|  |
| --- |
| 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](mailto:10415270@MACKENZISTA.COM.BR)  RA 10415977 – ANA VITÓRIA SILVA - 10415977[@MACKENZISTA.COM.BR](mailto:10415270@MACKENZISTA.COM.BR) |
| NOME DO PROFESSOR: **Prof. Dr. Felipe Albino dos Santos** |

Sumário

[Sumário 2](#_Toc165150140)

[1. Título: 4](#_Toc165150141)

[2. Introdução 4](#_Toc165150142)

[3. Objetivos: 4](#_Toc165150143)

[4. Metas: 5](#_Toc165150144)

[5. Cronograma: 6](#_Toc165150145)

[6. Fluxo Baseado em Pensamento Computacional em Contextos Organizacionais: 7](#_Toc165150146)

[A. Decomposição: 7](#_Toc165150147)

[B. Reconhecimento de padrões: 7](#_Toc165150148)

[C. Abstração: 7](#_Toc165150149)

[D. Design de Algoritmos: 7](#_Toc165150150)

[7. Organização e o contexto em que os dados foram gerados: 8](#_Toc165150151)

[8. Referências de aquisição do dataset: 9](#_Toc165150152)

[9. Dataset e Metadados 9](#_Toc165150153)

[A. Dataset: 9](#_Toc165150154)

[B. Descrição do Dataset: 9](#_Toc165150155)

[C. Tipo do Arquivo: 11](#_Toc165150156)

[D. Sensibilidade: 11](#_Toc165150157)

[E. Proprietário do Dado: 11](#_Toc165150158)

[F. Restrições de uso: 11](#_Toc165150159)

[10. Análise Exploratória: 12](#_Toc165150160)

[A. Descrição das Variáveis: 12](#_Toc165150161)

[E. Desafios e Limitações: 20](#_Toc165150162)

[F. Recomendações Preliminares: 21](#_Toc165150163)

[G. Conclusão da Análise Exploratória: 21](#_Toc165150164)

[H. Gráficos e Visualizações: 21](#_Toc165150165)

[I. Gráficos Electric\_Vehicle\_Population\_Data 22](#_Toc165150166)

[J. Gráficos Pontos de Carregamento 26](#_Toc165150167)

[11. Pipeline de Dados: 28](#_Toc165150168)

[12. Proposta Analítica: 29](#_Toc165150169)

[B. Apresentação do DATA.GOV: 29](#_Toc165150170)

[13. Medidas de Acurácia 31](#_Toc165150171)

[14. Analise e descrição dos resultados preliminares para um possível modelo de negócios. 36](#_Toc165150172)

[Gráfico 06: Quantidade de veículos do tipo PHEV por Estado 37](#_Toc165150173)

[Gráfico 07: 10 Cidades no Estado de WA com mais veículos elétricos 37](#_Toc165150174)

[Gráfico 08: Quantidade de veículos elétricos por ano 38](#_Toc165150175)

[Gráfico 09: Autonomia de Veículos BEV 39](#_Toc165150176)

[Gráfico 10: Autonomia de Veículos PHEV 39](#_Toc165150177)

[Gráfico 11: Quantidade de veículos elétricos por ano 40](#_Toc165150178)

[C. Modelo de negócio: 41](#_Toc165150179)

[15. Storytelling: 43](#_Toc165150180)

[15.1. Apresentação do Grupo 43](#_Toc165150183)

[15.2. Nome do Projeto 43](#_Toc165150184)

[15.3. Importância do Tema 43](#_Toc165150185)

[15.4. Desafios do Setor 43](#_Toc165150186)

[15.5. Abordagem Analítica 43](#_Toc165150187)

[15.6. Dados Disponíveis 44](#_Toc165150188)

[15.7. Análise Exploratória 44](#_Toc165150189)

[15.8. Resultados Pretendidos 44](#_Toc165150190)

[16. Artefatos do Projeto: 45](#_Toc165150191)

[A. Link Github: 45](#_Toc165150192)

[B. Link Projeto: 45](#_Toc165150193)

[C. Link Artefatos do Projeto: 45](#_Toc165150194)

**Tabelas**

|  |  |  |
| --- | --- | --- |
| **Tabela 01** | Cronograma ........................................................................................................................... | 3 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Título:

**Empresa:** Green Energy

**Core Business**: Infraestrutura de Carregamento para Automóveis Elétricos.

# 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.

# 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.

# 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.

# Cronograma:

Link:

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



*Tabela 1*

# Fluxo Baseado em Pensamento Computacional em Contextos Organizacionais:

## Decomposição:

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

## Reconhecimento de padrões:

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

## Abstração:

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

## Design de Algoritmos:

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

# 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.

# 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.

# Dataset e Metadados

## 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

## 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

## Tipo do Arquivo:

Os arquivos estão no formato CSV.

## Sensibilidade:

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

## Proprietário do Dado:

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

## Restrições de uso:

O DATA.GOV não restringiu sua utilização.

# 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..

## Descrição das Variáveis:

Nesta parte, serão apresentados detalhes abrangentes sobre as variáveis presentes nos datasest. 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.

1. **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

1. **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

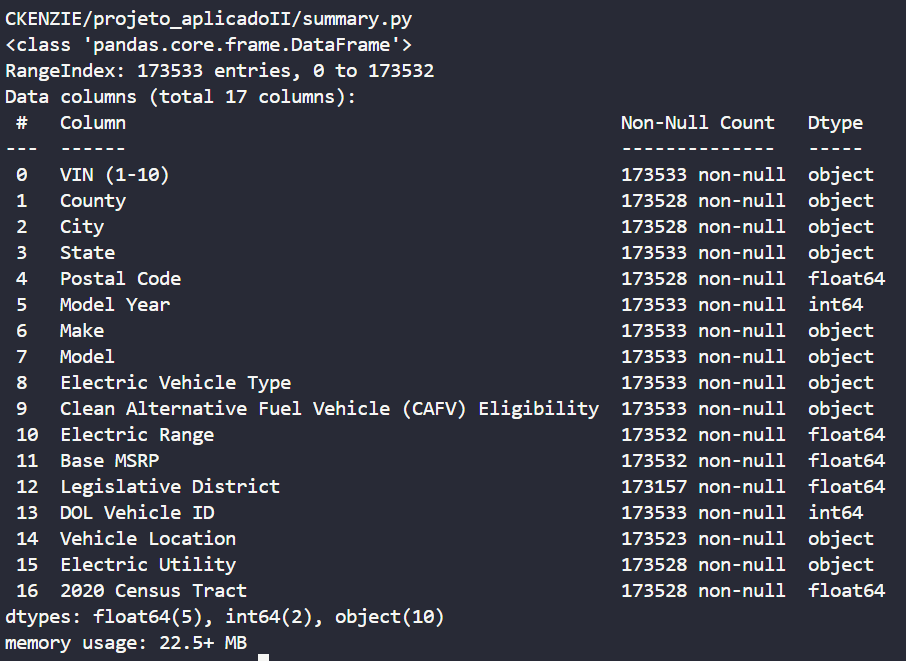
Modificações........: 01/03/2024 - Desenvolvimento

=============================================================

'''

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:**



1. **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

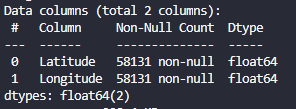
==========================================================

'''

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:**



1. **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}:")

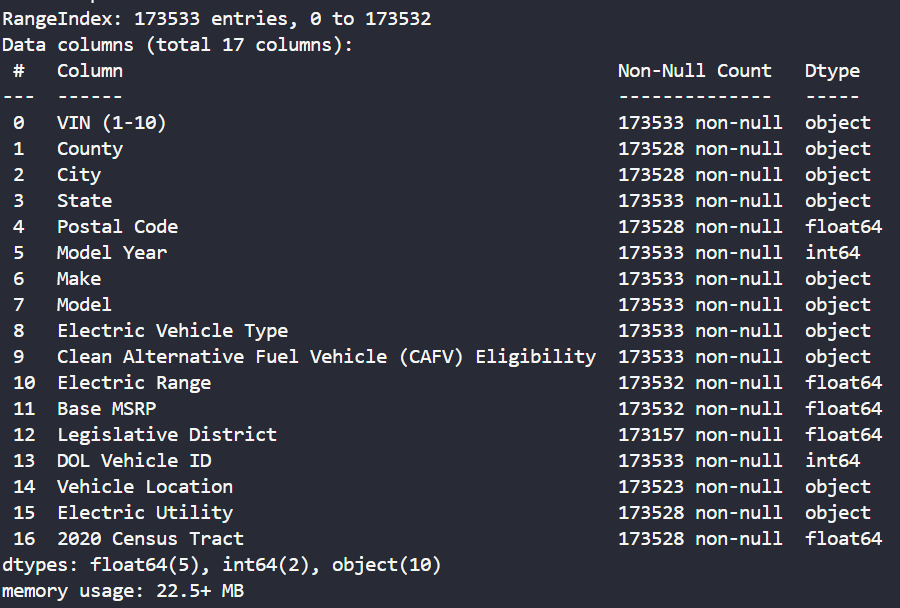
print(num\_registros\_por\_grupo)

print("\nGrupos formados:")

print(list(grupos.groups.keys()))

**Retorno:**

* **Quantidade de registros:** 173533
* **Quantidade de registros agrupados por WHO\_REGION:**



* **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', 'PORSCHE', 'RIVIAN', 'ROLLS ROYCE', 'SMART', 'SUBARU', 'TESLA', 'TH!NK', 'TOYOTA', 'VOLKSWAGEN', 'VOLVO', 'WHEEGO ELECTRIC CARS']**

1. **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

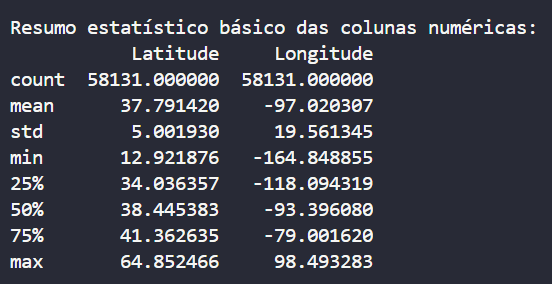
print(df[colunas\_numericas].describe(), "\n")

# Analisando a quantidade de registros

num\_registros = len(df)

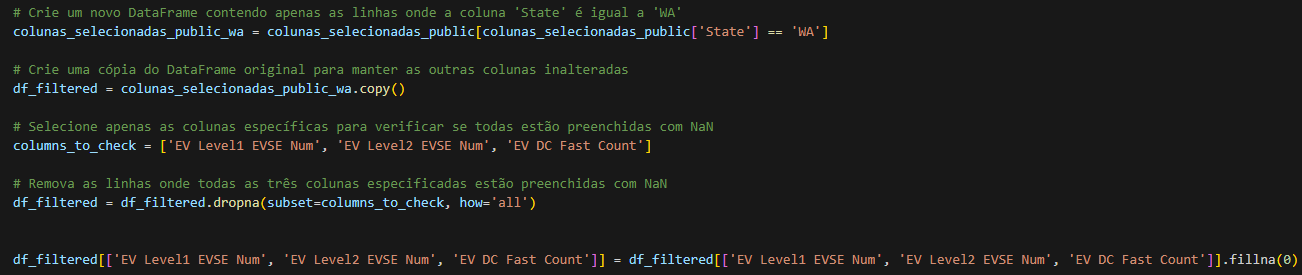
print(f"#### - Quantidade de registros: {num\_registros}\n")

**Retorno:**

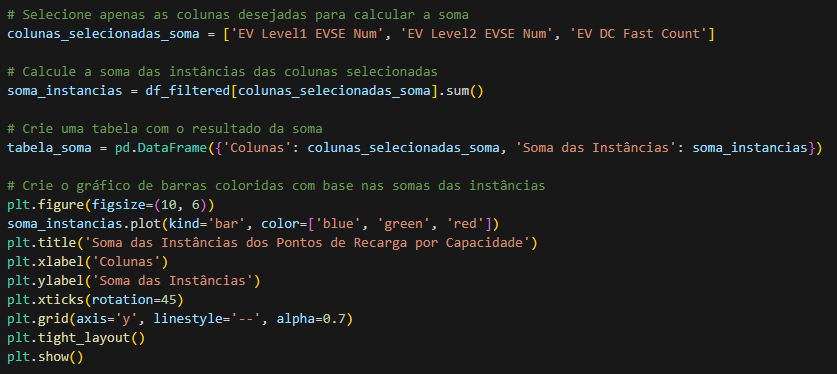


* 1. **Análise de Pontos de Recarga:**

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

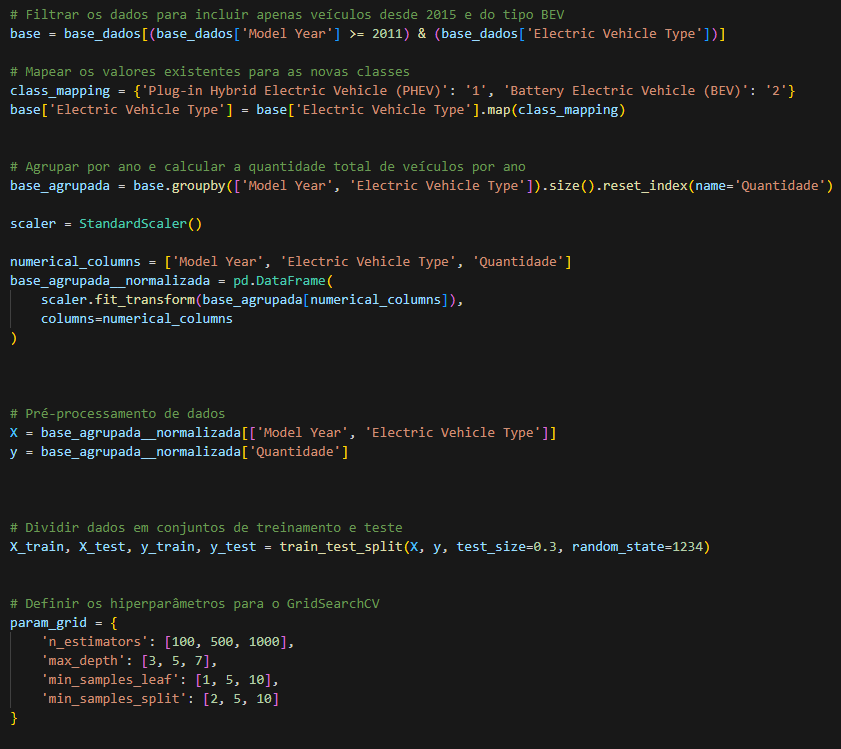


Calcula-se a soma das instâncias de diferentes tipos de pontos de recarga e cria-se um gráfico de barras para visualizar essa soma.

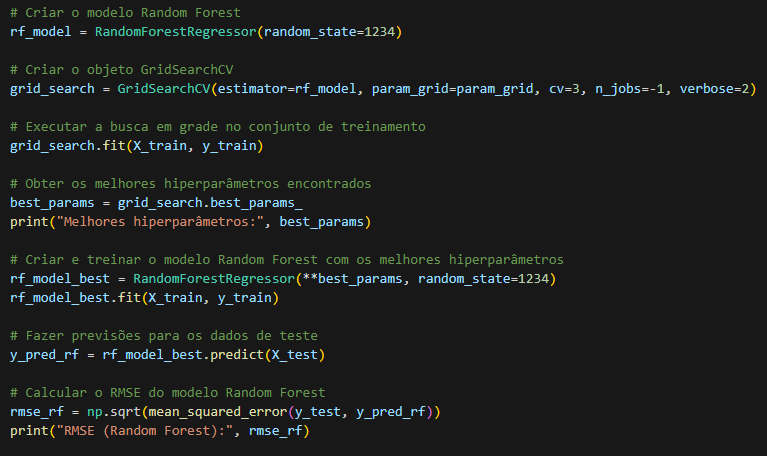


* 1. **Análise da Distribuição Temporal de Veículos:**

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



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



* 1. **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.

## 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.

## 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.

## 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.

## 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

## Gráficos Electric\_Vehicle\_Population\_Data

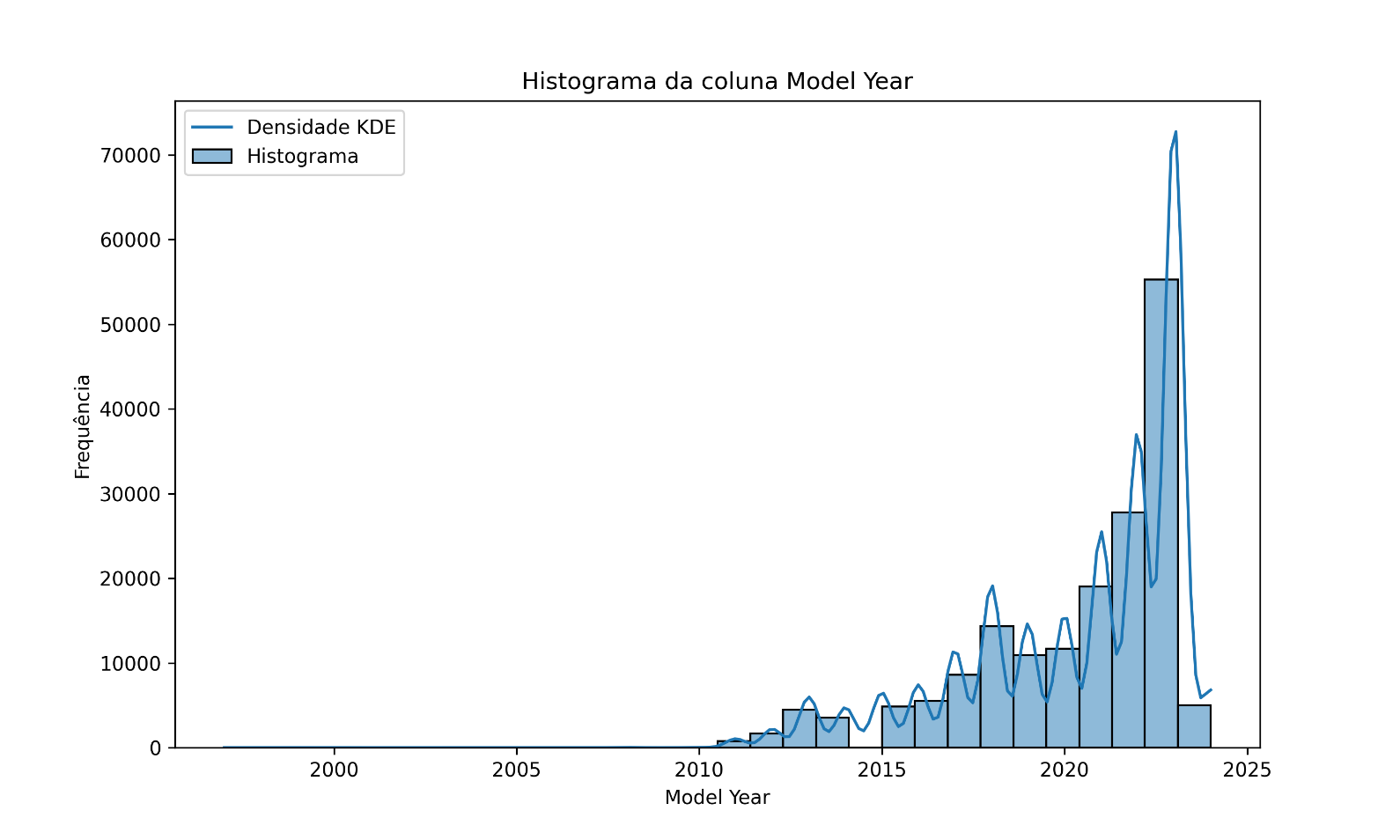
* ***Histograma do Ano do Modelo***

***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.*



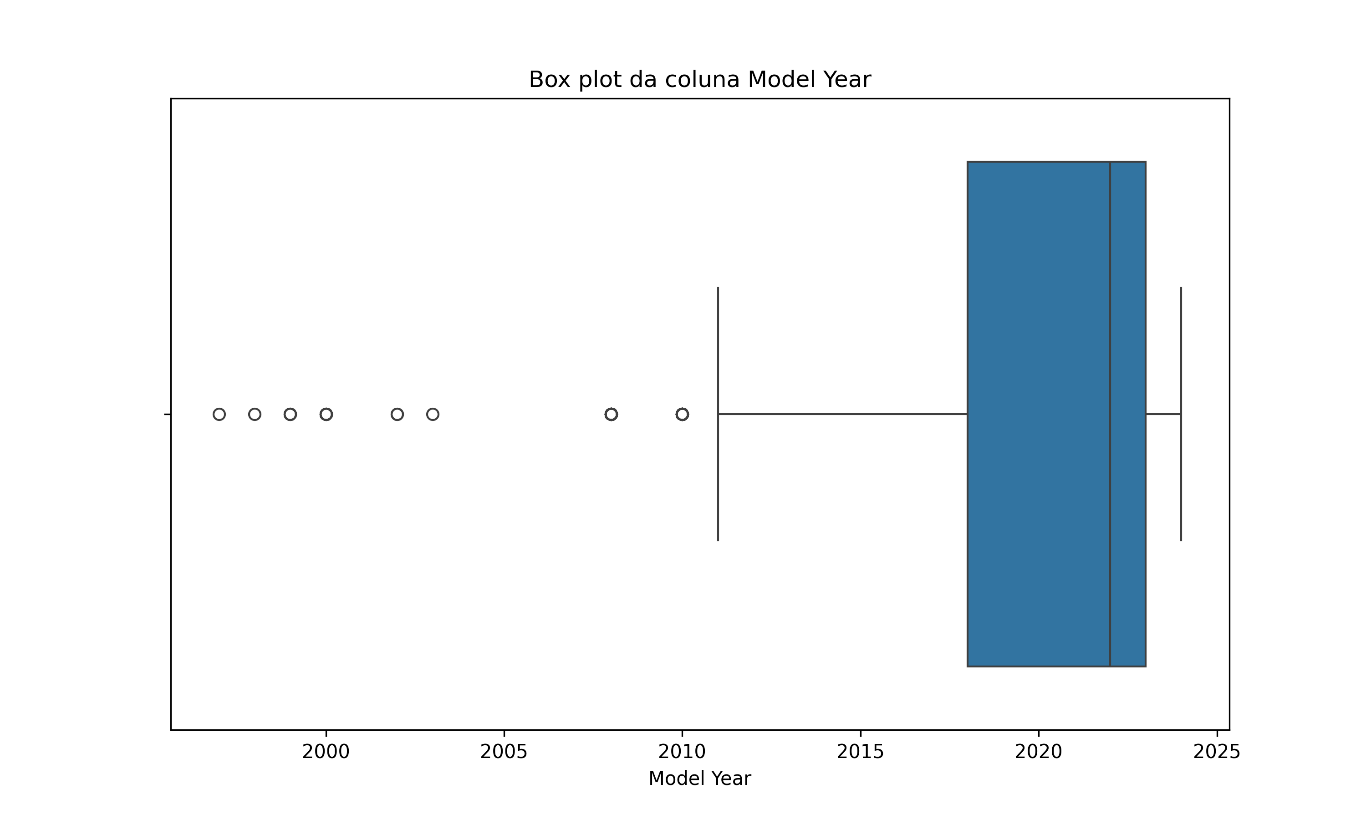
* ***Gráfico 02:*** *Boxplot do Ano do Modelo*

***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.*



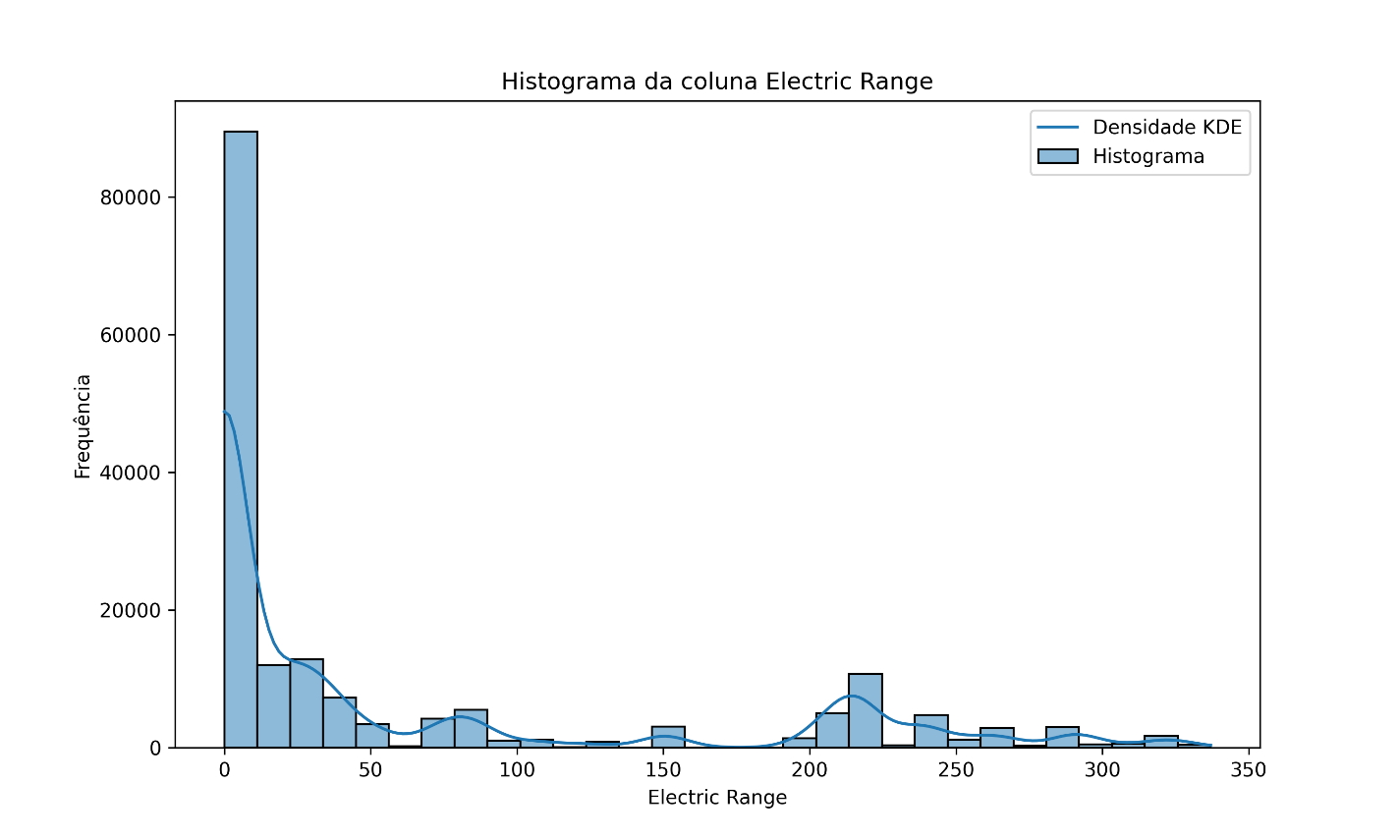
* ***Gráfico 03: Histograma do Alcance Elétrico***

***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.*



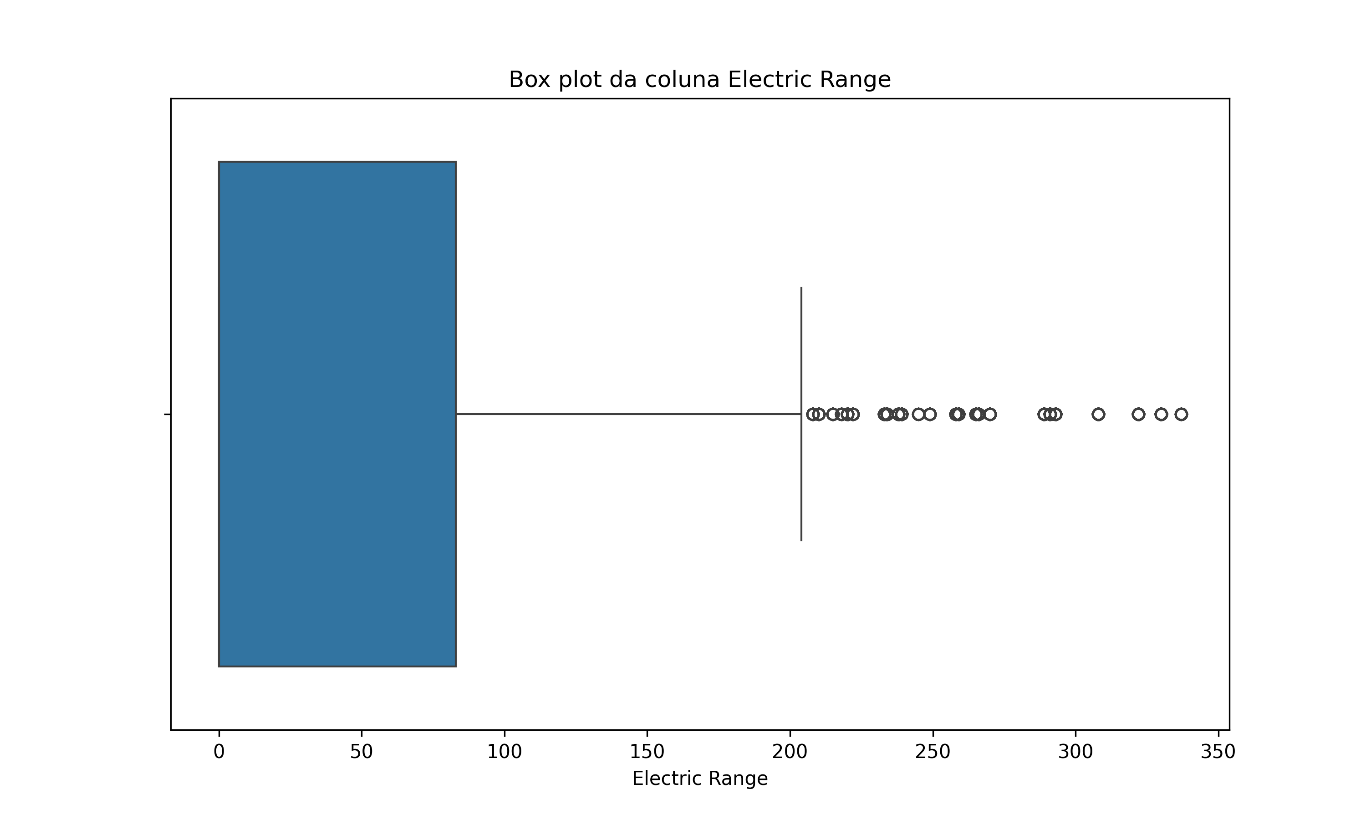
* ***Gráfico 04: Boxplot do Alcance Elétrico***

***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.*



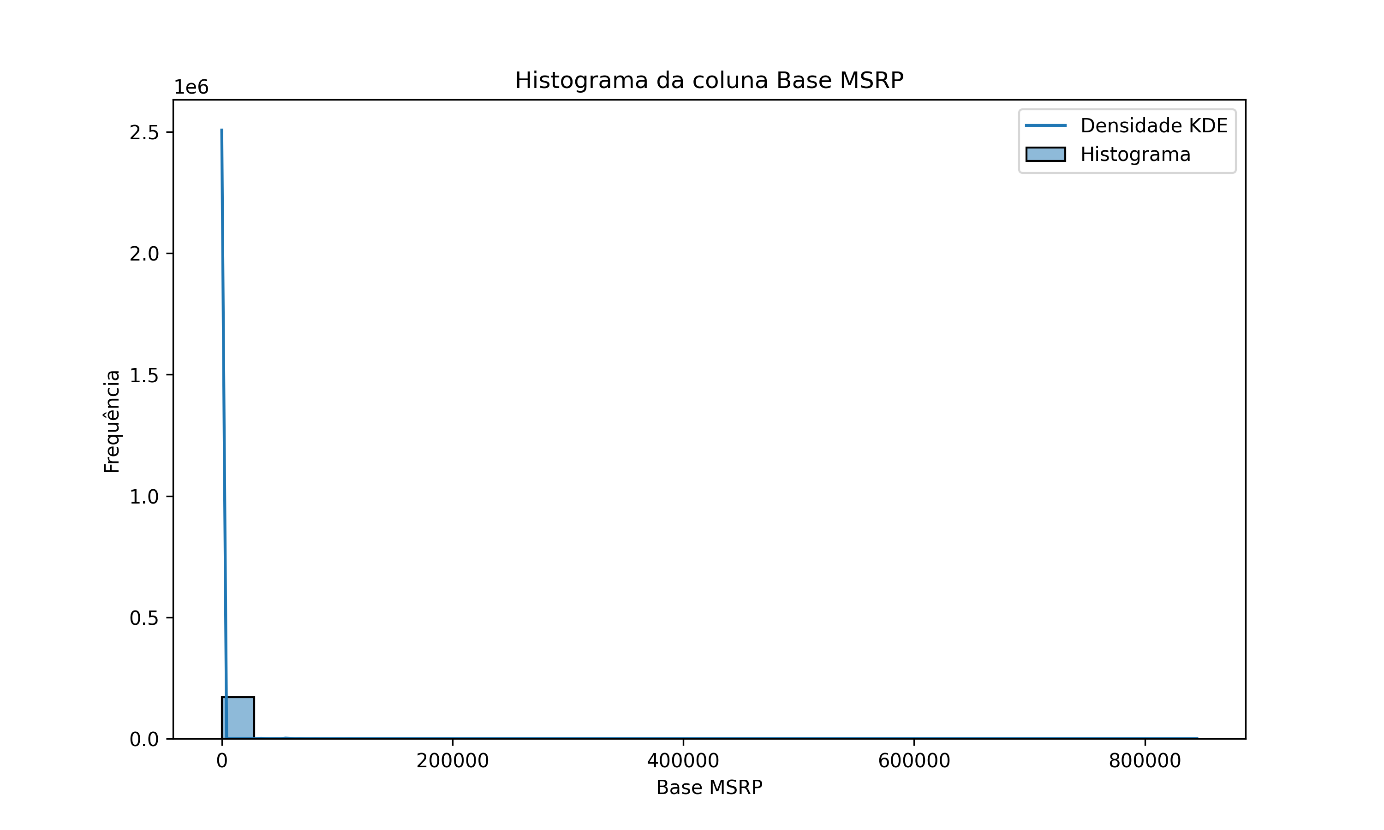
* ***Gráfico 05:*** *Histograma do Preço Sugerido pelo Fabricante (MSRP)*

***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.*



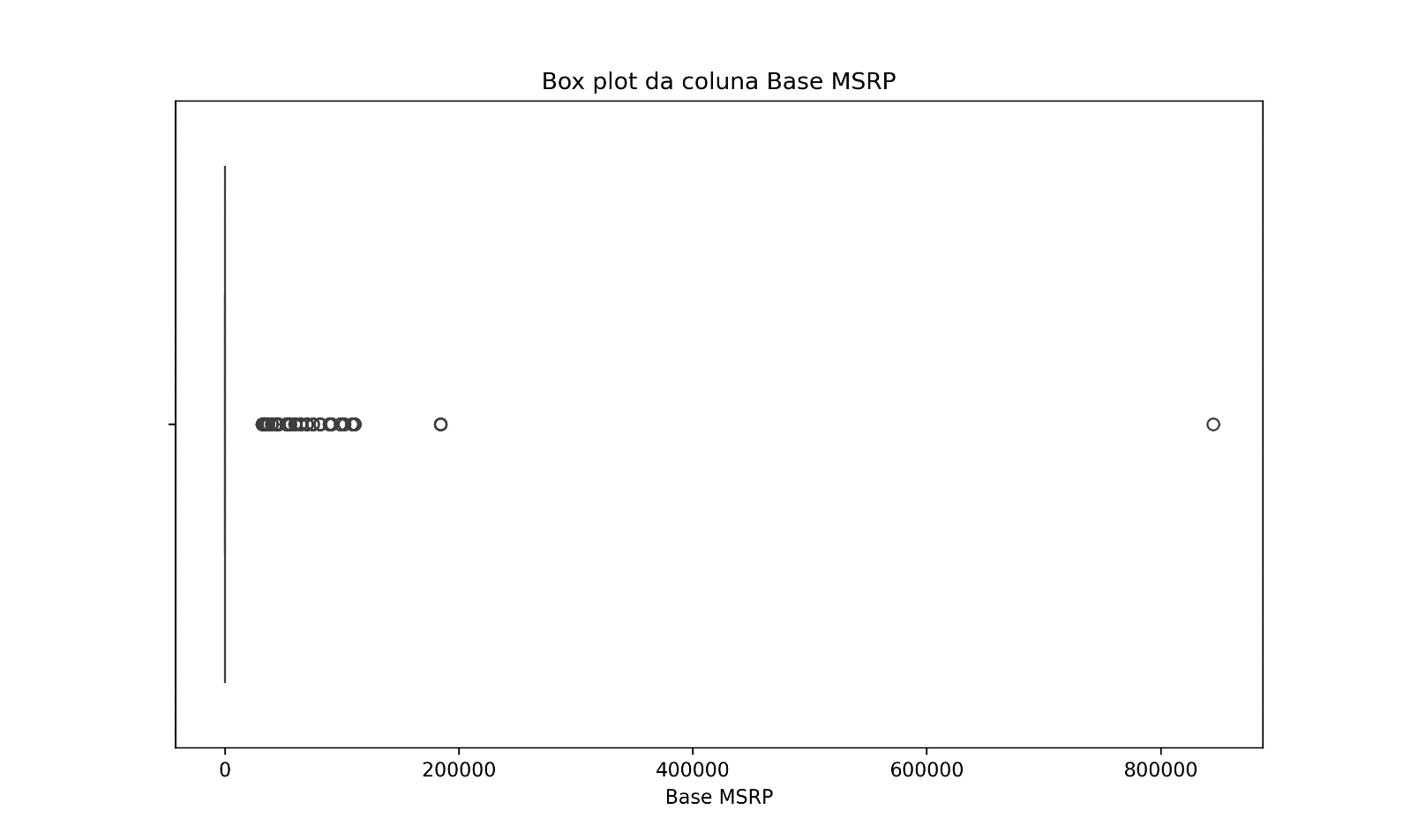
* ***Gráfico 06: Boxplot do Preço Sugerido pelo Fabricante (MSRP)***

***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.*



## Gráficos Pontos de Carregamento

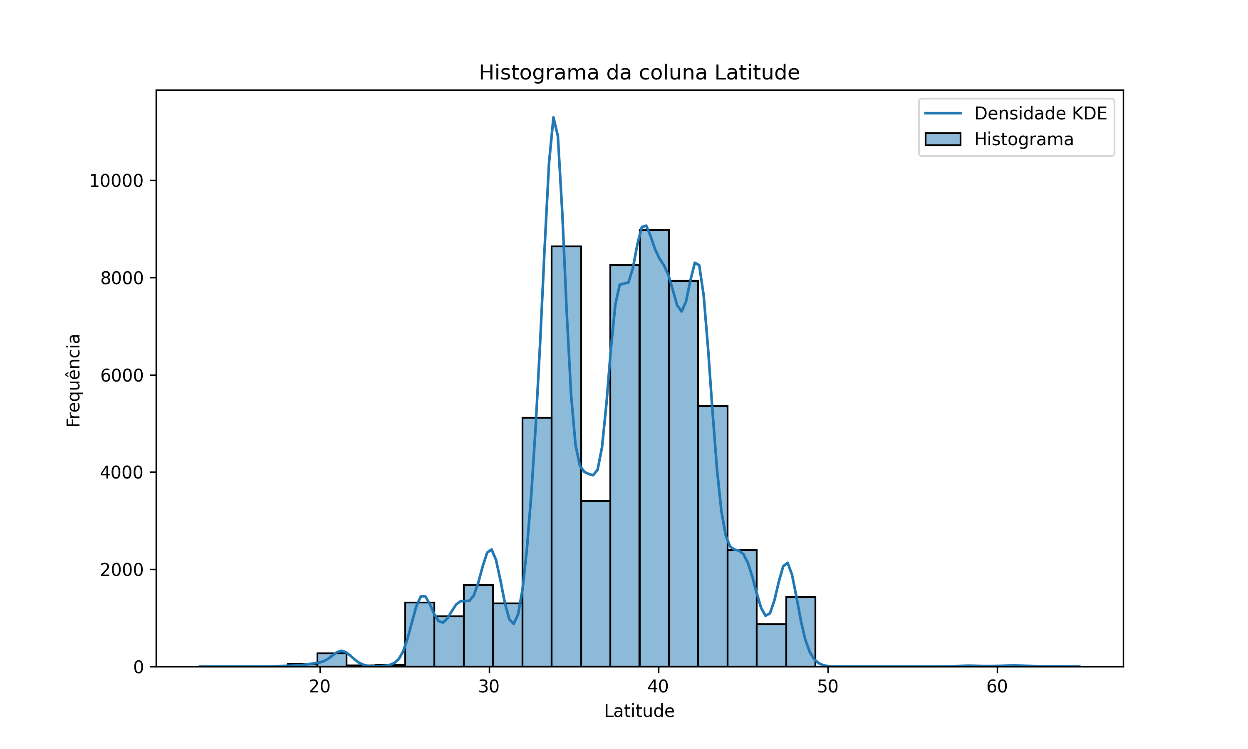
* ***Gráfico 01: Histograma Latitude***

**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.



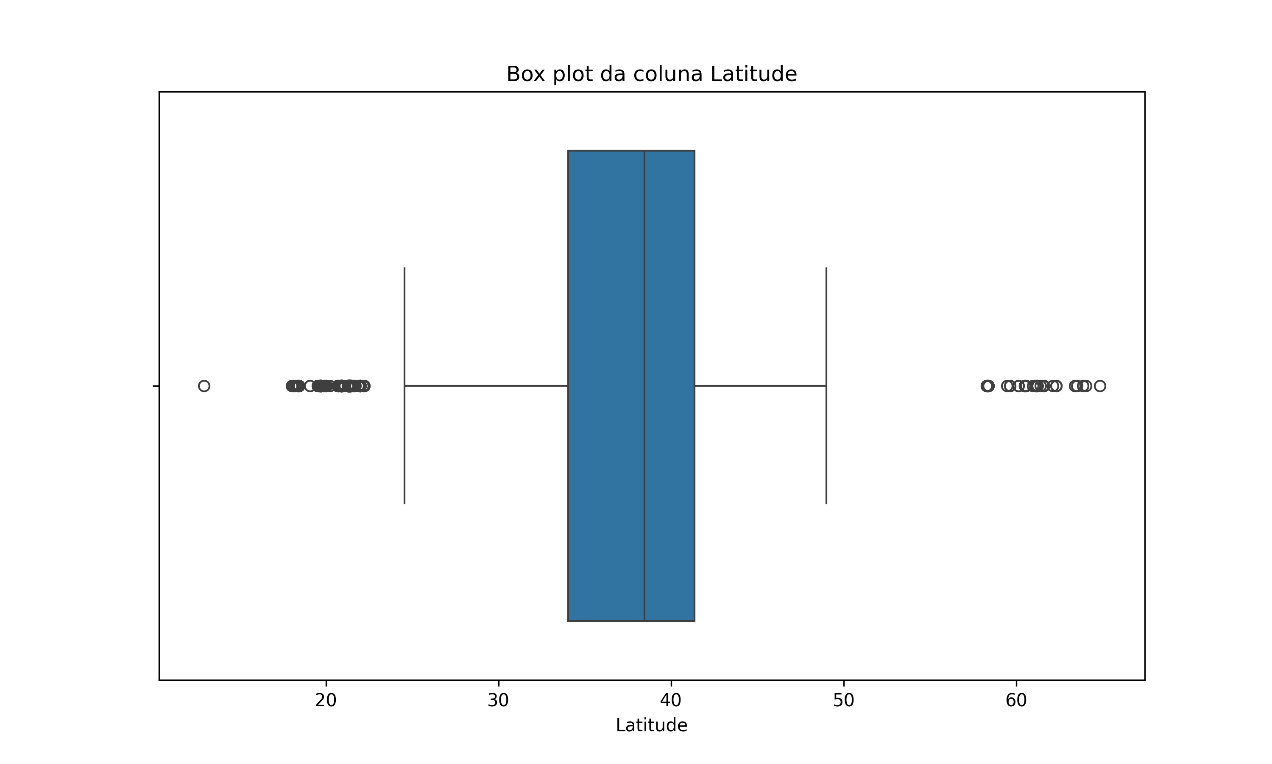
* ***Gráfico 02: Boxplot Latitude***

**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.



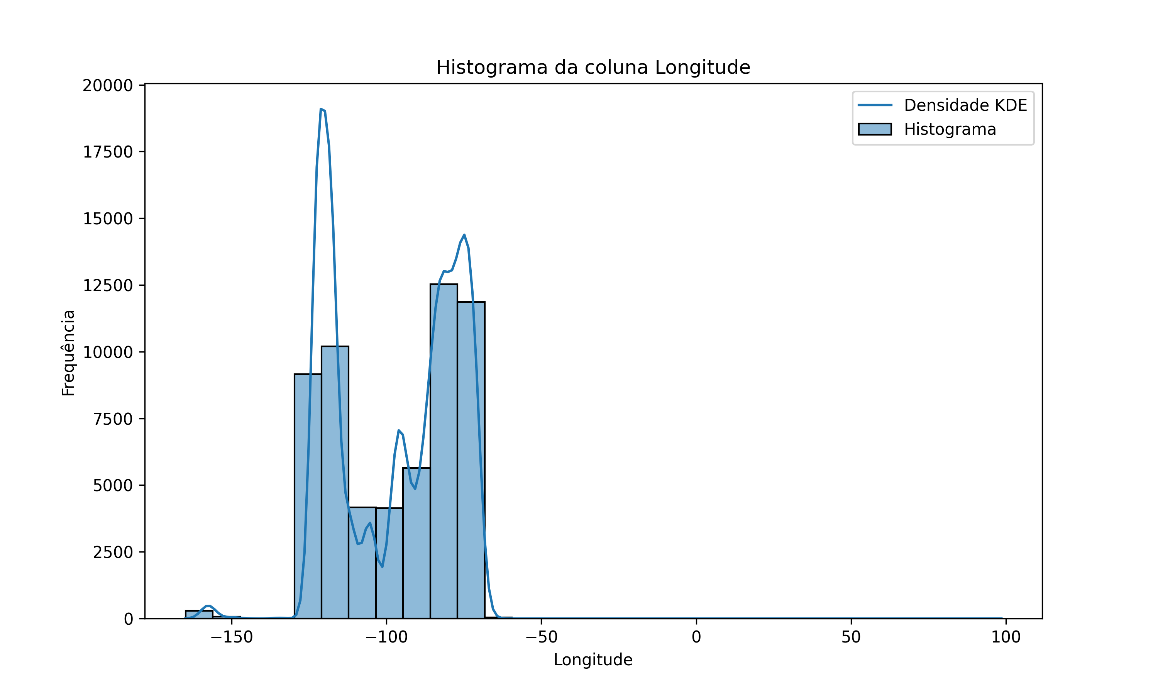
* ***Gráfico 03: Histograma Longitude***

**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.



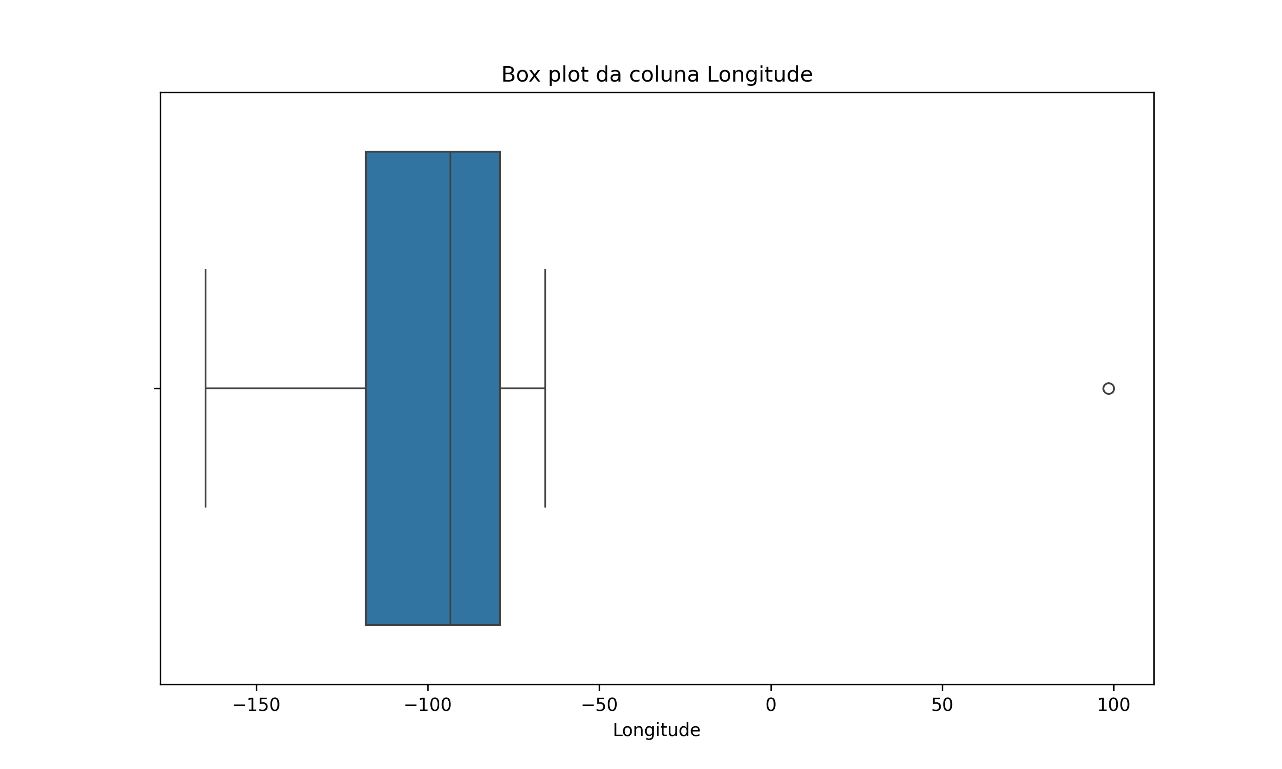
* ***Gráfico 04: Boxplot Longitude***

**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:**

# 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.

# 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.

## 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.

# Medidas de Acurácia

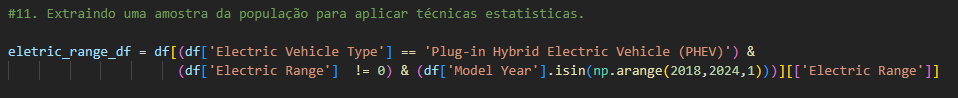
1. **Modelo de Regressão**

Para calcular a acurácia de um modelo de regressão linear, iremos utilizar o coeficiente de determinação R2 , que é a métrica comumente usada para avaliar o desempenho de modelos de regressão. O coeficiente varia de 0 a1, onde 1 indica um ajuste perfeito do modelo aos dados. No código você pode encontra-lo na linha model.score (X\_test, y\_test).

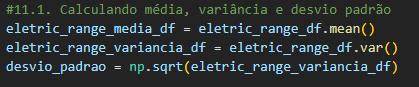


1. **Método Analítico Estatístico Preditivo**

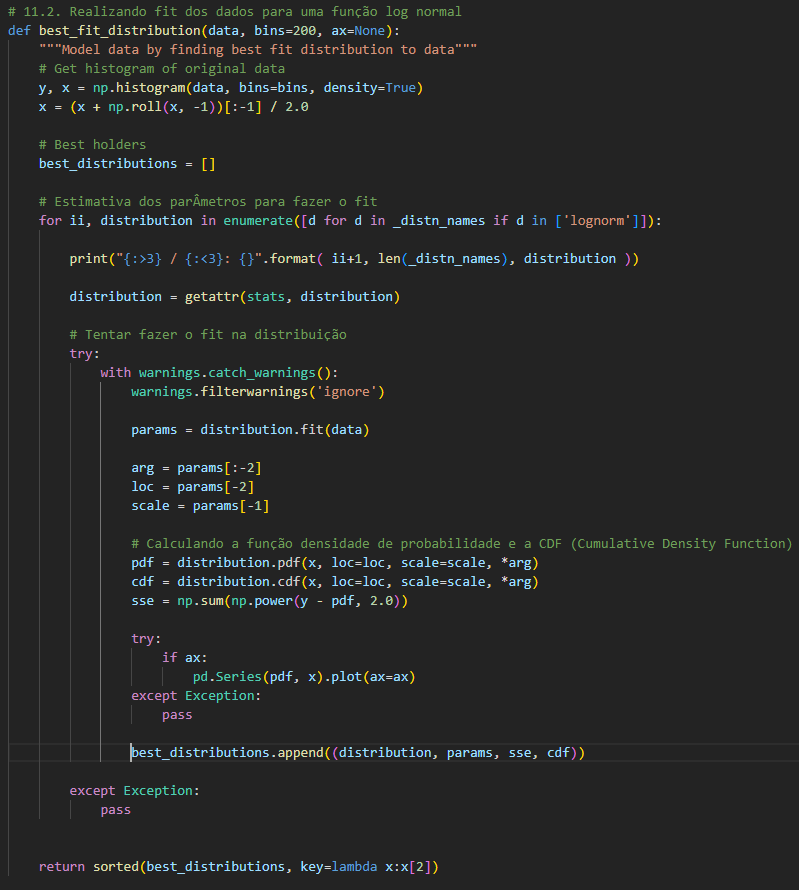
Além da análise exploratória dos dados, realizamos uma análise estatística preditiva mais específica e nichada extraindo uma amostra da população (base de dados completa) para analisar a autonomia de veículos do tipo PHEV nos período de 2018 a 2024.



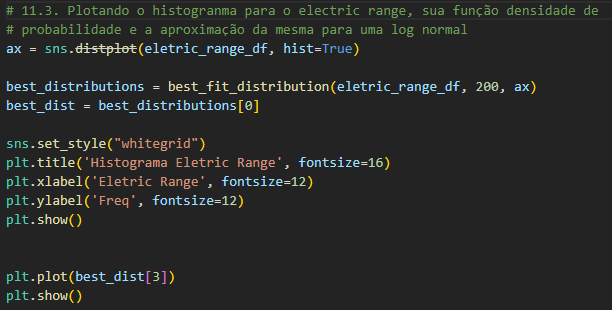
A princípio foram aplicados métodos analíticos como média, variância e desvio padrão



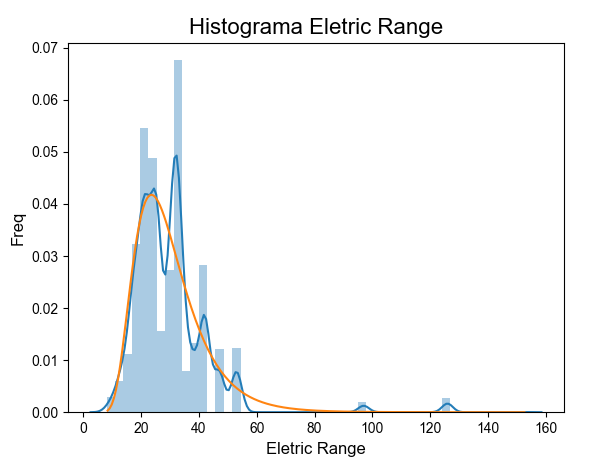
Construímos uma função que realiza o fit dos dados para uma log normal e extrai as funções de densidade de probabilidade



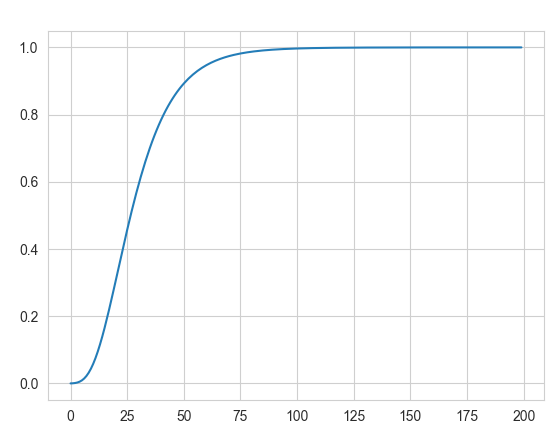
O próximo passo foi filtrar os dados de Electric Range para veículos do tipo PHEV e aplicar a análise da função de probabilidade, transcrita acima, para entendermos a probabilidade da existência de veículos dada uma certa autonomia. Resolvemos aproximar a função de densidade probabilidade para uma log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



Resultando nos gráficos abaixo:

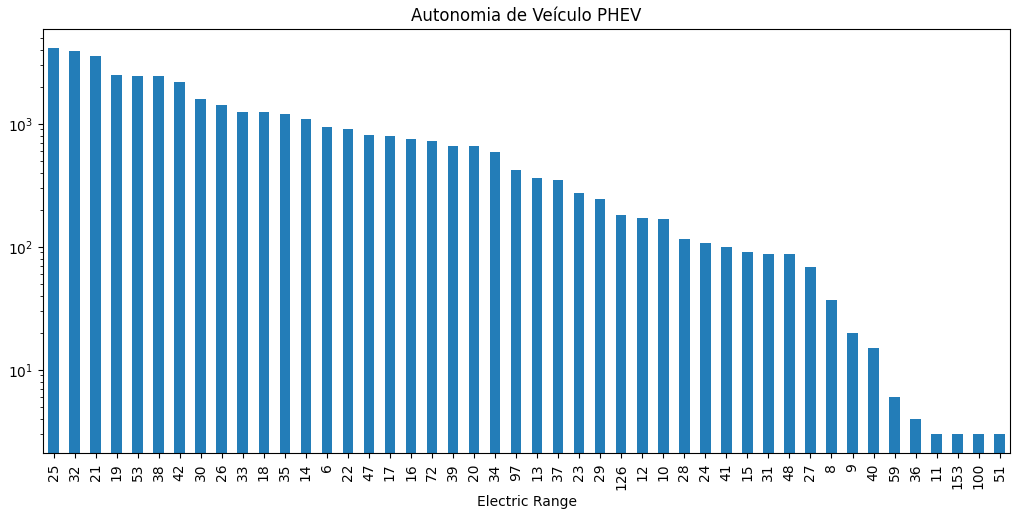


No gráfico abaixo, podemos ver que mais de 90% da autonomia dos veículos elétricos do tipo PHEV está abaixo da autonomia de 50.



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.



# Analise e descrição dos resultados preliminares para um possível modelo de negócios.

1. **Analise 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:

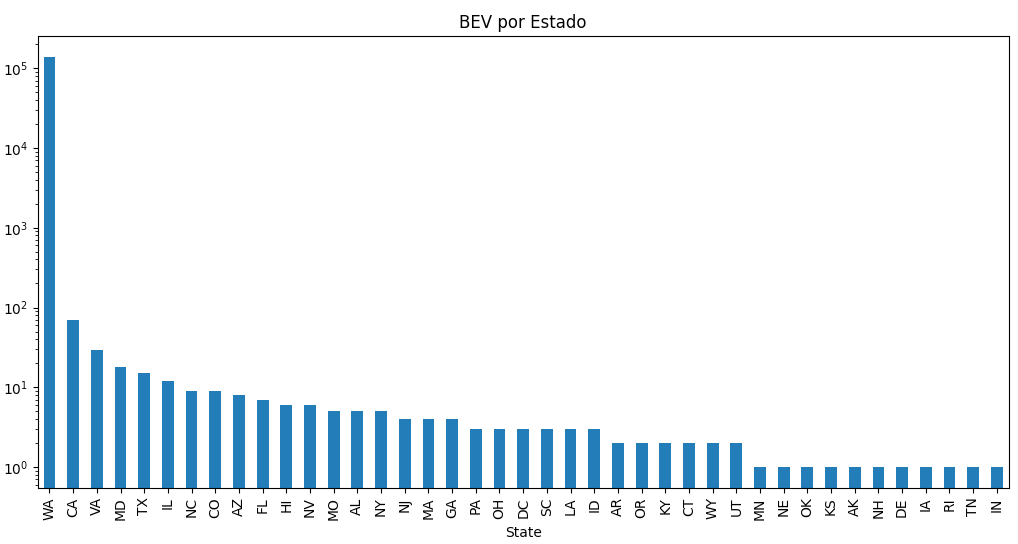
**Gráfico 05**: Quantidade de veículos do tipo BEV por Estado

**Tipo do Gráfico:** Barras

**Escala do Gráfico:** 10º a 106

**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 Battery 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

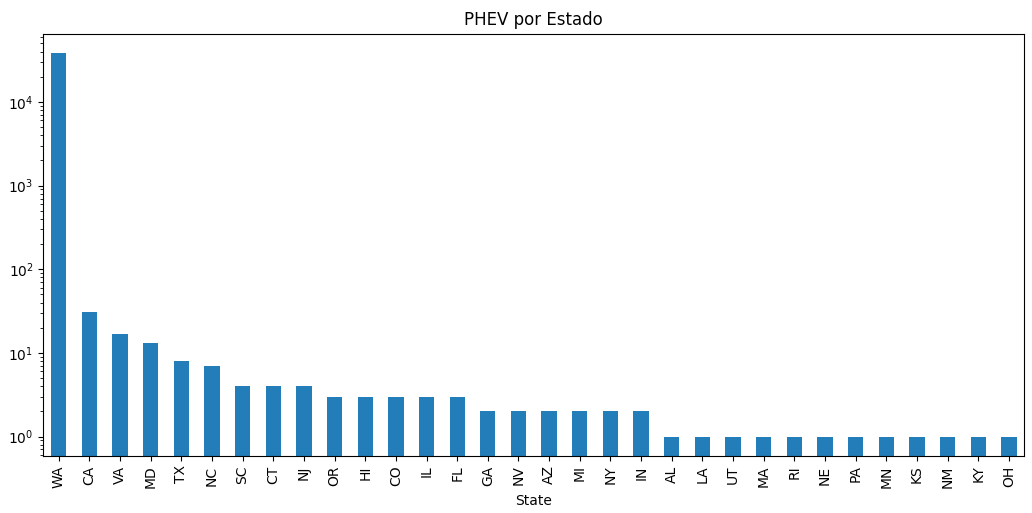
**Tipo do Gráfico:** Barras

**Escala do Gráfico:** 10º a 106

**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

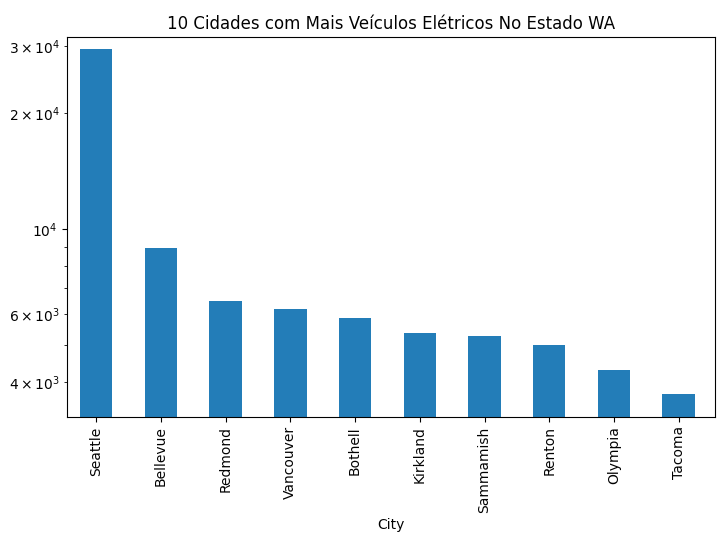
**Tipo do Gráfico:** Barras

**Escala do Gráfico:** 4x10³ a 4x104

**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 Seatle concentra uma quantidade expressiva desses veículos, vejamos:



## Gráfico 08: Quantidade de veículos elétricos por ano

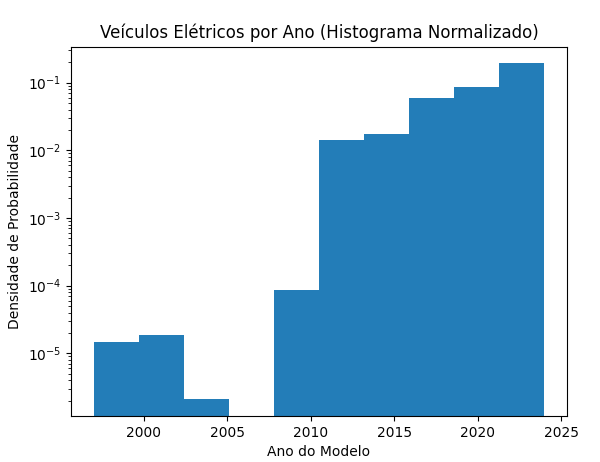
**Tipo do Gráfico:** Histograma

**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

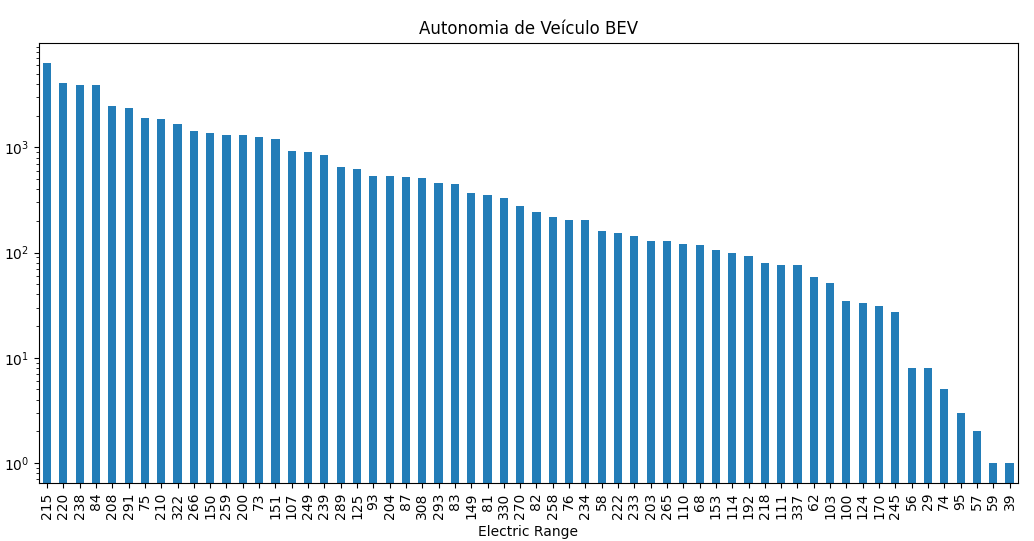
**Tipo do Gráfico:** Barras

**Escala do Gráfico:** 10º a 10³

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

**Eixo X:** Valores de autonomia dos veículos “Eletric 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

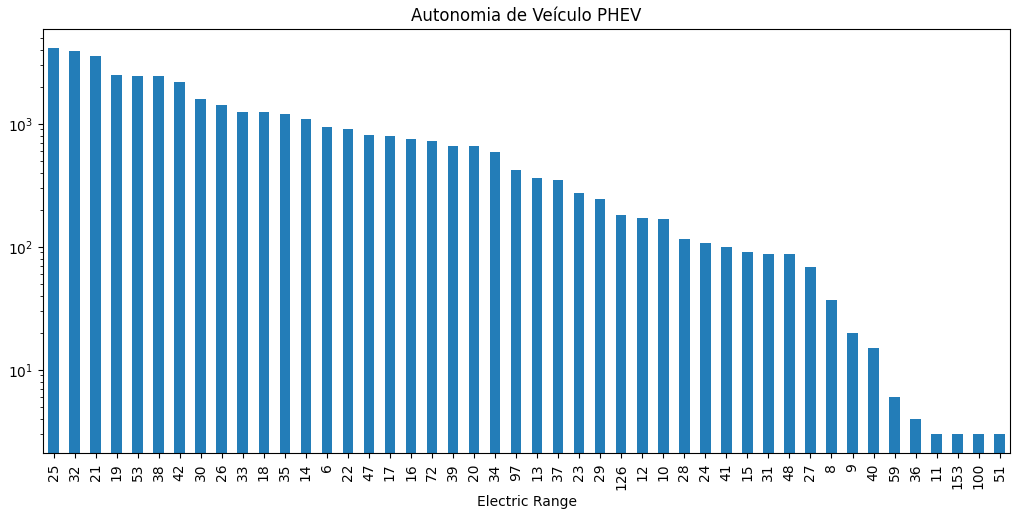
**Tipo do Gráfico:** Barras

**Escala do Gráfico:** 10º a 10³

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

**Eixo X:** Valores de autonomia dos veículos “Eletric Range”

**Resultados:** Replicamos a analise acima para veículos do tipo PHEV (híbridos) que, no geral, possuem uma autonomia expressivamente inferior aos veículos do tipo BEV, no entantoacabam 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:** Grafico de linhas

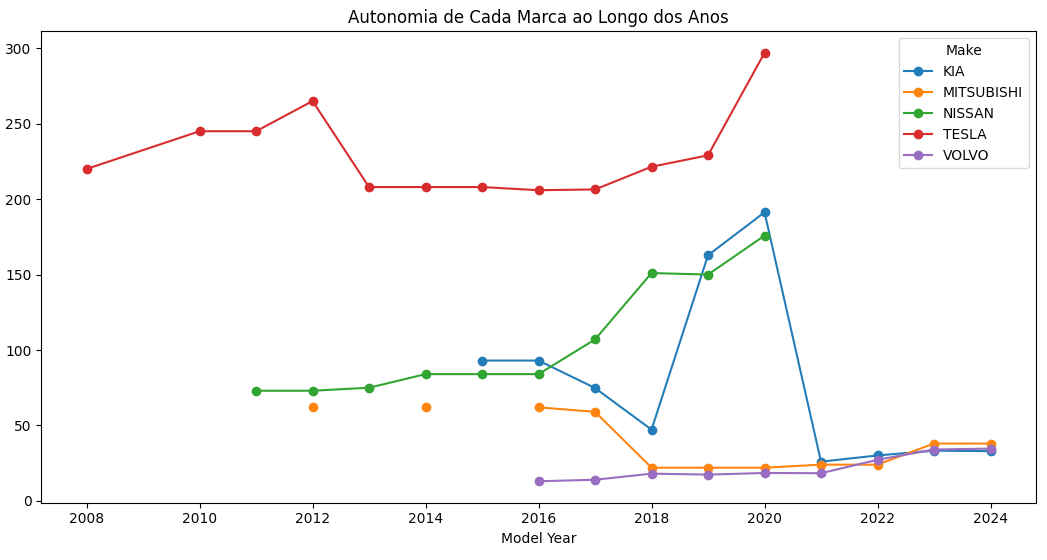
**Escala do Gráfico:** 0 a 300

**Eixo Y:** Eletric 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:



## 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 Seatle.
* 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.

# 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.



### Apresentação do Grupo

* + 1. Introdução aos Membros da Equipe
    2. Breve Histórico e Experiências Anteriores
    3. Motivação para Escolher o Tema

### Nome do Projeto

* + 1. Justificativa para o Nome Escolhido
    2. Significado e Impacto

### Importância do Tema

* + 1. Veículos Elétricos Cenário Atual/Futuro
    2. Importância e Contribuições

### Desafios do Setor

* + 1. Principais Desafios Das Empresas Do Setor
    2. Impacto Socioeconômico

### Abordagem Analítica

* + 1. Metodologia Utilizada
    2. Fontes e Qualidade Dos Dados
    3. Gaps e Oportunidade Identificadas

### Dados Disponíveis

* + 1. Fonte de Dados
    2. Qualidade e Confiabilidade dos Dados
    3. Limitações e Desafios na Coleta de Dados

### Análise Exploratória

* + 1. Primeiras Impressões e Descobertas
    2. Visualizações e Gráficos Relevantes
    3. Correlações e Tendências Identificadas

### Resultados Pretendidos

* + 1. Impacto Esperado da Análise
    2. Contribuições para a Empresa/Organização e Sociedade
    3. Passos Futuros e Recomendações

# Artefatos do Projeto:

## Link Github:

<https://github.com/meddavid/Mackenzie.git>

## Link Projeto:

<https://github.com/users/meddavid/projects/1>

## Link Artefatos do Projeto:

|  |  |
| --- | --- |
| **Scripts** | <https://github.com/meddavid/Mackenzie-Projeto-Aplicado-I/tree/1bcfc57e6e34b1c4621dc602b1d10a616addbc4a/99.%20Artefatos/02.%20scripts_python> |
| **Dataset** | <https://github.com/meddavid/Mackenzie-Projeto-Aplicado-I/tree/1bcfc57e6e34b1c4621dc602b1d10a616addbc4a/99.%20Artefatos/01.%20Dataset> |