



**UNIVERSIDADE FEDERAL DO CEARÁ  
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO  
CIÊNCIA DE DADOS**

**Relatório Sobre análise exploratória e visualização de dados**

**Professor:** Renan Gomes Vieira

**CIÊNCIA DE DADOS**

**Equipe:**

Eric de Araújo Albuquerque  
Francisco Lucas Farias da Silva  
Miguel Barbosa Farias  
Lucas Eduardo Mota

**CRATEÚS  
2025**

# SUMÁRIO

- 1.INTRODUÇÃO.....3
- 1.1.DESCRICÃO DAS COLUNAS.....3
- 1.2.OBJETIVOS.....3
- 2.ANÁLISE EXPLORATÓRIA DE DADOS.....4
- 3.VISUALIZAÇÃO DE DADOS.....5
- 4 DESAFIOS E APRENDIZAGENS.....7
- 5 CONCLUSÃO.....9

# 1 - Introdução

Este conjunto de dados abrangente contém 500.000 de entradas para veículos usados, projetados especificamente para treinar modelos de previsão de preços de alta precisão. Os dados foram gerados sinteticamente usando um script Python que estabelece correlações realistas entre os atributos de um veículo e seu preço de mercado. Inclui 25 das marcas de automóveis mais comuns, abrangendo uma ampla gama de modelos e especificações.

DATASET: Vehicle Price Prediction

Fonte: Kaggle - <https://www.kaggle.com/datasets/metawave/vehicle-price-prediction>

Última atualização: 2025

Tamanho aproximado: 500.000 registros

Formato: CSV (vehicle\_price\_prediction.csv)

## 1.1 - Descrição de colunas

### Principais colunas:

COLUNA	DESCRIÇÃO
make	Marca do veículo (ex: Toyota, BMW, Ford)
model	Modelo do veículo
year	Ano de fabricação
transmission	Tipo de câmbio (manual/automático)
mileage	Quilometragem percorrida
fuel_type	Tipo de combustível (gasolina, diesel, híbrido, elétrico)
engine_size	Tamanho do motor (litros)
owner_count	Quantidade de donos anteriores
accident_history	Histórico de acidentes (Major/Minor/N.A)
price	Preço atual do veículo (em USD)

## 1.2 - Objetivos

O objetivo desta análise é investigar quais fatores contribuem de forma significativa para a valorização ou desvalorização de um veículo, permitindo insights valiosos para compradores, vendedores, concessionárias e modelos de machine learning focados em precificação. A partir desses dados, exploraremos relações importantes, como impacto da quilometragem, influência da marca, efeito do tipo de

combustível, relevância do número de donos anteriores, entre outros aspectos fundamentais.

## 2. Análise exploratória de dados

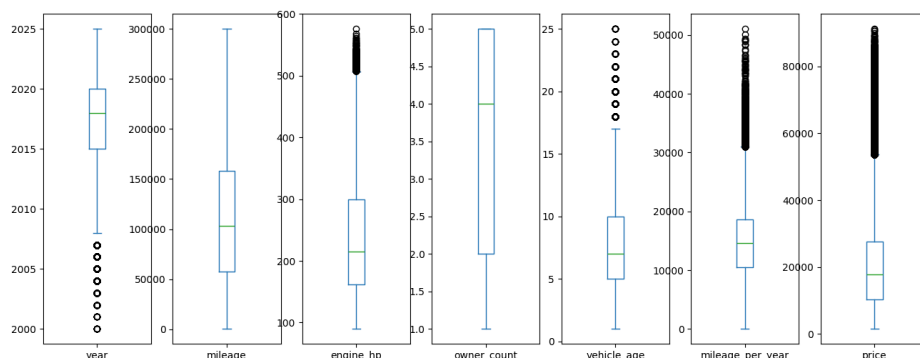
Foram feitos os seguintes passos para a exploração de dados:

- **Visualização dos dados em tabela**
  - Em python foi feito a leitura do arquivo de dataset para criar uma tabela que ajudaria a entender e visualizar melhor os dados e assim escolher as colunas que acharíamos mais apropriadas para a visualização de dados.
- **Remoção de colunas**
  - Foram removidas colunas que notamos desnecessárias e que não seriam utilizadas em nossas visualizações.
- **Observação da quantidade de dados**
  - Foi verificado a quantidade de colunas e linhas que estão no dataset para serem utilizadas.
- **Verificação de valores faltantes**
  - Foi identificado os dados das colunas que estavam faltando, com isso foi notado que somente a coluna **accident\_history** apresentava dados **NaN**, com isso foi feita uma análise para ver o que essa coluna continha, percebemos que nela havia o preenchimento apenas com dois tipos de valores, sendo eles: **Minor** e **Major**, ao perceber isso em primeira mão sugerimos preencher as colunas com o valor **NÃO HOUVE ACIDENTES**, mas decidimos que não há como saber se realmente o valor seria esse, a partir disso preenchemos os valores faltantes dessa colunas com **N.A.**
- **Descrição dos dados numérico**
  - Foi utilizado no código em python o comando **.describe()** para visualizar todos os dados que são numéricos da base de dados, com isso recebemos informações dos **quartis, média, valor máximo, valor mínimo e desvio padrão** das seguintes colunas: **year, mileage, engine\_hp, owner\_count, vehicle\_age, mileage\_per\_year, price.**
- **Verificação de dados duplicados**

Queríamos verificar se na base continha dados duplicados e notamos que não teve.
- **Demonstrações de gráficos para aprendizado**

Fizemos algumas análises dos dados para aprendermos mais sobre ele, em primeira mão fizemos um **plot** dos dados pegos pela função **.describe()** com isso conseguimos ver se o gráfico continha outliers e vimos que não continha pois muitos haviam dados fora do limite máximo e mínimo.

Fizemos um gráfico para verificar os dez carros mais caros e os dez carros mais baratos, também foi feito um gráfico de barras para verificar a quantidade de carros baseado no ano de fabricação e por fim para a nossa análise primária fizemos outro gráfico de barras observando o preço médio por ano do veículo.



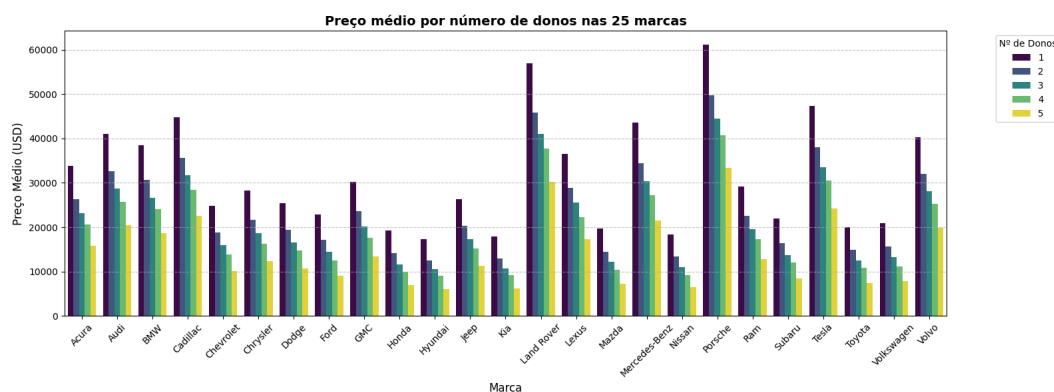
### 3. Visualização de dados

#### Análise do Impacto do Número de Donos no Preço por Marca

Como o número de proprietários anteriores influencia o preço médio dos veículos nas principais marcas do mercado ?

#### Propósito e Benefícios

- Identificar o padrão de desvalorização baseado no histórico de propriedade por marca
- Compreender a percepção de valor do consumidor em relação a veículos com múltiplos donos
- Estabelecer métricas de precificação que considerem o fator "número de proprietários"
- Orientar estratégias de compra e venda baseadas na sensibilidade de cada marca à rotatividade de donos



## **Ações Recomendadas Para Compradores**

- Priorizar veículos com menor número de donos em marcas premium onde este fator impacta significativamente o preço
- Avaliar o trade-off entre preço e número de donos em marcas onde a diferença é menos acentuada
- Solicitar histórico completo de propriedade como parte do processo de due diligence

## **Para Revendedores e Concessionárias**

- Destacar veículos com único dono como valor agregado, especialmente em marcas de luxo
- Desenvolver certificações de procedência para veículos com histórico de propriedade limitado
- Estabelecer políticas de precificação diferenciadas baseadas no número de proprietários anteriores

## **Para Avaliadores e Financiadoras**

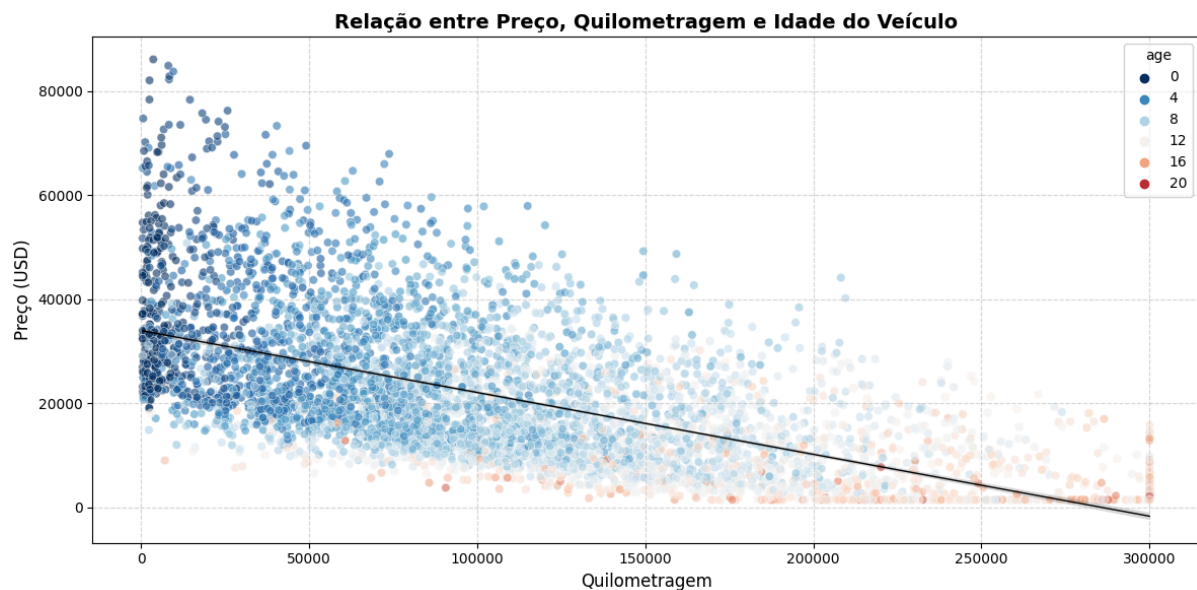
- Incorporar o fator "número de donos" nos modelos de valuation para marcas sensíveis a este critério
- Ajustar limites de financiamento considerando a desvalorização por rotatividade de propriedade
- Desenvolver scores de risco baseados na estabilidade de propriedade do veículo

## **Análise da Relação entre Preço, Quilometragem e Idade dos Veículos**

Como a quilometragem e a idade do veículo se relacionam com seu preço de mercado, e qual o impacto relativo de cada fator na desvalorização?

## **Propósito e Benefícios**

- Compreender a dinâmica de desvalorização baseada nos dois principais fatores: idade e uso
- Identificar padrões de mercado na relação entre quilometragem acumulada e valor residual
- Estabelecer métricas de depreciação para diferentes perfis de uso e idade
- Otimizar estratégias de compra e venda baseadas em evidências de correlação multivariada



## Ações Recomendadas

### Para Compradores

- Priorizar veículos com baixa quilometragem relativa à idade para melhor custo-benefício
- Evitar veículos com quilometragem excessivamente alta mesmo que recentes
- Considerar a combinação idade/quilometragem como indicador mais preciso que cada fator isoladamente

### Para Vendedores e Concessionárias

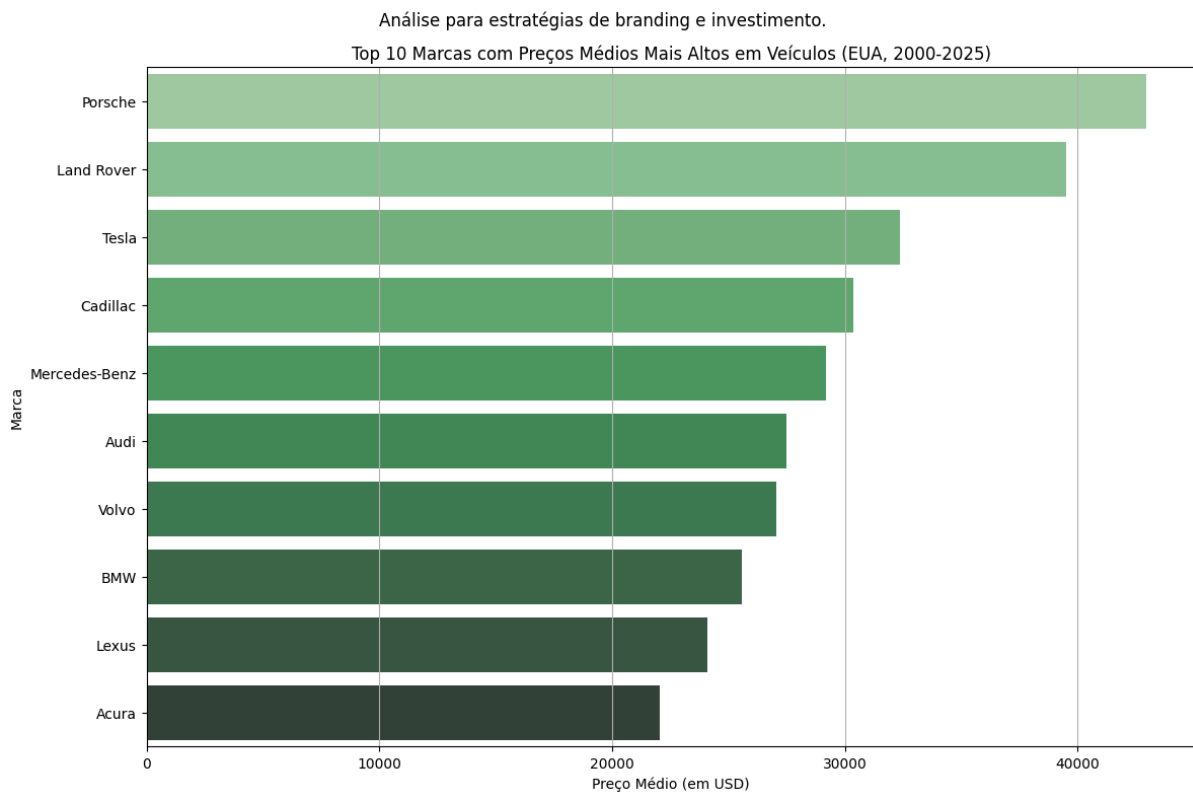
- Destacar a relação favorável idade-quilometragem como argumento de venda
- Desenvolver tabelas de precificação que considerem a interação entre os dois fatores
- Segmentar estoque por perfis de uso (baixa km/idade avançada vs alta km/idade reduzida)

## Análise de Marcas de Veículos por Preço Médio (2000-2025)

**Quais marcas de veículos possuem os preços médios mais altos no mercado (2000-2025)?**

### Propósito e Benefícios

- Guiar investidores e concessionárias a identificar marcas de alto valor para maximizar retornos
- Otimizar estratégias de precificação com base no posicionamento real de mercado das marcas
- Auxiliar na tomada de decisão para aquisição e revenda de veículos



## Ações Recomendadas

### Estratégias Imediatas

- Priorizar investimento em marcas de luxo identificadas no topo do ranking (Porsche, Land Rover, etc.)
- Desenvolver estratégias de marketing para associar a imagem da concessionária com marcas premium

### Planejamento de Longo Prazo

- Estabelecer parcerias com marcas de alto valor para ampliar portfólio
- Criar programas de certificação para veículos de marcas premium que justifiquem preços elevados
- Diversificar estoque com foco em marcas que mantêm melhor valor de revenda

## 4 Desafios e aprendizagens

Nossos desafios vieram de como explorar os dados e como tratá-los da maneira correta para serem utilizados, com isso tivemos que seguir passos comuns de toda exploração, procurando e pesquisando para saber quais são os passos padrões a se seguir.

Também tivemos desafios na elaboração das perguntas, basicamente todas elas eram mais voltadas ao preço pois ele seria a maior influência que seria repassada para clientes futuros, como as perguntas eram mais voltadas ao preço



ficamos empacados em elaborar novas perguntas que exemplifica-se o preço de formas diferentes comparando com atributos diferentes.

## **5. Conclusão**

A análise identificou padrões relevantes de preço no mercado de veículos usados, especialmente no recorte por marca e tipo de combustível. Os achados reforçam oportunidades 10 estratégicas para diferentes stakeholders do setor automotivo, ao mesmo tempo em que revelam limitações que podem orientar estudos futuros. Reitera-se o papel da análise de dados como vetor de competitividade e suporte decisório em um mercado altamente dinâmico.