



**UFOP**

Universidade Federal  
de Ouro Preto

**Universidade Federal de Ouro Preto  
Instituto de Ciências Exatas e Aplicadas  
Departamento de Computação e Sistemas**

**Espaço de dados veicular: coleta,  
processamento e análises**

**Igor Gabriel Souza Silva**

**TRABALHO DE  
CONCLUSÃO DE CURSO**

**ORIENTAÇÃO:  
Bruno Pereira dos Santos**

**Março, 2022  
João Monlevade–MG**

**Igor Gabriel Souza Silva**

**Espaço de dados veicular: coleta,  
processamento e análises**

Orientador: Bruno Pereira dos Santos

Monografia apresentada ao curso de Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

**Universidade Federal de Ouro Preto**

**João Monlevade**

**Março de 2022**

## SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

S586e Silva, Igor Gabriel Souza.  
Espaço de dados veicular [manuscrito]: coleta, processamento e análises. / Igor Gabriel Souza Silva. - 2022.  
66 f.: il.: color., gráf., tab., mapa.

Orientador: Prof. Dr. Bruno Pereira dos Santos.  
Monografia (Bacharelado). Universidade Federal de Ouro Preto.  
Instituto de Ciências Exatas e Aplicadas. Graduação em Sistemas de Informação .

1. Aquisição de dados. 2. Interface de programas aplicativos (Software). 3. Mobilidade social - Automóveis. 4. Sistemas de recuperação da informação. I. Santos, Bruno Pereira dos. II. Universidade Federal de Ouro Preto. III. Título.

CDU 004.775

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431



## FOLHA DE APROVAÇÃO

Igor Gabriel Souza Silva

**Espaço de dados veicular: coleta, processamento e análises**

Monografia apresentada ao Curso de Sistemas de Informação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação

Aprovada em 20 de Junho de 2022

### Membros da banca

Dr. Bruno Pereira dos Santos - Orientador (Universidade Federal de Ouro Preto)  
Dra. Tatiana Alves Costa (Universidade Federal de Ouro Preto)  
Dr. Luiz Carlos Bambirra Torres (Universidade Federal de Ouro Preto)

Bruno Pereira dos Santos, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 30/06/2022



Documento assinado eletronicamente por **Bruno Pereira dos Santos, PROFESSOR DE MAGISTERIO SUPERIOR**, em 30/06/2022, às 16:36, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0355141** e o código CRC **8247542E**.

*Este trabalho é dedicado aos meus pais, por todo suporte e apoio que me deram durante toda a trajetória acadêmica, a todos amigos e colegas, que desde o início me incentivaram e mantiveram uma parceria fundamental para o avanço na jornada, ao corpo discente e docente da Universidade Federal de Ouro Preto, que me transmitiram enorme conhecimento e a Deus, por nunca ter me abandonado nos momentos difíceis.*

# Agradecimentos

Agradeço aos docentes da Universidade Federal de Ouro Preto, por todo o apoio e aprendizado transmitido e ao meu professor orientador Bruno Pereira dos Santos, por toda a orientação e ensinamentos passados durante a elaboração deste trabalho.

*“Science is more than a body of knowledge; it is a way of thinking.”*

— Carl Sagan (1934 – 1996),  
*in: The Demon-Haunted World: Science as a Candle in the Dark.*

# Resumo

Elementos de mobilidade urbana passaram a ser uma dificuldade devido o contínuo aumento da população global. Simultaneamente, novas tecnologias para aprimorar os sistemas de mobilidade surgiram, devido a existência de uma maior quantidade de dados disponíveis em fontes internas e externas aos automóveis. Porém, existem restrições que podem limitar o acesso as informações durante a aquisição de dados que retratam os cenários dos trânsitos, resultando em uma escassez de conjuntos de dados relacionados a veículos. Nesta monografia, foi apresentado a criação de um conjunto de dados veicular, contendo dados intra e extra veiculares, com a finalidade de trazer um conjunto de dados público e com o máximo de informações disponíveis. Para a coleta de dados intra-veiculares, foi utilizado a interface *On-Board Diagnostic* (OBD) contida em veículos. Para a coleta de informações extra-veiculares, foi utilizado a *Application Programming Interface* (API) do *Google Maps*, *Here WeGo* e *Weather Channel*. Também foi realizada a rotulação e análise exploratória dos dados, com o objetivo de verificar o comportamento dos veículos em situações de condução diferentes e problemas existentes em dados relacionados a veículo. Na análise exploratória dos dados, foi possível identificar diferenças esperadas nas condições de condução dos automóveis e problemas que ocorrem em dados veiculares. Com o conjunto de dados desenvolvido, se torna possível diversos novos trabalhos futuros, além dos exemplos de análises demonstradas.

**Palavras-chaves:** OBD. Conjunto de dados. API. Intra-veicular. Extra-veicular.



# Abstract

Elements of urban mobility have become a difficulty due to the continuous increase in the global population. Simultaneously, new technologies to improve mobility systems have emerged, due to the existence of a greater amount of data available from internal and external sources to automobiles. However, there are restrictions that can limit access to information during data acquisition that depict traffic scenarios, resulting in a scarcity of vehicle-related datasets. In this monograph, the creation of a vehicular dataset was presented, containing intra and extra vehicular data, with the purpose of bringing a public dataset and with the maximum of available information. For the collection of intra-vehicular data, the OBD interface contained in vehicles was used. For the collection of extra-vehicular information, the API of Google Maps, Here WeGo and Weather Channel was used. Labeling and exploratory data analysis were also carried out, with the objective of verifying the behavior of vehicles in different driving situations and existing problems in vehicle-related data. In the exploratory data analysis, it was possible to identify expected differences in car driving conditions and problems that occur in vehicle data. With the data set developed, it becomes possible to several new future works, in addition to the examples of demonstrated analysis.

**Key-words:** OBD. Dataset. API. Intra-vehicular . Extra-vehicular.

# Lista de ilustrações

Figura 1 – Metodologia . . . . .	14
Figura 2 – Taxonomia dos dados veiculares . . . . .	18
Figura 3 – Conector OBD . . . . .	22
Figura 4 – Scanner ELM 327 . . . . .	23
Figura 5 – Desenvolvimento do conjunto de dados . . . . .	29
Figura 6 – Rota percorrida nos cenários 1, 2, 3, 4 e 8 . . . . .	36
Figura 7 – Correlação positiva entre variáveis . . . . .	38
Figura 8 – rastro de mobilidade . . . . .	39
Figura 9 – Distância de viagem . . . . .	39
Figura 10 – Velocidade (OBD) por cenário . . . . .	40
Figura 11 – km/l por cenário . . . . .	40
Figura 12 – Direção ociosa por cenário . . . . .	41
Figura 13 – Comportamento dos veículos . . . . .	42
Figura 14 – Dados incompletos . . . . .	43
Figura 15 – Dados conflitantes . . . . .	44
Figura 16 – Dados desiguais . . . . .	44
Figura 17 – Ponto fora da curva . . . . .	45
Figura 18 – Ponto fora da curva corrigido . . . . .	46

# Lista de tabelas

Tabela 1 – Evolução OBD . . . . .	21
Tabela 2 – Aplicativos de diagnóstico veicular . . . . .	24
Tabela 3 – Trabalhos relacionados . . . . .	26
Tabela 4 – Aplicativos de diagnóstico veicular . . . . .	30
Tabela 5 – Dados disponíveis para a coleta . . . . .	31
Tabela 6 – Links de acesso ao conjunto de dados . . . . .	35
Tabela 7 – Viagens por cenário . . . . .	36
Tabela 8 – Comparação entre tempo real da viagem e tempo das APIs . . . . .	42
Tabela 9 – Colunas do Dataset . . . . .	52
Tabela 9 – Colunas do Dataset . . . . .	53
Tabela 9 – Colunas do Dataset . . . . .	54
Tabela 9 – Colunas do Dataset . . . . .	55
Tabela 9 – Colunas do Dataset . . . . .	56
Tabela 9 – Colunas do Dataset . . . . .	57
Tabela 9 – Colunas do Dataset . . . . .	58
Tabela 9 – Colunas do Dataset . . . . .	59
Tabela 9 – Colunas do Dataset . . . . .	60
Tabela 9 – Colunas do Dataset . . . . .	61
Tabela 9 – Colunas do Dataset . . . . .	62

# Lista de abreviaturas e siglas

**ALDL** *Assembly Line Diagnostic Link*

**API** *Application Programming Interface*

**CAN** *Controller Area Network*

**CSV** *Comma-separated values*

**ECU** *Engine Control Unit*

**GPS** *Global Positioning System*

**ITS** *Intelligent Transportation Systems*

**OBD** *On-Board Diagnostic*

**PID** *Parâmetro ID*

**RPM** *Rotação por Minuto*

**VBS** *Visual Basic Scripting*

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	O problema de pesquisa	13
1.2	Objetivo Geral	13
1.3	Objetivo Específico	14
1.4	Metodologia	14
1.5	Organização do trabalho	15
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>16</b>
2.1	<b>ARCABOUÇO TÉCNICO INTRODUTÓRIO</b>	<b>16</b>
2.1.1	Sistemas de Transporte Inteligentes	16
2.1.2	Espaço de dados veicular	17
2.1.3	Fontes de dados veiculares	17
2.1.3.1	Fontes intra-veiculares	18
2.1.3.2	Fontes extra-veiculares	18
2.1.4	Unidade de Controle Eletrônico	19
2.1.5	Rede de Área Controlada	19
2.1.6	<i>On-boarding Diagnostic</i>	20
2.1.6.1	OBD-I	21
2.1.6.2	OBD-II	21
2.1.6.2.1	Parâmetro ID	22
2.1.7	<i>Scanner</i>	22
2.1.7.1	ELM 327	23
2.1.8	Torque PRO	23
2.1.9	<i>Application Programming Interface</i>	24
2.1.10	Ferramentas e serviços utilizados	25
2.1.11	conjunto de dados	25
<b>2.2</b>	<b>TRABALHOS RELACIONADOS</b>	<b>26</b>
<b>3</b>	<b>DESENVOLVIMENTO</b>	<b>29</b>
3.0.1	Criação dos dados	29
3.0.2	Aquisição dos dados	31
3.0.3	Preparação dos dados	32
3.0.4	Uso dos Dados	33
<b>4</b>	<b>RESULTADOS</b>	<b>35</b>
4.1	conjunto de dados desenvolvido	35

<b>4.2</b>	<b>Análise do conjunto de dados</b> . . . . .	<b>37</b>
4.2.1	Condições dos veículos nos cenários . . . . .	37
4.2.2	Problemas de dados . . . . .	43
4.2.2.1	Dados incompletos . . . . .	43
4.2.2.2	Dados conflitantes . . . . .	43
4.2.2.3	Dados desiguais . . . . .	44
4.2.2.4	Dados incertos . . . . .	44
4.2.2.5	Ponto fora da curva . . . . .	45
<b>5</b>	<b>CONCLUSÃO</b> . . . . .	<b>47</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>48</b>
	<b>APÊNDICES</b>	<b>50</b>
	<b>APÊNDICE A –</b> . . . . .	<b>51</b>

# 1 INTRODUÇÃO

Ao longo do tempo, governos e indústrias automotivas têm impulsionado o que se chama de Sistemas de Transporte Inteligentes (do inglês *Intelligent Transportation Systems (ITS)*). Segundo (LORCH et al., 2006), a Arquitetura do Sistema de Transporte Inteligente japonês previa a comunicação entre os veículos e toda as infraestruturas inteligentes existentes nas estradas, como por exemplo sensores, semáforos, entre outros. O ITS constitui um conjunto de tecnologias que são capazes de trazer inovações para melhorar a administração de todo o ITS. Segundo (CUNHA et al., 2017), o ITS não é proposto somente para melhorar as condições do tráfego de automóveis, mas também tem a intenção de tornar o setor de transportes mais seguro, mais sustentável e eficiente, evitando os problemas causados pelos congestionamentos dos tráfegos urbanos e efeitos dos problemas climáticos sobre o tráfego. Assim, se torna possível que os utilizadores possuam mais informações a respeito dos veículos e trânsito, com um sistema de transporte mais seguro, eficiente e inteligente. É notável que tem sido uma tendência cada vez maior, a adição de capacidades de comunicação e sensoriamento de contexto em equipamentos diversos como, por exemplo relógios, sistemas elétricos, geladeiras, entre outros. Além desses objetos de menor porte, também equipamentos com um porte maior como, por exemplo os carros, estão ganhando tais capacidades, para que, por exemplo, seja possível obter dados sobre a situação contextual e status dos veículos, através de diagnósticos gerados por diversos sensores.

Segundo (RETTORE et al., 2019), dados relacionados a veículos são classificados entre intra ou extra veículos. Dados intra-veiculares são gerados por um conjunto de sensores que estão a bordo do veículo e que são capazes de captar o contexto e interações entre o veículo, motoristas/passageiros, e o ambiente ao redor. Já os dados extra-veiculares correspondem ao subconjunto de dados obtidos por meio de sensores (reais ou virtuais) que tentam descrever o comportamento dos motoristas e do ambiente ao seu redor através de diversas fontes individuais ou fundidas.

Os dados intra veiculares podem ser coletados através de fontes de dados intra veiculares como, por exemplo da interface de acesso a *Engine Control Unit (ECU)* dos veículos, que pode ser acessada por meio da porta OBD<sup>1</sup> existente em veículos. A ECU é a entidade responsável por centralizar as informações produzidas por diversos sensores distribuídos dentro de veículos modernos. Exemplos desses sensores são Rotação por Minuto (RPM), velocidade do veículo, posição dos pedais, emissões de CO<sub>2</sub> entre outras.

---

<sup>1</sup> OBD é um sistema de autodiagnóstico que fica disponível na maioria dos veículos atuais. Esta tecnologia será explicada na Seção 2.1.6.

Uma das formas de coletar as informações extra-veiculares é por meio de [API](#) que são disponibilizadas por provedores de dados como, por exemplo o *Google Directions* e *Here We Go*. Apesar dessa facilidade de coleta de dados e da existência de conjuntos de dados com informações veiculares, poucos são os conjuntos de dados que possuem o acesso público para à comunidade de pesquisa em [ITS](#) e associados. Isso se deve a diversos fatores como, por exemplo privacidade, segurança, diversidade e incentivo à coleta e divulgação desse tipo de informação. Um conjunto de dados (do inglês *dataset*) é caracterizado por ser um conjunto de dados tabulares, os quais podem ser entradas utilizadas nos procedimentos de análise de dados e aprendizado de máquina comumente utilizados em [ITS](#). Tipicamente conjunto de dados são planilhas retangulares e sem formatações, em que cada linha representa um registro, muitas vezes esses registros estão associados a rótulos que identificam, por exemplo, o contexto daqueles dados. Neste sentido, é apresentado a obtenção dos registros intra e extra veiculares via [OBD](#) e [APIs](#), tratamento e análise de dados veiculares, para a criação de um conjunto de dados de acesso público, com dados provindos de diagnósticos de sensores intra veiculares e extra-veiculares.

A principal motivação para sustentar o projeto da monografia, é devido a ausência de conjunto de dados veiculares que possuem o acesso público para os usuários. Grande parte dos conjuntos de dados veiculares existentes não ficam disponíveis devido alguma questão de segurança ou porque nos dias de hoje, dados podem ser comercializados. Com o conjunto de dados proposto, é possível preencher a lacuna da ausência de conjuntos de dados e é possível realizar diversos trabalhos como, por exemplo utilizações no campo da aprendizagem de máquina, análises dos dados veiculares, previsões de comportamentos veiculares, identificação do tipo de veículo que foi usado nas viagens, auxílio no desenvolvimento de ferramentas de manipulação de dados e mobilidade, entre diversos outros usos.

## 1.1 O problema de pesquisa

Nos Sistemas de Transporte Inteligentes tipicamente vale-se de dados oriundos ou não dos veículos para a tomada de decisões, melhoria da eficiência de trânsito, segurança, entre outros. Entretanto, dados de livre acesso à comunidade de pesquisa são escassos. Este trabalho visa mitigar a problemática ao produzir dados reais bem como caracterizá-los e rotulá-los.

## 1.2 Objetivo Geral

O presente trabalho possui o objetivo de realizar a criação de um conjunto de dados de acesso público, com dados obtidos através da interface [OBD](#) presente em veículos



modernos e com dados obtidos através da [API](#) do *Google Directions*, *Weather Channel* e *Here WeGo*.

### 1.3 Objetivo Específico

Este trabalho possui os seguintes objetivos específicos:

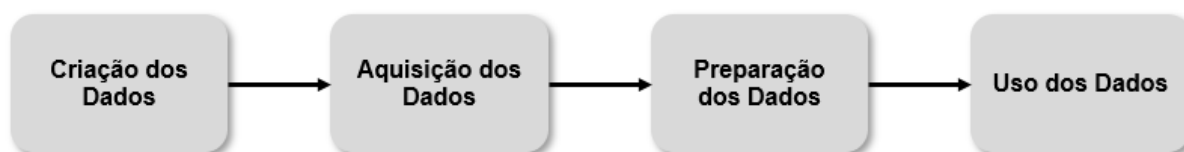
- Rotular os dados obtidos;
- Analisar os dados para diferentes condições de condução do veículo e para possíveis problemas existentes nas informações;

### 1.4 Metodologia

Nesta Seção, serão apresentadas e descritas as etapas para o desenvolvimento do trabalho e atingimento dos objetivos propostos.

Para a elaboração do referencial teórico desta pesquisa foi efetuada uma revisão bibliográfica em artigos científicos e pesquisas, que construíram ou utilizaram em algum momento do desenvolvimento do trabalho, algum conjunto de dados que possui dados veiculares ou sobre o trânsito, com o objetivo de analisar os conjuntos de dados existentes na literatura e verificar as necessidades e dificuldades para se obter um conjunto de dados veicular. Para a coleta de dados e criação do conjunto de dados proposto, foi definido 4 etapas, conforme a [Figura 1](#) que serão descritas nesta seção e detalhadas no desenvolvimento do trabalho.

Figura 1 – Metodologia



Fonte: Produzido pelo autor

A primeira etapa de criação dos dados tem como objetivo a geração de dados intra veiculares através do deslocamento de veículos em cenários que o automóvel percorre durante os dias e dados extra veiculares através da [API](#) do *Google Directions*, do *Here WeGo* e do *Weather Channel*. A segunda etapa, consiste na aquisição dos dados intra e extra veiculares que foram gerados na etapa anterior e que ficam disponíveis para a coleta. Esta coleta, foi realizada através do adaptador [OBD](#) presente em veículos e através de requisições nas [APIs](#) definidas. A etapa seguinte, consiste na preparação dos dados

que é onde ocorre a rotulação de todos os dados que foram coletados, de acordo com as características dos cenários que foram feitas as coletas, com o objetivo de facilitar a identificação das viagens, auxiliar as análises exploratórias e concluir a criação do conjunto de dados. A última etapa definida, consiste na utilização das informações existentes no conjunto de dados criado, gerando gráficos e tabelas para exibir a análise exploratória dos dados e possíveis problemas com as informações.

## 1.5 Organização do trabalho

O restante deste trabalho é organizado como se segue. O Capítulo 2 apresenta o arcabouço técnico introdutório e os trabalhos relacionados. No Capítulo 3 é apresentado as etapas que foram definidas para o desenvolvimento do trabalho. O Capítulo 4 apresenta os resultados obtidos e análises realizadas. No Capítulo 5 é apresentada a conclusão da monografia e os trabalhos futuros.

## 2 REVISÃO BIBLIOGRÁFICA

Neste Capítulo será apresentado inicialmente um arcabouço técnico introdutório que descreve os conceitos mais importantes empregados neste projeto. Estes são indispensáveis para a compreensão do trabalho. Na sequência, será apresentado os trabalhos da literatura que são correlatos a este.

### 2.1 ARCABOUÇO TÉCNICO INTRODUTÓRIO

#### 2.1.1 Sistemas de Transporte Inteligentes

Segundo (CUNHA et al., 2017), o crescimento desordenado das cidades tem causado graves problemas socioeconômicos e estruturais para a população, que auxiliam para o aumento das desigualdades sociais. Assim, serviços e recursos devem ser providos de maneira a lidar e diminuir esses problemas. Como exemplo desses problemas, temos a má ocupação do espaço urbano que auxilia para gerar diversos problemas de mobilidade. Dessa forma, com o passar do tempo, as adversidades referentes ao trânsito também têm aumentado graças ao maior número de automóveis em locomoção durante os dias e ao maior acúmulo de pessoas em cada região. Segundo o (IBPT, 2020), o número de automóveis em circulação tem aumentado a cada ano.

Com o aumento do número de veículos, os fabricantes e governos buscam tecnologias para aprimorar a competência, mobilidade e segurança no trânsito. Esse aprimoramento objetiva o aperfeiçoamento para tomada de decisão, melhoria da segurança e eficiência no transporte de pessoas e produtos. Nesse contexto, surge o ITS que é um sistema integrado de pessoas, estradas e veículos, desenvolvido para auxiliar e melhorar a segurança rodoviária, eficiência, conforto e a conservação ambiental por meio da realização de um tráfego mais suave, reduzindo o congestionamento do tráfego (ANDERSEN; SUTCLIFFE, 2000). Segundo (CUNHA et al., 2017), o ITS tem como objetivo aperfeiçoar a segurança e mobilidade dos transportes, aprimorar a produtividade das pessoas e reduzir os efeitos nocivos gerados no trânsito. Esse aperfeiçoamento é atingido por meio da inclusão de tecnologias de comunicação nos automóveis e na infraestrutura das cidades.

Para ocorrer a melhoria do transporte é necessário mais do que apenas a construção e desenvolvimento das infraestruturas físicas das cidades. Com a percepção de cidade inteligente convertendo as cidades em sociedades tecnológicas, trazendo maior praticidade e facilidade para a população, o ITS torna-se necessário entre todos. Em todas as cidades a mobilidade é algo fundamental: seja para frequentar à escola, mercado e trabalho, ou qualquer seja o destino, a população usufrui do sistema de transporte para se locomover

na cidade.

O ITS fornece serviços e soluções de transporte utilizando tecnologias de informação e telecomunicações de última geração (ANDERSEN; SUTCLIFFE, 2000). São diversas as tecnologias que fazem parte do ITS, como por exemplo o sistema de navegação dos veículos, sistemas de gerenciamento dos semáforos, radares de velocidade, sistemas de prevenção de colisões, informações do tráfego, informações meteorológicas, entre outros. Os serviços e tecnologias que o ITS disponibiliza, utilizam diversas quantias de dados e tecnologias de comunicação e possuem traços e particularidades próprias, que diferenciam dos serviços habituais (CUNHA et al., 2017).

### 2.1.2 Espaço de dados veicular

A quantidade de informações produzida nos ITSs são vastas, tendo diversos dados e fontes de coleta de dados. Dessa forma, essa monografia dará enfoque ao espaço de dados veicular que são as informações que ficam à disposição nos automóveis, abrangendo as informações do contexto que são coletadas pelo próprio automóvel, as informações referentes as ações do automóvel e usuários, e o espaço à volta do veículo.

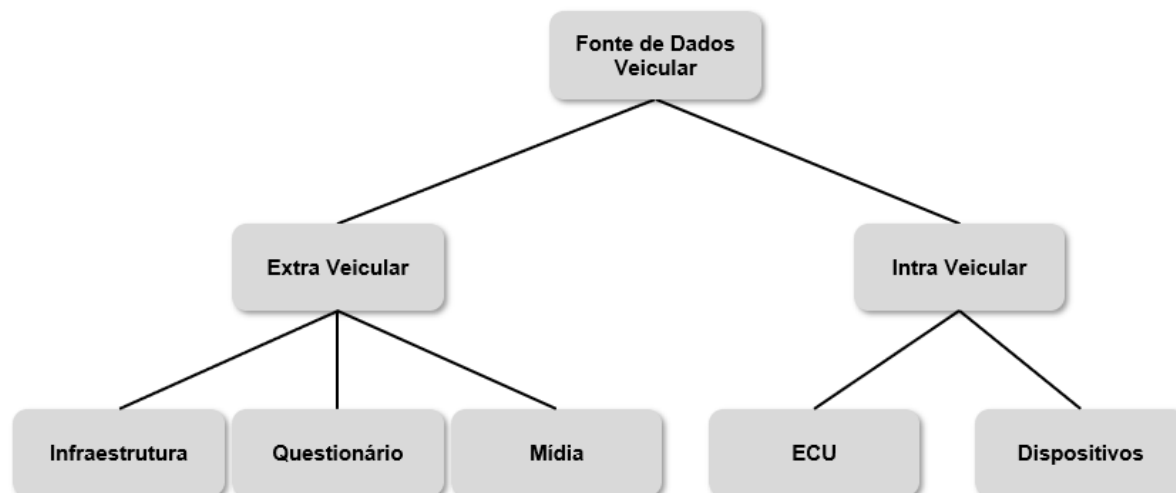
Segundo (RETTORE et al., 2019), o espaço de dados veicular pode ser separado em duas partes: a fonte de dados no veículo e os dados veiculares. A fonte de dados no veículo é responsável pela criação dos dados no automóvel, que consiste a etapa de identificação dos elementos do ambiente através de diversos sensores que o automóvel possui. Já a parte dos dados veiculares, inclui a obtenção de dados que é responsável por tornar disponível os dados através de algum meio, a preparação dos dados que consiste nos tratamentos que são feitos nos dados, o processamento de dados que é realizado por meio de alguma metodologia, e por último a utilização dos dados que pode suprir algum aplicativo ou ciclo.

### 2.1.3 Fontes de dados veiculares

No interior do universo do automóvel encontram-se diversas variedades de sensores, contendo as suas utilidades, e a capacidade de identificar determinados sinais e converte-los em informações. Os sensores veiculares são responsáveis por efetuar a leitura de alguma ocorrência, para que seja processada pela ECU presente no veículo.

Na Figura 2 são apresentados exemplos de fontes de dados que estão disponíveis para se obter informações sobre veículos. Primeiramente, para sugerir soluções para o ITS, se deve compreender os dados e as fontes de coleta existentes, como os elementos que são encarregados de obter e, em determinadas situações, disponibilizar o acesso aos dados a qualquer usuário. Nas próximas seções será feito o detalhamento das fontes de dados veiculares presentes na Figura 2.

Figura 2 – Taxonomia dos dados veiculares



Fonte: Produzido pelo autor

### 2.1.3.1 Fontes intra-veiculares

Os sensores intra-veiculares são aqueles que fazem parte do conjunto de sensores que detalham a interatividade que ocorre entre o automóvel, o motorista e passageiro. Qualquer sensor que esteja incluso em um automóvel para medir as condições do veículo, ou condutas dos usuários que estão a bordo, são considerados sensores intra-veiculares.

A [ECU](#) é um exemplo de fonte de dados intra-veicular que está presente nos veículos. Por meio dela é possível adquirir diferentes tipos de dados dos sensores presentes no automóvel, através de diagnósticos que são gerados, podendo trazer por exemplo as informações sobre a rotação do motor, consumo de CO<sub>2</sub>, combustível utilizado, velocidade do veículo, quilômetros por litro, coordenadas, entre outros dados. Outros exemplos de fontes de dados intra-veicular, são os dispositivos que ficam a bordo dos automóveis, por exemplo, dispositivos GPS, smartphones, dentre outros, os quais podem trazer dados sobre as coordenadas do veículo, a velocidade, entre outras informações. Os veículos-sonda também podem ser considerados uma fonte de dados intra-veicular, pois possui sensores responsáveis por realizar medições das situações de estradas, clima, entre outros dados.

### 2.1.3.2 Fontes extra-veiculares

Os sensores extra-veiculares são aqueles que fazem parte do conjunto de sensores que descrevem o espaço à volta do veículo e as atitudes dos motoristas. Um exemplo de fonte de dados extra-veiculares são questionários de estilo de condução, que podem fornecer informações sobre comportamentos, atitudes, distrações, acidentes, estado de estradas, dentre outros. Um segundo exemplo de fonte de dados extra-veiculares são os

sistemas de controle do tráfego, fornecendo informações sobre o fluxo de veículos, estradas, semáforos, entre outros.

É um desafio ter acesso a grande parte dos dados extra-veiculares como, por exemplo, dados de *loops* indutivos, câmeras de trânsito, rastros de usuários nas vias de transporte, dentre outros. Esses dados geralmente são controlados por entidades privados ou governamentais, portanto são inacessíveis em parte ou em sua totalidade, por questões de privacidade ou de mercado (SANTOS et al., 2018). Desse modo, um exemplo de fonte de dados extra-veicular para aquisição de informações são as mídias sociais, podendo a maioria vir a ser acessada por meio de API. Dessa forma, informações sobre o estado do trânsito, acidentes, situação das estradas, condições climáticas, dentre diversas outras, podem ser adquiridas através dessa fonte.

#### 2.1.4 Unidade de Controle Eletrônico

A ECU são a parte fundamental para todo o sistema embarcado automotivo, pois faz parte da eletrônica dos automóveis e é o cérebro onde todas as decisões são tomadas, baseadas na leitura das informações recebidas dos sensores utilizados em cada sistema. Além de gerenciar os sistemas mecânicos do veículo, a ECU também é responsável pelo gerenciamento dos sistemas que não são mecânicos, pelos sistemas críticos e pelos utilitários do automóvel (OSS, 2018).

Devido a impossibilidade dos sensores para a aquisição de informações mais corretas e para o funcionamento dos automóveis modernos, cada vez mais o número de sensores automotivos tem aumentado. Todas as informações que são coletadas pelos sensores automotivos são levadas para a ECU por meio da *Controller Area Network* (CAN), que pode ser acessada por meio da porta OBD presente em automóveis. As informações que a ECU recebe, são transformadas em digitais, para que algoritmos realizem os processamentos e tomadas de decisão.

Como citado em (LANDHÄUSSER et al., 2014), após receber os sinais elétricos transmitidos pelos sensores, a ECU processa esses dados para gerar sinais de controle para os atuadores.

#### 2.1.5 Rede de Área Controlada

Uma CAN é uma rede que integra um sistema responsável pela transmissão em tempo real de informações. Segundo (BORTH, 2016), CAN é um protocolo de comunicação serial criado para aplicações na indústria automotiva. Quando foi criado, tinha como objetivo substituir e simplificar arranjos elétricos que usavam diversos fios e cabos por apenas um barramento de comunicação. Segundo (SOUZA; CAMPOS, 2017), a CAN garante comunicação confiável, segura e eficiente entre sensores, atuadores, controladores

e outros dispositivos de controle em tempo real. Como características, a rede CAN possui alta confiabilidade, segurança e velocidade para a troca de informações das unidades de comando dos automóveis.

### 2.1.6 *On-boarding Diagnostic*

O OBD é um sistema de diagnóstico incluído no sistema embarcado dos veículos. Sua função é controlar de forma constante os componentes relacionados as emissões de gases nocivos (BLASCO, 2014). Esse sistema surgiu devido a necessidade de um regulamento para que fosse feito o controle de emissão nos veículos, evitando um excesso de poluição que estava acontecendo nos automóveis na década de 80. Assim surgiu o sistema de diagnóstico OBD, em que o próprio veículo examina se os sensores responsáveis por monitorar as emissões de gases do automóvel estão dentro das medidas corretas. Se houver alguma anormalidade o sistema emite um relatório com o status dos subsistemas do veículo mostrando ao motorista que será necessário algum reparo. (RETTORE et al., 2016) diz que uma analogia útil é supor que o OBD é a linguagem que usamos para falar sobre o estado de um veículo, por meio de um dispositivo de comunicação (CAN).

Outro benefício do OBD foi a padronização dos sistemas de diagnósticos dos automóveis, pois devido a existência de diversas montadoras com sistemas de diagnósticos diferentes, o valor das manutenções e ferramentas de diagnósticos eram bem maiores (OSS, 2018). Nas primeiras versões do OBD, apenas era apresentado uma luz que indicava algum mau funcionamento caso estivesse ocorrendo algum problema, não era indicado a origem do erro que estava ocorrendo, ficando com o mecânico a responsabilidade de localizar a falha através do seu conhecimento e sensores do veículo, podendo a manutenção ter um custo maior nesses casos.

Antes da padronização do OBD, a *General Motors* desenvolveu o *Assembly Line Diagnostic Link* (ALDL) que foi um sistema de diagnóstico de bordo. O ALDL utilizava pelo menos quatro conectores de ferramentas de diagnóstico distintos. Seus *layouts* de conectores mudavam a cada geração de ECU, então não houve nenhum padrão definido para o protocolo (OSS, 2018).

Ocorreram diversas evoluções na OBD até que se chegasse na versão usada atualmente. Hoje em dia é utilizado para diversas aplicações, disponibilizando uma maior quantidade de dados, um padrão de porta de comunicação digital que pode disponibilizar informações em tempo real, a possibilidade de realizar correções de falhas de dentro do veículo e o padrão de códigos dos diagnósticos de problemas.

### 2.1.6.1 OBD-I

Com a regulamentação do OBD-I foi esperado que os fabricantes de veículos desenvolvessem sistemas de controle de emissão de gases que fossem efetivos, porém não atendeu as expectativas. Segundo (OSS, 2018), o OBD-I foi uma experiência decepcionante de padronizar as emissões dos automóveis, pois ocorreu muita fragmentação da estruturação deste protocolo e muitos desencontros entre os fabricantes, não sendo possível executar com sucesso o objetivo de verificar os dados de emissão registrados na ECU dos veículos, reprovando veículos que não estivessem dentro das normas. Houve muita dificuldade para obter os dados das emissões de forma padronizada.

### 2.1.6.2 OBD-II

O OBD-II foi desenvolvido com o objetivo de solucionar os problemas e dificuldades que ocorreram com o OBD-I. Foi feita a evolução para o OBD-II visando obter uma padronização, um sistema mais fácil de usar e que possuísse maior potência em relação ao seu predecessor. A Tabela 1 abaixo apresenta características e diferenças entre o OBD-I e a OBD-II.

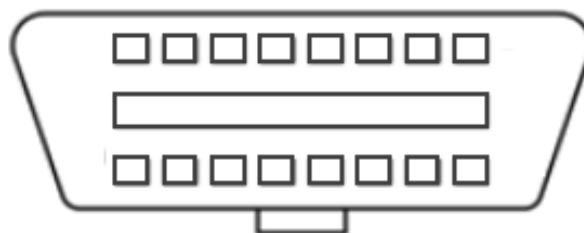
Tabela 1 – Evolução OBD

Parâmetros	OBD-I	OBD-II
Data	Antes de Janeiro de 1996	A partir de Janeiro de 1996
Função	Recolher informações dos diagnósticos e notificar possíveis defeitos no motor	Recolher informações dos diagnósticos e notificar possíveis defeitos no motor
Interface	Própria do fabricante	Interface global
Aplicação	Não teve muita popularidade	Maior popularidade e foi implantado desde a sua criação
Benefícios	Combustível e energia consumidos pelo automóvel	Realiza e avalia cálculos e algoritmos para solucionar os problemas que ocorrerem

No ano de 1996 foi definido como obrigatório nos Estados Unidos e na Europa. Essa definição veio com o objetivo de diminuir os custos dos reparos em automóveis nas oficinas, o controle de emissão de gases, possibilidade de o próprio motorista consiga adquirir um equipamento para obter o diagnóstico do veículo e também aumentar a fama do reparo eletrônico de automóveis. No Brasil foi definido como obrigatório o OBD-II a partir do ano de 2010, anteriormente os reparos em veículos eram mais caros e cada oficina acessava os dados e informações dos automóveis de maneira diferente.



Figura 3 – Conector OBD



Fonte: Produzido pelo autor

Com a evolução para **OBD-II**, foi definido um tipo universal de conector de diagnóstico e os seus pinos, além da organização das mensagens. Devido essa padronização que ocorreu, um único aparelho pode consultar o computador de bordo dos veículos e adquirir diagnósticos veiculares. A Figura 3 representa o tipo universal do conector **OBD** e seus pinos.

#### 2.1.6.2.1 Parâmetro ID

O **OBD-II** conta com Parâmetro ID (**PID**), os quais são códigos utilizados na requisição de informações de um automóvel, sendo utilizado como instrumento de diagnósticos. Os diagnósticos gerados pela **OBD-II** são ordenados através de procedimentos de operação e códigos de **PID**. Segundo (**OSS, 2018**), os **PIDs** são regulamentados pela **OBD-II** e ficam salvos na memória como um vetor, sendo cada posição do vetor um código hexadecimal que condiz com a sua representação e significado da informação que deseja ser passada. A **OBD-II** regulamenta diversos **PIDs**, mas os fabricantes não são obrigados a implementar todos existentes. Devido a isso, não são todos os automóveis que possuem suporte aos **PIDs** existentes, podendo existir alguns **PIDs** personalizados que não existem no padrão **OBD-II**.

#### 2.1.7 Scanner

O *scanner* é um aparelho responsável por realizar a leitura dos **PIDs** regulamentados pela **OBD-II**. Com o começo da utilização da **OBD-II** nos veículos, ocorreu a chegada de diversos instrumentos de diagnósticos para se acessar as informações veiculares pagando um preço bem mais baixo. Esses instrumentos possuem a competência de criar interfaces e realizar a comunicação com o protocolo **OBD** para verificar as informações do veículo. Um *scanner* pode se conectar com alguma porta **OBD** e fazer a comunicação entre o **OBD** e algum computador ou smartphone, através de conexão via cabo, Wi-fi ou bluetooth. Existem diversos tipos de *scanners* automotivos como, por exemplo o ELM 327, CARRORAMA Multilaser, Autel Maxiscan Ms309, entre outros. Neste trabalho será dado o enfoque ao ELM 327.

### 2.1.7.1 ELM 327

O ELM 327 é um microcontrolador responsável por realizar a tradução da interface OBD dos automóveis. Ao realizar a conexão do ELM 327 a interface OBD do veículo, deve ser feita o pareamento entre o *scanner* e o bluetooth de algum smartphone. Para realizar a leitura das informações e códigos de falhas do veículo, se deve acessar o *scanner* por algum aplicativo no smartphone. A Figura 4 apresenta um *scanner* ELM 327.

Figura 4 – Scanner ELM 327



Fonte: Magazine Luiza (2022)

### 2.1.8 Torque PRO

Para que seja possível realizar a conexão com o *scanner* ELM 327 conectado na porta OBD, é necessário a utilização de um aplicativo de diagnóstico. Durante o desenvolvimento do trabalho, foi utilizado o aplicativo Torque PRO para realizar a conexão entre a OBD e o smartphone.

O Torque PRO é um aplicativo de diagnóstico que quando conectado a um *scanner* OBD, permite a identificação de erros e informações na ECU do veículo. O aplicativo pode se comunicar com diversos sensores do sistema de gerenciamento do motor e de componentes eletrônicos do automóvel. Ao realizar a conexão entre o aplicativo e o *scanner*, deve ser inserido informações a respeito do veículo, para seguir com a geração de diagnósticos. Para a seleção do aplicativo, foi comparado características e pontos negativos de vários aplicativos de diagnóstico veicular, como demonstrado na Tabela 2 abaixo.

Tabela 2 – Aplicativos de diagnóstico veicular

Aplicativo	Pontos Positivos	Pontos Negativos
Torque PRO	<ul style="list-style-type: none"> <li>- Atividade em tempo real;</li> <li>- Histórico de falhas;</li> <li>- Facilidade para gerar relatórios;</li> <li>- Diagnósticos veiculares.</li> </ul>	<ul style="list-style-type: none"> <li>- Alto consumo de energia do celular;</li> <li>- Versão PRO deve ser comprada.</li> </ul>
OBD Car Doctor PRO	<ul style="list-style-type: none"> <li>- Desenho gráfico com os dados obtidos;</li> <li>- Atividade em tempo real;</li> <li>- Diagnósticos veiculares.</li> </ul>	<ul style="list-style-type: none"> <li>- Conexão bluetooth instável;</li> <li>- Dificuldade de geração de relatórios;</li> <li>- Versão PRO deve ser comprada.</li> </ul>
Carista OBD-II	<ul style="list-style-type: none"> <li>- Ajuste nos faróis;</li> <li>- Ajuste no som dos sensores traseiros;</li> <li>- Diagnósticos veiculares.</li> </ul>	<ul style="list-style-type: none"> <li>- Conexão bluetooth instável;</li> <li>- Erros de conexão ao alterar configurações.</li> </ul>

### 2.1.9 Application Programming Interface

Uma [API](#) é uma conexão entre computadores ou *softwares* para que seja possível utilizar serviços através de outros *softwares* que não desejam se relacionar com as particularidades internas de como um software funciona, mas somente com as suas funcionalidades. Diferentemente de uma interface de usuário que realiza a conexão entre *softwares* e usuários, uma [API](#) realiza a conexão entre *softwares*.

Segundo ([SCHMITT, 2013](#)), uma [API](#) é uma interface que pode estar conectada a diferentes sistemas e aplicativos, no entanto, para o usuário isto é imperceptível pelo fato de estar rodando por trás de tudo, enquanto o usuário usufrui de um aplicativo ou site, a sua [API](#) pode estar conectada a diversos outros sistemas e aplicativos, sem que o usuário perceba. Hoje é possível ter acesso a [APIs](#) que fornecem informações a respeito de veículos e do ambiente em que eles estão percorrendo. Para o desenvolvimento do trabalho, também foi realizada a coleta de informações através das seguintes [APIs](#):

- *HERE Maps (HERE Routing API e HERE Traffic API)*: [APIs](#) do *HERE We Go* que fornece informações de roteamento realizando o cálculo entre dois ou mais locais para diversas regiões do mundo, além de disponibilizar informações a respeito de como o tráfego está se movendo ao longo de estradas específicas.
- *Weather Channel*: Fornece informações meteorológicas a respeito das condições climáticas do local em que o usuário definir. A [API](#) fornece informações sobre temperatura, precipitação, vento, pressão barométrica, umidade, visibilidade, radiação ultravioleta e outros elementos de observações meteorológicas relacionados.

- *Google Directions (API Distance Matrix)*: Fornece dados referentes a distância e tempo de viagem estimado entre dois pontos, com base na rota recomendada pelo *Google Directions*.

### 2.1.10 Ferramentas e serviços utilizados

- *PowerBI*: é um programa desenvolvido pela *Microsoft* e que tem como função fornecer painéis com visualizações interativas, para que os usuários realizem análises de dados.
- *Scikit-Mobility*: é uma biblioteca desenvolvida para ser utilizada com a linguagem de programação *Python* realizando análises de mobilidade humana. Segundo ([PAP-PALARDO et al., 2019](#)), o *Scikit-Mobility* é eficiente e fácil de usar, pois estende o *pandas*, uma biblioteca *Python* popular para análise de dados. Além disso, o *Scikit-Mobility* fornece ao usuário muitas funcionalidades, desde a visualização de trajetórias até a geração de dados sintéticos, desde a análise de padrões estatísticos até a avaliação do risco de privacidade relacionado à análise de conjuntos de dados de mobilidade.
- *MobVis*: é uma biblioteca desenvolvida para ser utilizada com a linguagem de programação *Python*, realizando análises utilizando informações de mobilidade e extraíndo métricas.
- *Power Query*: é uma ferramenta de preparação e transformação de dados que possibilita o tratamento dos dados antes que se utilize eles.
- *Visual Basic Scripting (VBS)*: é uma linguagem de programação utilizada para criar *scripts* e tarefas que podem ser executadas no sistema operacional.
- *Kaggle*: é uma comunidade online onde se encontra diversos conjuntos de dados para a utilização em análises e aprendizado de máquina. Ela que autoriza os usuários a publicarem e buscarem conjuntos de dados para o uso.

### 2.1.11 conjunto de dados

Um conjunto de dados corresponde a um grupo de informações ordenadas, formando tabelas em que cada linha corresponde a um registro do conjunto. Esse grupo de dados pode ser adquirido através observações, análises e medições.

As informações de um conjunto de dados são habitualmente utilizadas para iniciativas de ciência de dados, treinamentos de algoritmo de aprendizado de máquina que vai antecipar alguma informação e para visualizar relatórios e gráficos que auxiliam em tomadas de decisões.

## 2.2 TRABALHOS RELACIONADOS

Os trabalhos relacionados utilizaram algum conjunto de dados relacionados ao espaço de dados veicular para a evolução da pesquisa e dos objetivos que haviam proposto, portanto o foco é a comparação entre os *datasets* veiculares usados na literatura e o conjunto de dados produzido neste trabalho. A Tabela 3 coloca em perspectiva este trabalho em relação aos trabalhos relacionados. Nesta comparação, os critérios de acesso público, dados intra-veiculares, dados extra-veiculares, cenários fixos e cenários aleatórios foram utilizados.

Tabela 3 – Trabalhos relacionados

Trabalho	Acesso		Fonte de dados		Cenários	
	Público	Restrito	Intra	Extra	Fixos	Aleatórios
(BERNARDI et al., 2018)		✓	✓		✓	
(AZADANI; BOUKERCHE, 2020)	✓		✓		✓	
(SANTOS et al., 2018)		✓		✓		✓
(RETTORE et al., 2018)		✓	✓		✓	✓
(UPPOOR et al., 2013)	✓		✓			✓
(KONG et al., 2018)		✓	✓			✓
(RETTORE et al., 2020)		✓		✓		✓
(SU et al., 2022)	✓			✓	✓	
(TEMPELMEIER et al., 2021)		✓	✓	✓		✓
(DAVIS; RAINA; JAGANNATHAN, 2020)	✓		✓	✓		✓
<i>Dataset</i> proposto	✓		✓	✓	✓	✓

Para construção do conjunto de dados utilizado em (BERNARDI et al., 2018), foi reunido os dados do barramento CAN de um conjunto de veículos, conectando um *scanner* a porta OBD dos automóveis e acessando através do aplicativo Torque Pro, os autodiagnósticos gerados pelo sistema de monitoramento de bordo. No total foram gerados três conjuntos de dados diferentes, cada um representando um tipo de cenário diferente. No trabalho apresentado, foi divulgado que o acesso ao conjunto de dados criado seria público, porém as informações não estão mais disponíveis.

Em (AZADANI; BOUKERCHE, 2020), foi utilizado dados providos da rede de área do controlador do veículo que são coletados através da porta OBD e fornecem as informações quase em tempo real sobre o automóvel para que fosse possível realizar as classificações desejadas. Foram avaliadas as informações providas do conjunto de dados de acesso público e de comportamento naturalístico de direção. Os dados foram coletados em um único trajeto fixo que foi definido e no mesmo horário em cada dia.

Em (SANTOS et al., 2018), para que fosse possível gerar um conjunto de dados e utilizar no desenvolvimento do trabalho, foi selecionado algumas contas de usuários do Twitter que abordam assuntos referentes ao tráfego e através de APIs da própria rede social, foi realizada a coleta dos dados extra veiculares. Devido a dificuldade para se encontrar um conjunto de dados relacionados ao trânsito, o conjunto de dados desenvolvido

não possui o acesso livre.

Para a construção do conjunto de dados utilizado em (RETTORE et al., 2018), foi realizada a coleta de dados intra-veiculares pela porta OBD de veículos em locomoção por diferentes rotas pré-definidas (cenários fixos) e também em locomoção para diversos fins nas rotinas diárias (cenários aleatórios). Os dados foram utilizados apenas no trabalho e não há acesso ao conjunto de dados.

O conjunto de dados apresentados em (UPPOOR et al., 2013) é reunido informações de diferentes fontes de dados e ferramentas de simulação, de modo a abranger os aspectos específicos necessários para uma caracterização adequada trânsito. O conjunto de dados é obtido considerando uma topologia de estrada realista, comportamentos microscópicos de motoristas e fluxos macroscópicos de tráfego, em cenários aleatórios. Todo o conjunto de dados utilizado no trabalho está disponível gratuitamente.

Para a construção do conjunto de dados utilizado em (KONG et al., 2018), foi realizado a coleta de informações intra-veiculares de diversos veículos. Foi gerado várias trajetórias de viagem dos automóveis em vários cenários aleatórios. No momento, o conjunto de dados que foi gerado, não possui acesso público para usuários.

Em (RETTORE et al., 2020), na geração do conjunto de dados utilizado no trabalho foi realizado a coleta de informações extra-veiculares através de *tweets* de usuários selecionados, extraindo um conjunto de palavras relacionadas a eventos incidentes, como congestionamento, acidente, construção, evento planejado, perigo na estrada, veículo desativado, trânsito, congestionamento, carro e clima. Para que fosse coletado o máximo possível de informações, foi feita a coleta de dados do *HERE WeGo*, *Tripadvisor* e *Bing Maps*.

Em (SU et al., 2022), foi realizado a coleta de dados extra-veiculares em cenários que foram pré-definidos pelos autores para a construção de um banco de dados com informações de fluxo de tráfego. Os dados apresentados no estudo estão com o acesso disponível mediante solicitação razoável do autor correspondente.

Os experimentos realizados no artigo (TEMPELMEIER et al., 2021) utilizaram um conjunto de dados de tráfego que possui dados intra e extra veiculares, que foram coletados de diversas fontes em cenários distintos. Os autores informam que os dados utilizados neste artigo não podem ser compartilhados devido a restrições legais.

Para a construção do trabalho proposto em (DAVIS; RAINA; JAGANNATHAN, 2020), os autores utilizaram conjuntos de dados de tráfego temporais que foram coletados por meio de fonte de dados extra e intra veiculares. Para a coleta dos dados, não foi definida nenhuma rota fixa, foi feita a coleta apenas em rotas aleatórias. Os dados coletados estão publicamente disponíveis para serem utilizados.

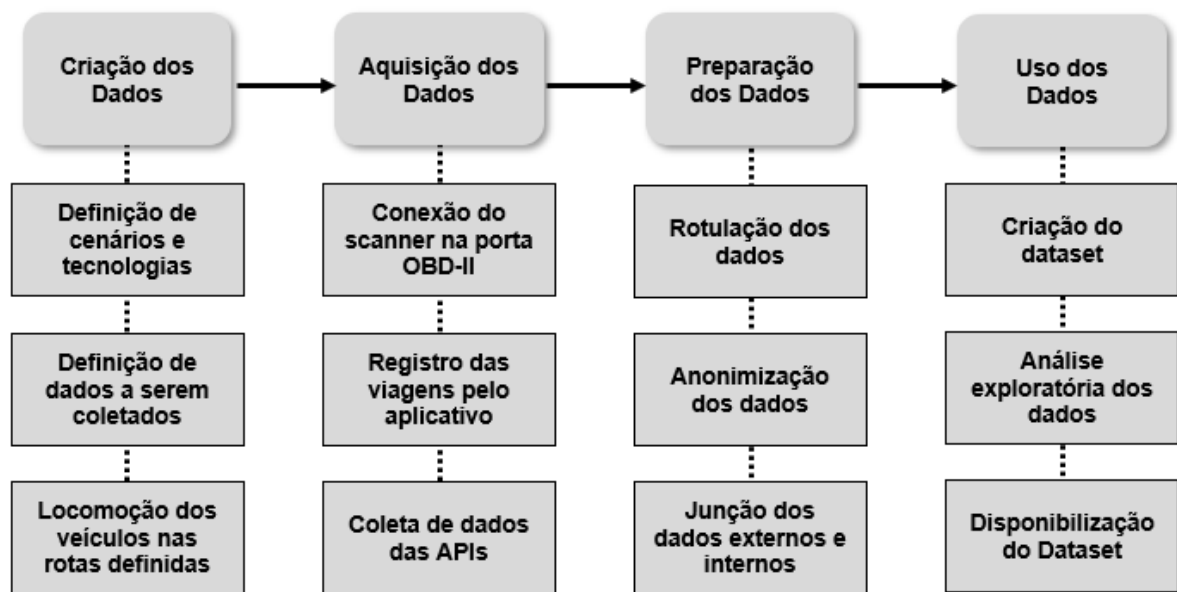
O conjunto de dados proposto nesse trabalho será desenvolvido com dados intra e

extra-veiculares coletados em diversos carros em locomoção por cenários pré-definidos e em cenários aleatórios. O conjunto de dados será publicado no *Kaggle* e no *Google Drive* para que tenha o acesso público por tempo ilimitado, para que qualquer usuário possa utilizar gratuitamente.

## 3 DESENVOLVIMENTO

Este Capítulo descreve os métodos utilizados para o atingimento dos objetivos definidos na Seção 1.2, os quais são resumidos na Figura 5. Na sequência será feito o detalhamento de todas as etapas que foram determinadas para que fosse possível a coleta de dados e criação do conjunto de dados proposto no trabalho.

Figura 5 – Desenvolvimento do conjunto de dados



Fonte: Produzido pelo autor

### 3.0.1 Criação dos dados

A criação dos dados é uma etapa essencial para o início do desenvolvimento de um conjunto de dados, pois ela é a partir dessa etapa que as próximas poderão ser realizadas. Para iniciar a criação dos dados é realizada a definição das fontes dos dados e os cenários para a realização das coletas de informações, além de definir todas as tecnologias e ferramentas que serão usadas para que se tenha dados disponíveis para a aquisição. Nessa etapa, as fontes de dados e sensores, estarão operando para que haja informações para coletar. Abaixo será feita a descrição do passo a passo que foi seguido para a criação dos dados e início do desenvolvimento do conjunto de dados.

- Definição de cenários: foi realizada a definição de cenários para se realizar a criação e coleta de dados intra-veiculares através da **OBD-II**, através da locomoção dos automóveis. Com o objetivo de se ter o máximo de cenários diários de locomoção de



veículos, foi definido oito cenários levando em consideração a tendência do fluxo de carros no período do dia que seriam feitas as coletas. A tendência de fluxo de veículos foi categorizada como alto no momento do dia que é esperado que tenha uma maior quantidade de veículos nas vias. Foi categorizado como tendência de fluxo médio, o momento que é esperado que tenha uma quantidade média de veículos nas vias e sem chances de ocorrer congestionamentos. Também foi definido como tendência de fluxo baixo, os momentos que são esperados que tenham poucos veículos percorrendo as vias. Para a realização da coleta de dados intra-veiculares, foi definido o número de carros que seriam utilizados na criação dos dados, sendo utilizado um Fox 1.6 fabricado no ano de 2012, um Fiat Uno 1.0 fabricado no ano de 2010 e um Ford KA 1.5 fabricado no ano de 2015. Também foi definido o número de motoristas que estariam dirigindo alguns dos veículos, sendo ambos do sexo masculino e com idade entre 20 e 30 anos. A Tabela 4 abaixo apresenta detalhadamente as características de cada cenário de locomoção dos automóveis para a coleta de informações, número de motoristas e número de veículos que foram utilizados no trabalho.

- Definição de meios de criação e coleta dos dados: Nesta etapa, foi feita a definição das fontes de dados que seriam utilizadas para a geração de informações e coleta dos dados. Foi determinado que os dados intraveiculares seriam coletados por meio de um *scanner* ELM 327. Para o registro das viagens no veículo, foi definido que seria utilizado o aplicativo Torque PRO em um smartphone. Em seguida, foi determinado que os dados extra-veiculares seriam coletados por meio da [API do Google Directions](#)<sup>1</sup>, [HERE WeGo](#)<sup>2</sup> e [Weather Channel](#)<sup>3</sup>.

Tabela 4 – Aplicativos de diagnóstico veicular

Cenário	Descrição	Horário (Intervalo)	Fluxo	Carros	Motoristas
1	Principal avenida em dia de semana	8:00 até 9:00 e 16:00 até 19:00	Alto	2	1
2	Principal avenida ao sábado	9:00 até 11:00	Alto	2	1
3	Principal avenida ao sábado	19:30 até 21:00	Baixo	2	1
4	Principal avenida ao domingo	9:00 até 21:00	Baixo	2	1
5	Ruas aleatórias em dia de semana	indeterminado	-	3	2
6	Ruas aleatórias no final de semana	indeterminado	-	3	2
7	Estradas / Rodovias	12:00 até 17:00	Baixo	2	1
8	Principal avenida em dia de semana	10:00 até 11:30 e 14:00 até 16:00	Médio	2	1

- Definição dos dados a serem coletados: Após a definição dos meios de coletas, foi realizada a verificação dos dados que estão disponíveis em cada fonte escolhida e buscado o máximo de informações possíveis em cada fonte que foi definida, para que haja uma maior quantidade de dados disponíveis para usuários do conjunto de dados

<sup>1</sup> Disponível em: <<https://developers.google.com/Directions/documentation/distance-matrix?hl=pt-br>>

<sup>2</sup> Disponível em: <[https://developer.here.com/documentation/traffic-api/dev\\_guide/index.html](https://developer.here.com/documentation/traffic-api/dev_guide/index.html)>

<sup>3</sup> Disponível em: <<https://weather.com/swagger-docs/call-for-code>>

e uma possibilidade maior de utilizações para se fazer com cada dado. A Tabela 5 apresenta uma lista não exaustiva dos dados que estão disponíveis para a coleta em cada fonte intra e extra-veicular, além da unidade que o dado é representado. A lista completa de dados está disponível no Apêndice A.

Tabela 5 – Dados disponíveis para a coleta

Dado	Descrição	Unidade/Saída	Fonte
Engine RPM	Rotação do motor	Rotação por minuto	OBD
Speed(OBD)	Velocidade do veículo registrada na OBD	Quilômetros por hora	OBD
Fuel used	Gasolina utilizada na viagem	Litros	OBD
$CO_2$	$CO_2$ produzida pelo veículo	Gramas por quilômetro	OBD
Latitude	Posição atual do veículo	Graus	OBD
Longitude	Posição atual do veículo	Graus	OBD
DurationHereWeGo	Duração da viagem segundo o Here WeGo	Minutos	API Here WeGo
DurationGoogleMaps	Duração da viagem segundo o Google Maps	Minutos	API Google Maps
WeatherDescriptions	Descrição do tempo no momento	Texto	API Weather Channel

- Locomoção dos veículos nas rotas definidas: Nesta etapa foi realizada a locomoção dos veículos em cada um dos cenários que foram definidos, no máximo de vezes possíveis. Com essa locomoção, os sensores veiculares realizam as coletas das informações disponíveis no momento, a cada 1 segundo. Após a coleta de informações, os dados ficam disponíveis para a aquisição através da porta **OBD** dos automóveis.

### 3.0.2 Aquisição dos dados

A etapa de aquisição dos dados é a fase em que se obtém as informações que ficam disponíveis após a coleta pelos sensores. Os dados internos do veículo ficam disponíveis para aquisição por meio de **PIDs** na porta **OBD**. Para isso se deve conectar o microcontrolador ELM 327 na porta **OBD**, registrar as viagens realizadas e assim adquirir os dados que lá estiverem. Para os dados externos dos veículos, deve ser realizado as requisições através das **APIs** definidas, para ter como retorno os dados disponibilizados. A seguir, será feito a descrição do passo a passo que foi realizado para a aquisição dos dados.

- Conexão do *scanner* ELM 327: Para que fosse possível realizar o registro das informações do veículo durante as viagens nos cenários definidos, foi feita a conexão do *scanner* ELM 327 na porta **OBD** existente abaixo do volante dos automóveis.

Após a conexão do *scanner* no automóvel, o aparelho fica disponível para se conectar através do bluetooth do smartphone.

- Registro das viagens: Com a conexão entre smartphone e *scanner* realizada, é feito o acesso ao aplicativo do Torque PRO que é responsável pelo registro das viagens e aquisição das informações que estão disponíveis na OBD. No início da locomoção em cada cenário, foi iniciado o registro das *logs* pelo aplicativo, realizando a aquisição do máximo de dados disponíveis nos sensores do veículo no momento. Ao término da locomoção, foi paralisado o registro das informações internas do veículo. Com o registro da viagem concluído, o aplicativo Torque PRO gera um arquivo em formato *Comma-separated values* (CSV) e o salva no smartphone, sendo um arquivo CSV para cada viagem realizada.
- Aquisição de dados das APIs: Na etapa de adquirir dados das APIs selecionadas foi utilizado a linguagem VBS e o *Power Query* para ser feita a criação de um meio para realizar requisições em cada uma das APIs. Momentos antes de ser feita a locomoção do veículo nos cenários fixos (cenários 1, 2, 3, 4 e 8) que foram definidos, era executado o *script* desenvolvido, que foi responsável por enviar a requisição para a API, receber o retorno que ela disponibiliza e armazenar os dados para serem utilizados posteriormente. Após cada uma dessas requisições serem feitas e os dados de cada API adquiridos, os mesmos eram salvos em um arquivo CSV, sendo um para cada retorno da API.

### 3.0.3 Preparação dos dados

A etapa de preparação dos dados é a fase em que ocorre o processo de rotulação e manipulação dos dados que foram adquiridos, para que seja possível os dados serem facilmente compreendidos e utilizados posteriormente. A seguir, será realizada a descrição dos passos realizados para a preparação das informações que foram coletadas na etapa de aquisição dos dados:

- Rotulação dos dados: a rotulação dos dados ocorreu com o objetivo de buscar a melhor maneira de adequar e fixar os dados para serem utilizados na criação do conjunto de dados. Nesta etapa foi realizado a identificação de todos os conjuntos de dados que haviam sido coletados na etapa de aquisição de dados, de acordo com características do cenário que foi feito a locomoção do veículo. Foi adicionado informações a respeito do veículo que estava sendo utilizado no momento da coleta daquele dado, o motorista que estava dirigindo, o cenário que foi realizada a coleta e uma identificação para separar cada uma das viagens que foram realizadas.
- Anonimização de dados: a anonimização é um método em que os dados pessoais são modificados de forma que alguma informação crítica não possa mais ser identificada

ou localizada. É comum a ocorrência da anonimização de dados em casos que alguma informação vá ser disponibilizada para diversos usuários, para que seja possível manter a segurança e credibilidade. Uma das razões para a pouca disponibilização de *datasets* veiculares é devido as informações geográficas que possuem. Segundo (ALVES, 2021), os procedimentos de anonimização têm sua relevância definida, pois garante a maior privacidade dos titulares de dados pessoais caso aconteça algum compartilhamento de seus dados e, também por garantir aos responsáveis pelo tratamento dos dados uma maior liberdade nas operações de tratamento.

Durante a coleta de dados veiculares para criação de um conjunto de dados, os pontos iniciais e finais de coleta geralmente são localizados na residência ou locais frequentados pelo usuário que está realizando a coleta. Para alcançar o objetivo de manter o conjunto de dados proposto no trabalho com o acesso público e manter a privacidade dos envolvidos na coleta, foi realizada a anonimização dos dados. O método utilizado para a anonimização, foi a remoção do início e fim de cada viagem que foi realizada em ruas aleatórias, para proteger a privacidade das pessoas que estavam executando os experimentos. Outro método possível para realizar a anonimização, seria a adição de ruído nos valores, porém não foi realizado pois poderia modificar a capacidade de localização da trajetória.

- Junção dos dados externos e internos: a fase de junção dos conjuntos de dados existentes, ocorreu com o objetivo de incrementar mais informações veiculares aos conjuntos. Foi selecionado cada uma das viagens que foram realizadas nos cenários fixos (de 1 a 4) e as informações que foram coletadas através de fonte de dados externos. Com a criação da relação entre as viagens realizadas e cada um dos dados extra-veiculares, foi realizado a junção dos dados intra-veiculares e dados extra-veiculares, formando um único conjunto de dados que representava todas as informações referente ao momento que estava sendo feita a locomoção dos automóveis.

### 3.0.4 Uso dos Dados

O uso dos dados consiste na etapa em que foi utilizado todas as informações que foram adquiridas e rotuladas nas etapas anteriores. O primeiro uso para todos os dados rotulados, foi a criação do conjunto de dados proposto nesse trabalho. Na sequência, foi realizada uma análise exploratória do conjunto de dados criado, com o objetivo de demonstrar os dados coletados. Abaixo será feito o detalhamento dos passos para o uso dos dados coletados.

- Criação do conjunto de dados: para a criação do conjunto de dados final que foi proposto neste trabalho, foi realizado a junção dos conjuntos de dados que haviam sido adquiridos e rotulados nas etapas anteriores, de todos os cenários que foram

definidos para a coleta das informações. Para os conjuntos de dados serem mesclados, foi utilizado funções e funcionalidades existentes no *Power Query* e em bibliotecas do *Python*. Para tanto, foi necessário realizar ajustes nos dados como, por exemplo alinhamentos, uniformização das unidades de medidas, marcações de tempo, entre outros.

- **Análise exploratória dos dados:** A análise exploratória dos dados é um passo fundamental para todos os projetos de análise de dados ou ciência de dados. Essa análise consiste em examinar e aprofundar nas características existentes em um conjunto de dados, na maioria das vezes utilizando modos de visualização de dados, como representações gráficas e Tabelas. Com a análise, é possível determinar o melhor modo de controlar os dados, fazendo que seja mais simples para os usuários encontrarem padrões, localizar avarias e testar hipóteses. Para a realização da análise exploratória, foi utilizado ferramentas como o *Power BI* e bibliotecas do *Python*, como por exemplo o *Pandas*, *Scikit-Mobility* e *MobVis*.
- **Publicação do conjunto de dados:** A fase de publicação do conjunto de dados, ocorreu com o foco principal em atingir o objetivo de manter o conjunto de dados proposto no trabalho com o acesso público. Para isso, foi selecionado o *Kaggle* e *Google Drive*. Em seguida foi realizado o envio do arquivo gerado nas etapas anteriores. Juntamente com o conjunto de dados, foi feito o envio das informações que o conjunto de dados possui, para que auxilie os usuários.

## 4 RESULTADOS

Este Capítulo apresenta os resultados obtidos a partir da metodologia proposta e do desenvolvimento que foi realizado. Inicialmente será apresentado o conjunto de dados criado, em seguida será demonstrado possíveis análises para se realizar com as informações do conjunto de dados e possíveis problemas que podem existir em dados veiculares coletados de sensores. Vale ressaltar que as análises apresentadas neste capítulo são exemplos de análises para se realizar com os dados, outras análises são possíveis de realizar com as outras informações contidas no conjunto de dados.

### 4.1 conjunto de dados desenvolvido

Com todos os dados obtidos através da porta **OBD** dos automóveis, foi possível realizar a rotulação das viagens feitas indicando o cenário em que foi realizado a coleta, o motorista que estava dirigindo o veículo no momento, qual o automóvel utilizado e a identificação única para cada viagem específica. Com a posse de todas as viagens veiculares rotuladas, foi possível realizar a junção dos dados que foram coletados em cada **OBD** com as viagens que foram feitas em cenários fixos. Após a junção das informações internas e externas do automóvel, foi feita a junção de todas as viagens para gerar o conjunto de dados final proposto no trabalho. O conjunto de dados desenvolvido foi publicado no *Kaggle* e no *Google Drive* para que o acesso ao mesmo seja livre. A Tabela 6 apresenta os links de acesso ao conjunto de dados desenvolvido.

Tabela 6 – Links de acesso ao conjunto de dados

Plataforma	Link
Kaggle	< <a href="https://www.kaggle.com/datasets/igorsilva02/vehicular-dataset">https://www.kaggle.com/datasets/igorsilva02/vehicular-dataset</a> >
Google Drive	< <a href="https://drive.google.com/drive/folders/1_62vmSg\YcskG3wFhjPJ5sbTen1ABIUuy?usp=sharing">https://drive.google.com/drive/folders/1_62vmSg\YcskG3wFhjPJ5sbTen1ABIUuy?usp=sharing</a> >

Com o passo a passo definido na etapa de criação e aquisição dos dados, foi realizado a captura de 355 viagens veiculares em diversas situações. A Tabela 7, apresenta o resumo sobre o número de viagens presentes no conjunto de dados por cada cenário.



## 4.2 Análise do conjunto de dados

Esta Seção apresenta exemplos de utilizações para os dados do conjunto de dados, analisando as condições e situação dos veículos em viagens, realizando comparações e demonstrando as diferenças do comportamento do automóvel nos cenários definidos

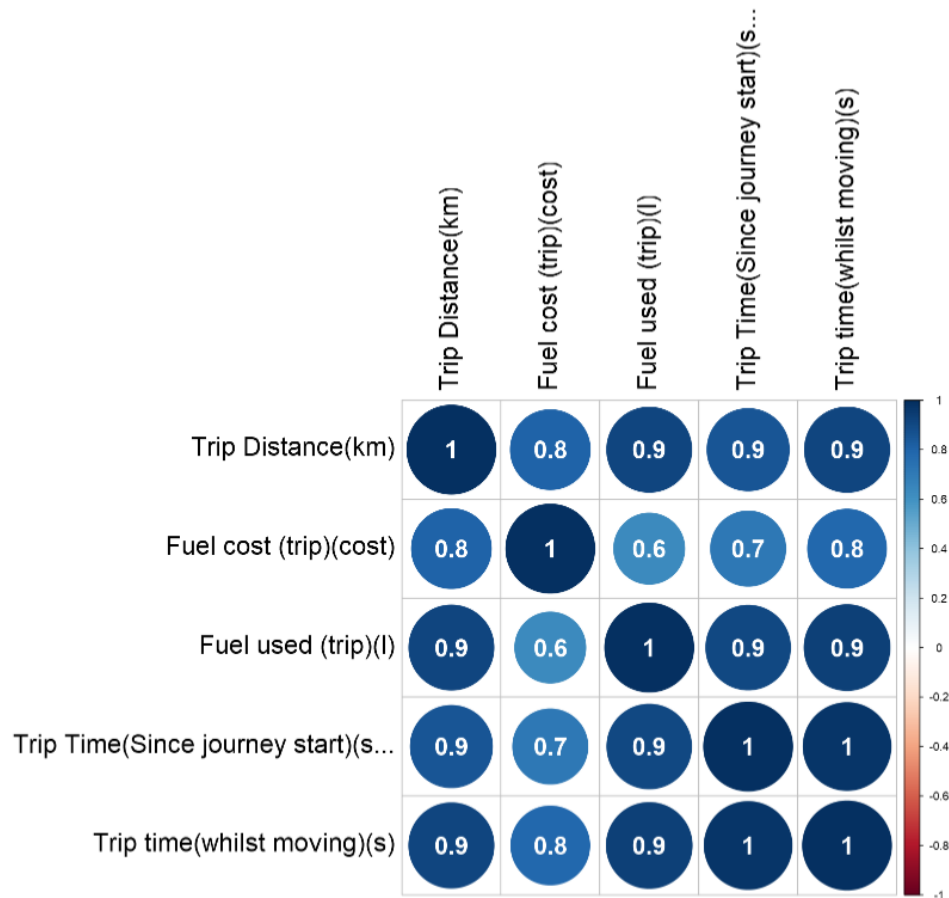
### 4.2.1 Condições dos veículos nos cenários

Com as informações presentes no dataset, é possível identificar o comportamento dos veículos em cada um dos cenários que foram definidos na Tabela 4. Os cenários 1, 2, 3, 4 e 8 representam rotas fixas nas duas principais avenidas da cidade, os cenários 5 e 6 rotas aleatórias do cotidiano e o cenário 7 rotas percorridas apenas em estradas e rodovias. Com essa separação dos cenários, sendo diferenciados pelo tipo de rota e pela tendência de fluxo de veículos em cada horário, as visualizações serão apresentadas para demonstrar as diferenças no comportamento dos automóveis em cada cenário e entre eles. Para os dados extra-veiculares, será apresentado comparações entre valores provindos de fontes externas e internas dos veículos.

Através dos dados é possível obter informações gerais sobre as condições do veículo para cada cenário como, por exemplo a velocidade média, quilometragem por litros, média de  $CO_2$ , gasolina utilizada, tempo da viagem, entre outras informações. Algumas das variáveis coletadas são apresentadas na Figura 7. Também é possível identificar correlações entre as variáveis que foram coletados. A medida de correlação é utilizada quando se deseja verificar se dois ou mais itens possuem algum tipo de associação. Quando o valor de uma variável aumenta e o da outra variável também tende a seguir esse aumento, a correlação é caracterizada como positiva. Quando o valor de uma variável começa a diminuir e o da outra variável segue diminuindo, a correlação é caracterizada como negativa. Com o valor de correlação que dois itens possuem, é possível ter um valor previsto para o outro item. Na Figura 7 é possível verificar a correlação entre algumas das variáveis que o conjunto de dados possui. As variáveis seguem um comportamento esperado e possuem uma correlação positiva como, por exemplo a distância, o tempo, o combustível utilizado e o custo do combustível gasto nas viagens seguirem cada um aumentando enquanto o outro também aumenta. Com isso é possível compreender que quando uma dessas variáveis aumentar o seu valor, a outra também seguirá o mesmo comportamento.



Figura 7 – Correlação positiva entre variáveis



Fonte: Produzido pelo autor

Utilizando a biblioteca *Python MobVis* (NOVAIS, 2022), é possível gerar visualizações e extrair métricas com os dados veiculares do conjunto de dados. A Figura 8 que foi gerada utilizando essa biblioteca, é possível verificar as coordenadas geográficas ao longo do tempo, que o veículo percorreu em um cenário aleatório. A escala de tempo é representada pela cor. Com isso é possível ver a evolução do movimento ao longo do tempo.

Figura 8 – rastro de mobilidade



Fonte: Produzido pelo autor

Além dos rastros de movimento do veículo, é possível verificar a distância percorrida entre os pontos geográficos durante a viagem. A Figura 9 representa a distância dos pontos geográficos contidos na viagem em metros, demonstrando que ao longo do tempo ocorreu em maior número a locomoção do automóvel pela distância de 6,83 metros e indicando que durante essa viagem aleatória, o veículo não ficou parado por muitos momentos, já que não é alto o número de ocorrências da distância de viagem em zero metros.

Figura 9 – Distância de viagem

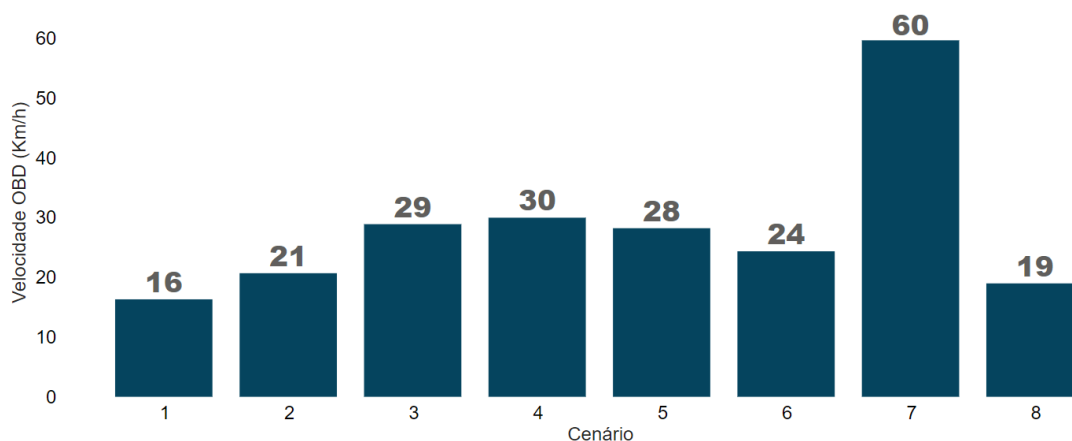


Fonte: Produzido pelo autor

O comportamento esperado em um veículo, varia de acordo com a situação em que

ele está no momento, sendo esperada reações e desempenhos diferentes dependendo se o automóvel está em uma rodovia, uma rua movimentada ou uma rua sem trânsito. Para verificarmos a variação do comportamento dos veículos, na Figura 10 é possível comparar a velocidade média dos veículos em cada cenário, tendo um menor valor nos cenários que possuem uma tendência maior de fluxo de veículos. Para o cenário de locomoção em rodovias e estradas, a diferença esperada também fica clara, devido os automóveis se locomoverem mais rapidamente.

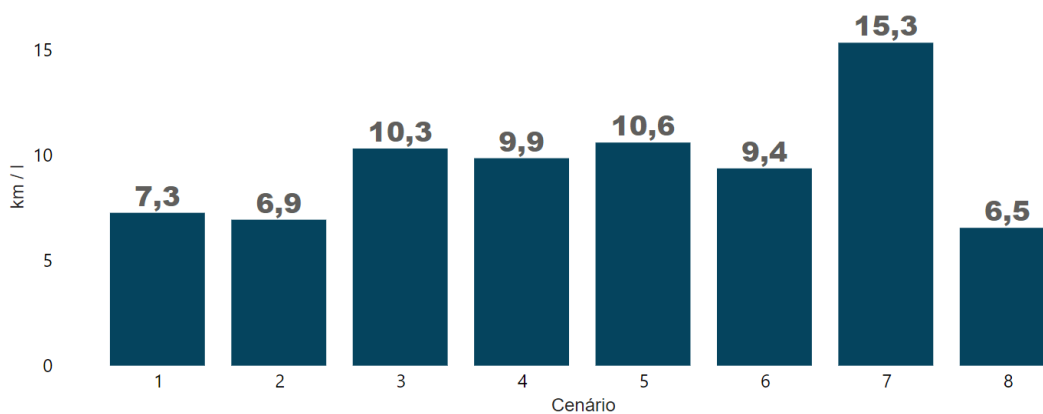
Figura 10 – Velocidade (OBD) por cenário



Fonte: Produzido pelo autor

Conforme a Figura 11, é possível verificar que em cenários que a tendência de fluxo de automóveis é alta e média, os veículos percorrem em média um valor bem menor de quilômetros por litro comparando com os cenários em que a tendência de fluxo é baixa, pois a velocidade média de locomoção é menor e o veículo fica impossibilitado de percorrer maiores distâncias.

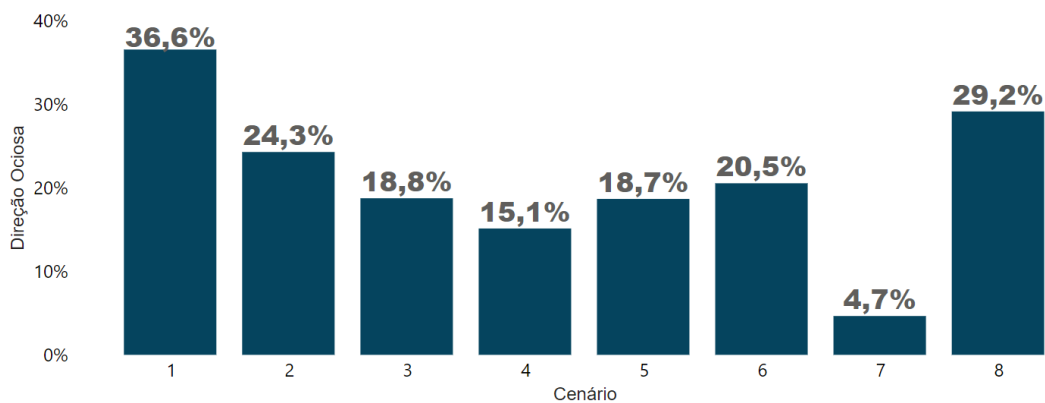
Figura 11 – km/l por cenário



Fonte: Produzido pelo autor

Também é possível verificar na Figura 12 que nesses cenários de médio e alto fluxo de veículos, a porcentagem média de direção ociosa, que indica o tempo médio que o motor de um automóvel ficou em funcionamento enquanto ele não estava em movimento, é bem maior que em viagens realizadas em rodovias.

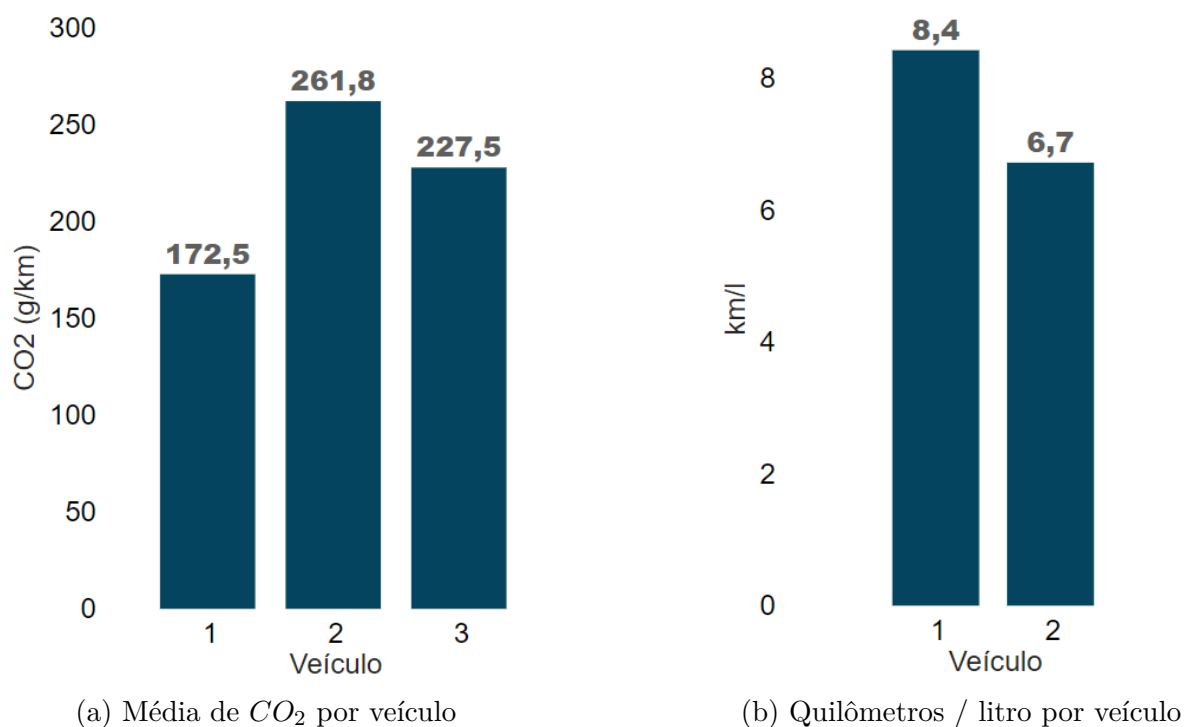
Figura 12 – Direção ociosa por cenário



Fonte: Produzido pelo autor

Cada veículo possui as suas especificações, ficha técnica e diferentes tecnologias empregadas na fabricação, devido a isso, mesmo em cenários similares é esperado um comportamento diferente entre automóveis distintos. Na Figura 13a, pode-se verificar uma média diferente entre a quantidade de gramas de  $CO_2$  que é emitida por quilômetro percorrido por cada veículo usado. Também pode-se verificar na Figura 13b a média de quilômetros por litro que o veículo 1 e 2 percorre no cenário 1 e 3, tendo uma diferença devido serem distintos.

Figura 13 – Comportamento dos veículos



Assim como exibido na Figura 10 e Figura 12, se um automóvel está se locomovendo em um ambiente com maior quantidade de trânsito, ocorrerá diferenças nas condições e desempenho do veículo comparado com um ambiente com menor quantidade de trânsito. Além do fluxo de automóveis nos ambientes, outras condições externas ao veículo podem interferir no seu comportamento e no tempo gasto na viagem, como por exemplo as condições climáticas. As ferramentas que realizam o cálculo da previsão de tempo entre pontos geográficos analisam diversas condições para prever o tempo da viagem, como por exemplo o estado do trânsito em tempo real, o tipo de rua que o carro terá que percorrer, a velocidade média a cada momento do dia, entre outros pontos. A Tabela 8 tem como objetivo realizar uma comparação entre o tempo médio que foi estimado pelas APIs do *Google Directions* e *Here WeGo* com o tempo real que foi gasto em 5 viagens selecionadas, sendo do cenário 1, 2, 3, 4 e 8. É possível verificar que o tempo previsto pela API do

Tabela 8 – Comparação entre tempo real da viagem e tempo das APIs

Cenário	Tempo Viagem	Tempo Here WeGo	Tempo Google Maps
1	00:06:42	00:05:00	00:05:00
2	00:06:15	00:06:00	00:06:00
3	00:04:25	00:05:00	00:05:00
4	00:03:35	00:04:00	00:04:00
8	00:07:21	00:06:00	00:06:00

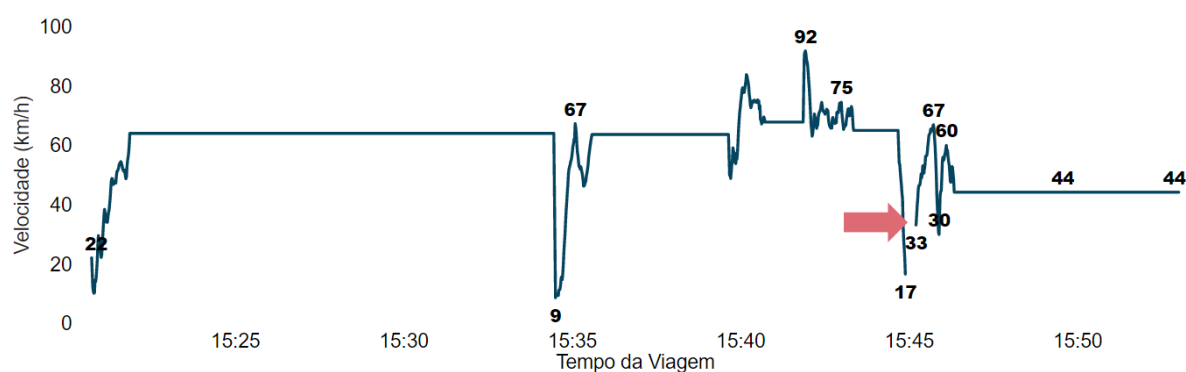
## 4.2.2 Problemas de dados

Em informações coletadas através de sensores, ocorre problemas devido os sensores serem uma fonte de dados inerentemente imperfeitas. Alguns dos dados presentes no conjunto de dados que foram coletados em sensores intra-veiculares, também estavam sujeitos a adversidades. Foi realizado uma análise exploratória nos dados, para demonstrar possíveis problemas que foram localizados no conjunto de dados. Com esses problemas, fica claro que dependendo da utilização do conjunto de dados, os dados devem ser tratados antes do uso dos mesmos. A seguir, é mostrado os conjuntos de problemas existentes nos dados de sensores intra-veicular

### 4.2.2.1 Dados incompletos

Os dados incompletos são aqueles que em algum momento estão sem nenhum valor (dados em branco), levando a valores e resultados incorretos. Esse problema pode vir a ocorrer, por exemplo, através de algum erro na leitura dos sensores ou através de problemas de comunicação. A Figura 14 apresenta valores referente a uma das viagens presentes no conjunto de dados, em que em determinado momento há uma interrupção da coleta da velocidade do veículo segundo o *Global Positioning System (GPS)* e após um tempo, retorna a coletar novamente. Uma das formas existentes para se solucionar o problema dos dados incompletos é realizar a substituição do valor faltante pela média, mediana ou moda da velocidade naquela viagem.

Figura 14 – Dados incompletos



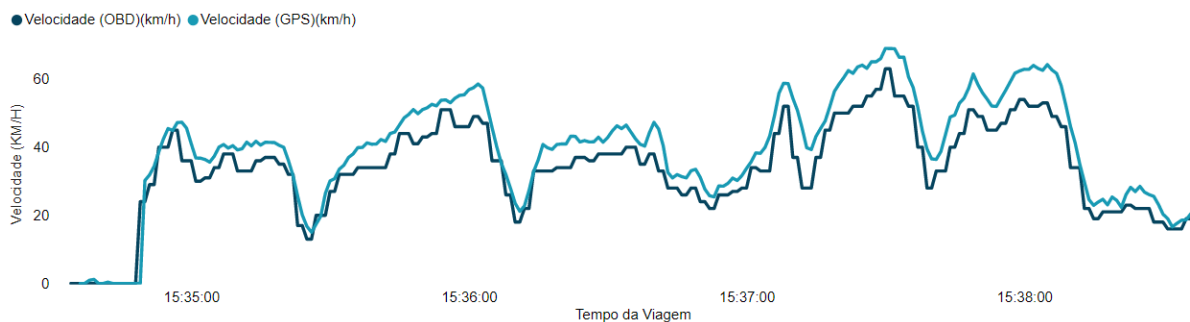
Fonte: Produzido pelo autor

### 4.2.2.2 Dados conflitantes

Um conflito de informação ocorre quando duas ou mais variáveis que deveriam representar a mesma informação, acabam apresentando valor divergente entre elas. Esses erros acontecem nas leituras realizadas pelos sensores e acabam gerando conflitos e dúvidas nas análises das informações. Segundo (YAGER, 1987), uma forma de solucionar conflitos entre dados, é utilizando a regra de combinação de Dempster. A Figura 15 representa a

ocorrência de conflitos de informações durante uma viagem, quando feita a comparação entre a velocidade do veículo segundo a **OBD** e a velocidade do veículo segundo o **GPS**. Por exemplo, por volta do horário de 15:37:00 a velocidade indicada pela **OBD** está em 28 quilômetros por hora e a indicada pelo **GPS** está em 39 quilômetros por hora, gerando uma divergência na informação.

Figura 15 – Dados conflitantes

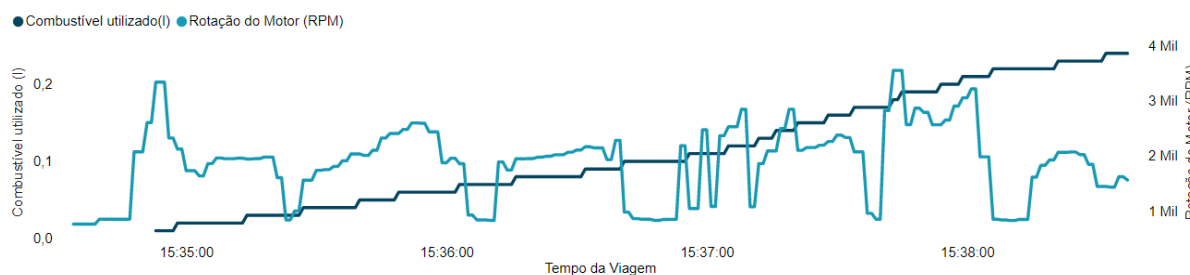


Fonte: Produzido pelo autor

#### 4.2.2.3 Dados desiguais

Alguns dos dados que estão disponíveis para serem coletados pelos sensores veiculares são considerados desiguais devido verificarem diferentes características em elementos e unidades distintas. Pode ser feito o estudo dos elementos que são desiguais para entender as atitudes dos mesmos, mas de forma separada e distinta. Na Figura 16, é apresentado um exemplo de uma dessemelhança entre sensores veiculares, como o combustível gasto na viagem (l) e a **RPM**.

Figura 16 – Dados desiguais



Fonte: Produzido pelo autor

#### 4.2.2.4 Dados incertos

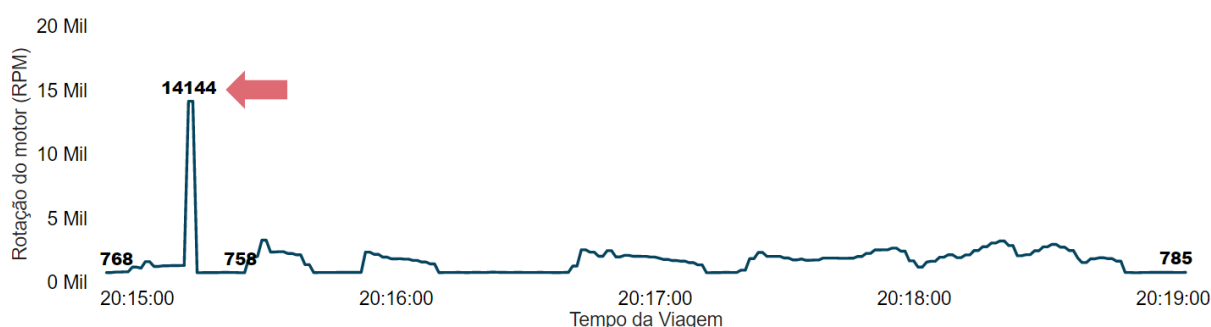
Todas as informações providas de sensores ou alguma fonte de dado externa podem possuir um nível de incerteza. Para a solução do problema de incerteza, pode ser utilizado inferência estatística, que obtém conclusões através de amostras. Como demonstrado na

Tabela 8, as informações não possuem total confiança do valor que foi passado, havendo uma diferença entre a média de tempo previsto para percorrer o cenário e o tempo médio real gasto naquele cenário.

#### 4.2.2.5 Ponto fora da curva

O ponto fora da curva é considerado aqueles dados que possuem valor incorreto, podendo ser gerado por alguma falha que ocorreu nos sensores. Esses dados incorretos podem fazer com que cálculos e análises estatísticas venham a ter valores incorretos e incertos. A Figura 17 apresenta número de RPM do motor do veículo durante uma viagem, porém é possível perceber um valor constando 14144 entre outros valores que estão bem abaixo. O veículo durante alguns momentos da viagem está com as RPM por volta de 1200, mas no segundo seguinte o valor altera para 14144 e em seguida volta para um valor abaixo de 1000. Esses acontecimentos são os pontos fora da curva.

Figura 17 – Ponto fora da curva

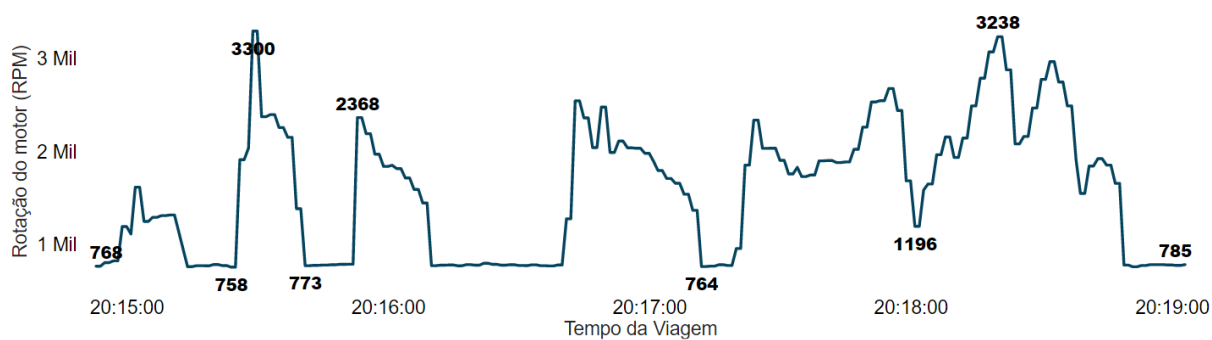


Fonte: Produzido pelo autor

Existem diversas formas de soluções para os pontos fora da curva, como por exemplo a exclusão do valor discrepante, métodos de clusterização para realizar a correção dos valores, análise e investigação do motivo de geração de pontos fora da curva, entre outros. Ao realizar a remoção do valor considerado ponto fora da curva, a visualização da RPM do motor sofre uma alteração, assim como demonstrado na Figura 18 abaixo, que representa a mesma viagem da Figura 17 que foi apresentada acima.



Figura 18 – Ponto fora da curva corrigido



Fonte: Produzido pelo autor

## 5 CONCLUSÃO

No presente estudo foi apresentado um novo conjunto de dados veiculares de acesso público, com dados extraídos da [API](#) do *Google Directions*, *Weather Channel* e *Here WeGo*, além de dados obtidos pela interface [OBD](#) contida em veículos. Com as etapas definidas no desenvolvimento, foi possível realizar a criação do conjunto de dados, com as informações rotuladas. Após, foi realizada análises nos dados do *dataset*, comparando resultados por cenários, comportamentos dos veículos e verificando possíveis problemas nos dados veiculares.

Com as análises das condições do veículo, foi possível observar a diferença do comportamento dos automóveis em situações parecidas, tendo diferenças no desempenho devido as características que cada veículo possui. Foi possível verificar as condições dos veículos em cada cenário de coleta, sendo possível verificar diferenças que são esperadas dependendo do cenário que o automóvel está percorrendo. Também foi possível identificar diferenças entre as informações provindas de diferentes fontes de dados veiculares, que deveriam conter a mesma informação. Além do mais, foi possível verificar possíveis problemas que podem existir em dados provindos de fonte de dados veiculares, por algum possível erro ou problema.

Com o *dataset* desenvolvido e as análises presentes neste trabalho, se torna possível a existência de diversos trabalhos futuros. Como possível implicação do *dataset* em trabalhos futuros, se tem a análise das condições do veículo e dos cenários, assim como foi demonstrado nos resultados. Outro uso para o *dataset* é em trabalhos relacionados aos possíveis problemas existentes em informações coletadas de fontes de dados veiculares, como demonstrado na monografia, sendo feita a tentativa de entendimento da origem dos problemas e realizando o tratamento dos dados. Também como trabalho relacionado, é possível a utilização dos dados veiculares para auxiliar no desenvolvimento de ferramentas e bibliotecas que realizem a manipulação de dados geográficos ou de alguma outra informação sobre veículos. Uma outra utilização possível para o conjunto de dados, é a tentativa de identificação do motorista que está dirigindo o automóvel em cada cenário, sem visualizar a coluna que identifica o motorista e tendo como base apenas os comportamentos e condições do automóvel no momento que cada um dos motoristas está dirigindo o carro. Dentre os exemplos de trabalhos futuros citados, o *dataset* no geral pode ser útil para desenvolvimentos de trabalhos na área de dados veiculares, devido à escassez de conjuntos de dados públicos existentes.

# Referências

- ALVES, D. V. Técnicas de anonimização de dados pessoais e a lei n. 13.709/2018. 2021. Citado na página 33.
- ANDERSEN, J.; SUTCLIFFE, S. Intelligent transport systems (its)-an overview. *IFAC Proceedings Volumes*, Elsevier, v. 33, n. 18, p. 99–106, 2000. Citado 2 vezes nas páginas 16 e 17.
- AZADANI, M. N.; BOUKERCHE, A. Performance evaluation of driving behavior identification models through can-bus data. In: IEEE. *2020 IEEE Wireless Communications and Networking Conference (WCNC)*. [S.l.], 2020. p. 1–6. Citado na página 26.
- BERNARDI, M. L. et al. Driver and path detection through time-series classification. *Journal of Advanced Transportation*, Hindawi, v. 2018, 2018. Citado na página 26.
- BLASCO, V. *Sistema de diagnóstico OBD II*. [S.l.]: Obtenido de Sistema de diagnóstico OBDII: <http://www.electronicar.net/IMG...>, 2014. Citado na página 20.
- BORTH, T. F. Analisando os impactos do uso do protocolo can fd em aplicações automotivas: estudo de caso. 2016. Citado na página 19.
- CUNHA, F. et al. Sistemas de transporte inteligentes: Conceitos, aplicações desafios. *Livro de Minicursos do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC'17)*, 2017. Citado 3 vezes nas páginas 12, 16 e 17.
- DAVIS, N.; RAINA, G.; JAGANNATHAN, K. A framework for end-to-end deep learning-based anomaly detection in transportation networks. *Transportation research interdisciplinary perspectives*, Elsevier, v. 5, p. 100112, 2020. Citado 2 vezes nas páginas 26 e 27.
- IBPT. *Novo estudo do IBPT revela aumento no número de veículos em circulação*. 2020. <https://ibpt.com.br/novo-estudo-do-ibpt-revela-aumento-no-numero-de-veiculos-em-circulacao-2/>. Citado na página 16.
- KONG, X. et al. Mobility dataset generation for vehicular social networks based on floating car data. *IEEE Transactions on Vehicular Technology*, IEEE, v. 67, n. 5, p. 3874–3886, 2018. Citado 2 vezes nas páginas 26 e 27.
- LANDHÄUSSER, F. et al. Electronic control unit (ecu). In: *Diesel Engine Management*. [S.l.]: Springer, 2014. p. 272–277. Citado na página 19.
- LORCH, J. R. et al. The smart way to migrate replicated stateful services. In: *Proceedings of the 1st ACM SIGOPS/EuroSys European Conference on Computer Systems 2006*. [S.l.: s.n.], 2006. p. 103–115. Citado na página 12.
- NOVAIS, L. *Um framework para visualização de métricas de mobilidade*. 2022. <https://github.com/lucNovais/MobVis#computer-installation>. Citado na página 38.

- OSS, M. Leitura obd2 através de smartphone. Universidade Tecnológica Federal do Paraná, 2018. Citado 4 vezes nas páginas 19, 20, 21 e 22.
- PAPPALARDO, L. et al. scikit-mobility: A python library for the analysis, generation and risk assessment of mobility data. *arXiv preprint arXiv:1907.07062*, 2019. Citado na página 25.
- RETTORE, P. H. et al. Driver authentication in vanets based on intra-vehicular sensor data. In: IEEE. *2018 IEEE Symposium on Computers and Communications (ISCC)*. [S.l.], 2018. p. 00078–00083. Citado 2 vezes nas páginas 26 e 27.
- RETTORE, P. H. et al. Vehicular data space: The data point of view. *IEEE Communications Surveys & Tutorials*, IEEE, v. 21, n. 3, p. 2392–2418, 2019. Citado 2 vezes nas páginas 12 e 17.
- RETTORE, P. H. et al. Towards intra-vehicular sensor data fusion. In: IEEE. *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. [S.l.], 2016. p. 126–131. Citado na página 20.
- RETTORE, P. H. et al. Road data enrichment framework based on heterogeneous data fusion for its. *IEEE Transactions on Intelligent Transportation Systems*, IEEE, v. 21, n. 4, p. 1751–1766, 2020. Citado 2 vezes nas páginas 26 e 27.
- SANTOS, B. P. et al. Enriching traffic information with a spatiotemporal model based on social media. In: IEEE. *2018 IEEE Symposium on computers and communications (ISCC)*. [S.l.], 2018. p. 00464–00469. Citado 2 vezes nas páginas 19 e 26.
- SCHMITT, P. R. M. *Aplicação web utilizando api google maps*. Dissertação (B.S. thesis) — Universidade Tecnológica Federal do Paraná, 2013. Citado na página 24.
- SOUZA, A. G. de; CAMPOS, G. L. Rede can veicular: levantamento bibliográfico e apresentação de conceitos iniciais. *ForScience*, v. 5, n. 1, 2017. Citado na página 19.
- SU, Z. et al. A traffic event detection method based on random forest and permutation importance. *Mathematics*, MDPI, v. 10, n. 6, p. 873, 2022. Citado 2 vezes nas páginas 26 e 27.
- TEMPELMEIER, N. et al. Mining topological dependencies of recurrent congestion in road networks. *ISPRS International Journal of Geo-Information*, Multidisciplinary Digital Publishing Institute, v. 10, n. 4, p. 248, 2021. Citado 2 vezes nas páginas 26 e 27.
- UPPOOR, S. et al. Generation and analysis of a large-scale urban vehicular mobility dataset. *IEEE Transactions on Mobile Computing*, IEEE, v. 13, n. 5, p. 1061–1075, 2013. Citado 2 vezes nas páginas 26 e 27.
- YAGER, R. R. On the dempster-shafer framework and new combination rules. *Information sciences*, Elsevier, v. 41, n. 2, p. 93–137, 1987. Citado na página 43.

# Apêndices

# APÊNDICE A –

Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
Id	Identificação da viagem	Inteiro	Rotulação
GPSTime	Tempo no GPS	Data e hora	OBD
DeviceTime	Tempo no Dispositivo	Data e hora	OBD
Longitude	Longitude	Graus	OBD
Latitude	Latitude	Graus	OBD
GPSSpeed(Meters/second)	Velocidade no GPS	Metros por segundo	OBD
HorizontalDilutionofPrecision	Grau de degradação do nível de acuracidade da posição horizontal	Graus	OBD
Altitude	Altitude	Metros	OBD
Bearing	Orientação	Graus	OBD
G(x)	Aceleração no eixo X	m/s <sup>2</sup>	OBD
G(y)	Aceleração no eixo Y	m/s <sup>2</sup>	OBD
G(z)	Aceleração no eixo Z	m/s <sup>2</sup>	OBD
G(calibrated)	Valor de G combinado	m/s <sup>2</sup>	OBD
0-100kphTime(s)	Desempenho	Segundo	OBD
0-100mphTime(s)	Desempenho	Segundo	OBD
0-200kphTime(s)	Desempenho	Segundo	OBD
0-30mphTime(s)	Desempenho	Segundo	OBD
0-60mphTime(s)	Desempenho	Segundo	OBD
1/4MileTime(s)	Desempenho	Segundo	OBD
1/8MileTime(s)	Desempenho	Segundo	OBD

Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
100-0KphTime(s)	Desempenho	Segundo	OBD
100-200KphTime(s)	Desempenho	Segundo	OBD
40-60MphTime(s)	Desempenho	Segundo	OBD
60-0MphTime(s)	Desempenho	Segundo	OBD
60-120MphTime(s)	Desempenho	Segundo	OBD
60-130MphTime(s)	Desempenho	Segundo	OBD
60-80MphTime(s)	Desempenho	Segundo	OBD
80-100MphTime(s)	Desempenho	Segundo	OBD
80-120KphTime(s)	Desempenho	Segundo	OBD
AbsoluteThrottle	Posição de aceleração		OBD
Position B(%)	absoluta		
AccelerationSensor(Total)(g)	Sensores de Aceleração (Total)	g	OBD
AccelerationSensor (Xaxis)(g)	Sensores de Aceleração (Eixo X)	g	OBD
AccelerationSensor (Yaxis)(g)	Sensores de Aceleração (Eixo Y)	g	OBD
AccelerationSensor (Zaxis)(g)	Sensores de Aceleração (Eixo Z)	g	OBD
AcceleratorPedal	Posição do pedal do		
PositionD(%)	acelerador	Porcentagem	OBD



Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
AcceleratorPedalPositionE(%)	Posição do pedal do acelerador	Porcentagem	OBD
ActualEngine%Torque(%)	Torque real do motor	Porcentagem	OBD
AirFuelRatio(Comanded):(1)	Relação de combustível de ar (comandada)		OBD
AirFuelRatio(Measured):(1)	Relação de combustível de ar (mensurado)		OBD
AmbientAirTemp(°C)	Temperatura do ar ambiente	Graus Celsius	OBD
AndroidDeviceBatteryLevel(%)	Nível da bateria do dispositivo Android	Porcentagem	OBD
AverageTripSpeed(WhilstMovingOnly)(km/h)	Velocidade média de viagem (apenas em movimento)	km/h	OBD
AverageTripSpeed(WhilstStoppedorMoving)(km/h)	Velocidade média de viagem (enquanto parado ou em movimento)	km/h	OBD
BarometricPressure(fromVehicle)(psi)	Pressão do ar medida pela ECU	psi	OBD
CatalystTemperature(Bank1Sensor 1)(°C)	Temperatura do catalisador	Graus Celsius	OBD
ChargeAirCoolerTemperature(CACT)(°C)	Temperatura do Resfriador de Ar de Carga	Graus Celsius	OBD

Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
CommandedEquivalence	Razão de Equivalência Comandada	Lambda	OBD
Ratio(lambda)			
CostPerMile/km (Instant)(R\$/km)	Custo por milha / km (instantâneo)	R\$/Km	OBD
CostPerMile/km (Trip)(R\$/km)	Custo por milha / km (Viagem)	R\$/Km	OBD
CO2(Average)(g/km)	CO2 (Média)	g/Km	OBD
CO2(Instantaneous)(g/km)	CO2 (Instantâneo)	g/Km	OBD
DistanceToEmpty (Estimated)(km)	Distância para acabar a gasolina (Estimativa)	km	OBD
DistanceTravelledSince CodesCleared(km)	Distância percorrida desde que os códigos foram apagados	km	OBD
DistanceTravelled WithMIL/CEL(km)	Distância percorrida com MIL / CEL	litros por km	OBD
DPFBank1Delta Pressure(psi)	Pressão DPF do banco 1	psi	OBD
DPFBank1InletPressure(psi)	Pressão de Entrada do banco 1	psi	OBD
DPFBank1Inlet Temperature(°C)	Temperatura DPF	Graus Celsius	OBD

Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
DriversDemand Engine%Torque(%)	Torque demandada pelo motorista	Porcentagem	OBD
EGRCommanded(%)	Recirculação de gases de escape (comandado)	Porcentagem	OBD
EGRError(%)	Erro de recirculação de gases de escape	Porcentagem	OBD
EngineCoolantTemperature(°C)	Temperatura do líquido de arrefecimento do motor	Graus Celsius	OBD
EngineKW(AtTheWheels)(KW)	Quilowatt do motor (nas rodas)	KW	OBD
EngineLoad(%)	Carga do motor	Porcentagem	OBD
EngineLoad(Absolute)(%)	Carga do motor (absoluta)	Porcentagem	OBD
EngineOilTemperature(°C)	Temperatura do óleo do motor	Graus Celsius	OBD
EngineReferenceTorque(Nm)	Torque de referência do motor	Newton-metro	OBD
EngineRPM(RPM)	Rotações por minuto do motor	RPM	OBD
EthanolFuel(%)	Combustível etanol	Porcentagem	OBD
EvapSystem VapourPressure(Pa)	Pressão de Vapor do Sistema Evap	Pascal	OBD

Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
ExhaustPressureBank1 (psi)	Banco de Pressão de Escape	psi	OBD
FuelCost(Trip)(Cost)	Custo do combustível na viagem	R\$	OBD
FuelFlowRate/Hour(1/hr)	Taxa de fluxo de combustível por hora	Litros por hora	OBD
FuelFlowRate/Minute (cc/min)	Taxa de fluxo de combustível por minuto	Litros por minuto	OBD
FuelLevel (FromEngineECU)(%)	Nível de combustível (da ECU do motor)	Porcentagem	OBD
FuelPressure(psi)	Pressão do Combustível	psi	OBD
FuelRailPressure(psi)	Pressão do trilho de combustível	psi	OBD
FuelRailPressure (RelativeToManifold Vacuum)(psi)	Pressão do trilho de combustível (relativa ao vácuo do coletor)	psi	OBD
FuelRate (DirectFromECU)	Taxa de combustível (diretamente da ECU)		OBD
FuelRemaining (CalculatedFrom VehicleProfile)(%)	Combustível Restante (Calculado a Partir do Perfil do Veículo)	Porcentagem	OBD
FuelTrimBank1 LongTerm(%)	Banco de Combustível 1 (Longo prazo)	Porcentagem	OBD

Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
FuelTrimBank1	Banco de Combustível 1	Porcentagem	OBD
ShortTerm(%)	(Curto prazo)		
FuelUsed(Trip)(l)	Gasolina utilizada na viagem	Litros	OBD
GPSAccuracy(m)	GPS Accuracy	Metros	OBD
GPSAltitude(m)	Altitude GPS	Metros	OBD
GPSBearing(°)	Orientação GPS	Graus	OBD
GPSLatitude(°)	Latitude GPS	Graus	OBD
GPSLongitude(°)	Longitude GPS	Graus	OBD
GPSvsOBDSpeed	Diferença de velocidade		
Difference(KM/h)	GPS vs OBD	km/h	OBD
Horsepower			
(AtTheWheels)(hp)	Potência (Nas Rodas)	Cavalo vapor	OBD
HybridBatteryCharge(%)	Carga da bateria híbrida	Porcentagem	OBD
IntakeAirTemperature(°C)	Temperatura do ar de admissão	Graus	OBD
IntakeManifoldPressure(psi)	Pressão do Coletor de Admissão	psi	OBD
KilometersPerLitre	Quilômetros por litro		
(Instant)(kpl)	(instantâneo)	km/l	OBD
KilometersPerLitre	Quilômetros por litro		
(LongTermAverage)(kpl)	(Média de longo prazo)	km/l	OBD

Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
LitresPer100Kilometer (Instant)(l/100km)	Litros por 100 Quilômetros (Instantâneo)	litros por 100km	OBD
LitresPer100Kilometer (LongTermAverage) (l/100km)	Litros por 100 Quilômetros (Média de longo prazo)	litros por 100km	OBD
MassAirFlowRate	Taxa de fluxo de massa de ar	Gramas	OBD
MilesPerGallon(Instant) (mpg)	Milhas por Galão (Instantâneo)	Milhas por galão	OBD
MilesPerGallon (LongTermAverage)(mpg)	Milhas por Galão (Média de longo prazo)	Milhas por galão	OBD
PercentageOfCityDriving (%)	Porcentagem de Condução Urbana	Porcentagem	OBD
PercentageOfHighwayDriving (%)	Porcentagem de Condução Rodoviária	Porcentagem	OBD
PercentageOfIdleDriving (%)	Porcentagem de condução ociosa	Porcentagem	OBD
RelativeAcceleratorPedalPosition (%)	Posição relativa do pedal do acelerador	Porcentagem	OBD
RelativeThrottlePosition (%)	Posição relativa do pedal do acelerador	Porcentagem	OBD
RunTimeSinceEngineStart (s)	Tempo de execução desde a(s) partida(s) do motor	Segundo	OBD

Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
Speed(GPS)(KM/h)	Velocidade no GPS	km/h	OBD
Speed(OBD)(KM/h)	Velocidade no OBD	km/h	OBD
ThrottlePosition (Manifold)(%)	Posição do acelerador (Múltiplo)	Porcentagem	OBD
TimingAdvance(°)	Tempo avançado	Graus	OBD
Torque(Nm)	Torque	Newton-metro	OBD
TripAverageKPL(kpl)	Média de quilômetros por litro	km/l	OBD
TripAverageLitres/100KM (l/100KM)	Média de quilômetros por litro por 100km	km/l	OBD
TripAverageMPG(mpg)	Média de galão por milhas	Milhas por galão	OBD
TripDistance(KM)	Distância de viagem	km	OBD
TripDistance	Distância da viagem	km	OBD
(StoredInVehicleProfile)(KM)	(armazenada no perfil do veículo)		
TripTime (SinceJourneyStart)(s)	Tempo de viagem (desde o início da viagem)	Segundo	OBD
TripTime (WhilstMoving)(s)	Tempo de viagem (em movimento)	Segundo	OBD
TripTime (WhilstStationary)(s)	Tempo de viagem (enquanto parado)	Segundo	OBD
TurboBoost& VacuumGauge(psi)	Turbo Boost e medidor de vácuo	psi	OBD

Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
Voltage(ControlModule)(V)	Voltagem (Módulo de Controle)	Volts	OBD
Voltage(OBDAdapter)(V)	Voltagem (Adaptador OBD)	Volts	OBD
VolumetricEfficiency (Calculated)(%)	Eficiência Volumétrica (Calculada)	Porcentagem	OBD
Temperature	Temperatura	Graus Celsius	API Weather Channel
WeatherDescriptions	Descrição do tempo	Texto	API Weather Channel
WindSpeed	Velocidade do vento	km/h	API Weather Channel
WindDegree	Direção do vento	Inteiro	API Weather Channel
WindDir	Direção do vento	Texto	API Weather Channel
Pressure	Pressão atmosférica	Pascal	API Weather Channel
Precip	Precipitação	mm	API Weather Channel
Humidity	Humidade	kg/m <sup>3</sup>	API Weather Channel
CloudCover	Cobertura de nuvens	Porcentagem	API Weather Channel
FeelsLike	Sensação Térmica	Graus Celsius	API Weather Channel
UvIndex	Índice de UV	Inteiro	API Weather Channel
Visibility	Visibilidade no momento (Entre 0 e 10)	Inteiro	API Weather Channel
IsDay	Define se é dia ou não	Texto	API Weather Channel
HereMapsSpeed	Velocidade prevista na via	km/h	API Here WeGo
Uncapped	(Pode ser acima do permitido)		
HereMapsSpeed	Velocidade prevista na via	km/h	API Here WeGo



Tabela 9 – Colunas do Dataset

<b>Dado</b>	<b>Descrição</b>	<b>Unidade / Saída</b>	<b>Fonte</b>
HereMapsFreeFlow	Velocidade de referência da via quando não há tráfego	km/h	API Here WeGo
HereMapsJamFactor	Quantidade de tráfego na via (Valor de 0 a 10, quanto maior o número, maior o tráfego)	Inteiro	API Here WeGo
HereMapsDuration	Tempo estimado para a viagem	Minutos	API Here WeGo
HereMapsDistance	Distância a ser percorrida na viagem	km	API Here WeGo
GoogleMapsDistance	Distância a ser percorrida na viagem	km	API Google Maps
GoogleMapsDuration	Tempo estimado para a viagem	Minutos	API Google Maps
Hour	Hora no momento da coleta	Hora	Rotulação
Driver	Motorista que está dirigindo o veículo	Inteiro	Rotulação
Vehicle	Veículo que está sendo utilizado	Inteiro	Rotulação
Scenario	Cenário de coleta de dados	Inteiro	Rotulação