



Bootcamp Data Analytics 2025

Challenge Correlação e Regressão



Desafio 4: Venda de Veículos Usados

Dificuldade: Média.

Você trabalha em uma empresa de revenda de veículos usados. O sucesso da empresa depende fortemente da precificação adequada dos veículos. Caso o preço seja muito alto, o carro não vende. Se for muito baixo, a empresa perde dinheiro.

Como parte da equipe de dados, você recebeu uma base de dados com informações sobre carros vendidos nos últimos anos, com o objetivo de entender quais fatores mais impactam no preço de venda.

A empresa espera que você identifique as variáveis mais relevantes e proponha uma análise baseada em correlações e modelos preditivos simples.

Base de Dados: desafio_carros_usados.csv

- **id:** Identificador único do veículo
- **make:** Marca do carro (ex: Ford, Toyota)
- **model:** Modelo do carro
- **year:** Ano de fabricação
- **price:** Preço de venda do carro
- **mileage:** Quilometragem (km rodados)
- **engine_size:** Tamanho do motor (em litros)
- **fuel_type:** Tipo de combustível (gasolina, diesel, elétrico)
- **transmission:** Tipo de transmissão (manual, automática)
- **doors:** Número de portas
- **color:** Cor do carro
- **tax:** Taxa anual de imposto veicular

- **mpg:** Milhas por Galão(indicador de eficiência de combustível)
- **sold_date:** Data de venda do veículo

Tarefa:

1. Análise de Correlação

- Calcule a correlação entre as variáveis numéricas e o preço do carro (**price**).
- Quais variáveis estão mais correlacionadas com o preço?
- Quais estão menos correlacionadas?

2. Análise das 5 Variáveis Mais Correlacionadas

Para as cinco variáveis com maior correlação com o preço:

1. Plote **histograma** e **boxplot** de cada variável.
2. Plote o **scatterplot** (gráfico de dispersão), com **price** no eixo Y e a variável no eixo X.
3. Faça uma **regressão linear simples**, utilizando cada variável como preditora (X) e o preço como resposta (Y). Interprete os coeficientes e o R^2 .