



CURSO: Tecnologia Em Ciência De Dados

POLO DE APOIO PRESENCIAL: Jundiaí e Higienópolis

SEMESTRE: 1/2024

COMPONENTE CURRICULAR / TEMA: **PROJETO APLICADO II {TURMA 02A} 2023/3**

NOME DO GRUPO – MacGyver

RA 10415058 – EDUARDO DAVID - 10415058@MACKENZISTA.COM.BR
RA 10415270 – FELIPE JOSÉ DA CUNHA - 10415270@MACKENZISTA.COM.BR
RA 10415636 – NATÁLIA FRANÇOZO - 10415636@MACKENZISTA.COM.BR
RA 10415977 – ANA VITÓRIA SILVA - 10415977@MACKENZISTA.COM.BR

NOME DO PROFESSOR: **Prof. Dr. Felipe Albino dos Santos**



Sumário

Sumário	2
1. Título:	4
2. Introdução	4
3. Objetivos:	4
4. Metas:	5
5. Cronograma:	6
6. Fluxo Baseado em Pensamento Computacional em Contextos Organizacionais:	7
A. Decomposição:	7
B. Reconhecimento de padrões:	7
C. Abstração:	7
D. Design de Algoritmos:	7
7. Organização e o contexto em que os dados foram gerados:	8
8. Referências de aquisição do dataset:	9
9. Dataset e Metadados	9
A. Dataset:	9
B. Descrição do Dataset:	9
C. Tipo do Arquivo:	11
D. Sensibilidade:	11
E. Proprietário do Dado:	11
F. Restrições de uso:	11
10. Análise Exploratória:	12
A. Descrição das Variáveis:	12
E. Desafios e Limitações:	20
F. Recomendações Preliminares:	21
G. Conclusão da Análise Exploratória:	21
H. Gráficos e Visualizações:	21
I. Gráficos Electric_Vehicle_Population_Data	22
J. Gráficos Pontos de Carregamento	26
11. Pipeline de Dados:	28
12. Proposta Analítica:	29
B. Apresentação do DATA.GOV:	29
13. Medidas de Acurácia	31
14. Análise e descrição dos resultados preliminares para um possível modelo de negócios.	36
Gráfico 06: Quantidade de veículos do tipo PHEV por Estado	37
Gráfico 07: 10 Cidades no Estado de WA com mais veículos elétricos	37
Gráfico 08: Quantidade de veículos elétricos por ano	38
Gráfico 09: Autonomia de Veículos BEV	39



Gráfico 10: Autonomia de Veículos PHEV	39
Gráfico 11: Quantidade de veículos elétricos por ano	40
C. Modelo de negócio:	41
15. Storytelling:	43
15.1. Apresentação do Grupo	43
15.2. Nome do Projeto	43
15.3. Importância do Tema	43
15.4. Desafios do Setor	43
15.5. Abordagem Analítica	43
15.6. Dados Disponíveis	44
15.7. Análise Exploratória	44
15.8. Resultados Pretendidos	44
16. Artefatos do Projeto:	45
A. Link Github:	45
B. Link Projeto:	45
C. Link Artefatos do Projeto:	45

Tabelas

Tabela 01 Cronograma	3
-----------------------------------	---



1. Título:

Empresa: Green Energy

Core Business: Infraestrutura de Carregamento para Automóveis Elétricos.

2. Introdução

Os veículos elétricos emergiram como uma resposta promissora aos desafios ambientais e sociais globais. Sua importância reside não apenas na redução das emissões de gases de efeito estufa e na melhoria da qualidade do ar nas áreas urbanas, mas também na diminuição da dependência de combustíveis fósseis e na diversificação das fontes de energia. Além disso, os veículos elétricos apresentam uma oportunidade única de impulsionar a transição para uma economia mais sustentável, criando novos empregos na indústria de energia limpa e estimulando a inovação tecnológica.

No entanto, a autonomia dos veículos elétricos ainda é um fator crucial para sua aceitação em massa. Embora os avanços na tecnologia tenham estendido significativamente a autonomia dos veículos elétricos nos últimos anos, ainda existe uma necessidade de expandir as redes de abastecimento para garantir uma experiência de condução conveniente e livre de preocupações para os seus proprietários. Isso implica investimentos contínuos em infraestrutura de carregamento, incluindo estações de carregamento rápido em áreas urbanas e rodovias, bem como soluções para carregamento em domicílio ou em locais de trabalho.

Uma rede robusta de abastecimento não só aumenta a confiança do consumidor na adoção de veículos elétricos, mas também desempenha um papel fundamental na redução das emissões de gases de efeito estufa e na promoção de uma mobilidade sustentável em todo o mundo.

3. Objetivos:

Este estudo tem como foco a análise e otimização da infraestrutura de carregamento de veículos elétricos no estado de WA (Washington) dos Estados Unidos. O objetivo principal é compreender a distribuição atual e as necessidades futuras dessa infraestrutura para suportar eficientemente o crescimento contínuo do mercado de veículos elétricos. Especificamente, buscaremos:

- Avaliar a distribuição geográfica da infraestrutura de carregamento de veículos elétricos no estado de WA.



- Identificar áreas com alta demanda de carregamento de veículos elétricos e baixa disponibilidade de estações de recarga.
- Identificar quais as Marcas de veículos elétricos mais vendidas.
- Analisar a evolução temporal da infraestrutura de carregamento em WA.
- Identificar padrões de crescimento e lacunas na infraestrutura de carregamento em relação ao aumento do número de veículos elétricos emplacados.
- Propor recomendações para otimizar a expansão da infraestrutura com base nas análises realizadas.

4. Metas:

Buscamos representar através da escolha das metas um plano estruturado para analisar e entender o desenvolvimento do mercado de veículos elétricos (BEVS) nos Estados Unidos, um setor em rápido crescimento e de grande importância para as estratégias de sustentabilidade e inovação tecnológica. As metas foram cuidadosamente selecionadas para abranger aspectos cruciais da dinâmica do mercado de BEVS, desde a aquisição de veículos até a infraestrutura de carregamento.

- Coletar e integrar dados de emplacamento de veículos elétricos em WA nos EUA de fontes confiáveis.
- Desenvolver uma metodologia robusta para analisar a distribuição e evolução da infraestrutura de carregamento.
- Realizar análises geoespaciais para mapear a cobertura atual e identificar lacunas na infraestrutura de carregamento.
- Utilizar técnicas de visualização de dados para comunicar eficazmente os resultados da análise.
- Produzir um relatório final com insights acionáveis e recomendações para stakeholders relevantes.



5. Cronograma:

Link:

<https://github.com/meddavid/Mackenzie-Projeto-Aplicado-II/blob/9a2bdc2681d80fe877677c40268520b7b067057d/01.%20ENTREGA%20ETAPA%2001/CRONOGRAMA%20-%20Projeto%20Aplicado%20II.xlsx>

EDT	Nome da tarefa	Duração	Início	Término	Nomes dos recursos	IIRR	TTRR
1	PROJETO - Green Energy	58 dias	Qui 08/02/24	Sáb 27/04/24		Qui 08/02/24	ND
1.1	FASE I – PREPARAÇÃO DE DATASET	19 dias	Qui 08/02/24	Ter 05/03/24		Qui 08/02/24	ND
1.1.1	Artefatos do projeto	1 dia	Qui 08/02/24	Qui 08/02/24		Qui 08/02/24	Qui 08/02/24
1.1.1.1	Link GITHUB	1 dia	Qui 08/02/24	Qui 08/02/24	Eduardo[50%]	Qui 08/02/24	Qui 08/02/24
1.1.1.2	Dataset	1 dia	Qui 08/02/24	Qui 08/02/24	Eduardo[50%]	Qui 08/02/24	Qui 08/02/24
1.1.2	Contexto do Estudo	1 dia	Sex 09/02/24	Sex 09/02/24		Sex 09/02/24	ND
1.1.2.1	Premissas do Projeto (definição de empresa, Core Business)	1 dia	Sex 09/02/24	Sex 09/02/24	Natalia	Sex 09/02/24	Sex 09/02/24
1.1.2.2	Objetivos	1 dia	Sex 09/02/24	Sex 09/02/24	Felipe	Sex 09/02/24	Sex 09/02/24
1.1.2.3	Metas	1 dia	Sex 09/02/24	Sex 09/02/24	Felipe	Sex 09/02/24	Sex 09/02/24
1.1.2.4	Cronograma de Atividades	1 dia	Sex 09/02/24	Sex 09/02/24	Eduardo[50%]	Sex 09/02/24	Sex 09/02/24
1.1.2.5	Pensamento Computacional em contextos organizacionais	1 dia	Sex 09/02/24	Sex 09/02/24	Eduardo[50%];Felipe;Natalia	Sex 09/02/24	ND
1.1.3	Referências de Aquisição do Dataset	1 dia	Qua 21/02/24	Qua 21/02/24		Qua 21/02/24	Qua 21/02/24
1.1.3.1	Origem dos Dados	1 dia	Qua 21/02/24	Qua 21/02/24	Eduardo;Felipe;Natalia	Qua 21/02/24	Qua 21/02/24
1.1.3.2	Limitação de Uso	1 dia	Qua 21/02/24	Qua 21/02/24	Eduardo;Felipe;Natalia	Qua 21/02/24	Qua 21/02/24
1.1.3.3	Período de Coleta	1 dia	Qua 21/02/24	Qua 21/02/24	Eduardo;Felipe;Natalia	Qua 21/02/24	Qua 21/02/24
1.1.4	Descrição da Origem	1 dia	Seg 04/03/24	Seg 04/03/24		Seg 04/03/24	Seg 04/03/24
1.1.4.1	Informações sobre a Organização que gerou os dados	1 dia	Seg 04/03/24	Seg 04/03/24	Eduardo;Felipe;Natalia	Seg 04/03/24	Seg 04/03/24
1.1.4.2	Contexto em que os dados foram gerados	1 dia	Seg 04/03/24	Seg 04/03/24	Eduardo;Felipe;Natalia	Seg 04/03/24	Seg 04/03/24
1.1.5	Descrição do Dataset	1 dia	Ter 05/03/24	Ter 05/03/24		Ter 05/03/24	Ter 05/03/24
1.1.5.1	Conteúdo	1 dia	Ter 05/03/24	Ter 05/03/24	Eduardo;Felipe;Natalia	Ter 05/03/24	Ter 05/03/24
1.1.5.2	Proposta	1 dia	Ter 05/03/24	Ter 05/03/24	Eduardo;Felipe;Natalia	Ter 05/03/24	Ter 05/03/24
1.1.5.3	Registro de Problemas ou Fenômenos	1 dia	Ter 05/03/24	Ter 05/03/24	Eduardo;Felipe;Natalia	Ter 05/03/24	Ter 05/03/24
1.2	Entrega Moodle - Etapa I	9 dias	Seg 11/03/24	Qui 21/03/24	Eduardo	Seg 11/03/24	ND
1.2.1	Definição da linguagem de programação usada no projeto.	1 dia	Seg 11/03/24	Seg 11/03/24	Eduardo;Felipe;Natalia;Ana	Seg 11/03/24	Seg 11/03/24
1.2.2	Análise exploratória da base de dados escolhida.	1 dia	Ter 12/03/24	Ter 12/03/24	Eduardo;Felipe;Natalia;Ana	Ter 12/03/24	Ter 12/03/24
1.2.3	Tratamento da base de dados (Preparação e treinamento).	1 dia	Ter 12/03/24	Ter 12/03/24	Eduardo;Felipe;Natalia;Ana	Ter 12/03/24	Ter 12/03/24
1.2.4	Definição e descrição das bases teóricas dos métodos.	1 dia	Ter 19/03/24	Ter 19/03/24	Eduardo;Felipe;Natalia;Ana	Ter 19/03/24	Ter 19/03/24
1.2.5	Definição e descrição de como será calculada a acurácia.	1 dia	Qua 20/03/24	Qua 20/03/24	Eduardo;Felipe;Natalia;Ana	Qua 20/03/24	Qua 20/03/24
1.2.6	BB Professor	1 dia	Qui 21/03/24	Qui 21/03/24	Eduardo;Felipe;Natalia;Ana	Qui 21/03/24	Qui 21/03/24
1.3	Entrega Moodle - Etapa II	1 dia	Sáb 27/04/24	Sáb 27/04/24	Eduardo	Sáb 27/04/24	Sáb 27/04/24
1.3.1	Método analítico definido na etapa anterior aplicado à base de dados.	1 dia	Sáb 27/04/24	Sáb 27/04/24	Eduardo;Felipe;Natalia;Ana	Sáb 27/04/24	Sáb 27/04/24
1.3.2	Medidas de acurácia, usando os métodos definidos na etapa anterior.	1 dia	Sáb 27/04/24	Sáb 27/04/24	Eduardo;Felipe;Natalia;Ana	Sáb 27/04/24	Sáb 27/04/24
1.3.3	Descrição dos resultados preliminares	1 dia	Sáb 27/04/24	Sáb 27/04/24	Eduardo;Felipe;Natalia;Ana	Sáb 27/04/24	Sáb 27/04/24
1.3.4	Apresentando um produto gerado	1 dia	Sáb 27/04/24	Sáb 27/04/24	Eduardo;Felipe;Natalia;Ana	Sáb 27/04/24	Sáb 27/04/24
1.3.5	Rascunhando um possível modelo de negócios.	1 dia	Sáb 27/04/24	Sáb 27/04/24	Eduardo;Felipe;Natalia;Ana	Sáb 27/04/24	Sáb 27/04/24
1.3.6	Esboço do storytelling.	1 dia	Sáb 27/04/24	Sáb 27/04/24	Eduardo;Felipe;Natalia;Ana	Sáb 27/04/24	Sáb 27/04/24
1.4	Entrega Moodle - Etapa III	25 dias	Qui 08/02/24	Qua 13/03/24	Eduardo	ND	ND
1.4.1	FASE IV - TBD	25 dias	Qui 08/02/24	Qua 13/03/24	Eduardo;Felipe;Natalia;Ana	ND	ND
1.4.1.1	BB Professor	1 dia	Qui 08/02/24	Qui 08/02/24	Eduardo;Felipe;Natalia;Ana	ND	ND
1.4.1.2	Atividades - TBD	1 dia	Qui 08/02/24	Qui 08/02/24	Eduardo;Felipe;Natalia;Ana	ND	ND
1.4.1.3	Entrega Moodle - Etapa IV	1 dia	Qua 13/03/24	Qua 13/03/24	Eduardo	ND	ND

Tabela 1



6. Fluxo Baseado em Pensamento Computacional em Contextos Organizacionais:

A. Decomposição:

Dividir o problema em partes menores: quantidade de marcas que produzem e vendem veículos elétricos, quantidade de veículos em circulação em Washington.

B. Reconhecimento de padrões:

Analisar relação entre marcas, comparar o consumo entre diferentes regiões, entre outros.

C. Abstração:

Construir uma análise exploratória sobre veículos: Utilizar dados e pesquisas atuais para criar análise.

D. Design de Algoritmos:

Criar um relatório para tomada de decisão: Com base nas análises, produzir relatório com recomendações para organizações.



7. Organização e o contexto em que os dados foram gerados:

Data.gov é o portal de dados abertos do governo federal dos Estados Unidos, que visa promover um governo mais transparente e responsável. Criado para aumentar a participação cidadã na gestão governamental, o site facilita o desenvolvimento econômico e apoia a tomada de decisões informadas tanto no setor privado quanto no público.

O portal implementa a Lei de Dados Governamentais OPEN (Título II da Lei de Fundamentos para a Elaboração de Políticas Baseadas em Evidências de 2018, Lei Pública 115-435). Esta lei transforma o Data.gov de uma política em um mandato legal, exigindo que agências federais disponibilizem suas informações online como dados abertos. Isso deve ser feito em formatos padronizados e legíveis por máquina, e incluir os metadados no catálogo Data.gov.

Data.gov colabora com várias agências federais para expandir a inclusão de seus datasets no portal conforme a implementação da lei prossegue. Adicionalmente, a lei determina que a Administração de Serviços Gerais (GSA) trabalhe conjuntamente com o Escritório de Gestão e Orçamento e o Escritório de Serviços de Informações Governamentais para criar um repositório online. Este repositório, disponível em resources.data.gov, oferece ferramentas, melhores práticas e padrões para promover a adoção de práticas de dados abertos em todo o governo federal.

Construído com software de código aberto, Data.gov permite que qualquer entidade, especialmente governos locais, estaduais e internacionais, utilize e adapte o código-fonte do site para suas próprias necessidades.

No contexto mais amplo, diversos estados, cidades e condados nos Estados Unidos têm lançado seus próprios portais de dados abertos. Ao integrar essas fontes de dados não federais, Data.gov enriquece seu catálogo, permitindo que pesquisas retornem datasets relevantes tanto de fontes federais quanto não federais.



8. Referências de aquisição do dataset:

Os dados têm origem no site oficial do governo dos Estados Unidos que apresenta um conjunto de dados que mostra os Veículos Elétricos de Bateria (BEVs) e os Veículos Elétricos Híbridos Plug-in (PHEVs) que estão atualmente registrados através do Departamento de Licenciamento (DOL) do Estado de Washington.

Este conjunto de dados destina-se ao acesso e uso público e foi atualizado em 17 de fevereiro de 2024.

9. Dataset e Metadados

A. Dataset:

Fonte: <https://catalog.data.gov/dataset/electric-vehicle-population-data>

Link para download: <https://github.com/meddavid/Mackenzie-Projeto-Aplicado-II/blob/eb68ff247f5a016aaa87133c838bd3a96dd27567/ARTEFATOS/DataSets.rar>

B. Descrição do Dataset:

Estamos utilizando dois conjuntos de dados, intitulados "Electric_Vehicle_Population_Data" e "Pontos_carregamento". Eles oferecem uma visão abrangente sobre a população de veículos elétricos e a infraestrutura de carregamento disponível. Vamos detalhar o que cada parte deste conjunto de dados representa e como os utilizaremos para análises diversas:

Identificação e Detalhes do Veículo: Cada entrada no conjunto de dados "Electric_Vehicle_Population_Data" começa com um Número de Identificação do Veículo (VIN), seguido por informações essenciais como marca, modelo, ano do modelo, e o tipo de veículo elétrico. Os tipos de veículos elétricos são categorizados principalmente como "Battery Electric Vehicle (BEV)" ou "Plug-in Hybrid Electric Vehicle (PHEV)", indicando se são totalmente elétricos ou híbridos. Esta seção é crucial para entender a variedade e popularidade de diferentes veículos elétricos no mercado.

Localização e Demografia: A localização geográfica é um aspecto fundamental deste conjunto de dados. Inclui detalhes como condado, cidade, estado e código postal. Além disso, há coordenadas geográficas precisas para cada veículo. Isso possibilita uma análise regional da



adoção de veículos elétricos, revelando padrões geográficos e potenciais lacunas na infraestrutura de suporte.

Elegibilidade Ambiental e Alcance Elétrico: Uma característica interessante é a indicação de se um veículo é classificado como um "Clean Alternative Fuel Vehicle (CAFV)" e seu alcance elétrico. Isso reflete a eficiência e o impacto ambiental dos veículos, essenciais para avaliar o progresso em direção a objetivos de sustentabilidade.

Aspectos Econômicos: O conjunto de dados inclui o preço base (MSRP) de cada veículo, embora muitos registros mostrem valores zerados, o que pode limitar análises econômicas.

Fornecedor de Energia: Cada entrada lista a companhia de energia elétrica associada ao veículo. Essa informação é valiosa para entender a relação entre a infraestrutura de energia e a adoção de veículos elétricos.

Dados do Censo: A inclusão de códigos do censo de 2020 abre possibilidades para análises demográficas detalhadas em relação à propriedade de veículos elétricos.

Pontos de Carregamento: O conjunto de dados "Pontos_carregamento" fornece localizações específicas de infraestrutura de carregamento através das coordenadas de latitude e longitude. Esta informação é crucial para mapear a acessibilidade ao carregamento e planejar melhorias na infraestrutura, facilitando a transição para a mobilidade elétrica. Metadados:

O conjunto de dados "**Electric_Vehicle_Population_Data**" carregado em Python como infra descrito, possui a seguinte estrutura e informações estatísticas:

Total de Entradas: 173.533

Total de Colunas: 17

Tipos de Dados: Objeto (strings), float64 e int64

O arquivo "**Pontos_carregamento.csv**" contém informações sobre pontos de carregamento para veículos elétricos e apresenta os seguintes detalhes estatísticos para as colunas de Latitude e Longitude:

Total de Entradas: 58.131

Total de Colunas: 2

Tipos de Dados: float64



C. Tipo do Arquivo:

Os arquivos estão no formato CSV.

D. Sensibilidade:

Os arquivos não possuem dados sensíveis e está compatível com a LGPD.

E. Proprietário do Dado:

Os arquivos são de domínio público e foi disponibilizado pelo DATA.GOV.

F. Restrições de uso:

O DATA.GOV não restringiu sua utilização.



10. Análise Exploratória:

A etapa de análise exploratória desempenha um papel fundamental na compreensão aprofundada dos datasets disponíveis. Esta seção visa enriquecer o conhecimento sobre as variáveis contidas nos conjuntos de dados, oferecendo uma perspectiva detalhada que complementa as informações já fornecidas nos metadados iniciais..

A. Descrição das Variáveis:

Nesta parte, serão apresentados detalhes abrangentes sobre as variáveis presentes nos datasets. Isso incluirá informações como o número total de registros para cada variável, os valores máximo e mínimo, a variância e o desvio padrão. Além disso, serão exploradas as distribuições das variáveis, destacando se elas seguem padrões conhecidos, como uma distribuição normal ou binomial. Também será avaliada a presença de dados ausentes e a identificação de valores discrepantes (outliers) que possam influenciar análises subsequentes.

01. Arquivo: Não aplicado

Caminho: Não aplicado

Desenvolvimento:

- `pip install pandas`
- `python.exe -m pip install --upgrade pip`
- `pip install matplotlib`
- `pip install seaborn`

Retorno:

- Successfully installed numpy-1.26.0 pandas-2.1.1 python-dateutil-2.8.2 pytz-2023.3.post1 six-1.16.0 tzdata-2023.3
- Installing collected packages: pyparsing, pillow, packaging, kiwisolver, fonttools, cycler, contourpy, matplotlib
- Successfully installed pip-23.2.1



02. Arquivo: summary.py

Caminho: .git\Projeto_aplicado_II\99. Artefatos\

Desenvolvimento:

```
'''
=====
Programa.....: SUMMARY
Autor.....: Eduardo David
Data.....: 01/03/2024
Descrição / Objetivo: Exibição de Análise de Metadados
Doc. Origem.....: Electric_Vehicle_Population_Data.csv
Solicitante.....: Professor Felipe Cunha
Uso.....: Projeto Apliado II
Modificações.....: 01/03/2024 - Desenvolvimento
=====
'''

import pandas as pd

# Caminho atualizado do arquivo CSV
file_path = "N:\\Drives compartilhados\\.....EDUARDO\\...
MACKENZIE\\...TERCEIRO SEMESTRE\\PROJETO-APLICADO-II\\01.
ENTREGA ETAPA 01\\Electric_Vehicle_Population_Data.csv"

# Carregando o arquivo CSV
df = pd.read_csv(file_path)

# Exibindo as primeiras linhas para uma visão geral
first_rows = df.head()

# Resumo da estrutura do conjunto de dados
structure = df.info()_
```

Retorno:

```
CKENZIE/projeto_aplicadoII/summary.py
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 173533 entries, 0 to 173532
Data columns (total 17 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   VIN (1-10)                               173533 non-null object
1   County                                   173528 non-null object
2   City                                    173528 non-null object
3   State                                   173533 non-null object
4   Postal Code                             173528 non-null float64
5   Model Year                             173533 non-null int64
6   Make                                    173533 non-null object
7   Model                                   173533 non-null object
8   Electric Vehicle Type                   173533 non-null object
9   Clean Alternative Fuel Vehicle (CAFV) Eligibility 173533 non-null object
10  Electric Range                           173532 non-null float64
11  Base MSRP                               173532 non-null float64
12  Legislative District                    173157 non-null float64
13  DOL Vehicle ID                         173533 non-null int64
14  Vehicle Location                       173523 non-null object
15  Electric Utility                       173528 non-null object
16  2020 Census Tract                      173528 non-null float64
dtypes: float64(5), int64(2), object(10)
memory usage: 22.5+ MB
```



03. Arquivo: summaryII.py

Caminho: .git\Projeto_aplicado_II\99. Artefatos\

Desenvolvimento:

```
'''
=====
Programa.....: SUMMARYII
Autor.....: Eduardo David
Data.....: 15/04/2024
Descrição / Objetivo: Exibição de Análise de Metadados
Doc. Origem.....: pontos_carregamento.csv
Solicitante.....: Professor Felipe Cunha
Uso.....: Projeto Apliado II
Modificações.....: 15/04/2024 - Desenvolvimento
=====
'''

import pandas as pd

# Caminho atualizado do arquivo CSV
file_path = "file_path = \"C:\\\\TEMP\\\\pontos_carregamento.csv\""

# Carregando o arquivo CSV
df = pd.read_csv(file_path)

# Exibindo as primeiras linhas para uma visão geral
first_rows = df.head()

# Resumo da estrutura do conjunto de dados
structure = df.info()_
```

Retorno:

```
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Latitude    58131 non-null    float64
1   Longitude    58131 non-null    float64
dtypes: float64(2)
```



04. Arquivos: Analytics.py

Desenvolvimento:

```
'''
=====
Programa.....: Analitics
Autor.....: Eduardo David
Data.....: 01/03/2024
Descrição / Objetivo: Exibição de Análise de Metadados
Doc. Origem.....: Electric_Vehicle_Population_Data.csv
Solicitante.....: Professor Felipe Cunha
Uso.....: Projeto Aplicado II
Modificações.....: 01/03/2024 - Desenvolvimento
=====
'''

import pandas as pd

# Caminho do arquivo CSV
file_path = "C:\\\\TEMP\\Electric_Vehicle_Population_Data.csv"

# Carregando o arquivo CSV
df = pd.read_csv(file_path)

# Exibindo as primeiras linhas para uma visão geral
print("Primeiras linhas do conjunto de dados:")
print(df.head(), "\n")

# Resumo da estrutura do conjunto de dados
print("Resumo da estrutura do conjunto de dados:")
df.info()

# Resumo estatístico básico das colunas numéricas
print("\nResumo estatístico básico das colunas numéricas:")
colunas_numericas = df.select_dtypes(include=[float,
int]).columns.tolist() # Automatizando a escolha de colunas numéricas
print(df[colunas_numericas].describe(), "\n")

# Analisando a quantidade de registros
num_registros = len(df)
print(f"#### - Quantidade de registros: {num_registros}\n")

# Analisando a quantidade de registros agrupados por uma categoria relevante
categoria = 'Make'
# Supondo que 'Make' seja uma coluna de categorização relevante
grupos = df.groupby(categoria)
num_registros_por_grupo = grupos.size()
```




```
print(f"#### - Quantidade de registros agrupados por {categoria}:")
print(num_registros_por_grupo)
print("\nGrupos formados:")
print(list(grupos.groups.keys()))
```

Retorno:

- Quantidade de registros: 173533
- Quantidade de registros agrupados por WHO_REGION:

```
RangeIndex: 173533 entries, 0 to 173532
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   VIN (1-10)                           173533 non-null object
1   County                               173528 non-null object
2   City                                 173528 non-null object
3   State                               173533 non-null object
4   Postal Code                          173528 non-null float64
5   Model Year                          173533 non-null int64
6   Make                                173533 non-null object
7   Model                               173533 non-null object
8   Electric Vehicle Type                173533 non-null object
9   Clean Alternative Fuel Vehicle (CAFV) Eligibility 173533 non-null object
10  Electric Range                       173532 non-null float64
11  Base MSRP                           173532 non-null float64
12  Legislative District                 173157 non-null float64
13  DOL Vehicle ID                     173533 non-null int64
14  Vehicle Location                    173523 non-null object
15  Electric Utility                    173528 non-null object
16  2020 Census Tract                   173528 non-null float64
dtypes: float64(5), int64(2), object(10)
memory usage: 22.5+ MB
```

- Grupos formados:

['ALFA ROMEO', 'AUDI', 'AZURE DYNAMICS', 'BENTLEY', 'BMW', 'CADILLAC', 'CHEVROLET', 'CHRYSLER', 'DODGE', 'FIAT', 'FISKER', 'FORD', 'GENESIS', 'GMC', 'HONDA', 'HYUNDAI', 'JAGUAR', 'JEEP', 'KIA', 'LAND ROVER', 'LEXUS', 'LINCOLN', 'LUCID', 'MAZDA', 'MERCEDES-BENZ', 'MINI', 'MITSUBISHI', 'NISSAN', 'POLESTAR', 'PORSCHÉ', 'RIVIAN', 'ROLLS ROYCE', 'SMART', 'SUBARU', 'TESLA', 'THINK', 'TOYOTA', 'VOLKSWAGEN', 'VOLVO', 'WHEEGO ELECTRIC CARS']



05. Arquivos: AnalyticsII.py

Desenvolvimento:

```
'''
=====
Programa.....: AnaliticsII
Autor.....: Eduardo David
Data.....: 15/04/2024
Descrição / Objetivo: Exibição de Análise de Metadados
Doc. Origem.....: pontos_carregamento.csv
Solicitante.....: Professor Felipe Cunha
Uso.....: Projeto Aplicado II
Modificações.....: 15/04/2024- Desenvolvimento
=====
'''

import pandas as pd

# Caminho do arquivo CSV
file_path = "C:\\TEMP\\pontos_carregamento.csv"

# Carregando o arquivo CSV
df = pd.read_csv(file_path)

# Exibindo as primeiras linhas para uma visão geral
print("Primeiras linhas do conjunto de dados:")
print(df.head(), "\n")

# Resumo da estrutura do conjunto de dados
print("Resumo da estrutura do conjunto de dados:")
df.info()

# Resumo estatístico básico das colunas numéricas
print("\nResumo estatístico básico das colunas numéricas:")
colunas_numericas = df.select_dtypes(include=[float,
int]).columns.tolist() # Automatizando a escolha de colunas numéricas
print(df[colunas_numericas].describe(), "\n")

# Analisando a quantidade de registros
num_registros = len(df)
print(f"#### - Quantidade de registros: {num_registros}\n")
```

Retorno:



Resumo estatístico básico das colunas numéricas:

	Latitude	Longitude
count	58131.000000	58131.000000
mean	37.791420	-97.020307
std	5.001930	19.561345
min	12.921876	-164.848855
25%	34.036357	-118.094319
50%	38.445383	-93.396080
75%	41.362635	-79.001620
max	64.852466	98.493283

B. Análise de Pontos de Recarga:

São realizadas operações de filtragem e limpeza em dados relacionados a pontos de recarga de veículos elétricos.

```
# Crie um novo DataFrame contendo apenas as linhas onde a coluna 'State' é igual a 'WA'
colunas_selecionadas_public_wa = colunas_selecionadas_public[colunas_selecionadas_public['State'] == 'WA']

# Crie uma cópia do DataFrame original para manter as outras colunas inalteradas
df_filtered = colunas_selecionadas_public_wa.copy()

# Selecione apenas as colunas específicas para verificar se todas estão preenchidas com NaN
columns_to_check = ['EV Level1 EVSE Num', 'EV Level2 EVSE Num', 'EV DC Fast Count']

# Remova as linhas onde todas as três colunas especificadas estão preenchidas com NaN
df_filtered = df_filtered.dropna(subset=columns_to_check, how='all')

df_filtered[['EV Level1 EVSE Num', 'EV Level2 EVSE Num', 'EV DC Fast Count']] = df_filtered[['EV Level1 EVSE Num', 'EV Level2 EVSE Num', 'EV DC Fast Count']].fillna(0)
```

Calcula-se a soma das instâncias de diferentes tipos de pontos de recarga e cria-se um gráfico de barras para visualizar essa soma.

```
# Selecione apenas as colunas desejadas para calcular a soma
colunas_selecionadas_soma = ['EV Level1 EVSE Num', 'EV Level2 EVSE Num', 'EV DC Fast Count']

# Calcule a soma das instâncias das colunas selecionadas
soma_instancias = df_filtered[colunas_selecionadas_soma].sum()

# Crie uma tabela com o resultado da soma
tabela_soma = pd.DataFrame({'Colunas': colunas_selecionadas_soma, 'Soma das Instâncias': soma_instancias})

# Crie o gráfico de barras coloridas com base nas somas das instâncias
plt.figure(figsize=(10, 6))
soma_instancias.plot(kind='bar', color=['blue', 'green', 'red'])
plt.title('Soma das Instâncias dos Pontos de Recarga por Capacidade')
plt.xlabel('Colunas')
plt.ylabel('Soma das Instâncias')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



C. Análise da Distribuição Temporal de Veículos:

Os dados são pré-processados e normalizados para análise temporal.

```
# Filtrar os dados para incluir apenas veículos desde 2015 e do tipo BEV
base = base_dados[(base_dados['Model Year'] >= 2015) & (base_dados['Electric Vehicle Type'])]

# Mapear os valores existentes para as novas classes
class_mapping = {'Plug-in Hybrid Electric Vehicle (PHEV)': '1', 'Battery Electric Vehicle (BEV)': '2'}
base['Electric Vehicle Type'] = base['Electric Vehicle Type'].map(class_mapping)

# Agrupar por ano e calcular a quantidade total de veículos por ano
base_agrupada = base.groupby(['Model Year', 'Electric Vehicle Type']).size().reset_index(name='Quantidade')

scaler = StandardScaler()

numerical_columns = ['Model Year', 'Electric Vehicle Type', 'Quantidade']
base_agrupada_normalizada = pd.DataFrame(
    scaler.fit_transform(base_agrupada[numerical_columns]),
    columns=numerical_columns
)

# Pré-processamento de dados
X = base_agrupada_normalizada[['Model Year', 'Electric Vehicle Type']]
y = base_agrupada_normalizada['Quantidade']

# Dividir dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1234)

# Definir os hiperparâmetros para o GridSearchCV
param_grid = {
    'n_estimators': [100, 500, 1000],
    'max_depth': [3, 5, 7],
    'min_samples_leaf': [1, 5, 10],
    'min_samples_split': [2, 5, 10]
}
```

Utilizando um modelo de regressão Random Forest, são feitas previsões da quantidade de veículos elétricos para anos futuros.



```
# Criar o modelo Random Forest
rf_model = RandomForestRegressor(random_state=1234)

# Criar o objeto GridSearchCV
grid_search = GridSearchCV(estimator=rf_model, param_grid=param_grid, cv=3, n_jobs=-1, verbose=2)

# Executar a busca em grade no conjunto de treinamento
grid_search.fit(X_train, y_train)

# Obter os melhores hiperparâmetros encontrados
best_params = grid_search.best_params_
print("Melhores hiperparâmetros:", best_params)

# Criar e treinar o modelo Random Forest com os melhores hiperparâmetros
rf_model_best = RandomForestRegressor(**best_params, random_state=1234)
rf_model_best.fit(X_train, y_train)

# Fazer previsões para os dados de teste
y_pred_rf = rf_model_best.predict(X_test)

# Calcular o RMSE do modelo Random Forest
rmse_rf = np.sqrt(mean_squared_error(y_test, y_pred_rf))
print("RMSE (Random Forest):", rmse_rf)
```

D. Visualização das Previsões:

Os resultados das previsões são acumulados e exibidos em um gráfico de linha para mostrar a tendência de crescimento da quantidade de veículos elétricos ao longo dos anos.

E. Desafios e Limitações:

É fundamental reconhecer e discutir quaisquer desafios e limitações encontrados durante a análise exploratória. Isso pode incluir a identificação de dados faltantes em grande quantidade, a presença de outliers que exigirão considerações adicionais e quaisquer variáveis que não sigam uma distribuição esperada. Essa seção abordará as complexidades encontradas e fornecerá um contexto importante para decisões posteriores no processo de análise.



F. Recomendações Preliminares:

Com base nas descobertas da análise exploratória, serão apresentadas recomendações preliminares. Isso pode envolver estratégias para lidar com dados faltantes, abordagens para tratar outliers e considerações sobre transformações de dados. As recomendações oferecerão diretrizes iniciais para orientar a próxima fase da análise, aproveitando os insights obtidos durante a exploração detalhada.

G. Conclusão da Análise Exploratória:

A análise exploratória é um estágio crítico para a compreensão aprofundada dos datasets, envolvendo a descrição minuciosa das variáveis, o uso de visualizações para destacar padrões, o reconhecimento de desafios e limitações, e a formulação de recomendações iniciais para abordar questões identificadas. Isso proporciona uma base sólida para análises subsequentes e tomadas de decisão informadas.

H. Gráficos e Visualizações:

Esta parte é dedicada à criação e apresentação de gráficos e visualizações que ajudarão na compreensão das características das variáveis. Isso inclui gráficos de barra, histogramas, box Plots e outras representações visuais adequadas à natureza dos dados.

Caminho: .git\Projeto_aplicado_I\99. Artefatos\03. GRAFICOS



I. Gráficos Electric_Vehicle_Population_Data

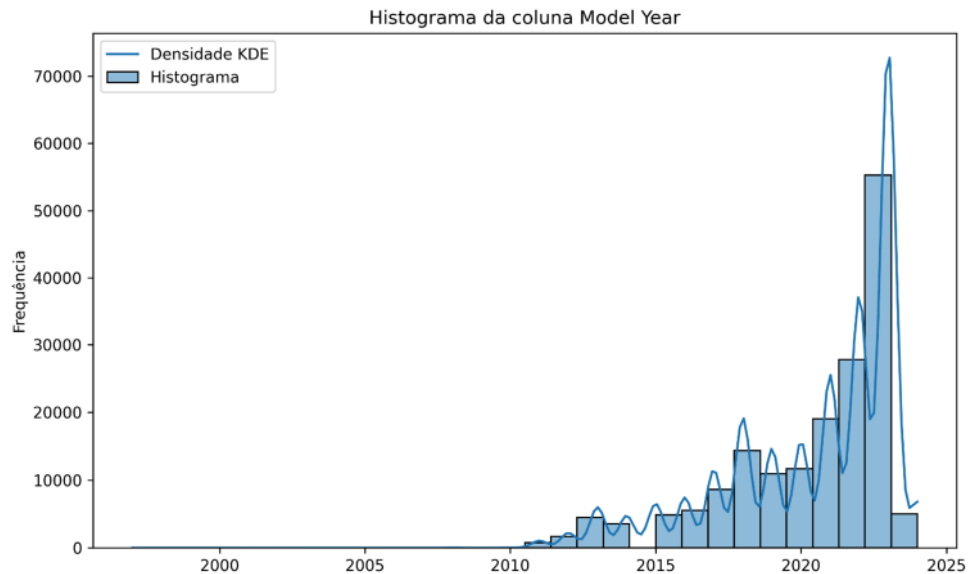
- **Histograma do Ano do Modelo**

Tipo do Gráfico: Histograma

Escala do Gráfico: Linear

Eixo Y: Frequência: Indica o número de veículos para cada ano de modelo.

Eixo X: Model Year Os anos de modelo dos veículos elétricos.



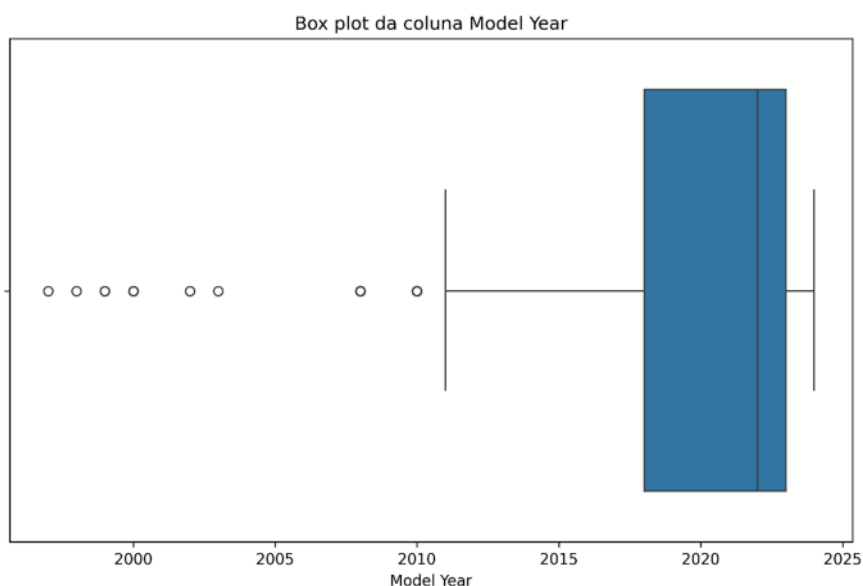
- **Gráfico 02: Boxplot do Ano do Modelo**

Tipo do Gráfico: Boxplot

Escala do Gráfico: Linear

Eixo Y: Model Year: Distribuição do ano do modelo dos veículos elétricos.

Eixo X: Não aplicável: Apenas uma variável, o ano do modelo.



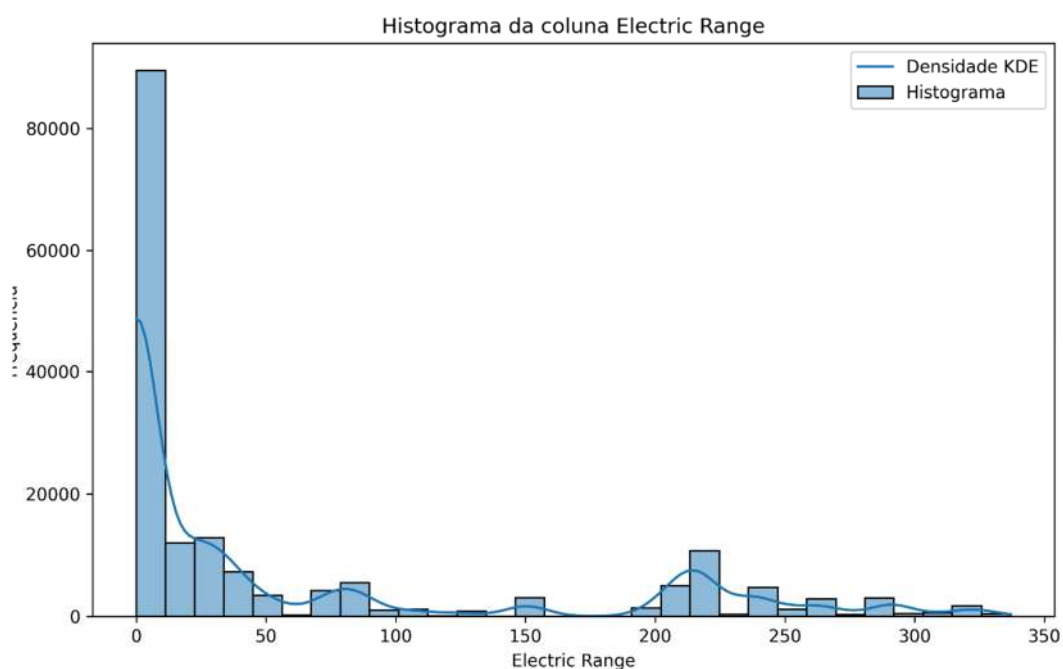
- **Gráfico 03: Histograma do Alcance Elétrico**

Tipo do Gráfico: Histograma

Escala do Gráfico: Linear

Eixo Y: Frequência: Número de veículos por alcance elétrico.

Eixo X: Electric Range: O alcance elétrico dos veículos em milhas.



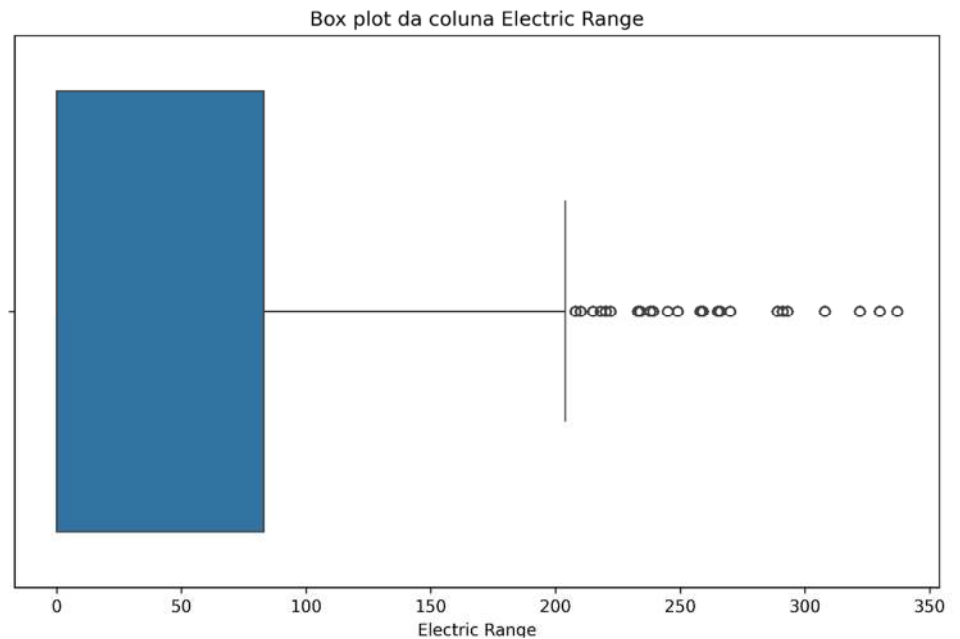
- **Gráfico 04: Boxplot do Alcance Elétrico**

Tipo do Gráfico: Boxplot

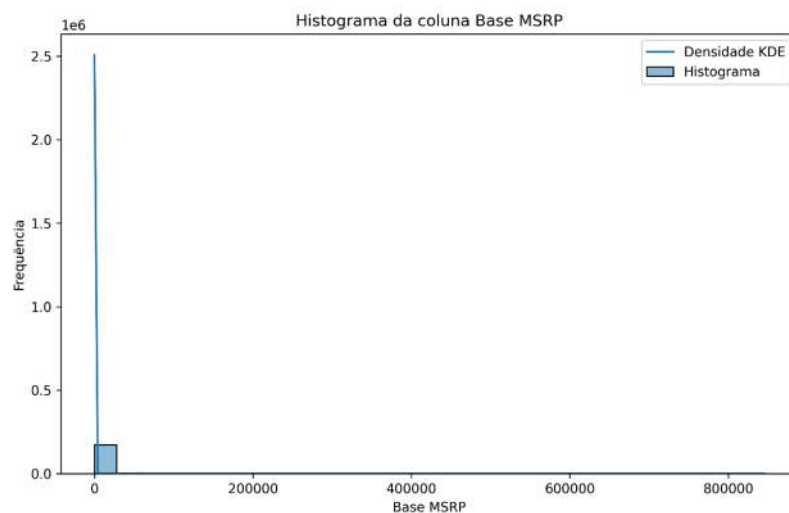
Escala do Gráfico: Linear

Eixo Y: Electric Range: Distribuição do alcance elétrico dos veículos.

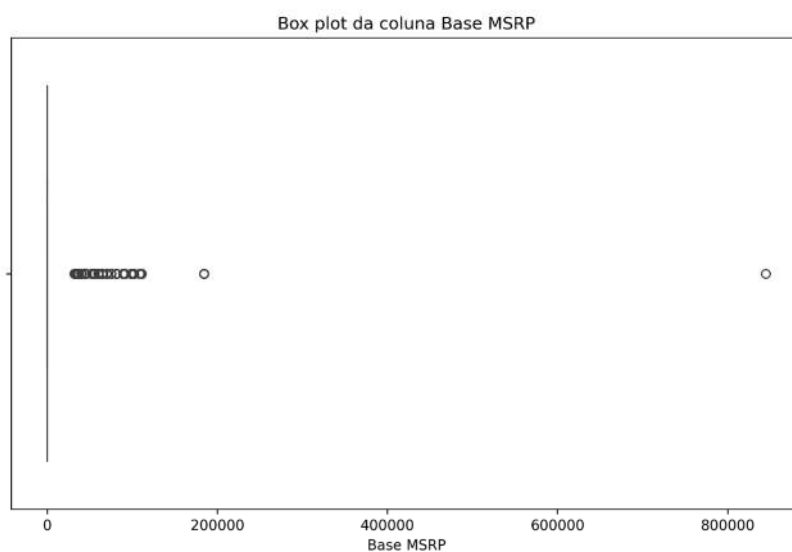
Eixo X: Não aplicável: Apenas uma variável, o alcance elétrico.



- **Gráfico 05: Histograma do Preço Sugerido pelo Fabricante (MSRP)**
Tipo do Gráfico: Histograma
Escala do Gráfico: Linear
Eixo Y: Frequência: Número de veículos para cada faixa de preço MSRP.
Eixo X: Base MSRP: O preço sugerido pelo fabricante.



- **Gráfico 06: Boxplot do Preço Sugerido pelo Fabricante (MSRP)**
Tipo do Gráfico: Boxplot
Escala do Gráfico: Linear
Eixo Y: Base MSRP: Distribuição dos preços sugeridos pelos fabricantes.
Eixo X: Não aplicável: Apenas uma variável, o preço MSRP.





J. Gráficos Pontos de Carregamento

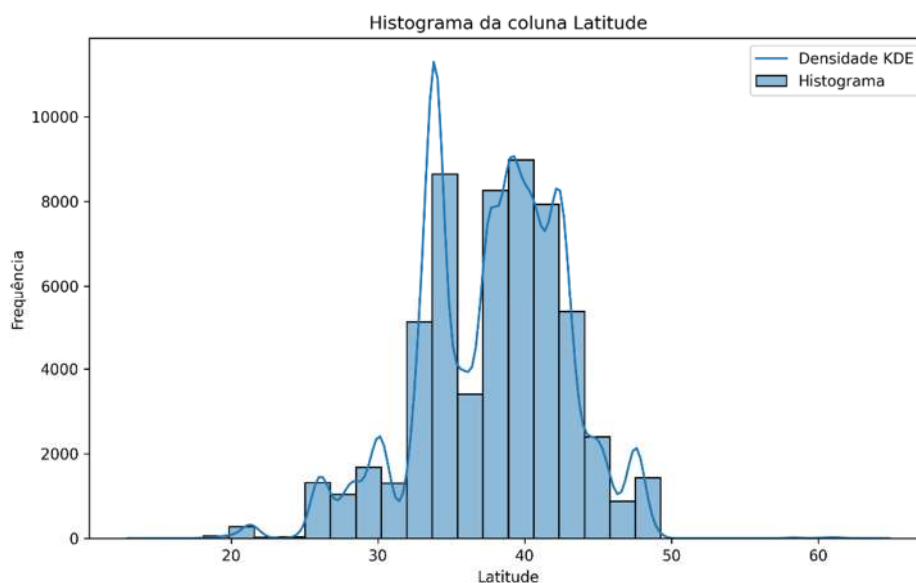
- **Gráfico 01: Histograma Latitude**

Tipo do Gráfico: Histograma

Escala do Gráfico: Linear

Eixo Y: Frequência: Representa o número de registros que caem dentro de cada intervalo de latitude.

Eixo X: Latitude: Mostra os intervalos de valores de latitude para os pontos de carregamento.



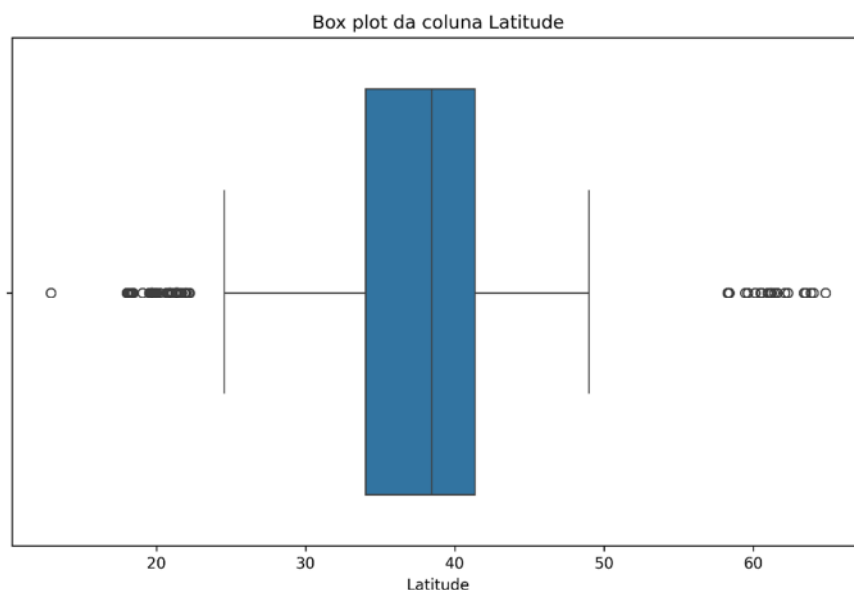
- **Gráfico 02: Boxplot Latitude**

Tipo do Gráfico: Boxplot

Escala do Gráfico: Linear

Eixo Y: Valores: Representa a distribuição de valores de latitude para os pontos de carregamento.

Eixo X: Não aplicável - Como é um boxplot de uma única variável, o eixo X serve apenas para rotular a variável analisada, que é a latitude.



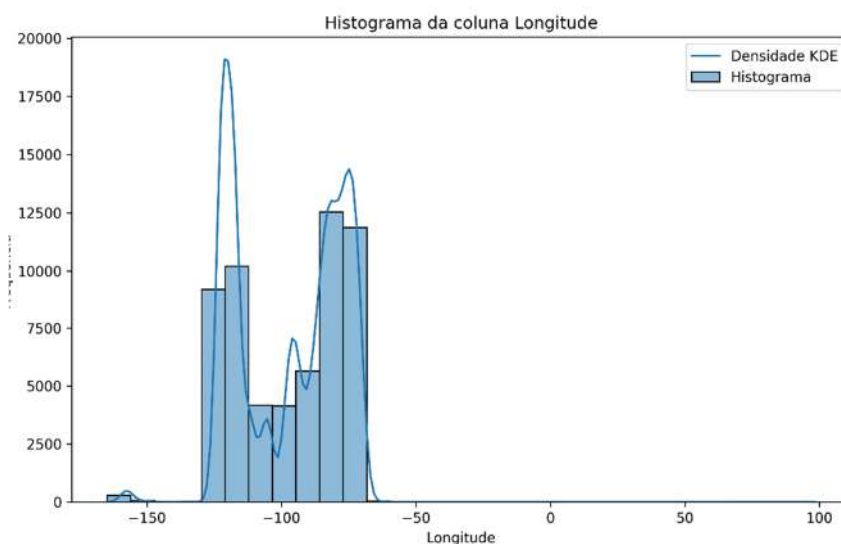
- **Gráfico 03: Histograma Longitude**

Tipo do Gráfico: Histograma

Escala do Gráfico: Linear

Eixo Y: Frequência - Indica o número de registros dentro de cada intervalo de longitude.

Eixo X: Longitude - Apresenta os intervalos de valores de longitude para os pontos de carregamento.



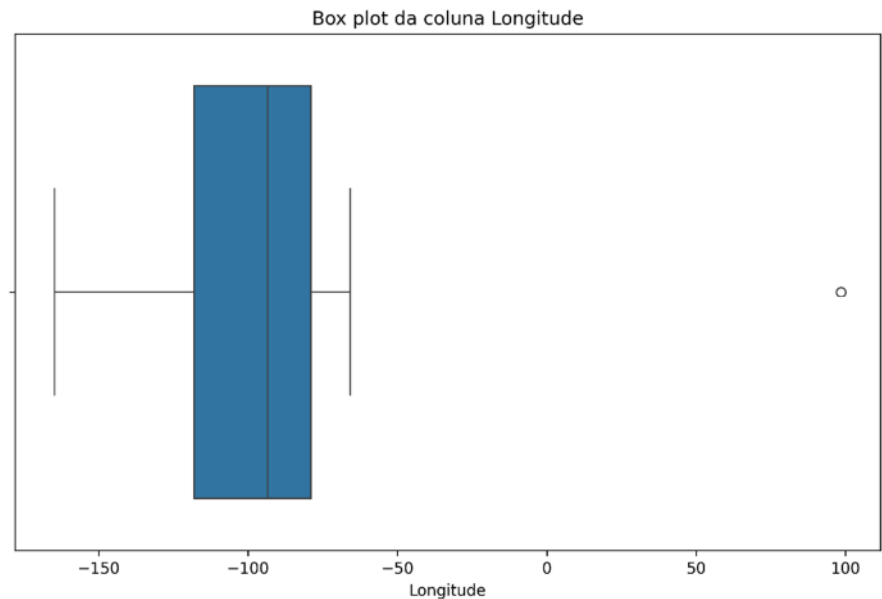
- **Gráfico 04: Boxplot Longitude**

Tipo do Gráfico: Boxplot

Escala do Gráfico: Linear

Eixo Y: Valores: Mostra a distribuição de valores de longitude para os pontos de carregamento.

Eixo X: Não aplicável: Semelhante ao boxplot da latitude, o eixo X serve para rotular a variável longitude.



Cada histograma dá uma visão da distribuição dos dados, permitindo identificar onde se concentram os pontos de carregamento. Os picos nos histogramas indicam as latitudes e longitudes mais comuns. Já os boxplots fornecem uma visão da dispersão dos dados, incluindo a mediana, os quartis e os possíveis outliers.

Código Fonte Gráficos:

Arquivo: Graphics_OF.py

Caminho: .git\Projeto_aplicado_I\99. Artefatos\02. scripts_python

Desenvolvimento:

11. Pipeline de Dados:

Coleta:

Processo de obtenção dos dados da OMS.

Limpeza:

Remoção de dados faltantes ou inconsistências;

Transformação de variáveis.

Análise Exploratória:

Estatísticas descritivas (médias, medianas, desvios padrão, etc.);

Visualizações gráficas (gráficos de barra, mapas de calor, etc.).

Técnicas de Análise:

Algoritmos e ferramentas a serem utilizados (e.g., Python, R);

Métodos de análise (regressão, clustering, análise de séries temporais).mapas de calo

Visualização:

Ferramentas e abordagens para visualizar os resultados (e.g., Tableau, Power BI, gráficos em Python/R).

Relatório Final:

Compilação dos resultados e recomendações em um formato acessível e compreensível.



12. Proposta Analítica:

Nesta proposta, vamos explorar o panorama dos veículos elétricos nos Estados Unidos, utilizando dados de registros e adoções desses veículos. Nosso principal objetivo é analisar como diferentes estados têm progredido na integração de veículos elétricos em suas frotas e identificar aqueles que ainda têm espaço para melhorias.

Utilizaremos técnicas básicas de análise para avaliar esses dados e criar representações visuais que tornem as informações mais acessíveis e fáceis de entender.

Esperamos que nosso estudo ofereça uma compreensão aprofundada sobre a origem, evolução e impacto dos veículos elétricos na sustentabilidade ambiental e na inovação tecnológica. Almejamos destacar as conquistas alcançadas e identificar áreas que necessitam de atenção e esforço adicional para promover efetivamente a transição energética.

B. Apresentação do DATA.GOV:

Breve histórico:

- Estabelecido como resposta à necessidade de transparência governamental, o portal Data.gov foi lançado em 2009. Desde sua criação, tem se consolidado como um ponto central para o acesso a dados governamentais nos Estados Unidos. Ampliando sua base de dados e capacidades continuamente, o portal lidera iniciativas para promover um governo aberto e responsável. Ao longo dos anos, Data.gov tem sido fundamental na promoção da inovação e no suporte à tomada de decisões baseadas em dados, estabelecendo padrões que facilitam o acesso público à informação governamental e incentivando a participação cidadã na gestão governamental.

Missão e objetivos:

- A missão fundamental do Data.gov é atuar como um facilitador de transparência e governança aberta em escala nacional, orientando e consolidando esforços das agências federais dos Estados Unidos para disponibilizar dados governamentais. Data.gov visa objetivos claros e impactantes: promover a acessibilidade e a reutilização de dados governamentais, fortalecer a participação cidadã e a inovação, e apoiar a tomada de decisões informadas tanto no setor público quanto no privado.



Importância global:

- Data.gov tem se estabelecido como um recurso fundamental para a transparência governamental, especialmente em momentos de crise, como durante a crise financeira e em resposta a emergências de saúde pública. O portal oferece acesso a dados cruciais que mobilizam recursos indispensáveis e informam a ação pública e privada. Ele estabelece parcerias sólidas com governos estaduais e locais, organizações internacionais e o setor privado, potencializando esforços conjuntos para promover práticas de dados abertos globalmente. Atualmente, sua relevância é amplificada na era digital, onde Data.gov desempenha um papel central no suporte à inovação e ao desenvolvimento sustentável através do acesso aberto e uso de dados governamentais.

Estrutura e funcionamento:

- Data.gov opera por meio de uma estrutura colaborativa, coordenando ações entre diversas agências federais nos Estados Unidos e mantendo uma plataforma centralizada através de seu portal online. Esta configuração assegura sua eficácia e abrangência em escala nacional. Os órgãos proeminentes dentro da gestão de Data.gov incluem a Administração de Serviços Gerais (GSA), o Escritório de Gestão e Orçamento (OMB) e o Escritório de Serviços de Informações Governamentais, todos essenciais na formulação de políticas e na promoção de práticas de dados abertos. Juntos, esses órgãos garantem que Data.gov mantenha um alinhamento consistente com seus objetivos de transparência e inovação governamental.

Iniciativas relativas aos Veículos Elétricos:

- Desde os primeiros sinais de reconhecimento da importância dos veículos elétricos para a sustentabilidade ambiental, organizações globais e governos têm sido fundamentais na definição de diretrizes e na alocação de recursos para promover esta tecnologia. Essas entidades têm sido essenciais não apenas na orientação, mas também na pesquisa e implementação de infraestruturas de carregamento. Projetos como incentivos fiscais e subsídios são cruciais para garantir uma adoção mais ampla dos veículos elétricos pelo mundo.



13. Medidas de Acurácia

A. Modelo de Regressão

Para calcular a acurácia de um modelo de regressão linear, iremos utilizar o coeficiente de determinação R^2 , que é a métrica comumente usada para avaliar o desempenho de modelos de regressão. O coeficiente varia de 0 a 1, onde 1 indica um ajuste perfeito do modelo aos dados. No código você pode encontra-lo na linha `model.score(X_test, y_test)`.

```
print(EV_DC_Fast_Count)
print(EV_DC_Fast_Count)
32.42251497005988

E analisar tendencias futuras usando graficos e estatisticas para analisar tendencias

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Carregar os dados
base_dados = pd.read_csv('D:/OneDrive - Instituto Presbiteriano Mackenzie/Mackenzie/Aulas/3 semestre/Projeto Aplicado II/Electric Vehicle Population Data.csv', index_col=False)

# Filtrar os dados para incluir apenas veiculos desde 2015 e do tipo BEV
base = base_dados[(base_dados['Model Year'] >= 2015) & (base_dados['Electric Vehicle Type'] == 'Battery Electric Vehicle (BEV)')]

# Agrupar por ano e calcular a quantidade total de veiculos por ano
base_agrupada = base.groupby('Model Year').size().reset(index=True, name='Quantidade')

# Pré-processamento de dados
X = base_agrupada['Model Year'].values.reshape(-1, 1)
y = base_agrupada['Quantidade'].values

# Dividir dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1234)

# Treinar o modelo
model = LinearRegression()
model.fit(X_train, y_train)

# Avaliar o modelo
print("R²:", model.score(X_test, y_test))
print("RMSE:", np.sqrt(mean_squared_error(y_test, model.predict(X_test))))

# Fazer previsões para um futuro "Model Year"
ano_futuro = 2030
previsao = model.predict([ano_futuro])

print(f"Previsão de vendas para {ano_futuro}: {previsao}")

# Criar um vetor de "Model Year" para as previsões
anos_futuro = np.arange(2023, 2030)
```



B. Método Analítico Estatístico Preditivo

Além da análise exploratória dos dados, realizamos uma análise estatística preditiva mais específica e nichada extraíndo uma amostra da população (base de dados completa) para analisar a autonomia de veículos do tipo PHEV nos período de 2018 a 2024.

```
#11. Extraíndo uma amostra da população para aplicar técnicas estatísticas.  
  
eletric_range_df = df[(df['Electric Vehicle Type'] == 'Plug-in Hybrid Electric Vehicle (PHEV)' &  
| (df['Electric Range'] != 0) & (df['Model Year'].isin(np.arange(2018,2024,1)))]
```

A princípio foram aplicados métodos analíticos como média, variância e desvio padrão

```
#11.1. Calculando média, variância e desvio padrão  
eletric_range_media_df = eletric_range_df.mean()  
eletric_range_variancia_df = eletric_range_df.var()  
desvio_padrao = np.sqrt(eletric_range_variancia_df)
```

Construímos uma função que realiza o fit dos dados para uma log normal e extrai as funções de densidade de probabilidade



```
# 11.2. Realizando fit dos dados para uma função log normal
def best_fit_distribution(data, bins=200, ax=None):
    """Model data by finding best fit distribution to data"""
    # Get histogram of original data
    y, x = np.histogram(data, bins=bins, density=True)
    x = (x + np.roll(x, -1))[:-1] / 2.0

    # Best holders
    best_distributions = []

    # Estimativa dos parâmetros para fazer o fit
    for ii, distribution in enumerate([d for d in _distn_names if d in ['lognorm']]):

        print("{:>3} / {:<3}: {}".format(ii+1, len(_distn_names), distribution))

        distribution = getattr(stats, distribution)

        # Tentar fazer o fit na distribuição
        try:
            with warnings.catch_warnings():
                warnings.filterwarnings('ignore')

                params = distribution.fit(data)

                arg = params[-2]
                loc = params[-2]
                scale = params[-1]

                # Calculando a função densidade de probabilidade e a CDF (Cumulative Density Function)
                pdf = distribution.pdf(x, loc=loc, scale=scale, *arg)
                cdf = distribution.cdf(x, loc=loc, scale=scale, *arg)
                sse = np.sum(np.power(y - pdf, 2.0))

                try:
                    if ax:
                        pd.Series(pdf, x).plot(ax=ax)
                except Exception:
                    pass

                best_distributions.append((distribution, params, sse, cdf))

        except Exception:
            pass

    return sorted(best_distributions, key=lambda x:x[2])
```



O próximo passo foi filtrar os dados de Electric Range para veículos do tipo PHEV e aplicar a análise da função de probabilidade, transcrita acima, para entendermos a probabilidade da existência de veículos dada uma certa autonomia. Resolvemos aproximar a função de densidade probabilidade para uma log normal, por efeitos de simplicidade.

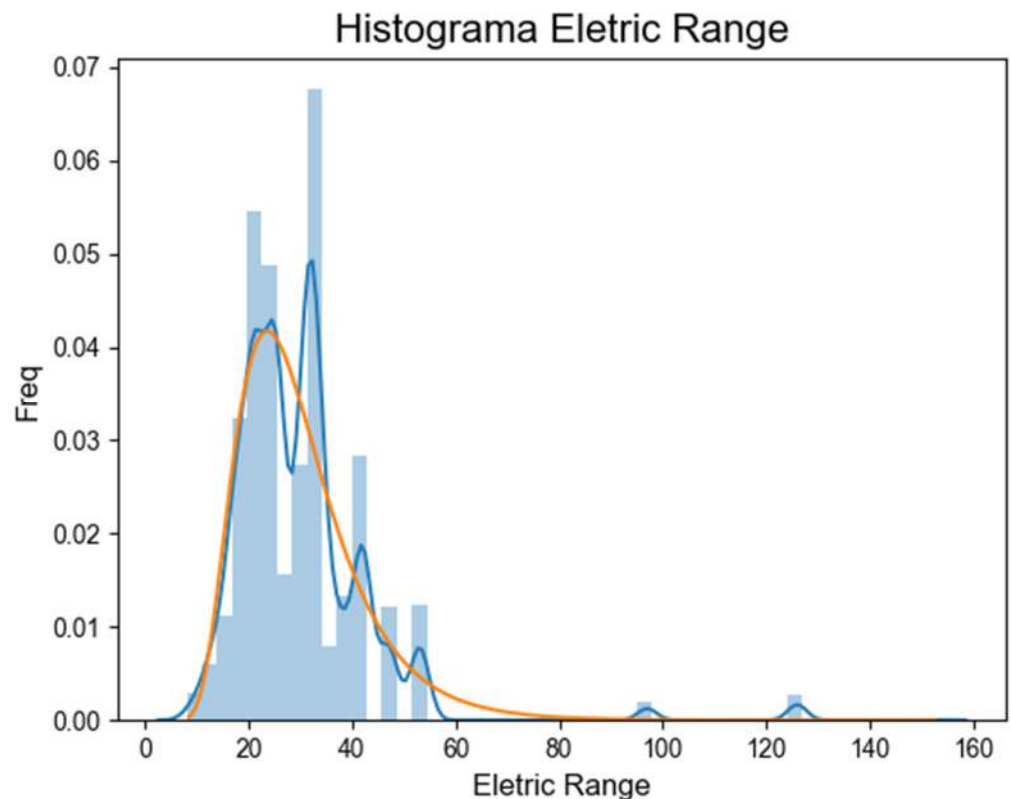
```
# 11.3. Plotando o histograma para o electric range, sua função densidade de
# probabilidade e a aproximação da mesma para uma log normal
ax = sns.distplot(eletric_range_df, hist=True)

best_distributions = best_fit_distribution(eletric_range_df, 200, ax)
best_dist = best_distributions[0]

sns.set_style("whitegrid")
plt.title('Histograma Eletric Range', fontsize=16)
plt.xlabel('Eletric Range', fontsize=12)
plt.ylabel('Freq', fontsize=12)
plt.show()

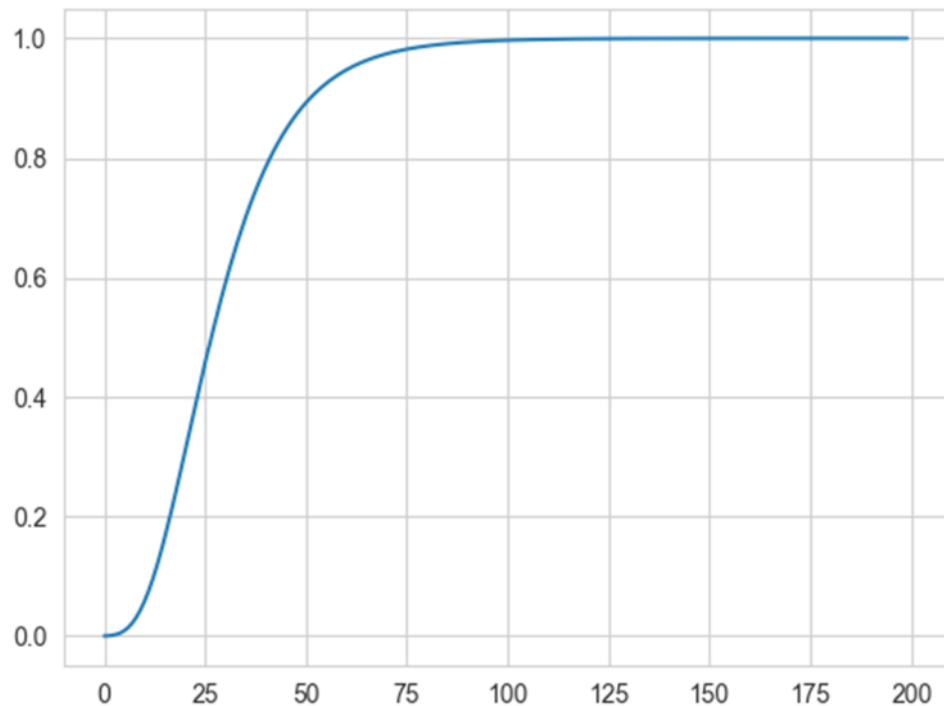
plt.plot(best_dist[3])
plt.show()
```

Resultando nos gráficos abaixo:



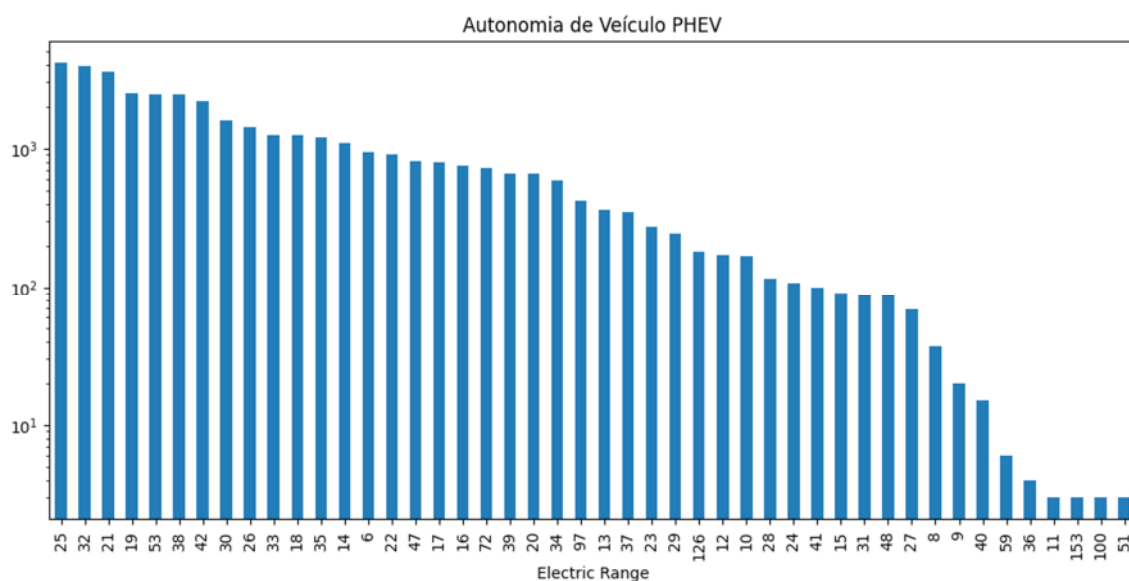


No gráfico abaixo, podemos ver que mais de 90% da autonomia dos veículos elétricos do tipo PHEV está abaixo da autonomia de 50.



Em resumo, as análises acima tratam com mais detalhe dessa amostra específica de veículos do tipo PHEV, modelos de 2018 a 2024, e auxiliam na tomada de decisão mais precisa e assertiva e sem a necessidade de se realizar sempre uma análise da base completa, visto que, na maioria das vezes são muito extensas e pesadas.

Usando este método pudemos, através de uma pequena amostra, chegar ao resultado bem próximo ao que se chegaria olhando para toda a população. Vejamos no gráfico abaixo, gerado através de toda a população de dados e que também podemos observar uma concentração de veículos com até 50 de autonomia.





14. Análise e descrição dos resultados preliminares para um possível modelo de negócios.

A. Análise de dados focada no modelo de negócio

Elaboramos análises complementares a fim de extrair algumas informações que auxiliem na estruturação de um possível modelo de negócios:

Gráfico 05: Quantidade de veículos do tipo BEV por Estado

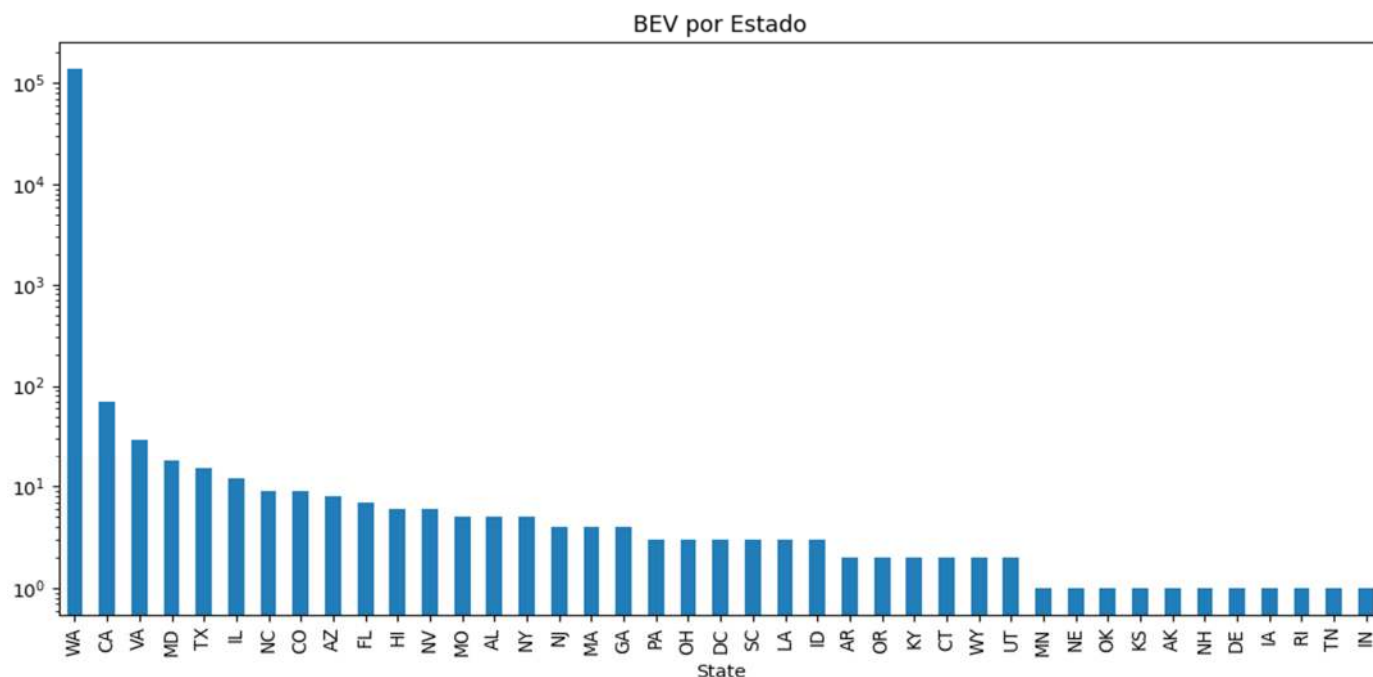
Tipo do Gráfico: Barras

Escala do Gráfico: 10^0 a 10^6

Eixo Y: Quantidade de Veículos BEV por região

Eixo X: Siglas dos Estados

Resultado: Avisando identificar qual a melhor localização para iniciar as instalações dos primeiros postos de Abastecimento Elétrico, filtramos em qual região está concentrada a maioria dos veículos elétricos dos Estados Unidos. Assim, geramos um gráfico com a quantidade de veículos do tipo Ba



attery Electric Vehicle (BEV) para cada Estado e verificamos que a maior concentração está em WA, vejamos:



Gráfico 06: Quantidade de veículos do tipo PHEV por Estado

Tipo do Gráfico: Barras

Escala do Gráfico: 10^0 a 10^6

Eixo Y: Quantidade de Veículos PHEV por região

Eixo X: Siglas dos Estados

Resultado: Geramos o mesmo gráfico acima, no entanto com veículos do tipo Plug-in Hybrid Electric Vehicle (PHEV) para cada Estado. A concentração se manteve no Estado de WA:

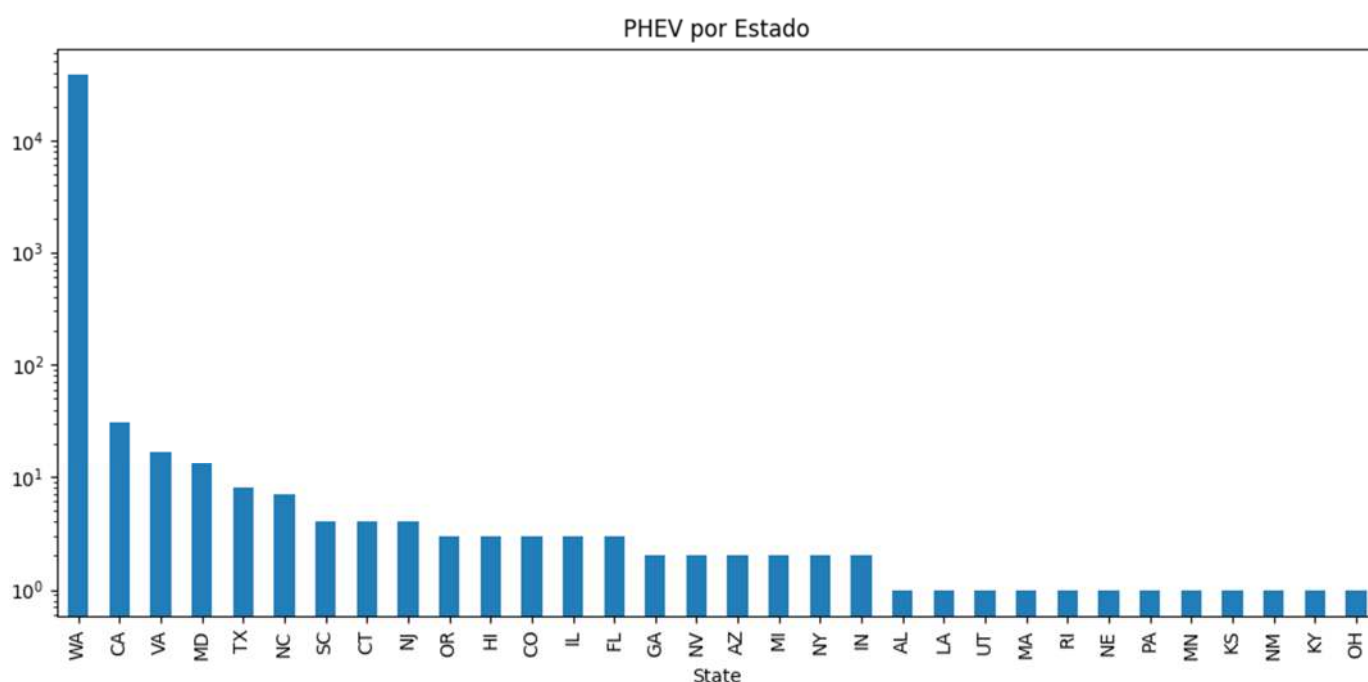


Gráfico 07: 10 Cidades no Estado de WA com mais veículos elétricos

Tipo do Gráfico: Barras

Escala do Gráfico: 4×10^3 a 4×10^4

Eixo Y: Quantidade de veículos por Cidade

Eixo X: Nomes das cidades

Resultado: Após obter a informação de que o Estado de WA concentra a maior quantidade de veículos elétricos do país, optamos por refinar a busca visando uma análise mais precisa. Assim, filtramos as 10 cidades localizadas no Estado WA com maior concentração de veículos elétricos (considerando ambos os tipos, BEV e PHEV), e percebemos que Seattle concentra uma quantidade expressiva desses veículos, vejamos:

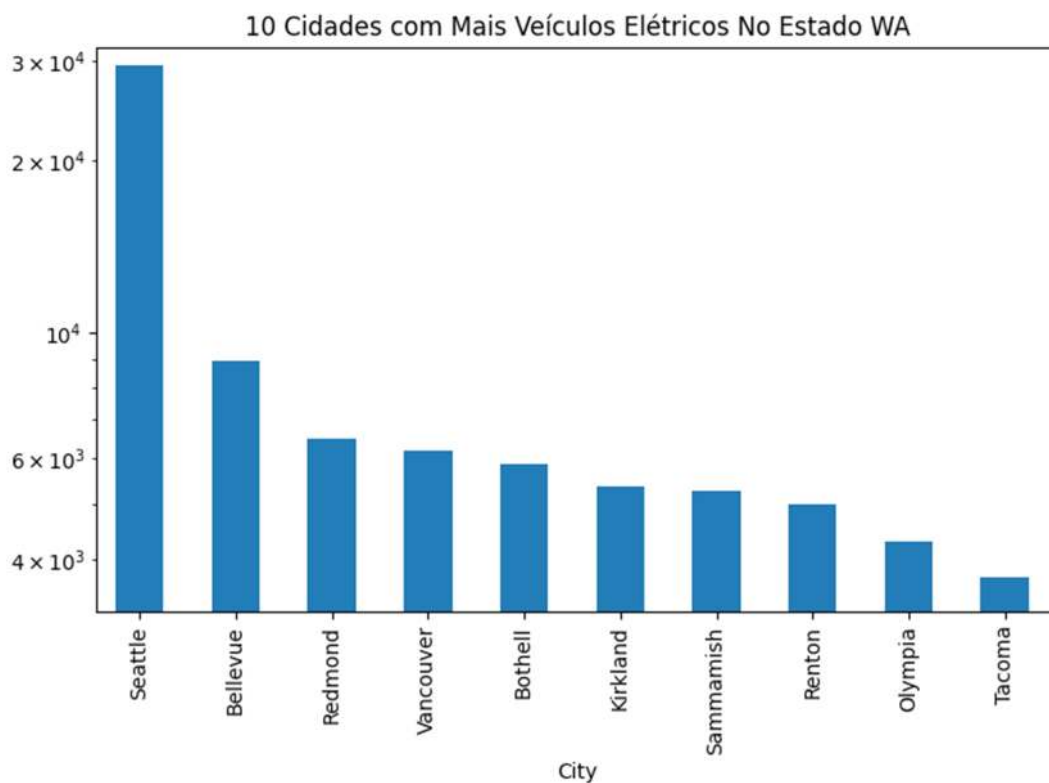


Gráfico 08: Quantidade de veículos elétricos por ano

Tipo do Gráfico: Histograma

Escala do Gráfico: 10^{-5} a 10^{-1}

Eixo Y: Densidade de probabilidade

Eixo X: Ano

Resultado: A fim de trazer mais segurança a criação de uma nova empresa, geramos o histograma abaixo que demonstra um crescimento na quantidade de veículos elétricos ao longo dos últimos anos:

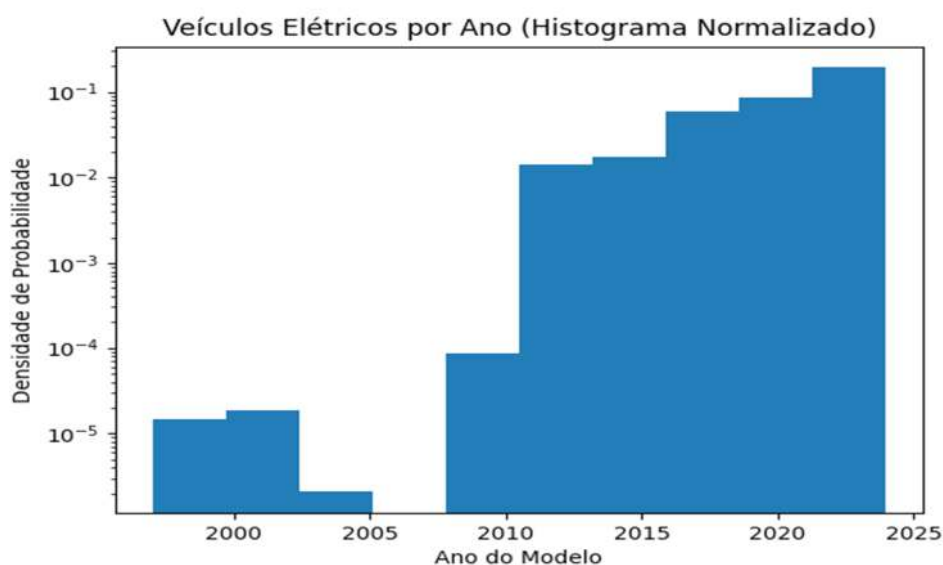




Gráfico 09: Autonomia de Veículos BEV

Tipo do Gráfico: Barras

Escala do Gráfico: 10^0 a 10^3

Eixo Y: Quantidade de veículos para cada valor de autonomia

Eixo X: Valores de autonomia dos veículos “Electric Range”

Resultado: Neste ponto, avaliamos a autonomia dos veículos listados em nosso banco de dados para o tipo BEV, em busca de entender a real necessidade desses consumidores e, futuramente, auxiliar a empresa na ampliação da rede de postos. Observamos que há veículos com autonomia que variam de 39 a 337, com uma concentração de veículos com autonomia de 215, 220 e 238.

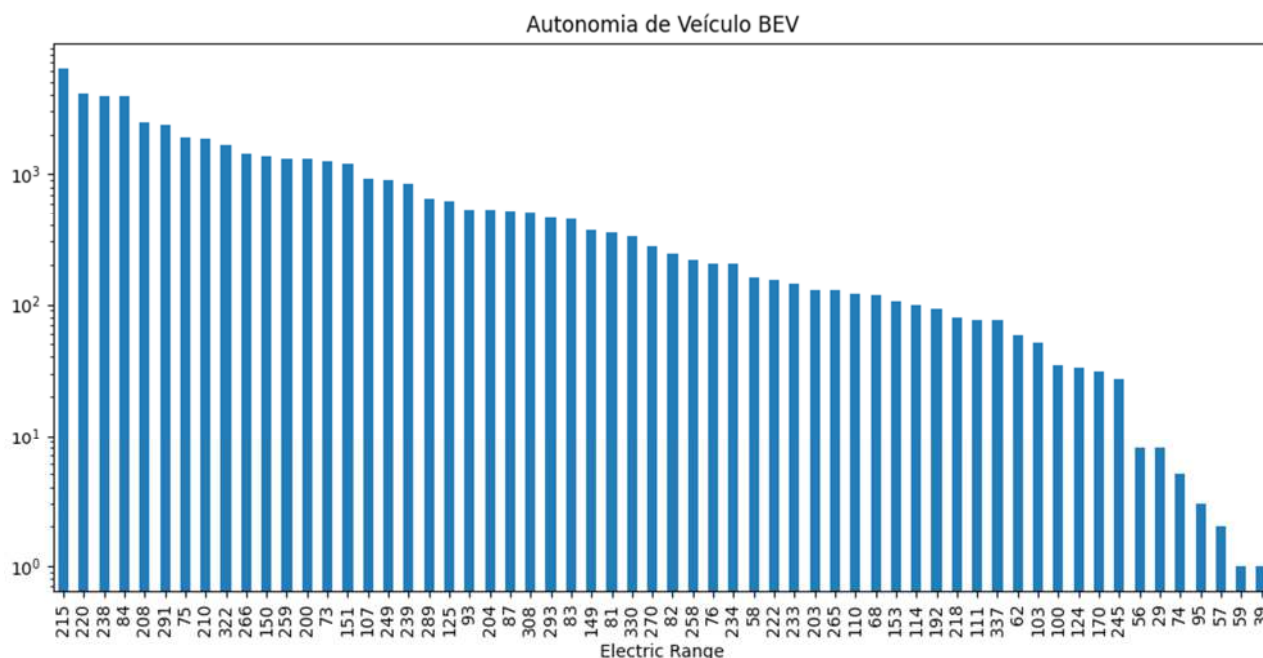


Gráfico 10: Autonomia de Veículos PHEV

Tipo do Gráfico: Barras

Escala do Gráfico: 10^0 a 10^3

Eixo Y: Quantidade de veículos PHEV para cada valor de autonomia

Eixo X: Valores de autonomia dos veículos “Electric Range”

Resultados: Replicamos a análise acima para veículos do tipo PHEV (híbridos) que, no geral, possuem uma autonomia expressivamente inferior aos veículos do tipo BEV, no entanto acabam atraindo o interesse do consumidor pela possibilidade de alternar entre carregamento elétrico e por combustível. Observamos que há veículos com autonomia que variam de 06 a 153, com uma concentração de veículos com autonomia de 25, 32 e 21, vejamos:

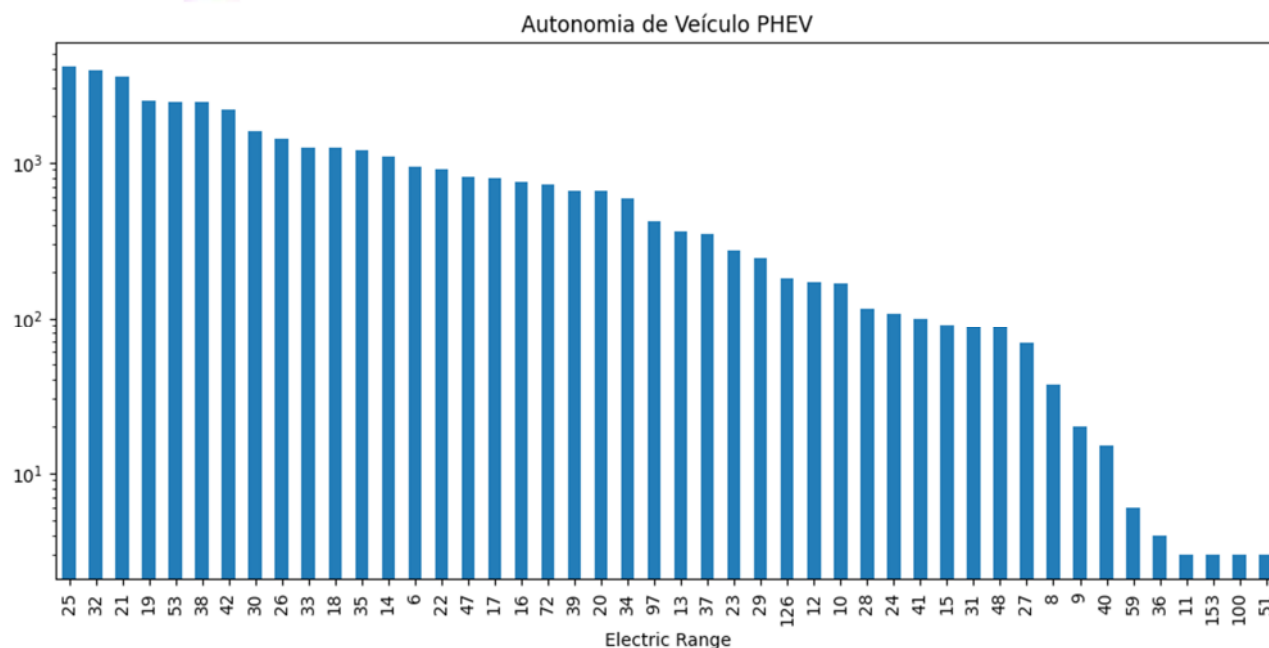


Gráfico 11: Quantidade de veículos elétricos por ano

Tipo do Gráfico: Gráfico de linhas

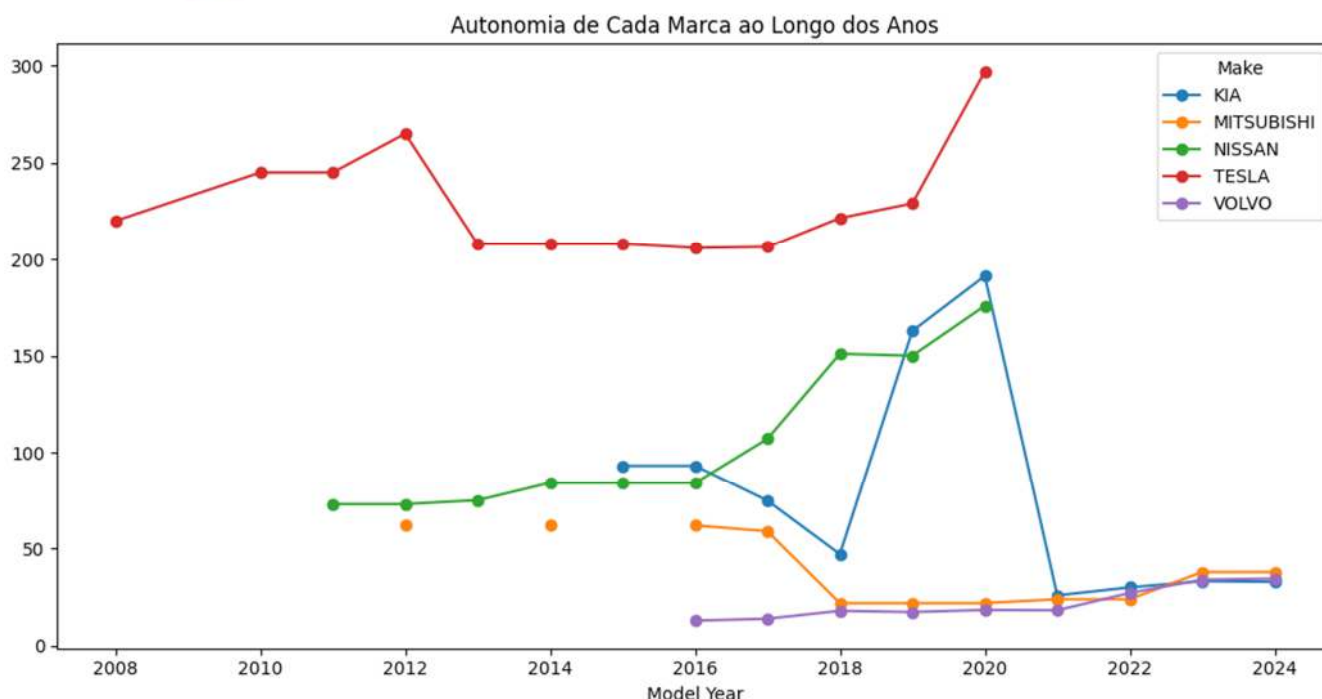
Escala do Gráfico: 0 a 300

Eixo Y: Electric Range (Autonomia)

Eixo X: Ano

Linhas: Marcas de veículo

Resultado: Uma forma de divulgar e impulsionar nossa empresa é buscar parcerias com marcas fabricantes de carros elétricos. Por isso, selecionamos 5 marcas para acompanhar sua evolução ao longo dos anos e entender melhor sua performance nesse mercado. Conforme esperado, verificamos que a Tesla, marca focada em veículos elétricos, se destaca de forma expressiva. A Nissan vem aumentando a autonomia de seus veículos, enquanto a Kia, por exemplo, teve seu melhor modelo em 2020, reduzindo a autonomia média nos últimos anos, sinalizando uma preferência, nos últimos anos, por fabricar veículos híbridos, visto que, como explicado anteriormente, apesar de possuírem uma menor autonomia, proporcionam maior liberdade de escolha ao consumidor:



C. Modelo de negócio:

Com base na análise de dados sobre veículos elétricos, identificamos um nicho de negócio promissor relacionado aos postos de carregamento para esses veículos. Essa oportunidade de expansão nos permite estruturar um modelo de negócios viável para atender às necessidades crescentes dos proprietários de carros elétricos.

- **Nome da empresa:** Green Energy
- **Core Business:** Infraestrutura de Carregamento para Automóveis Elétricos.
- **Público Alvo:** Consumidores de carros elétricos
- **Localização Estratégica:** Com base nas análises realizadas, considerando que a maior concentração de veículos elétricos se encontra no Estado WA, iniciaremos as instalações dos postos de abastecimento elétrico nesta região, mais especificamente na cidade de Seattle.
- **Proposta de Valor:** promover facilidade e acessibilidade com pontos de carregamento rápido que viabilizam a utilização de veículos elétricos para todos os tipos de viagens (curtas ou longas). Além de assumir um compromisso ambiental com um modelo de negócio sustentável.
- **Canais de Distribuição:** Os meios pelos quais a empresa pretende levar seu serviço ao mercado e os disponibilizar aos clientes são através de parcerias com montadoras de veículos elétricos, além de campanhas de Marketing Digital.



- **Relacionamento com o Cliente:** Promover programas de fidelidade com descontos e benefícios aos clientes recorrentes e manter disponível canais de atendimentos para dúvidas, sugestões e reclamações.



15. Storytelling:

O storytelling é uma ferramenta essencial para efetivamente comunicar e engajar o público. Por meio de narrativas bem construídas, transformamos dados, fatos e estatísticas em informações compreensíveis e relevantes. No contexto dos veículos elétricos, um tema de crescente importância devido à urgência climática e inovação tecnológica, uma narrativa precisa e envolvente é fundamental para capturar as complexidades e o potencial de impacto deste setor. Abaixo, apresentamos um esboço estruturado de storytelling que orientará nossa análise dos dados de veículos elétricos, assegurando que todos os aspectos chave sejam sistematicamente explorados e apresentados.

15.1. Apresentação do Grupo

- 15.1.1. Introdução aos Membros da Equipe
- 15.1.2. Breve Histórico e Experiências Anteriores
- 15.1.3. Motivação para Escolher o Tema

15.2. Nome do Projeto

- 15.2.1. Justificativa para o Nome Escolhido
- 15.2.2. Significado e Impacto

15.3. Importância do Tema

- 15.3.1. Veículos Elétricos Cenário Atual/Futuro
- 15.3.2. Importância e Contribuições

15.4. Desafios do Setor

- 15.4.1. Principais Desafios Das Empresas Do Setor
- 15.4.2. Impacto Socioeconômico

15.5. Abordagem Analítica

- 15.5.1. Metodologia Utilizada
- 15.5.2. Fontes e Qualidade Dos Dados
- 15.5.3. Gaps e Oportunidade Identificadas



15.6. Dados Disponíveis

15.6.1. Fonte de Dados

15.6.2. Qualidade e Confiabilidade dos Dados

15.6.3. Limitações e Desafios na Coleta de Dados

15.7. Análise Exploratória

15.7.1. Primeiras Impressões e Descobertas

15.7.2. Visualizações e Gráficos Relevantes

15.7.3. Correlações e Tendências Identificadas

15.8. Resultados Pretendidos

15.8.1. Impacto Esperado da Análise

15.8.2. Contribuições para a Empresa/Organização e Sociedade

15.8.3. Passos Futuros e Recomendações



16. Artefatos do Projeto:

A. Link Github:

<https://github.com/meddavid/Mackenzie-Projeto-Aplicado-II>

B. Link Projeto:

<https://github.com/meddavid/Mackenzie-Projeto-Aplicado-II>

C. Link Artefatos do Projeto:

Scripts	https://github.com/meddavid/Mackenzie-Projeto-Aplicado-II/tree/1bcfc57e6e34b1c4621dc602b1d10a616addbc4a/99.%20Artefatos/02.%20scripts_python
Dataset	https://github.com/meddavid/Mackenzie-Projeto-Aplicado-II/tree/1bcfc57e6e34b1c4621dc602b1d10a616addbc4a/99.%20Artefatos/01.%20Dataset

D. Link Youtube:

<https://youtu.be/QNk5DDfITro>