultando em congestionamentos e redes de transporte inadequadas que se tornam mais aparentes neste século (AUSTRÁLIA, 2014). Na Austrália, de acordo com o *Department of Transport and Main Roads* (TMR) (AUSTRÁLIA, 2022a) cerca de A\$30 bilhões serão investidos na infraestrutura de transportes até 2025 para mitigar os problemas de mobilidade nos sistemas do país.

Com o avanço dos computadores, a habilidade de simular diversos ambientes em poucos minutos ou horas fez com que a atividade se popularizasse. A capacidade de emular a variabilidade do tempo de um fenômeno de tráfego, torna a simulação uma ferramenta ideal na captura da complexidade dos sistemas reais. As simulações de tráfego podem ser divididas em macro, meso ou micro simulação, sendo a primeira indicada para medidas que influenciam grandes regiões de tráfego e a última para mudanças de influência local devido ao grau de agregação de seus dados (BARCELÓ, 2010; KOTUSEVSKI; HAWICK, 2009).

Para desenvolver um sistema viário estruturado e planejado a fim de garantir o bem-estar dos condutores com estradas não congestionadas e da população com sistemas eficientes, o TMR da Austrália, em parceria com a empresa *Urban Analytics & Complex Systems* (UACS), integra as simulações com modelos *Land Use and Transport Interaction* (LUTI) (TMR, 2023). Nesse sentido, esse estudo têm o intuito de contribuir para o último módulo do modelo LUTI que está sendo desenvolvido pelo TMR australiano: *Option Analysis*.

Segundo Lopes (2010), um modelo LUTI dinâmico modela o uso do solo e transportes de forma integrada, sendo a ferramenta adequada para o apoio à tomada de decisões no planejamento da mobilidade sustentável de alguma região de

interesse. Se as relações de dependência entre os dados forem consideradas desde o início, a ferramenta tem o potencial de produzir resultados eficazes para realização de previsões no tráfego.

Considerando isso, utilizou-se dados demográficos, estruturais e de tráfego fornecidos por modelagem da macrossimulação e alocação de tráfego produzida pelo TMR em conjunto com o modelo LUTI nas redes de tráfego australianas fornecidas. Este trabalho buscou estudar duas redes alternativas de tráfego desenvolvidas pelo departamento de transportes, a rede chamada Referências e a Modificada.

Além dessas, também foi fornecido uma rede para comparação chamada Base que representa o cenário atual das vias. Dados foram coletados da ferramenta computacional Emme da empresa Inro, e foram utilizados como entrada para desenvolver uma ferramenta de análise de custos de infraestrutura e custo benefício em conjunto com a UACS.

Para a tomada de decisão, foi criado um algoritmo com auxílio da ferramenta Alteryx de análise de dados, conectando os dados obtidos por meio da macrossimulação do Emme com os dados socioeconômicos do LUTI australiano. Partindo da premissa que é possível utilizar ferramentas de *machine learning* como regressão linear, método *spline* e floresta aleatória, para realização da previsão de custos com suas diversas possíveis variáveis complexas.

O *Options Analysis* utiliza dessas ferramentas de regressão e machine learning, além de métodos clássicos de análise de atendimento modal e de impactos ambientais baseados em emissões de tráfego. Conforme CREA (2020), o modelo criado por esse trabalho pode realizar um recorte da análise custo benefício definida pelo órgão, subsidiando a escolha de infraestruturas com menor custo com uma determinada confiança matemática e que supram as necessidades de capacidade das vias.

O TMR demonstrou em reuniões a importância do modelo, tendo em vista como a ferramenta pode fornecer de forma rápida e barata uma análise de alto nível sobre custos e comparação de cenários. A análise proposta pelo *Option Analysis* tem chamado a atenção do governo Australiano que vê o modelo como um grande avanço no processo de tomada de decisão da Austrália e do mundo.

## 1.1. OBJETIVOS

Para resolver a problemática sobre quais opções de infraestrutura devem ser adotadas na rede rodoviária australiana e seus custos de implementação, propõe-se os seguintes objetivos.

### 1.1.1. Objetivo Geral

Desenvolver um módulo que ligue os dados de saída do software Emme em um fluxo de trabalho em Alteryx Designer para análise de alternativas sobre investimentos em cenários futuros de infraestruturas viárias australianas.

### 1.1.2. Objetivos Específicos

- Definir os dados de saída da rede de simulação de transportes que serão novos dados de entrada para o modelo econômico;
- Criar modelo de previsão de custos no Alteryx conforme dados de entrada do modelo de macrossimulação Emme utilizando métodos de regressão e *machine learning*;
- Desenvolver método de avaliação de infraestruturas com utilização do modelo de custos e impactos viários;
- Avaliar diferentes cenários de futura infraestrutura com base nos custos e impactos obtidos.



## 2. FUNDAMENTAÇÃO TEÓRICA

Para desenvolver um módulo de integração entre softwares de simulação de tráfego e dados socioeconômicos, é preciso entender os principais conceitos relacionados a transportes e suas infraestruturas. Neste capítulo serão apresentados os principais conceitos relacionados aos objetivos do trabalho como, modelos de integração de uso do solo e transportes, macrossimulação, análise de dados, integração entre softwares e análise econômica em infraestrutura de transportes.

### 2.1. MODELOS DE INTEGRAÇÃO DE USO DO SOLO E TRANSPORTES

Pfaffenbichler (2003) afirma que métodos clássicos de previsão de uso do solo possuem diversas suposições feitas pelo modelo: as cidades são circulares, toda a população mora nas periferias e trabalha no centro e custos de transportes crescem uniformemente conforme se afasta. Enquanto os métodos de previsão de transporte assumem caracteres estáticos perante os diferentes tipos de usos do solo. No entanto, ambos formam um sistema dinâmico, integrado e são necessários para avaliar o desempenho de políticas urbanas sustentáveis (PFAFFENBICHLER, 2003).

Segundo Wegener (2003), os modelos LUTI integram o sistema de transporte e o uso do solo e vêm sendo desenvolvidos com objetivo de estimar padrões no desenvolvimento dessas áreas nas cidades do século 21. Além disso, são utilizados como alternativa moderna e eficaz na previsão de demanda associada ao modelo quatro etapas. Seu principal atrativo perante os modelos tradicionais é o fato de gerar suas próprias previsões conforme mudanças são feitas no sistema (MARTINEZ; ARAYA, 1998).

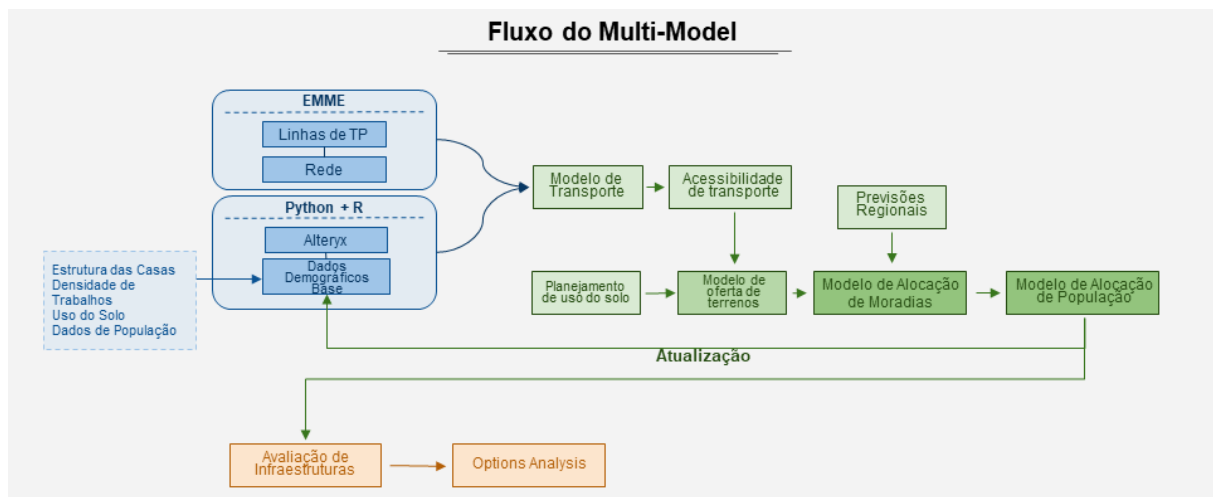
Modelos LUTI utilizam não só dados de transportes, mas, também, de uso do solo para compreender a situação atual, prever o desenrolar de situações futuras e auxiliar na tomada de decisões para a solução dos problemas existentes (WERNECK, 2015). A análise e a modelagem da interação entre esses dados são os aspectos mais delicados da integração entre as capacidades desses sistemas (PIETRANTONIO; STRAMBI; GUALDA, 1996).

Lopes (2010) analisa que modelos LUTI dinâmicos são a ferramenta mais indicada para análise de políticas de mobilidade urbana sustentáveis, pois as

incorpora com as dinâmicas de desenvolvimento urbano, produzindo o entendimento dos impactos das relações dinâmicas entre características de uso do solo e sistemas de transportes nas cidades. Acredita-se que, se as relações de dependência entre os dados forem consideradas desde o início, a ferramenta deve produzir resultados eficazes para realização de previsões no tráfego.

Este trabalho representa o último módulo do modelo LUTI do TMR australiano, utilizando seus dados de saída para realizar a análise de custos e benefícios dentro do seu escopo (Figura 1).

Figura 1 - Framework do Modelo LUTI do TMR



Fonte: Autor (2023).

Como parte do módulo de Avaliação de Infraestrutura, o *Options Analysis* (projeto desse trabalho) é uma parte separada do todo, e busca auxiliar o governo do estado de Queensland com a tomada de decisão em conjunto com o modelo de uso do solo e transportes. Além de dados e da integração com o modelo LUTI, os dados de macrossimulação de tráfego também foram utilizados advindos do software Emme.

## 2.2. MACROSSIMULAÇÃO DE TRANSPORTES

A modelagem de tráfego é uma área de estudo essencial para entender e prever o comportamento do fluxo de veículos em uma determinada rede de transporte. De acordo com Chowdhury, Santen e Schadschneider (2000), a modelagem de tráfego busca analisar e simular o fluxo de veículos, considerando fatores como

origem e destino das viagens, capacidade das vias, interações entre veículos, comportamento dos motoristas, entre outros.

Uma das abordagens mais comuns na modelagem e análise de tráfego é a macrossimulação, que envolve a análise de grandes áreas geográficas e longos períodos de tempo. Concentrado na simulação de tráfego em larga escala, o método permite avaliar o desempenho do sistema de transporte em termos de congestionamento, tempos de viagem e capacidade das vias ao considerar fatores como interações entre veículos e origem e destino das viagens (CHOWDHURY, et al. 2000; MAHMASSANI, 2001).

A utilização de técnicas de macrossimulação oferece uma compreensão abrangente dos padrões de tráfego e permite avaliar o desempenho de redes de transporte em diferentes cenários. Mahmassani (2001) destaca sua importância no estudo da capacidade das vias, tempos de viagem, congestionamentos e eficiência do sistema de transporte.

Para realização dessas técnicas, o software Emme, desenvolvido pela INRO Systems, é amplamente utilizado como uma ferramenta especializada. Este oferece recursos avançados para modelagem e simulação de tráfego, permitindo a criação de redes virtuais, análise de fluxo de veículos e avaliação de políticas de transporte (INRO Systems, 2021).

Ao combinar a macrossimulação de tráfego com sistemas de informações geográficas (SIG), é possível obter uma visão espacialmente explícita dos padrões de tráfego e das características das vias. Os SIG permitem a coleta, armazenamento, análise e visualização de dados geoespaciais relacionados ao tráfego (Longley *et al.*, 2015). A integração de ambos é essencial para uma análise abrangente e detalhada dos sistemas de transporte. O conjunto também desempenha um papel crucial na seleção de infraestruturas de transporte adequadas.

### 2.3. ANÁLISE DE DADOS

A análise de dados em estatística é uma disciplina essencial para a tomada de decisões informadas em uma ampla gama de setores, incluindo o setor de transportes. Segundo Cowie e Ison (2017), a análise estatística de dados de transporte é fundamental na compreensão dos padrões de deslocamento, na

identificação de tendências e na avaliação de políticas e projetos relacionados ao transporte.

De acordo com Lindner (2015), a análise de dados em estatística desempenha um papel crucial na modelagem e previsão de demanda de transporte, que é essencial para o planejamento e dimensionamento adequados de infraestruturas de transporte. Através da análise estatística de dados históricos de viagens, pode-se identificar padrões sazonais, tendências de crescimento, impactos de eventos especiais e comportamentos de viagem que são fundamentais para prever a demanda futura.

Além disso, a análise de dados em estatística também é utilizada na avaliação de impactos de políticas e projetos de transporte. Segundo McKinnon, Button e Nijkamp (2003), a análise estatística permite comparar indicadores-chave antes e depois da implementação de uma política ou projeto, permitindo uma avaliação precisa de seus efeitos sobre o desempenho do sistema de transporte, a economia e o meio ambiente.

Existem várias ferramentas poderosas disponíveis para auxiliar na coleta, organização e análise dos dados. Alguns dos principais softwares utilizados nessa área incluem o Alteryx, o R e o Python. O Alteryx, por exemplo, oferece recursos avançados para a manipulação e transformação de dados, permitindo a aplicação de técnicas estatísticas e a criação de modelos preditivos (Alteryx, 2023).

### **2.3.1 Análise, Coleta e Organização de dados**

Bussab e Morettin (2017) oferecem uma perspectiva abrangente sobre a coleta de dados, destacando sua importância como etapa fundamental em qualquer processo estatístico. Os autores enfatizam que a qualidade dos dados coletados é essencial para obter resultados estatisticamente significativos. Eles descrevem métodos de amostragem, enfatizando a necessidade de escolher adequadamente uma amostra representativa da população de interesse. Além disso, o livro aborda a criação de questionários e entrevistas, enfatizando a formulação de perguntas claras e objetivas para minimizar vieses na coleta de dados.

Além disso, Bussab e Morettin (2017) destacam a ética na coleta de dados, salientando a importância de obter o consentimento informado dos participantes e de garantir a privacidade das informações coletadas. Eles também abordam questões de

confiabilidade dos dados, ressaltando a necessidade de verificar a consistência dos dados coletados e a importância de eliminar valores atípicos ou discrepantes.

Após a coleta, a organização dos dados é essencial para facilitar sua análise. Gareth James, *et al.* (2013) ressaltam a importância da exploração de dados. Antes de aplicar métodos estatísticos avançados, é crucial entender os dados em sua forma bruta. Isso envolve a visualização e resumo dos dados para identificar padrões, tendências e possíveis outliers. A análise exploratória de dados ajuda a guiar a seleção de técnicas estatísticas apropriadas e a evitar conclusões precipitadas.

Nesse sentido, as técnicas de gerenciamento e estruturação de dados ganham destaque. Provost e Fawcett (2013) ressaltam que a organização adequada dos dados em tabelas ou bancos de dados é fundamental para garantir sua integridade e permitir a aplicação de análises estatísticas com a limpeza de dados. Enfatiza-se que os dados geralmente contêm erros, valores ausentes e inconsistências que podem distorcer as análises. Portanto, a primeira etapa no processo de gerenciamento de dados é identificar e corrigir esses problemas.

A análise exploratória de dados desempenha um papel crucial na identificação de padrões e tendências nos dados. Hoaglin *et al.* (2006), abordam a importância da análise de tendências e formas (shapes) nos dados. Eles destacam como os padrões ao longo do tempo ou em diferentes grupos podem ser cruciais para a compreensão do comportamento dos dados. A análise de tendências ajuda a identificar mudanças e desenvolvimentos ao longo do tempo, enquanto a análise das formas dos dados revela distribuições e variações subjacentes.

Além disso, Gareth James, *et al.* (2013), enfatiza que a relevância da modelagem estatística e do *machine learning* na análise de dados. Discute-se a aplicação de modelos de regressão, classificação, agrupamento e outros métodos para extrair informações significativas dos dados. São abordados tópicos como validação de modelos, seleção de recursos e avaliação de desempenho, destacando a importância de avaliar a qualidade das análises. Também destaca a flexibilidade das técnicas de análise de dados e como elas podem ser aplicadas em uma variedade de campos, desde ciências sociais e econômicas até ciências naturais e engenharia.

### 2.3.2 Alteryx como Ferramentas de Análise de Dados

Existem diversos softwares disponíveis para realizar análises estatísticas e facilitar o processo de análise de dados. Nesta seção, apresenta-se alguns dos principais softwares utilizados nessa área, incluindo o Alteryx, mencionado anteriormente.

Uma dessas ferramentas é o Microsoft Excel que oferece recursos para importar, manipular e visualizar dados, bem como a aplicação de fórmulas e funções estatísticas para realizar análises básicas (Microsoft, 2021). Apesar de suas limitações em lidar com grandes volumes de dados e em realizar análises estatísticas avançadas, o Excel ainda é amplamente utilizado em várias áreas de pesquisa e negócios devido à sua acessibilidade e facilidade de uso.

Por outro lado, o R é uma linguagem de programação estatística amplamente reconhecida e utilizada para a análise de dados. Com uma vasta coleção de pacotes estatísticos disponíveis, o R oferece uma ampla gama de funções e métodos para realizar análises avançadas, modelagem estatística e visualizações sofisticadas (R Core Team, 2021).

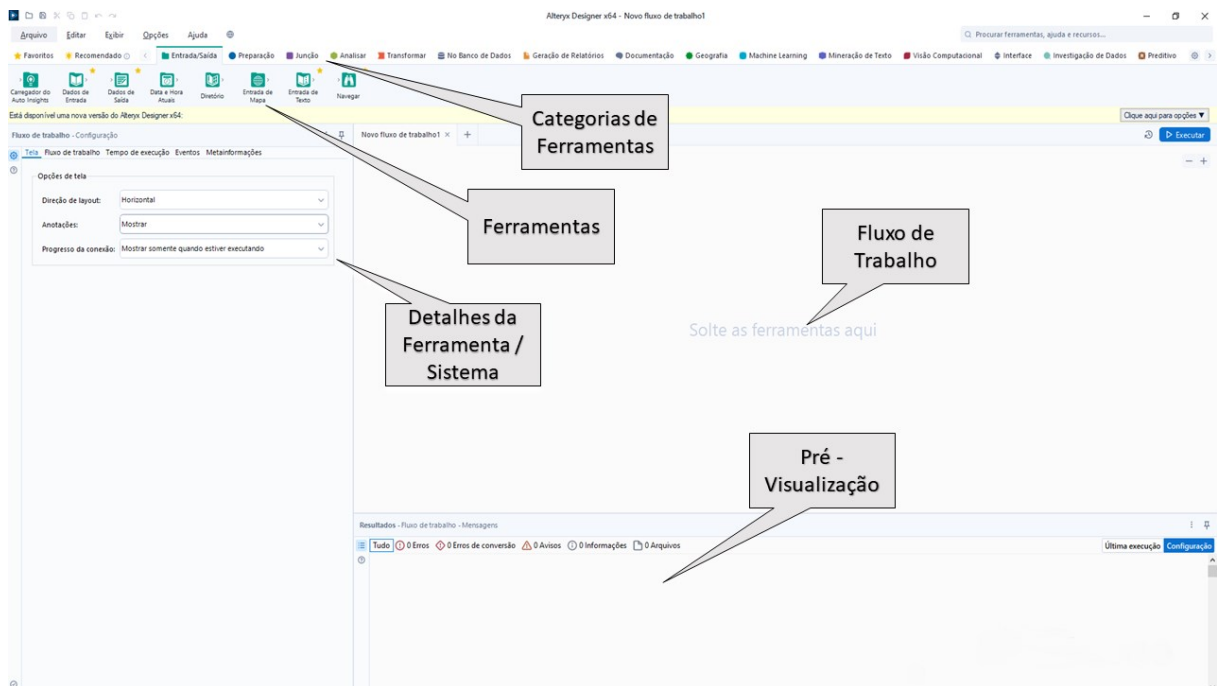
Ainda segundo R Core Team (2021), uma das vantagens do R reside na facilidade com que é possível criar gráficos bem elaborados e de alta qualidade para publicação, inclusive incorporando símbolos matemáticos e fórmulas, quando necessário. O R é amplamente adotado em ambientes acadêmicos e científicos devido à sua capacidade de lidar com análises complexas e reproduzíveis.

Da mesma forma, o Python tem ganhado popularidade como uma poderosa ferramenta de análise de dados. Com bibliotecas como Pandas e NumPy, ele oferece recursos para importar, manipular e analisar dados com eficiência (Python Software Foundation, 2021). Sua capacidade de realizar análises estatísticas avançadas, desenvolver modelos preditivos e construir aplicativos interativos faz do dele escolha popular entre pesquisadores e analistas de dados.

Utilizando de todas essas ferramentas, integrando-as e otimizando seus processos, o software Alteryx tem grande competitividade na área. Segundo a empresa Alteryx (2023), o software oferece uma interface intuitiva e recursos avançados para realizar tarefas complexas, como limpeza de dados, agregação, combinação de fontes e aplicação de técnicas estatísticas.

Uma das vantagens da ferramenta é a sua interface visual (Figura 2), que permite criar fluxos de trabalho arrastando e soltando ferramentas e conectando-as para realizar diferentes etapas da análise de dados. Essa abordagem permite aos usuários visualizarem e compreenderem facilmente as transformações e análises aplicadas aos dados, facilitando a exploração (Alteryx, 2023).

Figura 2 - Interface do Software Alteryx Designer



Fonte: Autor (2023).

Segundo Alteryx (2023), existem, no total, 28 categorias de ferramentas na interface do software (Quadro 1). Cada uma dessas categorias desempenha funções e tem propósitos distintos, abrangendo ferramentas para auxiliar na entrada e saída de dados, preparação, manipulação, análise e até mesmo documentação.

Quadro 1 - Conjunto de Categorias de Ferramentas do Software Alteryx

Categorias de Ferramentas	Descrição
Visão Computacional	Use o Designer para inserir, processar e analisar imagens e, depois, gerar a saída delas.
Conectar	Ferramentas para o Alteryx Connect.
Mineração de Texto	Analise os dados do texto.
Ferramentas de entrada/saída	Forneça entradas e saídas para os fluxos de trabalho.
Machine Learning	Crie modelos de aprendizado de máquina.
Preparação	Prepare os dados para análise posterior.
Junção	Combine fluxos de dados, agregando-os a um esquema.
Analisar	Separe os valores de dados em um esquema de tabela padrão.
Transformar	Sumarize ou reorganize os dados.
No Banco de Dados	Conecte uma base de dados para combinar e visualizar os dados.
Geração de Relatórios	Ajuda para a apresentação e organização dos dados.
Documentação	Melhore a apresentação de um fluxo de trabalho adicionando anotações e organizando ferramentas.
Geografia	Manipulação e processamento espacial de dados e edição de objetos geográficos.
Ferramentas de Interface	Projete elementos de interface de usuário para aplicativos e macros.
Investigação de Dados	Entenda os dados a serem usados em um projeto de análise preditiva.
Preditivo	Modelagem preditiva, comparação de modelos e ferramentas para teste de hipóteses
Teste AB	Realize experimentos de teste A/B.
Série Temporal	Plotagem de gráficos de séries temporais regulares e univariadas e ferramentas de previsão.
Agrupamento Preditivo	Agrupe registros ou campos em um número menor de grupos.
Prescritivo	Determine a melhor estratégia ou resultado para diferentes situações.
Conectores	Recupere ou envie dados para a nuvem ou Internet.
Endereço	Padronize listas de endereços e realize a geocodificação para o nível de código postal com 9 dígitos.
Análise Demográfica	Extraia dados utilizando o mecanismo do Allocate dentro do Alteryx.
Calgary	Recupere dados de contagem listados e realize análises em bases de dados de grande escala.
Desenvolvedor	Crie macros e aplicativos analíticos e execute programas externos.
Laboratório	Ferramentas que não são para uso em produção.
Exemplos de SDK	Exemplos de ferramentas de SDK

Fonte: Alteryx Designer (2023).



Este trabalho se concentra exclusivamente nas categorias de Ferramentas de Entrada/Saída, que são utilizadas para inserir e exportar dados, bem como em ferramentas como 'Navegar', que auxilia na compreensão e validação dos resultados; 'Preparação', que inclui a ferramenta 'Filtro' para a filtragem de dados; 'Criar Amostra', que é empregada para separar dados de treinamento e validação; e a ferramenta 'Fórmula', que possibilita a realização de alterações nos dados com base em fórmulas criadas pelo usuário. Além disso, também são abordadas as categorias 'Junção', usada para unir dados e realizar junções, e as 'Ferramentas Preditivas', que englobam os algoritmos de previsão.

#### 2.4. ALGORÍTMOS DE PREVISÃO

O *machine learning* é uma área da inteligência artificial que se concentra no desenvolvimento de algoritmos que podem aprender com dados e fazer previsões ou tomar decisões com base nesses dados. Ele se baseia em modelos estatísticos e matemáticos para identificar padrões em grandes conjuntos de dados. A ideia por trás do *machine learning* é que, ao aprender com exemplos e observações, os algoritmos possam ser treinados para reconhecer padrões e fazer previsões precisas para novos dados (JAMES *et al.*, 2017).

De acordo com Alpaydin (2010), seus principais métodos incluem o aprendizado supervisionado, não-supervisionado e por reforço. O aprendizado supervisionado, por exemplo, é utilizado quando se tem um conjunto de dados rotulados e se deseja prever novas observações. Já o aprendizado não-supervisionado é usado para encontrar padrões e estruturas nos dados, sem a necessidade de rótulos prévios. E o aprendizado por reforço é empregado em problemas que envolvem tomadas de decisão sequenciais, como jogos e robótica.

Os algoritmos não-supervisionados, diferentemente do aprendizado supervisionado, têm o objetivo de identificar regularidades e agrupamentos nos dados. Entre os principais algoritmos de aprendizado não-supervisionado estão a análise de componentes principais, o agrupamento hierárquico, o *K-Means* e os mapas auto-organizáveis. Esses são amplamente utilizados em áreas como reconhecimento de padrões, análise de imagens e processamento de linguagem natural, entre outras (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; ALPAYDIN, 2010).

Os algoritmos de aprendizado por reforço visam encontrar a melhor ação a ser tomada em um determinado estado para maximizar uma recompensa. Eles são frequentemente usados em jogos, robótica e controle de processos. O algoritmo *Q-learning* é um exemplo clássico do método. Outros algoritmos populares incluem a política de gradiente e o *Actor-Critic*. O aprendizado por reforço pode ser desafiador, pois envolve a tomada de decisões em um ambiente incerto e dinâmico, o que requer um equilíbrio entre a exploração e a exploração de informações (SUTTON; BARTO, 2018; FRANCOIS-LAVET; et al., 2018).

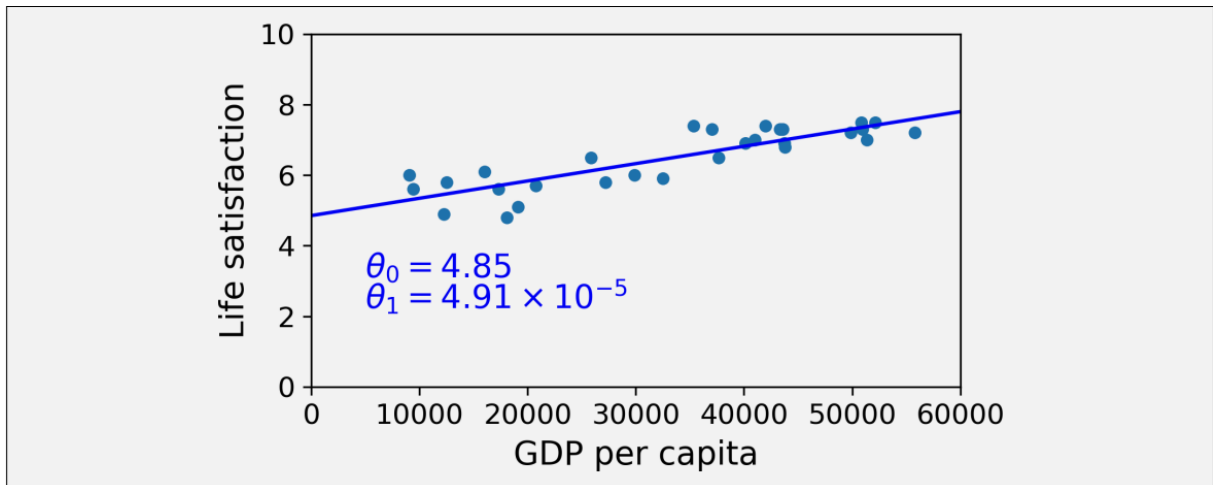
Este estudo trabalha apenas com o aprendizado supervisionado, devido a sua característica de ele ser treinado a partir de um conjunto de dados rotulados, ou seja, dados em que as saídas desejadas são conhecidas. Isso permite que o algoritmo possa aprender a prever novas saídas para dados de entrada não vistos anteriormente. Alguns exemplos de algoritmos de aprendizado supervisionado são a regressão linear, o método *Spline* e as árvores de decisão e sua variante Floresta Aleatória ou Random Forest (ALPAYDIN, 2010).

#### **2.4.1 Regressão Linear**

Para Hastie *et al.* (2009), a regressão linear é um método de aprendizado de máquina que visa modelar a relação entre uma variável dependente e uma ou mais variáveis independentes por meio de uma equação linear. Ela é amplamente utilizada na área de análise de dados para prever ou explicar a relação entre variáveis quantitativas. Segundo Alpaydin (2010), a regressão linear é um modelo simples e fácil de interpretar, o que a torna uma das técnicas mais utilizadas em análise de dados. O modelo assume que a relação entre as variáveis é linear e pode ser expressa por uma equação matemática simples.

Existem vários tipos de regressão linear, cada um adequado para diferentes situações. A regressão linear simples é usada quando há apenas uma variável independente, enquanto a regressão linear múltipla é utilizada quando há mais de uma variável independente. O processo de ajustar a linha de regressão é realizado por meio da minimização da soma dos quadrados dos resíduos entre os valores observados e os valores previstos pela equação de regressão (GÉRON, 2018). De acordo com Hastie *et al.* (2009), a minimização é feita através do método dos mínimos quadrados, que busca encontrar a linha que melhor se ajusta aos dados (Figura 3).

Figura 3 - Exemplo de Resultado para uma Regressão Linear



Fonte: GÉRON, A. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (O'Reilly Media, 2017), p. 100.

Segundo Géron (2017), apesar de sua simplicidade, a regressão linear é uma técnica poderosa para análise de dados e pode ser aplicada em uma ampla variedade de problemas. Ela pode ser usada para prever valores futuros de uma variável, avaliar a importância de uma variável em relação a outra, identificar possíveis valores atípicos nos dados, entre outras aplicações.

A Regressão Gama é um tipo de modelo linear generalizado, ou seja, um tipo de regressão simples ou múltipla que possui um comportamento não linear. Esse método utiliza variáveis contínuas e positivas quebrando assim algumas das limitações do modelo linear (Pérez, 2022).

No entanto, é importante ter em mente que a regressão linear tem suas limitações e pressupostos, e deve ser utilizada com cuidado e acompanhada de análises adicionais. Hastie, *et al.* (2009) cita que existem algumas limitações da regressão linear que devem ser consideradas ao usá-la para análise de dados, tais como:

- Pressupostos: A regressão linear pressupõe que os dados sigam uma distribuição normal e que a relação entre as variáveis seja linear. Se esses pressupostos não forem atendidos, os resultados da análise podem ser imprecisos ou inválidos.
- Valores extremos: A regressão linear é sensível a valores extremos nos dados, que podem afetar a precisão dos resultados. Portanto, é importante identificar e tratar esses valores antes de realizar a análise.

- Correlação não implica causalidade: A regressão linear pode ser usada para identificar a relação entre duas variáveis, mas isso não implica necessariamente uma relação causal. É preciso ter cuidado para não extrapolar os resultados da análise.
- *Overfitting*: A regressão linear pode ser ajustada excessivamente aos dados de treinamento, o que pode levar a uma baixa capacidade de generalização para novos dados. Para evitar isso, é importante usar técnicas de validação cruzada e regularização.
- Multicolinearidade: Se as variáveis independentes forem altamente correlacionadas entre si, isso pode levar a problemas de multicolinearidade na análise, onde é difícil identificar a importância relativa de cada variável.

#### 2.4.2 Floresta Aleatória (*Random Forest*)

A Floresta Aleatória (Random Forest) é uma técnica de aprendizado de máquina supervisionado utilizada para classificação, regressão e outras tarefas que se beneficiam de um conjunto de árvores de decisão. O método Random Forest constrói várias árvores de decisão independentes, cada uma treinada em um subconjunto aleatório dos dados de treinamento, e combina as previsões de cada árvore para chegar a uma previsão final (BREIMAN, 2001).

As árvores de decisão são construídas utilizando um conjunto aleatório de características em cada divisão, o que aumenta a variabilidade entre as árvores e reduz a correlação entre elas. Isso permite que o método Random Forest generalize melhor para novos dados e reduza a probabilidade de *overfitting*. Além disso, a técnica permite identificar a importância relativa de cada característica na previsão, o que pode ser útil para análises adicionais (CUTLER, *et al.*, 2012).

O método Random Forest tem sido aplicado em várias áreas, incluindo ciência de dados, bioinformática, finanças e marketing. É amplamente utilizado devido à sua facilidade de implementação, capacidade de lidar com grandes conjuntos de dados e boa capacidade de generalização. Liaw e Wiener (2002) demonstraram a aplicação do método em classificação e regressão utilizando a linguagem R. Cutler *et al.* (2012) aplicaram o método em ecologia para classificação de espécies.

Em resumo, o método é uma técnica de aprendizado de máquina amplamente utilizada que combina várias árvores de decisão independentes para melhorar a capacidade de generalização e identificar a importância relativa de cada característica na previsão. Suas aplicações incluem ciência de dados, bioinformática, finanças e marketing, entre outras áreas (CUTLER, *et al.*, 2012).

O método Random Forest tem suas limitações, assim como qualquer técnica de aprendizado de máquina. Algumas delas incluem:

- Tempo de treinamento: embora seja um método eficaz, o tempo de treinamento do Random Forest pode ser relativamente longo, especialmente para grandes conjuntos de dados (Louppe, 2014).
- Dificuldade de interpretação: como o método combina várias árvores de decisão, pode ser difícil interpretar as decisões tomadas pelo modelo, o que pode ser um desafio para algumas aplicações (Biau, 2012).
- Viés de classe: quando há um desequilíbrio nas classes de uma variável de resposta, o modelo pode ser tendencioso em relação à classe majoritária.

### 2.4.3 Método Spline

A previsão *spline* é uma técnica de análise de dados utilizada para suavizar e prever valores de uma série temporal. Essa técnica consiste em ajustar uma curva suave que se adapta aos dados, permitindo prever valores futuros. A principal vantagem da previsão *spline* é sua capacidade de modelar não linearidades e não estacionariedades em uma série temporal, o que a torna útil em uma variedade de aplicações, como finanças, medicina e engenharia (Jolliffe & Cadima, 2016).

Ela é baseada no método de interpolação, onde a curva suave é ajustada aos dados existentes e usada para prever valores futuros. A interpolação é feita utilizando um polinômio suave que passa pelos pontos de dados, o que torna a curva suave e adaptável aos dados. No entanto, a previsão *spline* tem suas limitações e pressupostos, como a suposição de que a série temporal é estacionária e que os dados são independentes e identicamente distribuídos (Marsh, Cormier, 2002).

Outra limitação é sua sensibilidade ao número de pontos de controle utilizados para ajustar a curva suave. Um número insuficiente de pontos de controle pode resultar em uma curva muito simples e subestimar a variação nos dados, enquanto um número excessivo de pontos de controle pode resultar em uma curva muito complexa e superestimar a variação nos dados (Ramsay & Silverman, 2005).

Apesar das limitações, a previsão *spline* continua sendo uma técnica útil de análise de dados em muitas áreas, especialmente quando os dados são não lineares e não estacionários. A escolha do número de pontos de controle e a avaliação da estacionariedade dos dados são importantes para garantir a eficácia da técnica (Hastie & Tibshirani, 1990).

## 2.5. ANÁLISE ECONÔMICA EM INFRAESTRUTURA DE TRANSPORTES

No início do desenvolvimento da economia dos transportes, a atenção estava voltada principalmente para a análise dos custos e benefícios dos diferentes modos de transporte. Essa abordagem se baseava na teoria econômica tradicional e buscava identificar a forma mais eficiente de alocação de recursos no setor de transportes (Rodrigue *et al.*, 2006).

Com o passar do tempo, a economia dos transportes expandiu seu escopo de pesquisa para incluir a análise de custo-benefício dos sistemas de transporte. Essa abordagem busca quantificar os custos e benefícios econômicos associados aos investimentos e políticas no setor de transportes, considerando não apenas os aspectos financeiros, mas também os impactos sociais e ambientais (Andriola, Bonatto, Dutra, 2019).

A análise de custo-benefício se tornou uma ferramenta essencial para avaliar projetos de infraestrutura de transporte e programas de políticas públicas. Ela permite comparar os custos de implementação e operação de determinada intervenção com os benefícios esperados, como redução de congestionamentos, economia de tempo e recursos, melhoria na segurança viária e aumento da acessibilidade. Essa abordagem considera não apenas os efeitos diretos, mas também os efeitos indiretos e de longo prazo sobre a economia, a sociedade e o meio ambiente (Boardman *et al.*, 2017).

Para realizar uma análise de custo-benefício abrangente, são utilizadas técnicas e métodos econômicos, como a valoração econômica de impactos

ambientais e a atribuição de valores monetários a benefícios intangíveis, como a redução de acidentes e o ganho de qualidade de vida. Essa abordagem permite tomar decisões informadas sobre investimentos em infraestrutura de transporte e políticas de transporte, considerando não apenas os custos imediatos, mas também os benefícios de longo prazo para a sociedade como um todo (Flyvbjerg *et al.*, 2003).

### **2.5.1 Análise Custo Benefício**

Segundo o Conselho Regional de Engenharia e Agronomia (CREA), a análise de custo-benefício (ACB) desempenha um papel fundamental na avaliação de projetos de infraestrutura. A ACB envolve diversos passos, como a identificação do projeto, a análise de demanda, o design técnico, estimativas de custo, cronograma de implementação e a avaliação de aspectos ambientais e impactos ambientais potenciais, impactos sociais e considerações sobre mudanças climáticas, seguindo um processo sistemático de avaliação da componente ambiental.

Devido à complexidade da análise de custo-benefício e à variedade de órgãos, dados e estudos envolvidos, este trabalho propõe uma análise inicial, abordando custos de construção, manutenção e economia no tempo gasto em viagens. Também inclui uma breve análise ambiental relacionada à poluição veicular e uma análise do atendimento modal, representando um recorte dentro do amplo escopo da ACB.

Os custos são classificados em custos de investimento, que englobam todos os investimentos em construção, projetos, estudos e equipamentos, e custos operacionais, que incluem manutenção e reposição de equipamentos (CREA, 2020).

Palma *et al.* (2011) destacam que as funções de custo desempenham um papel crítico na economia dos transportes, pois auxiliam na modelagem e compreensão dos custos associados à prestação de serviços de transporte. A função de custo fixo representa os custos que não variam com a quantidade de transporte, incluindo manutenção de infraestrutura, depreciação de ativos e custos administrativos. Essa função é, essencialmente, uma constante, independentemente do volume de serviços de transporte prestados.

A função de custo variável representa os custos que variam com o nível de atividade de transporte, como combustível, mão de obra e manutenção de veículos. À medida que a quantidade de transporte aumenta, os custos variáveis também

aumentam proporcionalmente. A função de custo fixo e a função de custo variável, quando combinadas, permitem calcular o custo total de operação de um sistema de transporte, que varia de acordo com o nível de atividade.

É proposto que, por meio da análise de dados históricos com o uso de *machine learning*, como visto na Seção 2.4 (Algoritmos de Previsão), seja possível identificar padrões complexos e fatores influentes que afetam o custo de projetos de infraestrutura. A capacidade de aprendizado contínuo desses modelos pode ajudar as equipes de construção a ajustar suas estratégias à medida que novos dados e desafios surgem, melhorando a eficiência e a tomada de decisões informadas em projetos de infraestrutura de transporte, o que pode substituir a análise do custo total.

De acordo com o CREA (2020), embora as análises preliminares que normalmente fazem parte do Estudo de Viabilidade Técnica e Ambiental não façam formalmente parte da Análise de Custo e Benefício, seus resultados devem ser resumidos e utilizados como a principal fonte de dados de entrada para a ACB abrangente.

Essa documentação deve abranger detalhes sobre a identificação do projeto, incluindo suas características pertinentes em cada alternativa técnica (cenário) analisada. A análise de demanda (ou estudo de mercado) deve levar em consideração a demanda atual, baseada em dados fornecidos e/ou coletados, assim como a demanda futura, que se baseia em modelos que consideram questões macroeconômicas e sociais (CREA, 2020).

Além disso, o projeto técnico, estimativas de custo e cronograma de implementação são baseados em estudos de engenharia, incluindo o desenvolvimento do projeto, definição de localização, design, cronograma e economia. A análise também considera a identificação de aspectos ambientais, potenciais impactos ambientais, impactos sociais e considerações relacionadas às mudanças climáticas, seguindo um processo sistemático de avaliação da componente ambiental (CREA, 2020).

#### 2.5.1.1. Economia de Tempo de Viagem

O reconhecimento de que o tempo é um recurso escasso afeta a demanda por bens e serviços no mercado, bem como a alocação de recursos financeiros



limitados. Isso sugere que o tempo desempenha um papel importante nas atividades de consumo e produção. O uso do tempo em atividades "não produtivas" envolve um custo de oportunidade que deve ser valorizado. Teorias de alocação de tempo fornecem um quadro natural para derivar uma medida teórica do valor do tempo de viagem (VTTS) (Palma, *et al.* 2011).

O tempo pode ser considerado como uma mercadoria, uma vez que pode gerar utilidade diretamente para o indivíduo quando é "gasto" em atividades específicas. Além disso, atua como um meio para o consumo de bens e serviços de mercado, da mesma forma que o dinheiro é um meio para a compra (e consequente consumo) desses bens e serviços. No entanto, o valor do tempo em uma atividade específica não é o mesmo que o valor do tempo em outra atividade (Palma, *et al.* 2011).

Hensher (1977) propôs uma abordagem alternativa para avaliar o valor das economias de tempo de viagem, inicialmente concentrando-se em viagens domésticas e internacionais de avião. Essa abordagem reconhece diversos componentes, incluindo custos de oportunidade e desconforto relativo. Ao contrário da abordagem tradicional, que valoriza o tempo economizado no trabalho com base no salário bruto e acrescenta um valor marginal.

Hensher (1997) expandiu ainda mais esse modelo, incorporando elementos adicionais, como efeito de produtividade, custos associados ao desconforto relativo, perda de tempo de lazer e a possibilidade de transferência de compensação entre empregador e trabalhador. A combinação desses componentes resulta na "Fórmula Hensher" para estimar o valor do tempo economizado em viagens de trabalho.

$$VTTS = (1 - r - pq)MP + \frac{1 - r}{1 - t}VW + \frac{r}{1 - t}VL + MPF \quad (1)$$

Onde  $r$  é a proporção do tempo de viagem economizado que é usado para atividades de lazer,  $p$  representa a proporção do tempo de viagem economizado à custa do trabalho realizado durante a viagem, em comparação com o tempo gasto no escritório,  $q$  reflete a produtividade relativa do trabalho realizado durante a viagem em comparação com o trabalho no escritório,  $MP$  corresponde ao produto marginal do trabalho, ou seja, a mudança adicional na produção decorrente da contratação de uma unidade adicional de trabalho.  $VL$  representa o valor do tempo livre para o trabalhador

em relação ao tempo de viagem,  $VW$  reflete o valor do tempo gasto em trabalho enquanto no escritório em relação ao tempo de viagem.  $MPF$  indica o valor da produção adicional gerada devido à redução da fadiga,  $t$  é a taxa de imposto pessoal do trabalhador, refletindo a inflação de  $rVL$  e  $(1 - r)VW$  para fins de compensação. Isso significa que um empregador deve compensar um trabalhador pelo tempo de viagem economizado, não em termos de aumento de renda, mas para levar em consideração o fato de que o aumento na utilidade do trabalhador não está sujeito a impostos.

Hensher (1977) procurou calcular o valor do tempo economizado em viagens relacionadas ao trabalho e, para isso, introduziu uma abordagem que diferencia o valor do tempo gasto em lazer ( $VL$ ) do valor do tempo gasto em atividades de trabalho ( $VW$ ).  $VL$  é a valoração tradicional do comportamento relacionado ao tempo economizado associado à troca de tempo de viagem por tempo de lazer. Essa abordagem é utilizada em viagens de negócios ou comerciais, nas quais a maior parte da viagem ocorre fora do horário normal de trabalho, refletindo um equilíbrio entre lazer e trabalho.

Como resultado, o valor do tempo economizado em algumas circunstâncias de trabalho pode ser menor do que a taxa média bruta de salário, levando em consideração a combinação de tempo de trabalho e tempo fora do trabalho. Essa abordagem tem sido aplicada em diversos estudos em vários países, como Suécia, Noruega, Holanda e Reino Unido, sendo amplamente aceita como o método oficial para avaliar o valor do tempo economizado.

Os principais componentes da fórmula incluem  $(1 - r)MP$ ,  $rVL$  e  $(1 - r)VW$ . É comum, em muitas aplicações, não diferenciar o valor do tempo economizado para o trabalhador com base em se o tempo economizado seria gasto no trabalho ou no lazer, assumindo implicitamente que o valor privado do tempo economizado ( $VP$ ) é o mesmo em ambos os casos ou que  $VW$  é igual a  $VL$ . Geralmente, dados sobre outros componentes, como  $p$ , não estão disponíveis (com  $q$  definido como 1,0 e a influência de  $t$  e  $MPF$  assumida como negligível), resultando na fórmula empírica simplificada:

$$VBTT = (1 - r)MP + rVP \quad (2)$$

De acordo com IPEA (2019), os resultados do valor da economia de tempo ( $r$ ) está entre 79% e 36% do valor do tempo laboral. Valor esse que está dentro da

margem estabelecida por Palma *et al.* (2011) para a Austrália em pesquisa de 2005, que constatou que o valor de tempo laboral é entre 72% e 28%.

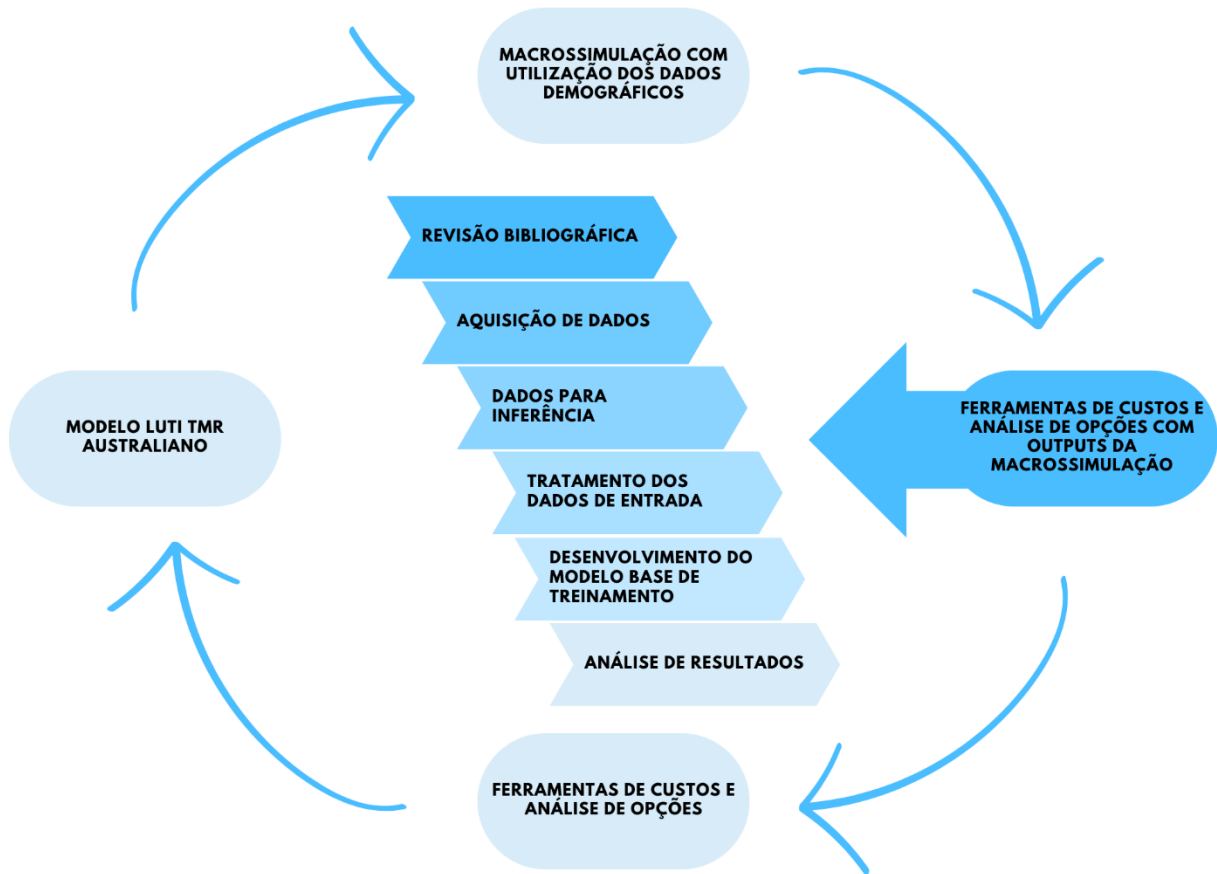
### 3. METODOLOGIA

A metodologia da pesquisa visa contribuir para a tomada de decisões relacionadas a investimentos em infraestrutura viária na Austrália, fazendo uso dos dados de saída dos softwares EMME, que realizam macrossimulação de tráfego, aplicados a um fluxo de trabalho no software Alteryx, voltado para análise de dados. Essa aplicação é direcionada ao módulo final do modelo LUTI do sudeste do estado de Queensland, Austrália.

Devido à natureza prática do estudo, os dados utilizados foram obtidos diretamente de fontes externas na Austrália, incluindo empresas privadas e órgãos do governo do estado de Queensland. Essa característica provou ser um desafio adicional na criação dos modelos, uma vez que a disponibilidade de dados é limitada devido a políticas de segurança das empresas. Questões como problemas de comunicação e atrasos na entrega de dados podem representar obstáculos que limitam a aplicação da metodologia.

A criação da metodologia envolveu seis estágios de desenvolvimento, incluindo revisão bibliográfica, desenvolvimento, validação e aplicação da ferramenta de custos e análise de opções da rede de transporte australiana (Figura 4). O objetivo deste estudo é analisar alternativas de investimento em cenários futuros de infraestrutura.

Figura 4 – Metodologia



Fonte: Autor (2023).

A revisão bibliográfica tem como objetivo compreender os principais métodos existentes para a previsão de custos relacionados à infraestrutura de transportes. Além disso, ela contextualiza o ambiente de trabalho como parte de um modelo LUTI, que utiliza dados de macrossimulação de transportes e métodos de análise de dados e custo-benefício para investimentos futuros em redes de tráfego.

Conforme mencionado anteriormente, a aquisição de dados foi principalmente realizada por meio de parcerias com empresas privadas que colaboram com o governo australiano, bem como a utilização de fontes de dados oficiais do governo. Essa etapa também envolve o estudo dos dados disponíveis para compreendê-los e definir os resultados da rede de simulação de transportes, que servirão como novos inputs para o modelo econômico. A etapa de aquisição, estudo, tratamento e filtragem de dados é exclusivamente voltada para a obtenção de dados, para que o modelo de previsão de custos possa realizar suas inferências.

O tratamento dos dados de entrada teve como objetivo identificar possíveis problemas nos dados utilizados para inferências e solucioná-los, contribuindo assim para o funcionamento eficiente da ferramenta de previsão de custos e para uma representação mais precisa dos dados econômicos de acordo com a realidade.

O penúltimo estágio do desenvolvimento do modelo de custos envolve a criação de um módulo no Alteryx Designer que integra os dados obtidos da macrossimulação de tráfego, já tratados nas etapas anteriores, e realiza o treinamento com os dados adquiridos, seguido pela validação correspondente.

Finalmente, a análise de resultados implementa métodos de avaliação das infraestruturas, tais como a economia de tempo de viagem, a análise de emissões e o custo por usuário da rede por hora. Essas análises fornecem uma perspectiva superficial, mas de grande importância para a avaliação das redes, além do cálculo bruto de custos realizado pelo módulo mencionado anteriormente.

### 3.1. AQUISIÇÃO DOS DADOS

Os dados, assim como os custos de construção em infraestrutura de transportes, foram obtidos por meio do histórico do Queensland *Transport and Roads Investment Program* (QTRIP) (AUSTRÁLIA, 2023a), que inclui obras de infraestrutura rodoviária e de transporte público, como vias exclusivas para ônibus, trens e bondes na região sudeste de Queensland, Austrália. Os dados incluem informações como o ano de projeto, o nome do projeto, o custo total da etapa do projeto e o tipo de trabalho, que pode ser categorizado em construção, interseções, estudos e trabalhos de manutenção.

Dados adicionais foram coletados de fontes oficiais do governo, como os *Local Government Infrastructure Plans* (LGIP) (Brisbane, 2020) de diversos governos regionais na região sudeste de Queensland, a fim de obter um modelo mais robusto e com maior quantidade de dados para o treinamento.

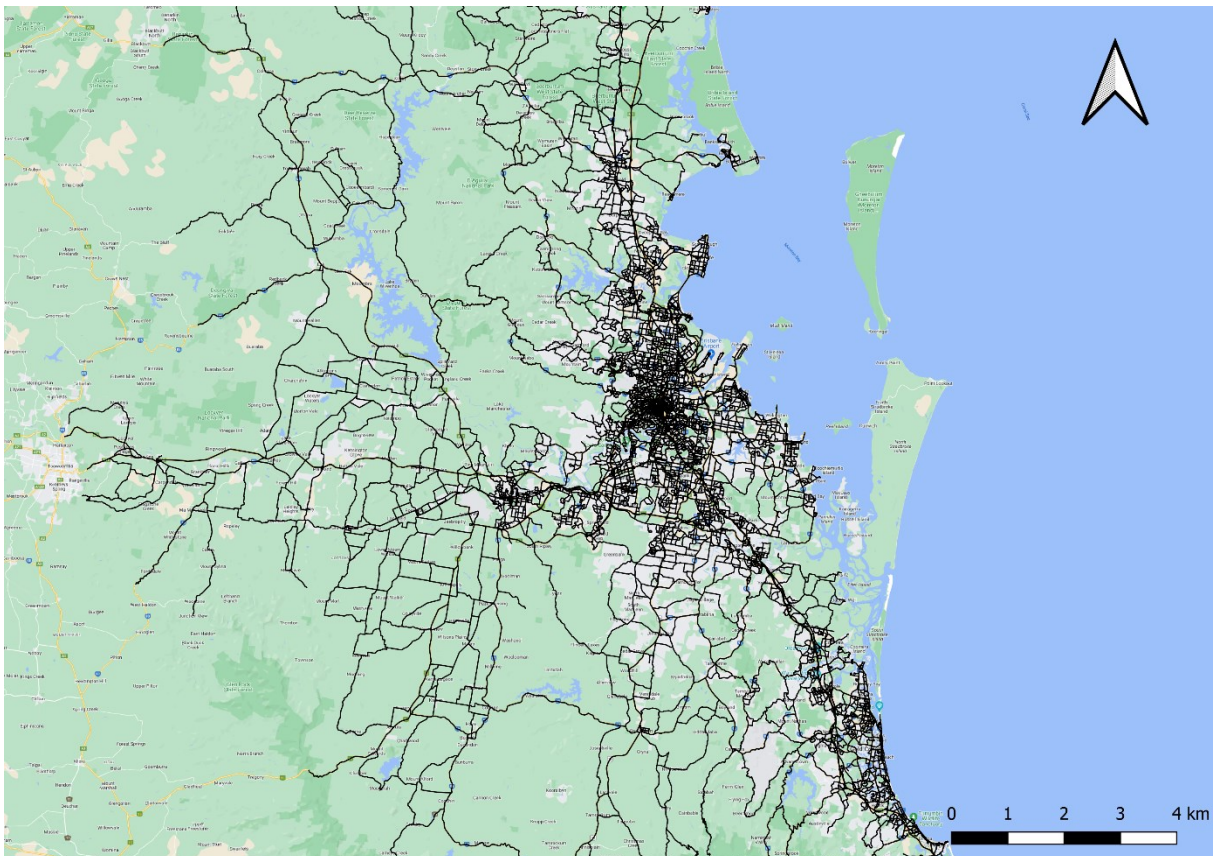
Para o desenvolvimento dos modelos, foi necessária a atualização dos dados de custo, uma vez que os valores de anos anteriores estavam defasados em relação à cotação atual da moeda australiana. Para realizar essa atualização, utilizou-se os dados de índice de preços ao produtor da ABS (AUSTRÁLIA, 2023b) e aplicou-se a equação 3 para atualização de valores para o último ano fiscal australiano, que corresponde a julho de 2022.

$$Va = Vi \left( \frac{Ia}{Ii} \right) \quad (3)$$

Sendo  $Va$  o valor atualizado,  $Vi$  o valor inicial,  $Ia$  a taxa/índice atual e  $Ii$  a taxa/índice inicial.

Com isso, a variável objetivo do treinamento do modelo de previsão de custos estava em totalidade de acordo com seus valores atualizados. A obtenção de informações extras para o treinamento dos modelos foi realizada utilizando de um arquivo shape contendo os links de toda região (Figura 5) e foi fornecido pelo TMR Australiano.

Figura 5 – Shape com vias do Sudeste de Queensland

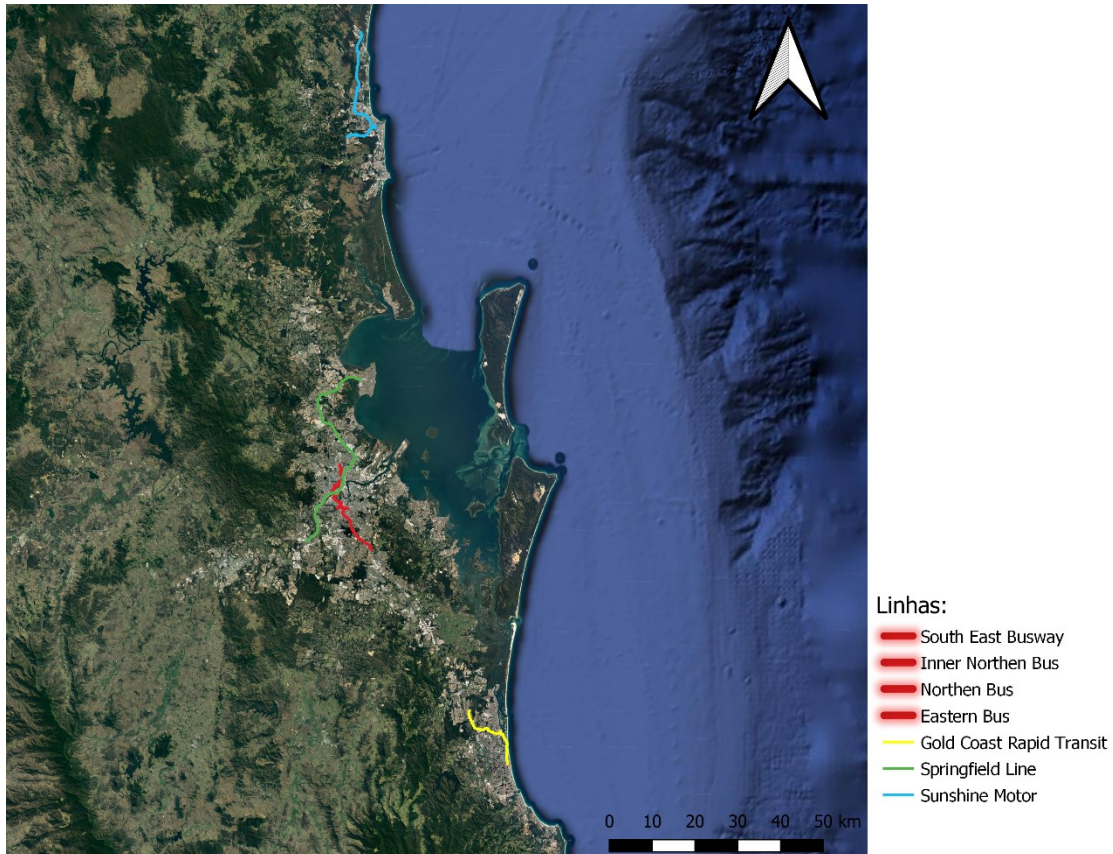


Fonte: Autor (2023).

Ao comparar o arquivo shape com imagens de satélite, e baseado no conhecimento prévio da região por parte do autor, foi possível realizar a separação dos links por projeto e ano da base de dados QTRIP. Assim, permitiu-se a obtenção

do comprimento total dos links (Figura 6) através da função \$length, função que realiza esse cálculo, e da ferramenta de estatística para somar todos os comprimentos das seções, ambos processos realizados no software de geoprocessamento QGIS.

Figura 6 – Links Utilizados Para Treinamento



Fonte: Autor (2023).

Os links tratam de sete projetos de transporte criados desde o início dos anos 2000, sendo quatro delas (South East, Inner Northern, Northern e Eastern) vias de ônibus, Springfield Line sendo de trem, Gold Coast Rapid Transit de bonde e Sunshine Motorway que é uma rodovia. Apesar da extensão total ser obtida, os empreendimentos são separados por fases, assim criou-se lista com 33 dados de treinamento dentre os quais, 12 de ônibus, oito de rodovias, sete de bonde e seis de trem.

Para os dados vindos de coleta das fontes governamentais de Queensland, o tratamento foi diferente, passando por um processo de coleta das variáveis, de velocidade, comprimento, modo e custo da obra. As amostras de treinamento para



novas infraestruturas totalizaram seis dados para o modo ferroviário e 58 para estradas.

Os dados das fontes governamentais de Queensland foram agregados ao fornecidos pelo TMR, formando a base de dados para treinamento do modelo com 12 dados para projetos de ônibus, 66 para rodovias, 7 para bonde e 12 para trem.

### 3.2. DADOS PARA INFERÊNCIA

Foram fornecidas duas redes de tráfego para a realização do custeio: uma rede chamada 'Referência', que fornece links da rede de transporte existente vários novos projetos de grande porte na região do Sudeste do estado de Queensland; e o cenário 'Modificado de Referência', que é semelhante ao primeiro, mas com a adição de alguns segmentos diferentes ou a remoção deles na categoria de ruas. O cenário modificado também altera o ano de alguns projetos de outros modos de transporte, como o trem, lançando-os antes ou depois do cenário de referência. O cenário atual sem modificações é chamado de Base.

Os dados são estruturados de maneira semelhante, com redes para os anos de 2021 (atuais), 2026, 2031, 2036, 2041 e 2046. Todas incluem as mudanças ocorridas em seu respectivo ano, além de todos os links dos anos anteriores. Portanto, em 2046, todas as informações de todos os anos anteriores estão presentes, juntamente com as mudanças ocorridas em 2046.

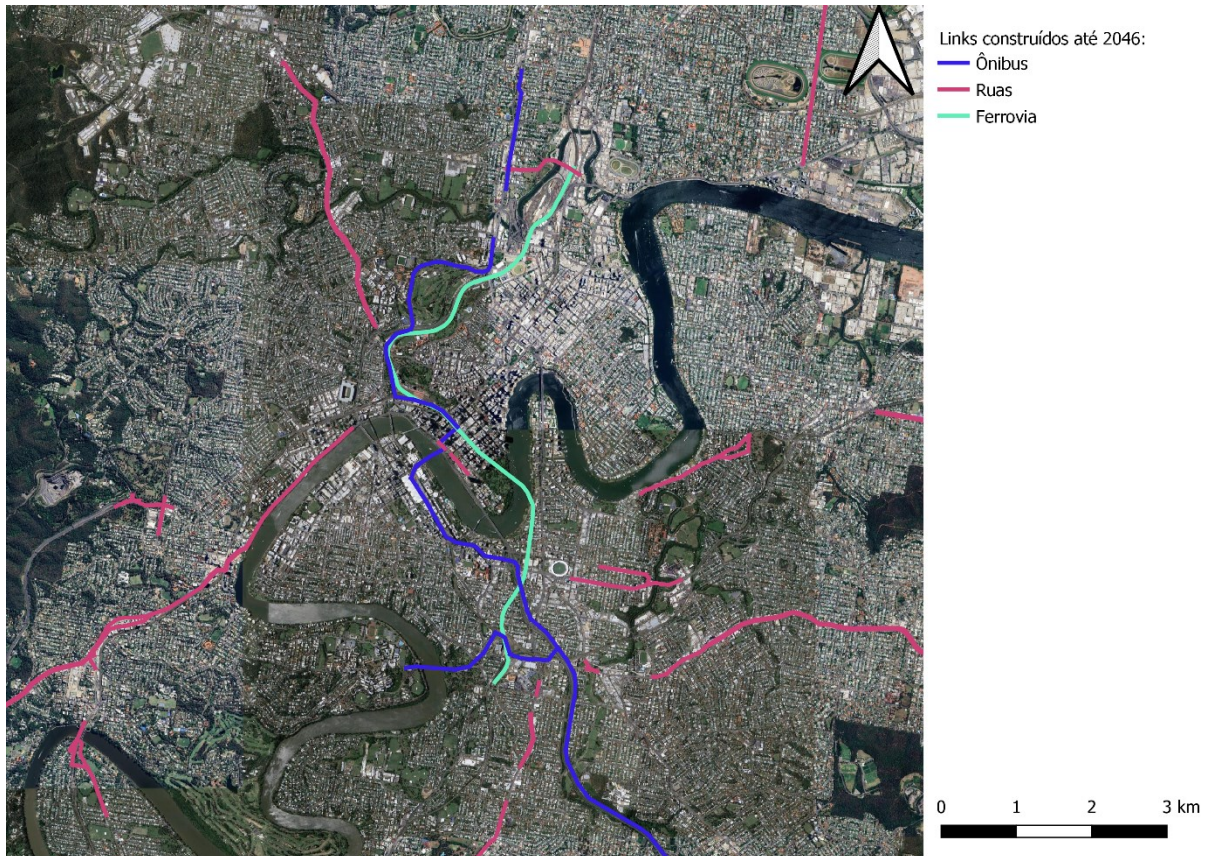
Além disso, em relação aos dados, as colunas que têm relevância para o modelo deste trabalho são as colunas A e B, que representam, respectivamente, os nós de origem e destino do conector. Também foram selecionados como pertinentes para a análise os dados referentes à velocidade média de simulação dos veículos nas vias, número de faixas, comprimento, modo da via e ação a ser tomada.

A variável de ação a ser tomada pode ser exemplificada como "a" para adicionar um link à rede, "m" para modificar características do link, como os modos utilizados na via, e "d" para deletar. Para o estudo, apenas links com "a" foram considerados, já que é a única ação relevante para a análise de custo de construção de infraestruturas.

Os modos de via incluem ruas, trens leves, trens pesados, faixas de pedestre, vias de ônibus, rotas de balsa e ciclovias. O foco do custeio é apenas nos custos de

construção de vias de tráfego para veículos e ônibus, além dos trilhos (Figura 7). Portanto, vias de transportes ativos e de barcos não serão custeadas.

Figura 7 - Parte da Rede de Dados para Realização da Inferência



Fonte: Autor (2023).

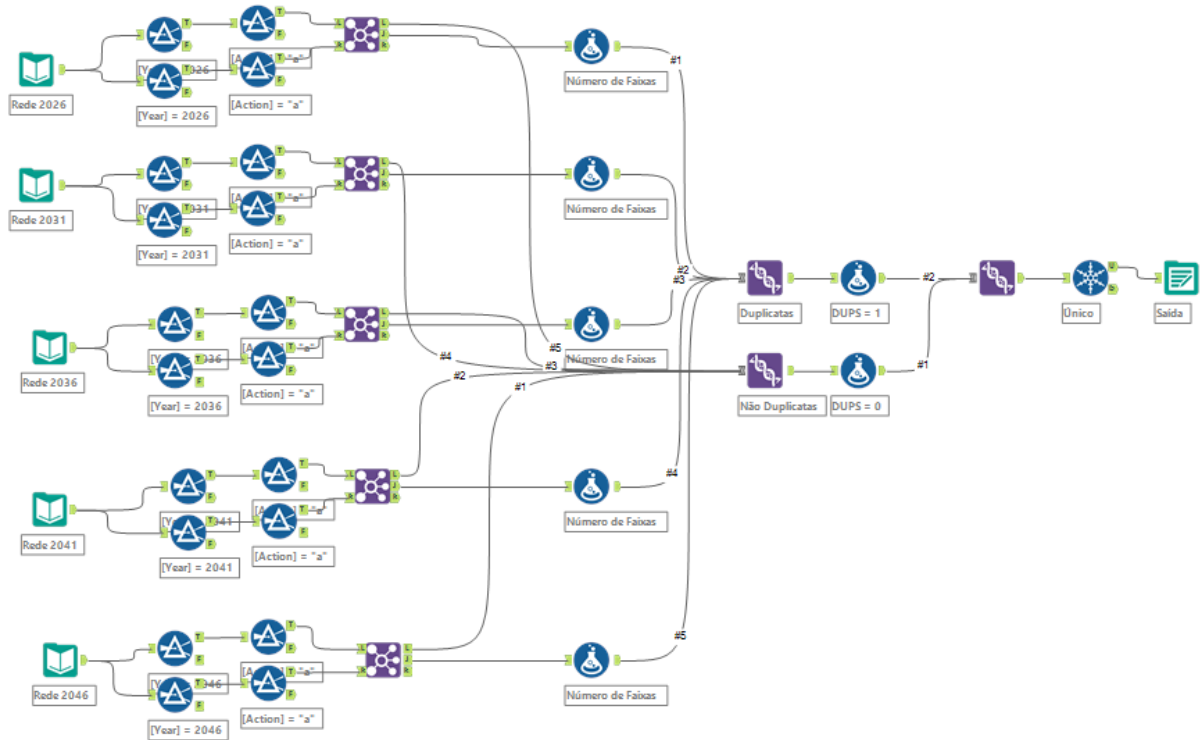
Ao todo são avaliados 2442 links no caso de Referência, desses, 39 são linhas de ônibus, 2152 são ruas, 40 são bondes e 210 trens. Já no caso modificado, a rede possui 2414 links, desses, 39 são ônibus, 2124 ruas, 40 bondes e 210 trens. Sendo assim a única diferença entre os cenários, os anos de aplicação entre alguns modos e alguns projetos de ruas que foram considerados ou não entre os cenários Referência e Modificado.

### 3.3. TRATAMENTO DOS DADOS DE ENTRADA

Com o modelo de treinamento construído, o próximo passo é filtrar os dados de infraestrutura da rede a serem custeados. Para tal um fluxo de trabalho no Alteryx foi criado solucionando alguns problemas averiguados enquanto estudou-se os dados

como a existência de “duplicatas” de dados nos arquivos *shapefile* fornecidos pelo modelo EMME (Figura 8).

Figura 8 - Fluxo de Trabalho para Seleção de Duplicatas



Fonte: Autor (2023).

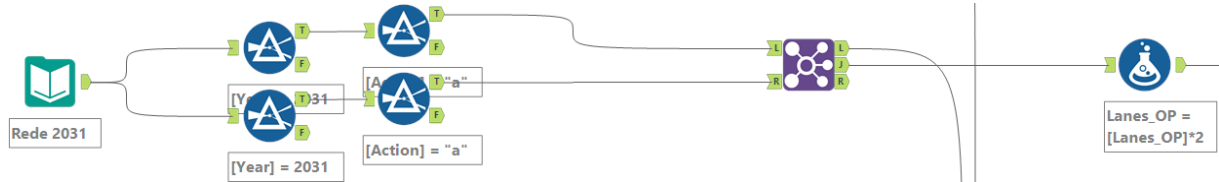
Devido à maneira como o software de macrossimulação funciona, uma via que permite movimentos em ambos os sentidos (mão dupla) é dividida em duas vias separadas com sentidos opostos. Portanto, vias com duas faixas, uma em cada sentido, acabam sendo separadas em duas vias diferentes.

Isso se torna um problema devido à economia de escala, que é um conceito econômico que descreve a redução dos custos médios de produção à medida que a quantidade produzida de um bem ou serviço aumenta (Loo, s.d). Construir uma estrada de duas vias é proporcionalmente mais barato do que construir duas estradas diferentes de uma via.

Para solucionar o problema, o fluxo de trabalho no Alteryx divide os dados em duas cópias de si mesmos e, em seguida, com auxílio da ferramenta Junção, agrupa os dois fluxos de dados em apenas um, combinando a coluna do nó A (origem) de um fluxo com a coluna do nó B (destino) do outro. Após isso, é executada a ferramenta Unique para eliminar qualquer duplicação de dados, adicionando um marcador de

duplicata àqueles duplicados e, por fim, incluindo o número de faixas da sua cópia no campo *Lanes* (Figura 9).

Figura 9 - Fluxo para o Ano 2031 na Seleção de Duplicatas



Fonte: Autor (2023).

Esse processo de adicionar o número de faixas se faz necessário, como mencionado anteriormente, porque, custear uma rua de quatro faixas é mais econômico do que custear duas ruas de duas faixas. Após o modelo de custos rodar, o valor dos links é dividido por dois, uma vez que está sendo custeado duas vezes o mesmo projeto: um conector de A até B e outro conector de B até A.

Vale ressaltar que a ferramenta Junção não agrupa as vias de mão única, portanto, no restante da ferramenta, é possível retirar essas vias, já que são marcadas como não duplicatas. Essas vias não necessitam de nenhum tratamento adicional e são custeadas normalmente, sem a multiplicação do número de faixas e divisão de custos.

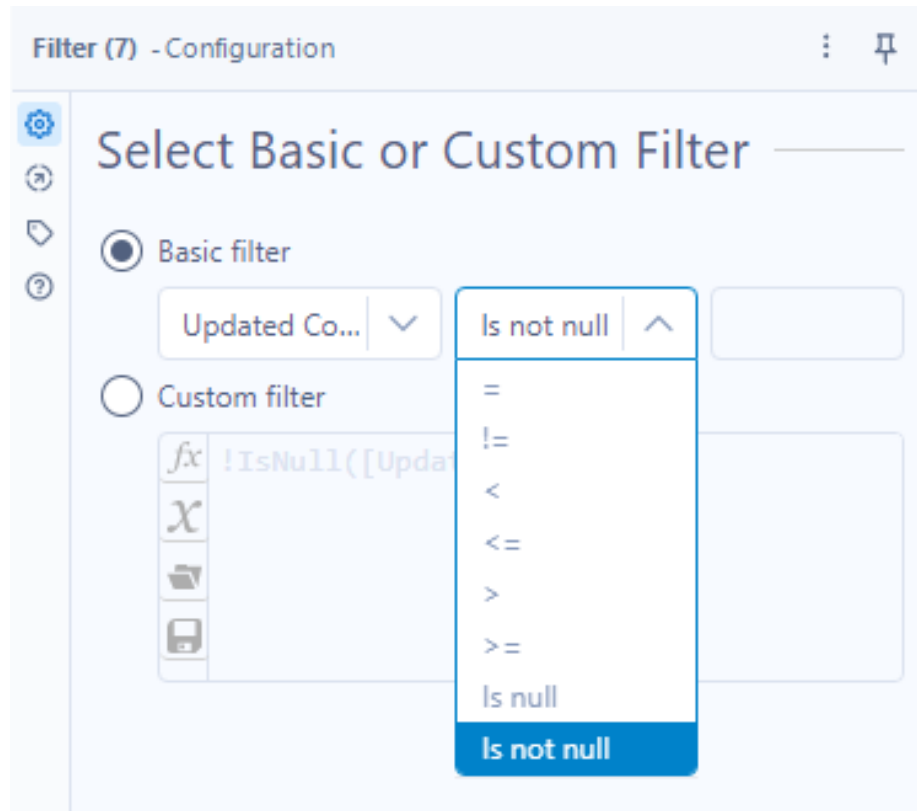
### 3.4. DESENVOLVIMENTO DO MODELO BASE DE TREINAMENTO

Para a construção de um modelo de previsão de custos, foi utilizado o software de análise de dados Alteryx Designer em seu módulo de predição, utilizando os modelos de *machine learning*/regressão mencionados anteriormente, como Regressão Linear, Regressão Gama, *Spline* e Floresta Randômica. Os campos mencionados na seção anterior foram utilizados como fonte de dados para o treinamento. A utilização dos diferentes métodos tem como objetivo identificar o melhor método que responde aos dados conforme as variáveis de treinamento selecionadas.

Como demonstrado previamente, a planilha contém diferentes tipos de trabalhos dentro de um projeto. Para obter amostras relevantes no processo de treinamento, é necessário realizar a filtragem dos dados. O Alteryx Designer permite essa ação por meio da ferramenta Filtro (Figura 10) dentro de um fluxo de trabalho,

possibilitando duas abordagens: a simplificada, na qual uma coluna da planilha de dados e uma condição são selecionadas, ou a customizável, na qual o usuário pode criar uma fórmula.

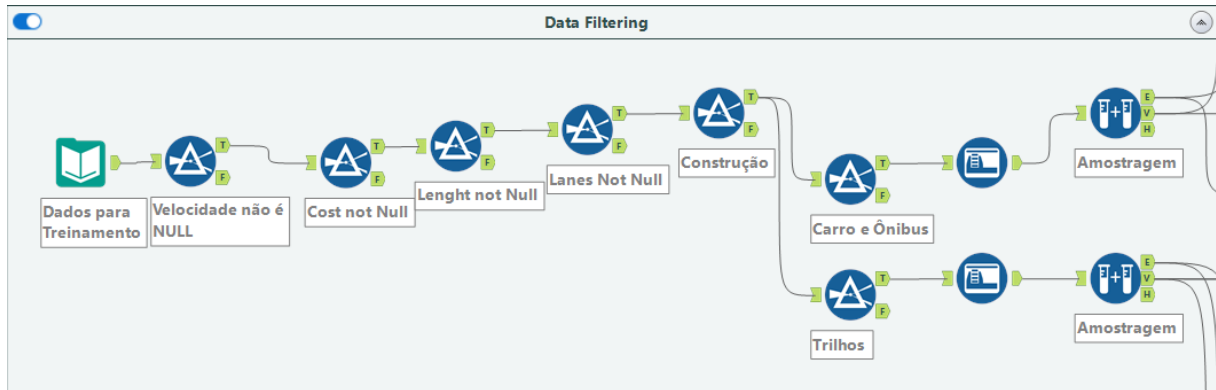
Figura 10 - Ferramenta Filtro do Alteryx Designer



Fonte: Autor (2023)

Para este modelo, apenas o filtro básico foi utilizado, selecionando as colunas referentes ao tipo de ação realizado no projeto (Seção 3.2), neste caso, construção, e o condicional igual. Também foram empregados filtros para remover valores nulos das colunas de custo, comprimento e número de faixas de tráfego. Após uma sequência de ferramentas aplicadas à fonte de dados, o fluxo de trabalho de filtragem foi encerrado, resultando em um total de 97 amostras de treinamento (Figura 11).

Figura 11 - Fluxo de Trabalho para Filtragem dos Dados



Fonte: Autor (2023).

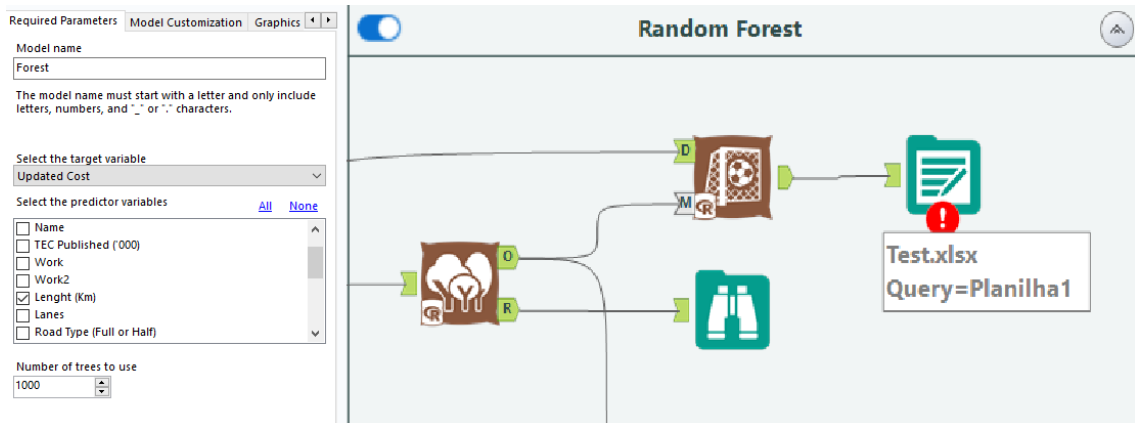
Após o processo de filtragem de dados, ocorre a separação em duas correntes de fluxo de trabalho, uma para os modos rodoviários (ônibus e estradas) e outra para os modos de trilho (bondes e trens pesados). Isso foi feito devido às grandes diferenças de custo entre os diferentes modais, o que pode alterar drasticamente e influenciar o modelo como um todo.

O último passo no preparo dos dados de treinamento é a separação dos dados em treinamento e validação, realizado pelo Criação de Amostras do Alteryx. A porcentagem de treinamento foi definida como 80%, um valor mais elevado devido à baixa quantidade de amostras, mas ainda dentro de uma margem considerada aceitável pelo TMR australiano.

Foram selecionados quatro modelos para avaliação dos custos de construção de projetos de infraestrutura: Floresta Randômica, *Spline*, Regressão Gama e Regressão Linear. A métrica do coeficiente de determinação ( $R^2$ ) foi o fator escolhido para avaliação e seleção dos modelos que a serem aplicados no desenvolvimento futuro, pois é uma métrica que estatisticamente representa a variação entre o esperado e o obtido.

Ao final deste processo, três métodos serão eliminados e apenas o melhor método avaliado será selecionado para a realização do trabalho. O método da Floresta Randômica é o único que requer dados adicionais além dos padrões que os algoritmos necessitam para rodar, sendo o número de árvores na floresta e a quantidade de galhos (Figura 12).

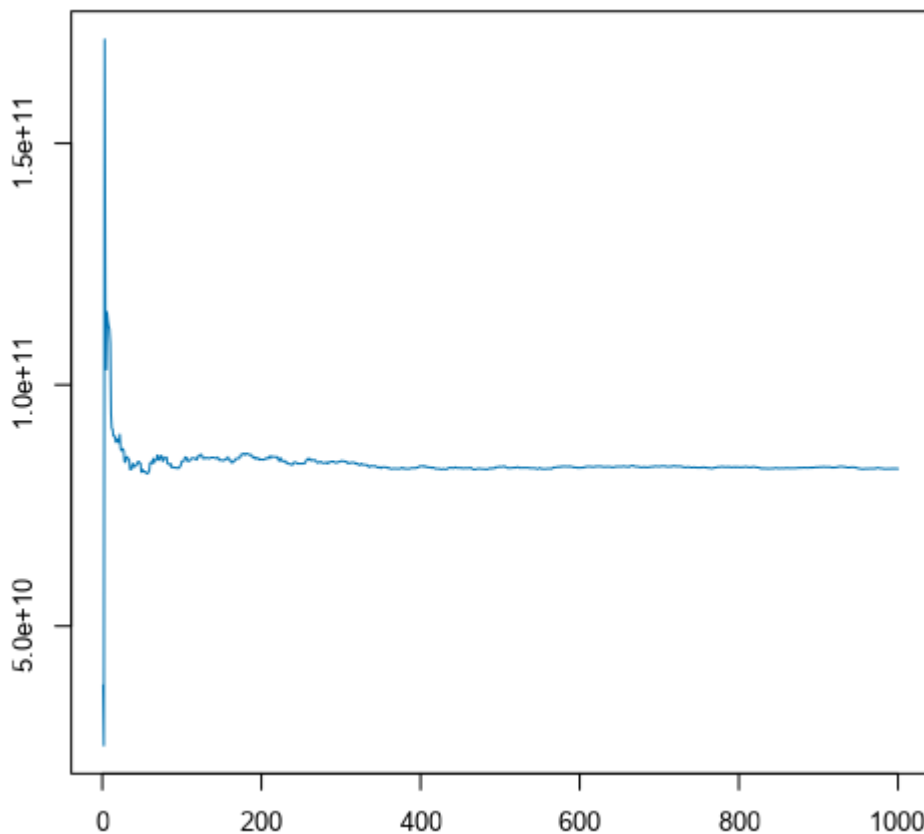
Figura 12 - Modelo com Ferramentas de Predição



Fonte: Autor (2023).

Neste estudo, com base na experiência empírica, foi definido um valor de 1000 árvores e 15 galhos (Figura 13), em detrimento do padrão de 500 árvores e 10 galhos.

Figura 13 - Erro de Acordo com o Número de Árvores



Fonte: Autor (2023).

Conforme observado, o modelo rapidamente atinge um erro baixo por volta de 100 árvores, todavia estabiliza apenas próximo das 400 onde a volatilidade do modelo

é muito baixa. É importante notar que a maior quantidade de árvores tende a resultar em um modelo melhor, porém com o custo computacional em troca. A melhoria tende ao infinito como mostrado na figura, porém por um processo longo em retorno.

Os quatro algoritmos possuem estrutura semelhante à figura acima. O modelo recebe os dados de treinamento, executa o algoritmo e gera o resultado, que é adicionado à ferramenta Score. Essa ferramenta, além do modelo, recebe os dados de validação e os aplica a essa amostragem, gerando assim os dados estatísticos para a análise por parte do usuário.

As variáveis selecionadas para utilização no treinamento foram o comprimento (Length), o número de faixas e a velocidade média dos veículos na simulação, com o objetivo de buscar o Custo (Updated Cost) na ferramenta Score, que aplica o modelo aos dados a serem previstos.

### 3.5. ANÁLISE DE RESULTADOS

Após o modelo concluir seus passos de filtragem, preparação, treinamento e inferência nos custos, inicia-se o processo de análise. Nesse sentido, o objetivo é agrupar os resultados dos custos de infraestrutura por governos locais (LGA) e anos, a fim de avaliar os investimentos futuros nas regiões administrativas do sudeste de Queensland. Isso fornecerá informações cruciais sobre regiões que podem estar recebendo investimentos inadequados ou até mesmo sendo negligenciadas em comparação com outras.

Uma métrica importante para a análise de custo-benefício é o custo total por ano e por período. Para calculá-lo, somam-se os custos de construção de infraestruturas, agrupando-os por categorias como ônibus, estradas e ferrovias, separando-os de acordo com os anos de implementação. É fundamental observar que, embora bondes e trens pesados sejam custeados separadamente, a categoria de ferrovias engloba ambos, conforme orientação do TMR.

É crucial ressaltar, que os valores previstos de custo também foram escalados de acordo com uma inflação de 3% ao ano, valor utilizado pelo governo Australiano para situações onde a precisão inflacionária não é requerida. Nesse caso, apenas uma representação dela é o suficiente utilizando-se então uma aplicação simples de juros compostos com o período em anos e sendo representado por  $t$  que é 3 para 2026, 8 para 2031, 13 para 2036, 18 para 2041 e 23 para 2046 (Equação 4).



$$FV = PV(1 + r)^t \quad (4)$$

Onde  $FV$  é o valor futuro,  $PV$  é o valor presente/inicial,  $r$  é a taxa e  $t$  é o número de períodos (anos).

Outra métrica relevante é a análise da economia de tempo de viagem, pois envolve tanto o capital público quanto o das pessoas físicas e jurídicas. Conforme discutido na seção 2.5.1.1 e na equação 2, se calcula o valor agregado à mudança viária, utilizando-se o valor de  $MP$  (produto marginal do trabalho) de 32,82 dólares australianos por hora (ABS, 2023), um  $VP$  (valor privado do tempo economizado) de 23,63 e um  $r$  (valor da economia de tempo em porcentagem) de 0,72, de acordo com os estudos mostrados no capítulo 2.5.1.1 para a Austrália.

No âmbito da coleta e análise de dados de transporte, essa abordagem envolve uma análise abrangente da mobilidade da população em uma determinada região. Isso inclui o monitoramento dos diferentes modos de transporte utilizados pelas pessoas, como automóveis, transporte público, trens, bicicletas ou a pé. A análise se estende à avaliação dos custos associados a esses deslocamentos, com o objetivo de calcular o custo por usuário na rede, considerando a segregação por modos de transporte.

Dividindo o custo total por modo por dia pelo fluxo de pessoas na rede por dia é possível obter o custo por usuário da rede, métrica que avalia se um cenário é, na prática, realmente mais caro que outros possíveis cenários. Essa abordagem fornece informações essenciais para o planejamento de transporte e a alocação de recursos, permitindo uma análise detalhada dos custos e da eficiência dos investimentos em transporte.

Outra análise realizada é a ambiental, que pode abranger uma grande gama de pesquisa. O presente trabalho busca apenas fazer uma inspeção superficial de poluentes com base no número e tipo de veículos utilizando as vias e, para tal, multiplica-se o número de veículos separados pelos tipos, carro, ônibus, caminhão, trem, bonde e motocicletas, pela média de gases poluentes emitidos por km, e também, pela quilometragem média dos veículos na via (Fórmula 5).

$$E_t = \sum_i (F_i \cdot D_i \cdot E_i) \quad (5)$$

Onde  $F$  é o fluxo de veículos do tipo  $i$ ,  $D$  é a distância média percorrida pela categoria  $i$ , e  $E$  é a emissão que o modo  $i$  emite por quilômetro.

Essa análise será superficial, pois não considera todos os fatores que podem afetar as emissões de tráfego, como o perfil de velocidade, a topografia, a idade da frota de veículos, as condições de tráfego, entre outros. Porém, devido à limitação de dados fornecidos, é uma alternativa válida, já que uma análise superficial pode ser útil para avaliar a ordem de magnitude das emissões e identificar tendências em um estudo estratégico. No entanto, não substitui uma análise mais completa e precisa que leve em consideração todos os fatores relevantes, para um estudo mais detalhado.

## 4. ANÁLISE DE RESULTADOS

Nesta seção, é realizada a análise dos resultados com o propósito de avaliar os custos previstos pelo modelo desenvolvido neste estudo, bem como os impactos de tráfego em diversos cenários de infraestrutura. Inicialmente, é apresentada uma visão geral dos resultados obtidos, com foco na métrica  $R^2$  aplicada aos diferentes modelos, como regressão linear, regressão gama, método *spline* e *random forest*. A partir dessa avaliação, é determinado qual método será adotado na continuação deste trabalho. Em seguida, realiza-se uma análise minuciosa dos resultados de previsão de custos para os cenários, aplicando as métricas específicas de análise de cenários.

### 4.1. Métodos Preditivos

Com a conclusão bem-sucedida da construção do modelo de previsão na Seção 3.4 e a definição precisa dos dados de treinamento após filtragem, o próximo passo envolveu a execução do processo no Alteryx Designer. É notável que, dentro do ambiente do software, o processo inteiro foi concluído em aproximadamente 15 minutos, resultando na extração dos dados finais para planilhas do Excel.

Todos os modelos utilizaram as mesmas variáveis no treinamento, sendo elas, o comprimento (*Length*), o número de faixas e a velocidade média dos veículos na simulação, com o objetivo de buscar o Custo (*Updated Cost*) na ferramenta *Score*, que aplica o modelo aos dados a serem previstos. O coeficiente de determinação  $R^2$  foi utilizado para auxiliar na decisão de qual modelo utilizar na continuidade do trabalho ao analisar a variação do valor estimado pelos modelos em função dos valores observados através de regressão linear simples. Esse método foi também utilizado por Pfaffenbichler (2003) e Lopes (2010) na análise do ajuste de seus modelos de demanda.

Nesse caso, a análise da reta de uma regressão apresenta, também, valores úteis como o intercepto que numa situação onde o modelo prediz valores de acordo com a realidade deve se aproximar de zero, e a inclinação da reta que deve ser o mais próximo de  $45^\circ$ , com coeficiente de  $X$  (valores observados) mais próximo de 1 nessa mesma situação (Hastie, *et al.* 2009). Pode-se assim verificar a qualidade dos

modelos averiguando os três coeficientes citados e tomar a decisão de qual modelo utilizar nas aplicações futuras (Tabela 1).

Tabela 1 – Resultados dos Coeficientes de Cada Método

Modelos	Real x Estimado			
	Análise da Regressão Simples	Coeficientes		Ajuste
	Parâmetros	$\alpha$ (Intercepto)	$\beta$ (Inclinação da Retas)	$R^2$ (Coeficiente de Determinação)
	Valores Ideais	0.000	1.000	1.000
Linear	Valores Obtidos	284.269,000	0,265	0,189
	Diferença	284.269,000	-0,735	-0,811
Gama	Valores Obtidos	320.713,000	0,202	0,113
	Diferença	320.713,000	-0,798	-0,887
Spline	Valores Obtidos	40.133,000	0,768	0,670
	Diferença	40.133,000	-0,232	-0,329
Random Forest	Valores Obtidos	28.220,000	0,820	0,745
	Diferença	28.220,000	-0,180	-0,255

Fonte: Autor (2023).

Com  $R^2$  médio de 0,7454 para os três estados analisados (estradas, vias para ônibus e trilhos), o método Random Forest destacou-se como o mais qualificado para o treinamento e para realizar inferências nos dados para predição. A ordem dos modelos por qualidade do coeficiente de determinação foi seguida pelo modelo *Spline*, com 0,6708, Regressão Linear, com 0,1891, e Regressão Gama, com 0,1131, sendo assim o valor que menos se aproximou de um.

De maneira geral, o intercepto e a inclinação da reta apresentaram comportamentos semelhantes ao do  $R^2$ , com o Random Forest mostrando-se o método mais apto para a continuidade do trabalho. Os valores de  $\alpha$  tendem a ser mais elevados para esse cenário devido aos altos custos de infraestrutura envolvidos.

Conforme explicado na Seção 2.4, métodos como o Random Forest generalizam melhor novos dados e reduzem a probabilidade de *overfitting*, sendo, portanto, muito eficazes na previsão de dados futuros com múltiplas variáveis. Os

métodos de regressão são mais apropriados para outros tipos de previsão, nos quais as variáveis independentes e dependentes possuem relações lineares ou gama, como sugere o padrão de seus nomes.

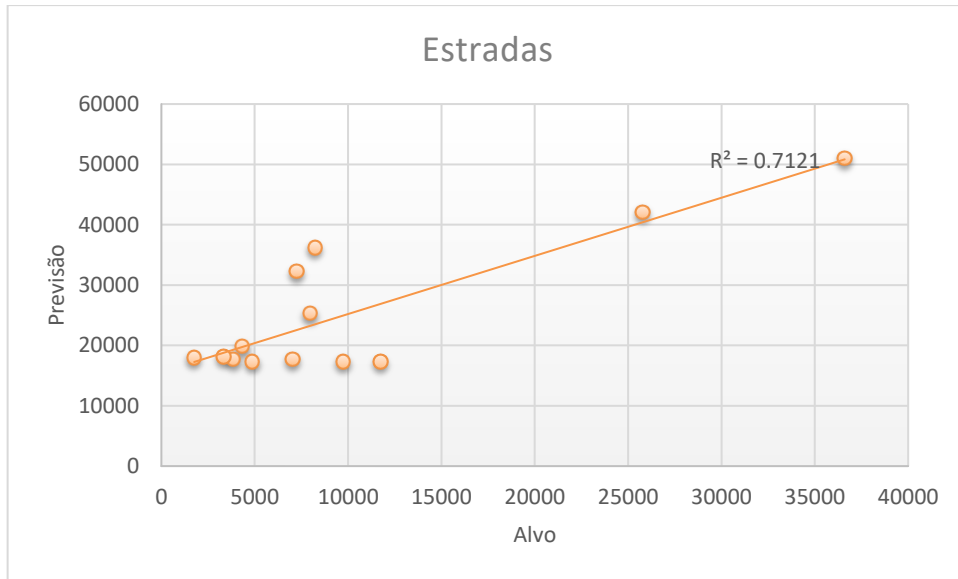
O método preferido mostrou-se especialmente eficiente na previsão de dados relacionados a estradas (Figura 15) e aos modos de trilhos (Figura 14). Isso pode ser explicado pelo número adicional de dados obtidos nos LGIP para essas categorias, enquanto, no caso dos ônibus (Figura 16), que não contou com amostras adicionais além das fornecidas pelo TMR Australiano, o desempenho foi inferior.

Figura 14 - Validação da Categoria Trilhos



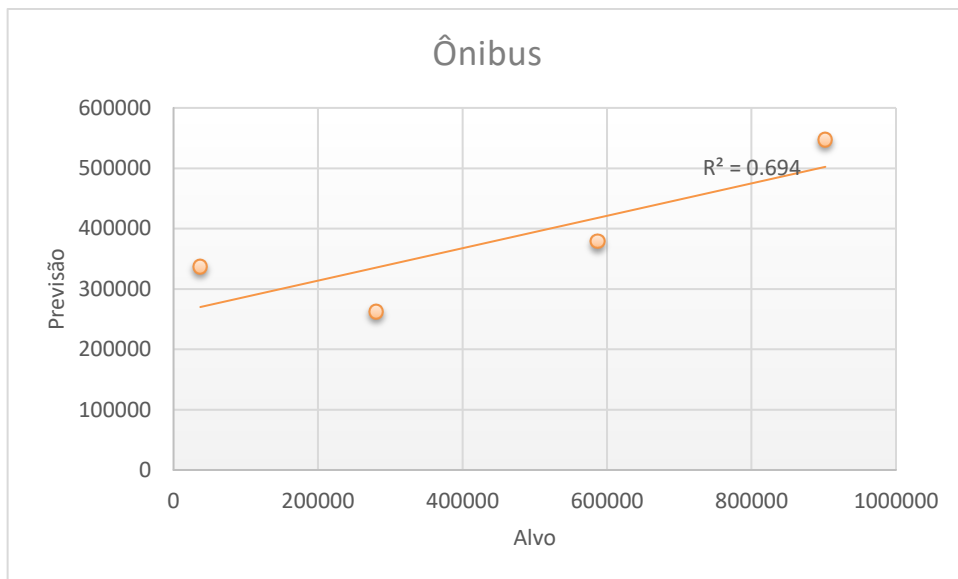
Fonte: Autor (2023).

Figura 15 - Validação da Categoria Estrada



Fonte: Autor (2023).

Figura 16 - Validação da Categoria Ônibus



Fonte: Autor (2023).

Os resultados apresentados nas figuras 14, 15 e 16 representam os valores mínimos atingidos pelo modelo durante o treinamento, o qual foi conduzido utilizando diversos conjuntos de amostras para validação. O desempenho do  $R^2$  variou, selecionando diferentes amostragens para validação, entre 0,71 e 0,95 para estradas, 0,69 e 0,75 para ônibus, e 0,83 e 0,91 para trilhos. Esses valores foram considerados satisfatórios, considerando a limitada quantidade de amostras disponíveis em todas as categorias, e estão em conformidade com as expectativas do TMR Australiano.

## 4.2. Inferências nos Cenários

Com o método de previsão de cenários definido, foi suficiente selecionar o modelo no fluxo de trabalho dentro do Alteryx Designer para realizar as inferências. Isso inclui a separação dos custos por LGA (Tabela 2) e a aplicação da taxa de 3% ao ano durante o período específico em que a obra será realizada. O workflow também é responsável por dividir por dois o valor total das vias que são consideradas duplicatas, conforme detalhado na seção 3.3.

Tabela 2 - Resultados da Previsão de Custos por LGA Cenário Referência

LGA	Ano	Estrada	Trilhos	Ônibus	Custo (3%/ano)
Brisbane	2026	574,710,725	\$4,145,953,974	\$ 2,988,174,221	\$ 7,708,838,920
	2031	\$ 792,090,502	\$ -	\$ 2,139,106,150	\$ 2,931,196,653
	2036	\$ 1,011,251,310	\$ -	\$ -	\$ 1,011,251,310
	2041	\$ 2,125,539,763	\$ 2,203,382,133	\$ -	\$ 4,328,921,895
	2046	\$ -	\$ -	\$ -	\$ -
	Total	\$ 4,503,592,300	\$ 6,349,336,106	\$ 5,127,280,371	\$ 15,980,208,778
Gold Coast	2026	\$ 2,057,417,810	\$ 4,953,038,710	\$ -	\$ 7,010,456,520
	2031	\$ 1,076,513,854	\$ -	\$ -	\$ 1,076,513,854
	2036	\$ 479,177,911	\$ 7,646,396,978	\$ -	\$ 8,125,574,889
	2041	\$ 2,048,163,265	\$ -	\$ -	\$ 2,048,163,265
	2046	\$ 130,816,665	\$ 3,713,745,348	\$ -	\$ 3,844,562,013
	Total	\$ 5,792,089,505	\$16,313,181,035	\$ -	\$ 22,105,270,541
Gympie	2026	\$ -	\$ -	\$ -	\$ -
	2031	\$ 1,414,057,476	\$ -	\$ -	\$ 1,414,057,476
	2036	\$ -	\$ -	\$ -	\$ -
	2041	\$ -	\$ -	\$ -	\$ -
	2046	\$ -	\$ -	\$ -	\$ -
	Total	\$ 1,414,057,476	\$ -	\$ -	\$ 1,414,057,476
Ipswich	2026	\$ 1,137,606,741	\$ -	\$ -	\$ 1,137,606,741
	2031	\$ 1,433,515,753	\$ -	\$ -	\$ 1,433,515,753
	2036	\$ 2,167,951,621	\$ -	\$ -	\$ 2,167,951,621
	2041	\$ 4,285,635,004	\$ 3,112,646,581	\$ -	\$ 7,398,281,585
	2046	\$ -	\$ -	\$ -	\$ -
	Total	\$ 9,024,709,118	\$ 3,112,646,581	\$ -	\$ 12,137,355,699
Logan	2026	\$ 1,555,565,184	\$ -	\$ 2,218,212,095	\$ 3,773,777,278
	2031	\$ 857,872,620	\$ -	\$ -	\$ 857,872,620
	2036	\$ 148,146,571	\$ -	\$ -	\$ 148,146,571
	2041	\$ 674,680,479	\$ 4,007,162,998	\$ -	\$ 4,681,843,477
	2046	\$ 1,207,167,831	\$ -	\$ -	\$ 1,207,167,831
	Total	\$ 3,343,332,695	\$ 4,007,162,998	\$ 2,218,212,095	\$ 9,568,752,801

LGA	Ano	Estrada	Trilhos	Ônibus	Custo (3%/ano)
	Total	\$ 4,443,432,684	\$ 4,007,162,998	\$ 2,218,212,095	\$ 10,668,807,777
Moreton Bay	2026	\$ 917,429,237	\$ -	\$ -	\$ 917,429,237
	2031	\$ 3,951,735,729	\$ -	\$ -	\$ 3,951,735,729
	2036	\$ 1,962,696,869	\$ -	\$ -	\$ 1,962,696,869
	2041	\$ 1,749,820,568	\$ -	\$ -	\$ 1,749,820,568
	2046	\$ 1,453,036,947	\$ -	\$ -	\$ 1,453,036,947
	Total	\$ 10,034,719,350	\$ -	\$ -	\$ 10,034,719,350
Redland	2026	\$ 180,153,581	\$ -	\$ -	\$ 180,153,581
	2031	\$ -	\$ -	\$ -	\$ -
	2036	\$ 26,723,668	\$ -	\$ -	\$ 26,723,668
	2041	\$ -	\$ -	\$ -	\$ -
	2046	\$ -	\$ -	\$ -	\$ -
	Total	\$ 206,877,248	\$ -	\$ -	\$ 206,877,248
Scenic Rim	2026	\$ 3,260,730	\$ -	\$ -	\$ 3,260,730
	2031	\$ -	\$ -	\$ -	\$ -
	2036	\$ -	\$ -	\$ -	\$ -
	2041	\$ 163,400,808	\$ -	\$ -	\$ 163,400,808
	2046	\$ -	\$ -	\$ -	\$ -
	Total	\$ 166,661,538	\$ -	\$ -	\$ 166,661,538
Somerset	2026	\$ -	\$ -	\$ -	\$ -
	2031	\$ -	\$ -	\$ -	\$ -
	2036	\$ -	\$ -	\$ -	\$ -
	2041	\$ -	\$ -	\$ -	\$ -
	2046	\$ -	\$ -	\$ -	\$ -
	Total	\$ -	\$ -	\$ -	\$ -
Sunshine Coast	2026	\$ 1,470,252,058	\$ -	\$ -	\$ 1,470,252,058
	2031	\$ 848,076,073	\$ -	\$ -	\$ 848,076,073
	2036	\$ 510,056,414	\$ -	\$ -	\$ 510,056,414
	2041	\$ 4,065,352,561	\$ 6,749,282,610	\$ -	\$ 10,814,635,171
	2046	\$ 1,609,571,724	\$ 14,897,690,066	\$ -	\$ 16,507,261,790
	Total	\$ 8,503,308,831	\$ 21,646,972,675	\$ -	\$ 30,150,281,506
NSW	2026	\$ -	\$ -	\$ -	\$ -
	2031	\$ 1,431,536,924	\$ -	\$ -	\$ 1,431,536,924
	2036	\$ -	\$ -	\$ -	\$ -
	2041	\$ -	\$ -	\$ -	\$ -
	2046	\$ -	\$ -	\$ -	\$ -
	Total	\$ 1,431,536,924	\$ -	\$ -	\$ 1,431,536,924

Fonte: Autor (2023).

Observa-se que todos os LGAs recebem algum tipo de obra de grande infraestrutura no período próximo de 23 anos, com exceção de Somerset. Esta região, apesar de possuir a maior área do sudeste de Queensland, é a que possui a menor



população, o que é um indicativo da falta de grandes investimentos no período estudado. Outro dado interessante é o investimento em Nova Gales do Sul (NSW), estado vizinho de Queensland; esse investimento se deve à parceria na construção de uma rodovia que cruza a cidade de Gold Coast e ultrapassa a divisa dos estados.

A média de investimentos por governo local foi de A\$10.286.423.991,28, sendo que cerca de 65% de todo o investimento foi feito nas três regiões mais povoadas da região, em ordem de investimentos: Sunshine Coast, Gold Coast e Brisbane. Como mencionado, os motivos são a grande aglomeração de pessoas, crescimento populacional que gera a necessidade de novas infraestruturas, além de possíveis decisões políticas. A categoria com menor investimento foi a de vias para ônibus, e a com maior investimento foi a de ferrovias (Tabela 3).

Tabela 3 - Custos Totais Previstos por Ano Cenário Referência

Year	Road	Rail	Bus	Estimated Cost (3%/y)
<b>2026</b>	\$ 7,896,396,065.89	\$ 9,098,992,683.01	\$5,206,386,315.91	\$ 22,201,775,064.81
<b>2031</b>	\$11,805,398,931.22	\$ -	\$2,139,106,150.33	\$ 13,944,505,081.55
<b>2036</b>	\$ 6,306,004,362.94	\$ 7,646,396,977.64	\$ -	\$ 13,952,401,340.58
<b>2041</b>	\$15,112,592,447.43	\$ 16,072,474,321.39	\$ -	\$ 31,185,066,768.82
<b>2046</b>	\$ 4,400,593,167.59	\$ 18,611,435,413.90	\$ -	\$ 23,012,028,581.49
<b>TOTAL</b>	\$45,520,984,975.07	\$ 51,429,299,395.94	\$7,345,492,466.24	\$ 104,295,776,837.26

Fonte: Autor (2023).

Devido ao alto custo de construção que as ferrovias possuem em comparação aos outros modos, com apenas 256,59 km de extensão, o modo ferroviário foi o que teve o maior custo de implementação com os projetos previstos pelo cenário Referência. Com 1.685,53 km de projetos, as estradas ficaram com a segunda colocação no ranking de investimentos, o que demonstra a diferença de valores de acordo com o tipo de infraestrutura.

As estradas para ônibus demonstraram o menor valor, com 14,74 km de extensão, e também apresentam uma característica interessante: é previsto investimentos nessa categoria apenas até 2031. Isso se deve à baixa quantidade de ruas exclusivas de ônibus na região, onde faixas para ônibus em ruas compartilhadas com outros tipos de veículos são mais comuns. Ou seja, faixas exclusivas de ônibus são majoritariamente incluídas na categoria de estradas desse estudo.

Em comparação, o cenário Modificado possui uma distribuição de projetos similar por LGA (Tabela 4). Visto que apenas projetos do modo rodoviário são diferentes, ver Seção 3.2, um fato que pode alterar os valores das categorias ferroviárias e de ônibus são os anos dos projetos, visto que alguns projetos do cenário de Referência podem estar adiantados ou atrasados em comparação à nova rede, assim alterando os valores de acordo com a inflação que é de 3% ao ano.

Tabela 4 - Custos Previstos por LGA Cenário Modificado

LGA	Ano	Estrada	Ferrovias	Ônibus	Custo (3%/ano)
Brisbane	2026	\$ 602,832,561	\$ 4,145,999,564	\$ 2,988,136,347	\$ 7,736,968,472
	2031	\$ 792,082,719	\$ -	\$ 2,139,136,493	\$ 2,931,219,212
	2036	\$ 1,011,245,889	\$ -	\$ -	\$ 1,011,245,889
	2041	\$ 2,125,562,585	\$ 2,203,384,913	\$ -	\$ 4,328,947,498
	2046	\$ -	\$ -	\$ -	\$ -
	TOTAL	\$ 4,531,723,754	\$ 6,349,384,477	\$ 5,127,272,839	\$ 16,008,381,070
Gold Coast	2026	\$ 2,057,407,414	\$ 4,953,025,676	\$ -	\$ 7,010,433,090
	2031	\$ 1,076,516,822	\$ -	\$ -	\$ 1,076,516,822
	2036	\$ 479,192,666	\$ 7,646,388,221	\$ -	\$ 8,125,580,887
	2041	\$ 2,048,164,483	\$ -	\$ -	\$ 2,048,164,483
	2046	\$ 130,815,477	\$ 3,713,741,755	\$ -	\$ 3,844,557,232
	TOTAL	\$ 5,792,096,862	\$ 16,313,155,652	\$ -	\$ 22,105,252,515
Gympie	2026	\$ -	\$ -	\$ -	\$ -
	2031	\$ 1,414,060,649	\$ -	\$ -	\$ 1,414,060,649
	2036	\$ -	\$ -	\$ -	\$ -
	2041	\$ -	\$ -	\$ -	\$ -
	2046	\$ -	\$ -	\$ -	\$ -
	TOTAL	\$ 1,414,060,649	\$ -	\$ -	\$ 1,414,060,649
Ipswich	2026	\$ 2,052,967,779	\$ -	\$ -	\$ 2,052,967,779
	2031	\$ 71,452,343	\$ -	\$ -	\$ 71,452,343
	2036	\$ 2,258,026,778	\$ -	\$ -	\$ 2,258,026,778
	2041	\$ 4,285,618,514	\$ 3,112,642,401	\$ -	\$ 7,398,260,915
	2046	\$ -	\$ -	\$ -	\$ -
	TOTAL	\$ 8,668,065,415	\$ 3,112,642,401	\$ -	\$ 11,780,707,816
Logan	2026	\$ 1,601,320,438	\$ -	\$ 2,218,165,767	\$ 3,819,486,206
	2031	\$ 857,873,280	\$ -	\$ -	\$ 857,873,280
	2036	\$ 148,146,534	\$ -	\$ -	\$ 148,146,534
	2041	\$ 674,651,593	\$ 4,007,161,347	\$ -	\$ 4,681,812,940
	2046	\$ 1,207,164,576	\$ -	\$ -	\$ 1,207,164,576
	TOTAL	\$ 4,489,156,421	\$ 4,007,161,347	\$ 2,218,165,767	\$ 10,714,483,536
Moreton Bay	2026	\$ 263,632,996	\$ -	\$ -	\$ 263,632,996
	2031	\$ 4,012,843,195	\$ -	\$ -	\$ 4,012,843,195
	2036	\$ 2,016,876,338	\$ -	\$ -	\$ 2,016,876,338

LGA	Ano	Estrada	Ferrovias	Ônibus	Custo (3%/ano)
	2041	\$ 1,680,860,362	\$ -	\$ -	\$ 1,680,860,362
	2046	\$ 1,453,036,375	\$ -	\$ -	\$ 1,453,036,375
	TOTAL	\$ 9,427,249,266	\$ -	\$ -	\$ 9,427,249,266
Redland	2026	\$ 180,152,575	\$ -	\$ -	\$ 180,152,575
	2031	\$ -	\$ -	\$ -	\$ -
	2036	\$ 26,723,688	\$ -	\$ -	\$ 26,723,688
	2041	\$ -	\$ -	\$ -	\$ -
	2046	\$ -	\$ -	\$ -	\$ -
	TOTAL	\$ 206,876,263	\$ -	\$ -	\$ 206,876,263
Scenic Rim	2026	\$ 3,260,775	\$ -	\$ -	\$ 3,260,775
	2031	\$ -	\$ -	\$ -	\$ -
	2036	\$ -	\$ -	\$ -	\$ -
	2041	\$ 163,400,700	\$ -	\$ -	\$ 163,400,700
	2046	\$ -	\$ -	\$ -	\$ -
	TOTAL	\$ 166,661,475	\$ -	\$ -	\$ 166,661,475
Somerset	2026	\$ -	\$ -	\$ -	\$ -
	2031	\$ -	\$ -	\$ -	\$ -
	2036	\$ -	\$ -	\$ -	\$ -
	2041	\$ -	\$ -	\$ -	\$ -
	2046	\$ -	\$ -	\$ -	\$ -
	TOTAL	\$ -	\$ -	\$ -	\$ -
Sunshine Coast	2026	\$ 2,195,124,932	\$ -	\$ -	\$ 2,195,124,932
	2031	\$ 999,956,964	\$ -	\$ -	\$ 999,956,964
	2036	\$ 510,054,757	\$ -	\$ -	\$ 510,054,757
	2041	\$ 4,590,414,169	\$ 6,749,313,601	\$ -	\$ 11,339,727,770
	2046	\$ 1,609,572,793	\$ 14,897,708,600	\$ -	\$ 16,507,281,393
	TOTAL	\$ 9,905,123,615	\$ 21,647,022,202	\$ -	\$ 31,552,145,816
NSW	2026	\$ -	\$ -	\$ -	\$ -
	2031	\$ 1,477,481,501	\$ -	\$ -	\$ 1,477,481,501
	2036	\$ -	\$ -	\$ -	\$ -
	2041	\$ -	\$ -	\$ -	\$ -
	2046	\$ -	\$ -	\$ -	\$ -
	TOTAL	\$ 1,477,481,501	\$ -	\$ -	\$ 1,477,481,501

Fonte: Autor (2023).

Nesse cenário, nota-se novamente que Somerset é o único LGA que não recebe grandes investimentos em infraestruturas de transporte. A proporção dos investimentos continua a mesma, com cerca de 66,4% de todo o investimento sendo alocado nas três maiores regiões: Sunshine Coast, Gold Coast e Brisbane. O valor total (Tabela 5) também foi muito próximo do esperado, devido às pequenas mudanças entre as duas previsões.

Tabela 5 - Custo Total por Ano Cenário Modificado

Ano	Estrada	Ferrovia	Ônibus	Custo (3%/ano)
2026	\$ 8,956,699,470.97	\$ 9,099,025,240.10	\$5,206,302,113.87	\$ 23,262,026,824.94
2031	\$10,702,267,474.25	\$ -	\$2,139,136,492.56	\$ 12,841,403,966.81
2036	\$ 6,450,266,649.47	\$ 7,646,388,221.21	\$ -	\$ 14,096,654,870.68
2041	\$15,568,672,404.94	\$16,072,502,262.18	\$ -	\$ 31,641,174,667.12
2046	\$ 4,400,589,220.77	\$18,611,450,355.22	\$ -	\$ 23,012,039,575.99
<b>TOTAL</b>	<b>\$46,078,495,220.40</b>	<b>\$51,429,366,078.72</b>	<b>\$7,345,438,606.43</b>	<b>\$ 104,853,299,905.55</b>

Fonte: Autor (2023).

Como anteriormente mencionado, apenas a categoria de estradas foi modificada, resultando em um acréscimo no custo de A\$557.510.245,33. O segundo cenário foi de suma importância para reforçar a validação do modelo, visto que os cenários de ferrovia e ônibus foram custeados de maneira similar, conforme esperado devido à igualdade de links para os modos.

Como o modelo previu ambos os cenários em diferentes execuções do Alteryx Designer, a proximidade de valores indica estabilidade do modelo, ou seja, uma precisão mais alta. Sendo assim, o modelo pode até estar prevendo valores fora da otimalidade, como visto anteriormente na seção 4.1, porém ele é consistente em prever esses valores.

### 4.3. Análise de Benefícios

Dentro da análise de custo-benefício deste trabalho, propôs-se inicialmente uma comparação direta entre os custos de construção dos diferentes cenários estudados. Conforme visto na seção anterior, o cenário com menor custo total foi o de Referência, apresentando uma diferença de A\$557.523.068,29, ou 0,534%, em relação ao Modificado.

Dessa forma, a análise de custo de infraestrutura não se mostra, nesse caso, como uma variável que define quão melhor é um cenário em relação ao outro, uma vez que seus custos são muito similares. Para os tomadores de decisão, outras variáveis podem ser muito mais impactantes do que essa pequena diferença de custo total, como impactos ambientais, custo por usuário, economia de tempo e diversos fatores que afetam o funcionamento e a fluidez do tráfego.

A segunda análise proposta é o custo do tempo de viagem economizado. Para tal, calcula-se o VBTT para ambos os cenários, que será o mesmo, pois os valores dos salários médios e porcentagem do tempo de viagem convertido em valor monetário também são iguais. Substituindo os valores na fórmula, obtém-se A\$26,20 por hora economizada por usuário.

O tempo médio de viagem por link da rede Base (atual e sem modificações), é de 0,019591 horas ou 1,17546 minutos. Utiliza-se esse valor como base de comparação para os tempos de viagem entre as simulações de ambos os cenários alternativos. O cenário de Referência obteve uma melhoria no tempo de viagem médio entre links, com 0,017397 horas ou 1,04382 minutos, representando uma melhoria de 0,13164 minutos. Já o cenário Modificado obteve uma melhoria para 0,0172527 horas ou 1,035162 minutos, garantindo 0,140298 minutos de melhoria em comparação à rede Base.

Multiplicando os tempos pelo VBTT calculado, chega-se ao valor economizado por viagem por link (Tabela 6). É importante ressaltar que a média de comprimento de links é de 0,489761 km, ou 489,761 metros. Os valores são calculados por link, ou seja, o valor economizado por um usuário a cada link viajado, já que os dados de tempo de viagem por cenário não chegaram até data da análise deste trabalho, apesar de terem sido solicitados via e-mail e por meio de conferência online múltiplas vezes.

Tabela 6 - Custo Economizado por Viagem

Custo Economizado por Viagem por Link				
Cenário	Velocidade (Km/h)	Economia de Tempo (h)	VBTT	Custo Economizado (A\$/h/usuário)
Referência	49,6639	0,002194	26,20	0,0574828
Modificado	51,2539	0,0023383	26,20	0,06126346

Fonte: Autor (2023).

Com uma diferença no custo economizado por viagem de A\$0,00378066 hora/usuário/link, o melhor cenário foi o Modificado. Apesar de o valor parecer pequeno, é importante lembrar que esse custo não se refere à totalidade da viagem do usuário, mas sim ao tempo que ele leva para percorrer o link de A a B ou 489,761

metros. Considerando que uma viagem consiste em um conjunto de diversos links, sendo esse o valor multiplicado pelo número de links utilizados e pelo número de usuários. Portanto, o custo total de uma viagem completa seria composto por várias parcelas desse valor, resultando em uma quantia mais significativa.

Pode-se atualizar o valor para uma variável de distância, utilizando do comprimento médio de links, assim chega-se a um custo economizado de viagem no cenário Referência de A\$0,11736 hora/usuário/km e de A\$0,1251 hora/usuário/km. Assim, concluindo novamente a melhoria do cenário Modificado de A\$0,00772 hora/usuário/km.

Infelizmente, os dados necessários para as análises de impactos ambientais e custo por usuário nas vias não foram enviados em tempo hábil. Apesar das múltiplas solicitações por diversos meios, não foi possível obter essas informações devido a fatores externos. A ausência desses dados inviabiliza a realização dessas análises, pois não é possível substituir ou estimar com precisão os números de frota e usuários.

Com base na análise de custo de construção, permite-se concluir que os stakeholders do governo devem avaliá-la para verificar se há orçamento disponível para as obras. A decisão de implementar um cenário em detrimento do outro dependerá dessa avaliação. No caso deste estudo, a diferença no valor é pequena em relação ao montante total, sugerindo que não deve ser um impedimento significativo para a implementação de um cenário em comparação ao outro.

Quanto ao custo do tempo economizado por viagem por link, conclui-se que o cenário Modificado apresentou uma simulação mais eficiente, tornando-o uma opção favorável em comparação ao Referência.

## 5. CONCLUSÃO

O *Option Analysis* é uma ferramenta que utiliza *machine learning* com o método Random Forest. Sua criação teve como objetivo principal integrar os dados de saída do software de macrossimulação Emme em um fluxo de trabalho no Alteryx Designer para a análise de alternativas de investimentos, com base nos custos de construção de transporte. Esses custos são derivados de dados fornecidos pelo Departamento de Transportes e *Main Roads* de Queensland, Austrália, e por empresas privadas parceiras do governo.

Para alcançar esse objetivo, foi proposto que, após a coleta e recebimento dos dados, eles fossem filtrados com base em sua relevância para o treinamento do modelo econômico. Essa relevância seria determinada pela correlação com a variável objetivo, que é o custo de construção da infraestrutura. As variáveis selecionadas foram velocidade de simulação, comprimento do link e modo (tipo de infraestrutura), utilizadas para treinar o modelo de *machine learning*.

Com os inputs selecionados, o processo de criação do modelo foi realizado. Como mencionado anteriormente, o método Random Forest foi escolhido. Para prever custos, o modelo precisou utilizar uma técnica adicional, somando o número de faixas de vias que possuem duplicatas dentro da simulação de tráfego, representando diferentes sentidos das vias, e depois dividindo por dois.

Foram propostas quatro análises para avaliar diferentes cenários (Referência e Modificado): a comparação direta dos custos de construção, a economia do custo de tempo de viagem na rede, o custo dos impactos ambientais com base na composição de tráfego da simulação e o custo da infraestrutura diluído por usuário da via. Essas análises permitem a comparação das diferentes redes de tráfego e a definição de um cenário mais favorável.

A execução dessa etapa depende da disponibilidade de dados provenientes das fontes mencionadas acima. Contudo, devido a atrasos no processo de entrega desses dados por empresas que trabalham com o TMR australiano, análises ambientais dependentes dos dados de composição de tráfego e a análise de custo por usuário não puderam ser realizadas, constituindo um grande limitante para o estudo.

O projeto é de grande relevância para auxiliar na tomada de decisão em relação a cenários de tráfego futuro. O TMR destacou, em reuniões, a importância do modelo, principalmente o de previsão de custos, evidenciando como a ferramenta pode fornecer uma análise de alto nível sobre custos e comparação de cenários de forma rápida e econômica.

Os cenários mostraram-se bastante parelhos ao utilizar a métrica de custo total de construção, com uma diferença de apenas A\$557.510.245,33, sendo o cenário Modificado o mais caro. No entanto, esse valor é irrisório quando comparado ao montante total dos cenários, representando pouco mais que 0,5% dos custos de ambos.

Outra análise realizada foi a de custo economizado por tempo de viagem. Com uma diferença de A\$0,00772 hora/usuário/km, o cenário Modificado foi o que mais economizou tempo de viagem. Embora esse valor pareça pequeno, é importante destacar que ele deve ser multiplicado pela quantidade de quilômetros percorridos pelos usuários durante sua viagem, bem como pelo número de usuários no sistema.

Como mencionado anteriormente, as análises de impactos ambientais e custo por usuário não puderam ser realizadas devido à falta de dados. Portanto, com base nos resultados obtidos, pode-se afirmar que, a menos que os A\$557.510.245,33 de custo adicional no cenário Modificado sejam considerados significativos, este cenário é o mais indicado com base nas análises realizadas neste estudo.

Apesar de dados importantes para uma análise custo benefício não terem sido utilizados, devido a atrasos na entrega por parte das empresas envolvidas, o método foi validado e se mostrou proficiente nas atividades de custeio e análise de economia de tempo.

Para trabalhos futuros, sugere-se incluir na análise de custos uma avaliação dos impactos ambientais, considerando fatores como tipo de terreno, desmatamento e possíveis danos à fauna e flora, além das emissões para a construção e para a frota simulada. A adição dos dados faltantes neste estudo também é crucial para melhorias futuras.



## REFERÊNCIAS

ALPAYDIN, Ethem. **Introduction to Machine Learning**. 2nd ed. Massachusetts: The MIT Press, 2010.

ALTERYX. Alteryx: Soluções para análise de dados. Disponível em: <https://www.alteryx.com/>. Acesso em: 21 maio 2023.

Alteryx. Ferramentas do Designer. Alteryx Help, acesso em 01 de novembro de 2023. Disponível em: <https://help.alteryx.com/current/pt/designer/tools.html>

Andriola, Cesar L., Bonatto, Andrey Z. E., Dutra, Diego L. **Análise Custo Benefício em Transportes: O Valor Estatístico da Vida para o Brasil**. In: Congresso de Pesquisa e Ensino em Transporte da ANPET, 33º, 2019, Balneário Camboriú. Universidade Federal do Rio Grande do Sul, Laboratório de Sistemas de Transportes.

AUSTRÁLIA, Australian Bureau of Statistics. **Average Weekly Earnings**, Australia. ABS, 2023. Disponível em: <https://www.abs.gov.au/statistics/labour/earnings-and-working-conditions/average-weekly-earnings-australia/may-2023>. Acesso em: 12 outubro 2023.

AUSTRÁLIA, Department of Transport and Main Roads. **South East Queensland Multi Model Project Report**. Australia, 2023; Documento não disponível publicamente.

AUSTRÁLIA, Queensland. Brisbane City Council. **Local Government Infrastructure Plan**. Brisbane, 2023. Disponível em: <https://www.brisbane.qld.gov.au/planning-and-building/planning-guidelines-and-tools/local-government-infrastructure-plan/lgip-components-extrinsic-material-and-supporting-documents>. Acesso em: 12 outubro 2023.

AUSTRÁLIA, Transports and Main Roads. **Queensland Transport and Roads Investment Program (2000 – 2020)**. Brisbane, Queensland. 2020.

AUSTRÁLIA. Australian Bureau of Statistics. **National, state and territory population**. 2022. Disponível em: <https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release>. Acesso em: 20 junho 2022.

AUSTRÁLIA. Australian Bureau of Statistics. **Population Projections, Australia**. 2018. Disponível em: <https://www.abs.gov.au/statistics/people/population/population-projections-australia/2017-base-2066#cite-window1>. Acesso em: 05 dez. 2022.

AUSTRALIA. Australian Bureau of Statistics. **Producer Price Indexes**, Australia. Canberra, 2023. Disponível em: <https://www.abs.gov.au/statistics/economy/price->

indexes-and-inflation/producer-price-indexes-australia/latest-release#data-downloads. Acesso em: 22 de novembro de 2023.

AUSTRÁLIA. Department of Transport and Main Roads. **QTRIP Interactive Tool**. Queensland, 2022. Disponível em: <https://www.tmr.qld.gov.au/QTRIPonline>. Acesso em: 05 dez. 2022.

AUSTRALIA. Office of the Auditor General. **Main Roads Projects to Address Traffic Congestion**. Australia, 2014. Disponível em: [https://audit.wa.gov.au/wp-content/uploads/2015/03/report2015\\_02-TrafficCongestion.pdf](https://audit.wa.gov.au/wp-content/uploads/2015/03/report2015_02-TrafficCongestion.pdf). Acesso em: 05 dez. 2022.

AUSTRÁLIA. Parliament of Australia. **Australia's future population**. 2010. Disponível em: [https://www.aph.gov.au/About\\_Parliament/Parliamentary\\_Departments/Parliamentary\\_Library/pubs/BriefingBook43p/futurepopulation#:~:text=According%20to%20the%20ABS%2C%20Australia%27s,62.2%20million%20people%20by%202101](https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook43p/futurepopulation#:~:text=According%20to%20the%20ABS%2C%20Australia%27s,62.2%20million%20people%20by%202101). Acesso em: 20 junho 2022.

BARCELÓ, Jaume. **Fundamentals of Traffic Simulation, International Series in Operations Research & Management Science**. Edição 2010. Springer Science Business Media, 2011.

BIAU, Gérard. **Analysis of a random forests model**. Journal of Machine Learning Research, Université Pierre et Marie Curie, Paris, França, 2012.

Boardman, A. E., Greenberg, D. H., Vining, A. R., Weimer, D. L. **Cost-benefit analysis: Concepts and practice**. 5<sup>th</sup> edition, Reino Unido, Cambridge University Press, 2017.

BRASIL. Ministério da Economia. Secretaria Especial de Produtividade, Emprego e Competitividade. **Guia Prático de Análise Custo-Benefício de Projetos de Investimento em Infraestrutura**. Brasília, 2020.

BREIMAN, L. **Random forests**. *Machine Learning*, Kluwer Academic Publishers, p. 5-32, Holanda, 2001.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 8. ed. São Paulo: Saraiva, 2017.

CHOWDHURY, D.; SANTEN, L.; SCHADSCHNEIDER, A. **Statistical Physics of Vehicular Traffic and Some Related Systems**. Physics Report, North-Holland, Países Baixos, 199-329, 2000.

Corporate Finance Institute. **What are Economies of Scale?**. Disponível em: <https://corporatefinanceinstitute.com/resources/knowledge/economics/economies-of-scale/>. Acesso em 28 de outubro de 2023.

Cowie, J., Ison S. **The Routledge Handbook of Transport Economics**. 1<sup>st</sup> edition, Reino Unido, Routledge. 2017.

CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. **Random forests**. In: ZHANG, Cha (Ed.), MA, Yungian (Ed.). **Ensemble Machine Learning: Methods and Applications**. Boston: Springer, 2012. p. 157-175.

DEPARTMENT OF TRANSPORTS AND MAIN ROADS. **Multimodel Framework**. Queensland: Departamento de Transportes e Estradas de Queensland. 2023. Documento não disponível para o público.

Ferrari, T.K, et al. **Estimativa do Valor da Vida Estatística e do Valor da Economia de Tempo em Viagens nas Rodovias Brasileiras com a Utilização de Pesquisa de Preferência Declarada**. Brasília: Ipea, 2019. (Texto para Discussão, n. 2533).

Flyvbjerg, B., Bruzelius, N., Rothengatter, W. **Megaprojects and risk: An anatomy of ambition**. Reino Unido, Cambridge University Press, 2003.

FRANCOIS-LAVET, Vincent, et al. **Introduction to Deep Reinforcement Learning**. Foundations and Trends in Machine Learning: Vol. 11, No. 3-4, 2018.

GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 1. ed. O'Reilly Media, 2017. Capítulo 4, p. 119-184.

HASTIE, T.; TIBSHIRANI, R. **Generalized additive models**. 1<sup>st</sup> ed. Estados Unidos, Chapman Hall CRC press, 1990.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2nd ed. New York: Springer, 2009.

Hensher, D.A., **Value of Business Travel Time**. Oxford: Pergamon Press, 1977. Reino Unido.

HOAGLIN, David C.; MOSTELLER, Frederick; TUKEY, John W. **Exploring Data Tables, Trends, and Shapes**. Estados Unidos. Wiley Interscience, 2006.

INRO. **What's New in EMME 4.6**. 2021. Disponível em: <<https://info.inrosoft.com/blog/whats-new-in-emme-4-6>>. Acesso em: 22 de novembro de 2023.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An Introduction to Statistical Learning**. 2<sup>nd</sup> edition, Estados Unidos, Nova Iorque, Springer, 2013.

JAMES, M. **Transport Infrastructure**. Parliament of Australia, 2014. Disponível em: [https://www.aph.gov.au/About\\_Parliament/Parliamentary\\_Departments/Parliamentary\\_Library/pubs/BriefingBook44p/Transport#:~:text=As%20Australia%20advances%20into%20the,general%20shipping%20delays%20and%20restrictions](https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook44p/Transport#:~:text=As%20Australia%20advances%20into%20the,general%20shipping%20delays%20and%20restrictions). Acesso em: 05 dez. 2022.

JOLLIFFE, I. T.; CADIMA, J. **Principal component analysis: a review and recent developments**. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, v. 374, n. 2065, p. 20150202, 2016.

KOTUSEVSKI, G.; HAWICK, K.A. **A review of traffic simulation software**. Computer Science, Institute of Information and Mathematical Sciences, Massey University. Auckland, Nova Zealândia, v. 13, p. 35-54, 2009.

LIAW, A.; WIENER, M. **Classification and regression by random forest**. R News, [S.l.], v.2, n.3, p.18-22, 2002. Disponível em: [https://cran.r-project.org/doc/Rnews/Rnews\\_2002-3.pdf#page=18](https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf#page=18). Acesso em: 23 abr. 2023.

LINDNER, A. **Análise de Dados em Estatística Desempenha um Papel Crucial na Modelagem e Previsão de Demanda de Transporte**. Dissertação (Mestrado em Ciências) – Escola de engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2015.

LONGLEY, P.A. et al. **Geographic Information Science and Systems**. 4. Ed. John Wiley & Sons, Inc, Estados Unidos, 2015.

LOPES, S, B. **Uma ferramenta para planejamento da mobilidade sustentável com base em modelo de uso do solo e transportes**. 2010. Tese (Doutorado em Engenharia de Transportes) – Escola de engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2010.

LOUPPE, G. **Understanding random forests: from theory to practice**. Ph.D. Thesis, Université de Liège, Liege, França 2014.

MAHMASSANI, Hani S. **Dynamic Network Traffic Assignment and Simulation Methodology for Advanced System Management Applications**. Networks and Spatial Economics, 267-292, 2001.

MARSH, L.C., CORMIER, D.R. **Spline Regression Models**. Sage University, Califórnia, 2002.

MARTÍNEZ, F. J.; ARAYA, C. **Land-use impacts of transport projects: user benefits, rents and externalities**. In: World Conference on Transport Research, Antwerp, July. 1998.

McKinnon, A. Button, K., Nijkamp, P. **Transport Logistics**. Reino Unido, Edward Elgar Publishing, 2003.

MICROSOFT. Microsoft Excel. Disponível em: <https://www.microsoft.com/pt-br/microsoft-365/excel>. Acesso em: 21 maio 2023.

PALMA, A. *et al.* **A Handbook of Transport Economics**. Edward Elgar Publishing, 2011. Reino Unido.

PÉREZ, Fernando L. **Modelos Lineares Generalizados**. Universidade Federal do Paraná, Departamento de Estatística, 2022. Disponível em:

<http://leg.ufpr.br/~lucambio/GLM/GLM.html>. Acesso em: 22 de novembro de 2023.

PFAFFENBICHLER, P. **The Strategic, Dynamic and Integrated Urban Land Use Transport Model MARS - (Metropolitan Activity Relocation Simulator) - Development, testing and application**. Ph. D. Thesis (Transport Planning and

Traffic Engineering). Institute for Transport Planning and Traffic Engineering - Vienna University of Technology, Viena, p. 11-46, 2003.

PIETRANTONIO, H., STRAMBI, O. e GUALDA, N. D. F. **Integração entre políticas de uso de solo e de transportes: dificuldades e necessidades**. 1996, Anais. Brasília: Anpet, 1996.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**. 1. Ed. Estados Unidos. O'Reilly Media, 2013.

PYTHON SOFTWARE FOUNDATION. **Python: Python Programming Language**. Disponível em: <https://www.python.org/>. Acesso em: 21 maio 2023.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, 2021. Disponível em: <https://www.R-project.org/>. Acesso em: 21 maio 2023.

RAMSAY, J. O.; SILVERMAN, B. W. **Functional data analysis**. 2<sup>nd</sup> edition, Estados Unidos, Springer Science & Business Media, 2005.

Rodrigue, J. P., Comtois, C., Slack, B. **The geography of transport systems**. 5<sup>th</sup> edition, Reino Unido, Routledge, 2006.

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning: An Introduction**. 2nd ed. Cambridge, MA: MIT Press, 2018.

WEGENER, M. **Overview of land-use and transport models**. In: CUPUM03 - THE 8th INTERNATIONAL CONFERENCE ON COMPUTERS IN URBAN PLANNING AND URBAN MANAGEMENT, 2003, Sendai. Anais...Center for Northeast Asian Studies, Tohoku University, 2003.

WERNECK, L. F. **Contribuição de um modelo LUTI para o planejamento integrado em transportes e uso do solo na região metropolitana de Belo Horizonte**. Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Minas Gerais, 2015.