

# EFICIÊNCIA AUTOMOTIVA

**Pitch | Avaliação do módulo 3**

Gabriella Regina Neres de Souza  
Data Science



# O PROBLEMA

Com o constante aumento no preço dos combustíveis, um desafio crescente no setor automotivo é encontrar formas de melhorar a eficiência e reduzir os custos com transporte.

## Público Alvo

O público alvo são os consumidores que enfrentam dificuldade em escolher veículos que equilibrem economia de combustível, desempenho e custo-benefício.

Ao mesmo tempo, fabricantes lidam com o desafio de desenvolver carros que atendam os consumidores sem comprometer a potência ou o design considerando a necessidade de reduzir custos e emissões.



**Referência:** Dado sobre consumo de combustível citado - [www.cnnbrasil.com.br](http://www.cnnbrasil.com.br)



# PROPOSTA DE SOLUÇÃO

O objetivo deste projeto é tratar, extrair e analisar informações de um dataset para entender se há um padrão ocorrido entre as features, prever quais são os principais focos que os fabricantes automotivos devem seguir para obter um resultado crescente e constante nas vendas e satisfazer os consumidores economicamente validando a eficiência dos veículos.

Foi realizado um processo de análise exploratória e preditiva utilizando o dataset MPG (Milhas por galão) da biblioteca seaborn que possui informações do gasto de combustível em relação a fatores importantes que possam contribuir com este gasto como potência, peso, aceleração entre outros.

## Metadado:

- **MPG:** Milhas por galão (alvo da regressão linear).
- **Cylinders:** Número de cilindros.
- **Displacement:** Volume do motor.
- **Horsepower:** Potência do motor.
- **Weight:** Peso do carro.
- **Acceleration:** Aceleração.
- **Model Year:** Ano do modelo.
- **Origin:** Local de fabricação.



# IMPORTAÇÃO E ANÁLISE INICIAL

## Informações iniciais:

- O dataset possui 398 entradas
- Duas colunas são strings (Nome e Origem)
- e as outras são floats
- Análise estatística descritiva - A média dos carros possuem gasto de 23 mpg

## Limpeza e tratamento de dados:

- Remoção de dados nulos encontrados na coluna 'Horsepower'
- Remoção da coluna 'Name'
- Tratamento da coluna 'Origin' para binários

```
# Importação das bibliotecas
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Carregando dataset seaborn
df = sns.load_dataset('mpg')

# Visualização os dados
df.head()
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	name
0	18.0	8	307.0	130.0	3504	12.0	70	usa	chevrolet chevelle malibu
1	15.0	8	350.0	165.0	3693	11.5	70	usa	buick skylark 320
2	18.0	8	318.0	150.0	3436	11.0	70	usa	plymouth satellite
3	16.0	8	304.0	150.0	3433	12.0	70	usa	amc rebel sst
4	17.0	8	302.0	140.0	3449	10.5	70	usa	ford torino

```
df.describe()
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year
count	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000
mean	23.445918	5.471939	194.411990	104.469388	2977.584184	15.541327	75.979592
std	7.805007	1.705783	104.644004	38.491160	849.402560	2.758864	3.683737
min	9.000000	3.000000	68.000000	46.000000	1613.000000	8.000000	70.000000
25%	17.000000	4.000000	105.000000	75.000000	2225.250000	13.775000	73.000000
50%	22.750000	4.000000	151.000000	93.500000	2803.500000	15.500000	76.000000
75%	29.000000	8.000000	275.750000	126.000000	3614.750000	17.025000	79.000000
max	46.600000	8.000000	455.000000	230.000000	5140.000000	24.800000	82.000000

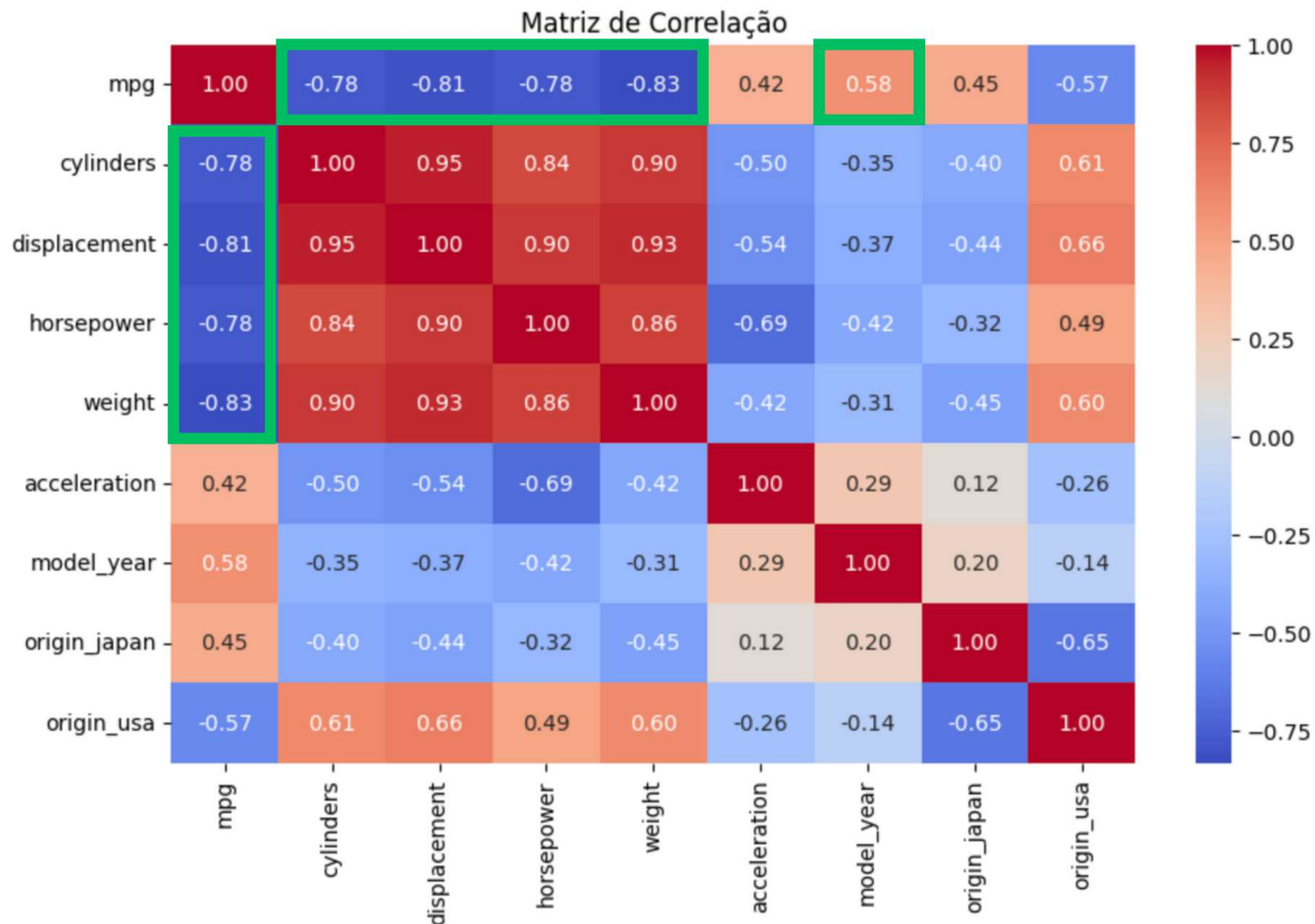




# MATRIZ DE CORRELAÇÃO

- O peso (weight) e a potência (horsepower) têm alta correlação negativa, ou seja, **carros mais pesados e mais potentes, consomem mais combustível.**
- A quantidade de cilindros e volume de motor também possui uma alta correlação negativa.
- O modelo do carro possui uma correlação moderada positiva, ou seja, **modelos mais recentes são mais eficientes.**

```
# Matriz de correlação
plt.figure(figsize=(10, 6))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Matriz de Correlação")
plt.show()
```

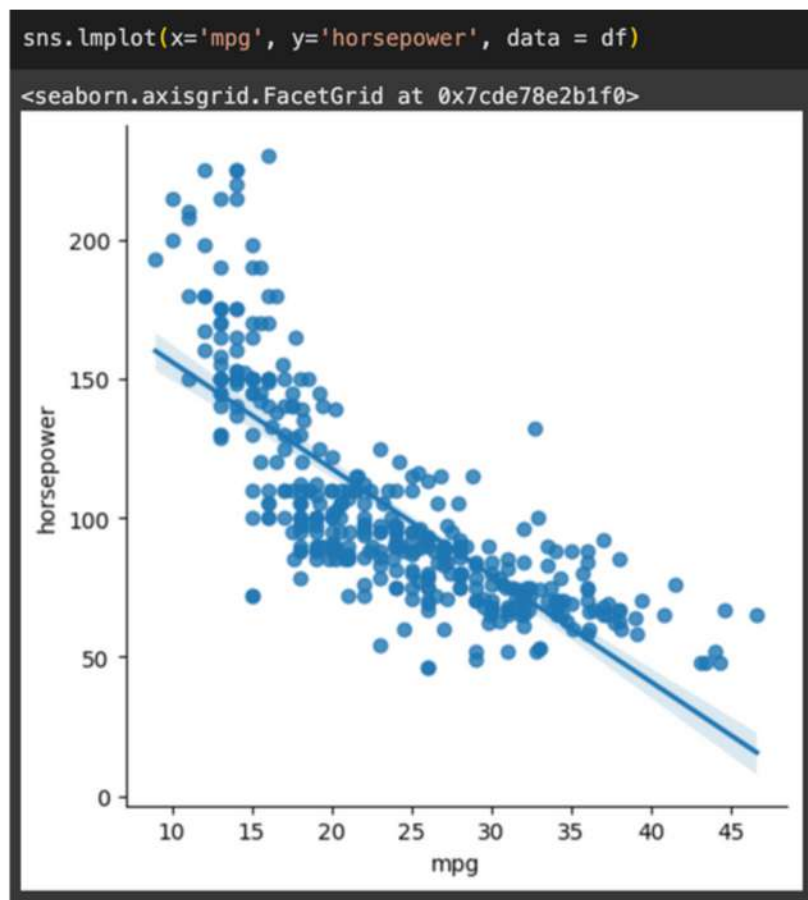


# VISUALIZAÇÃO DOS DADOS

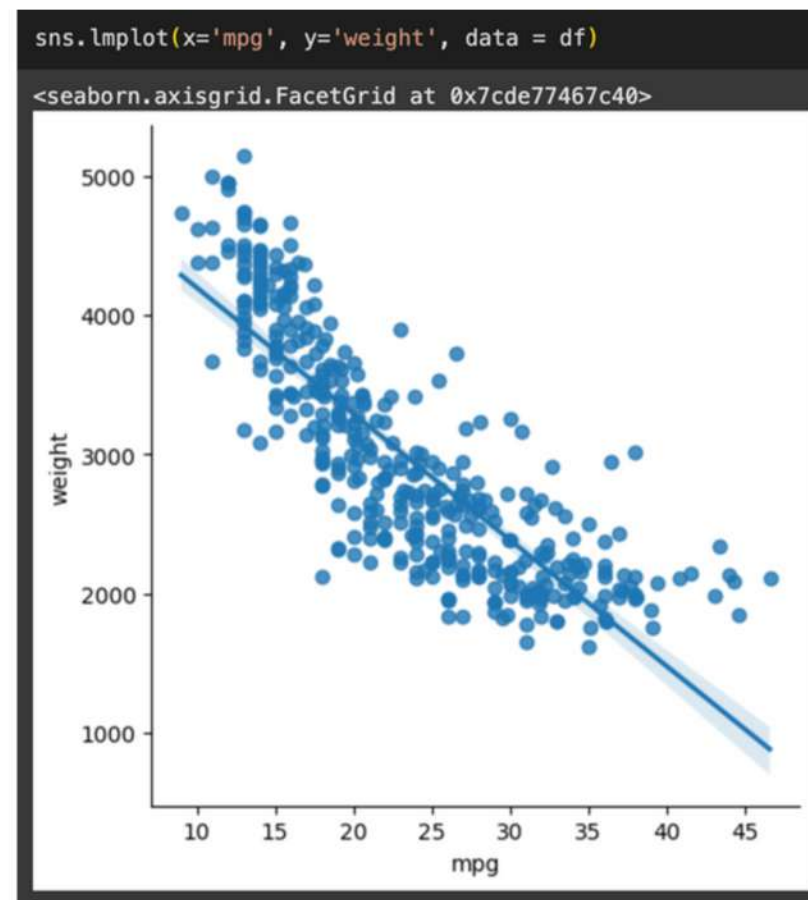
Utilizando `seaborn.pairplot` é possível visualizar relações lineares negativas entre MPG x Potência e MPG x Peso. Também há uma relação linear positiva entre MPG x Aceleração.

Apesar de analisar as variáveis com base no mpg, achei interessante visualizar o gráfico entre 'weight' e 'horsepower'. O que significa que o peso está diretamente ligada a potência de um carro, tendo um padrão linear positivo.

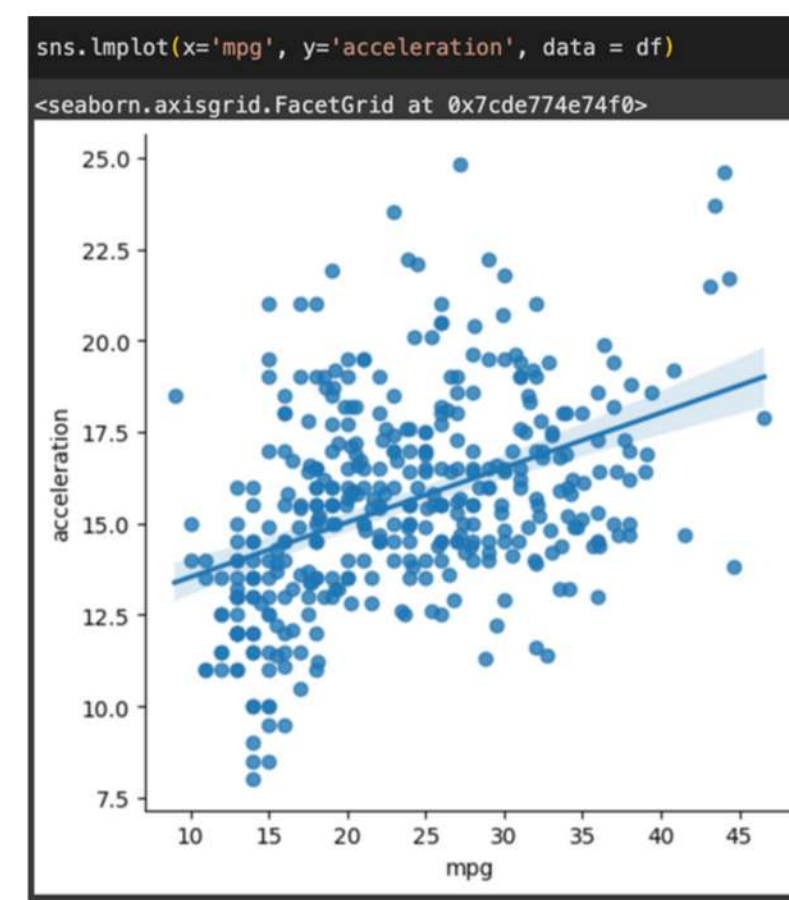
**MPG x Potência**



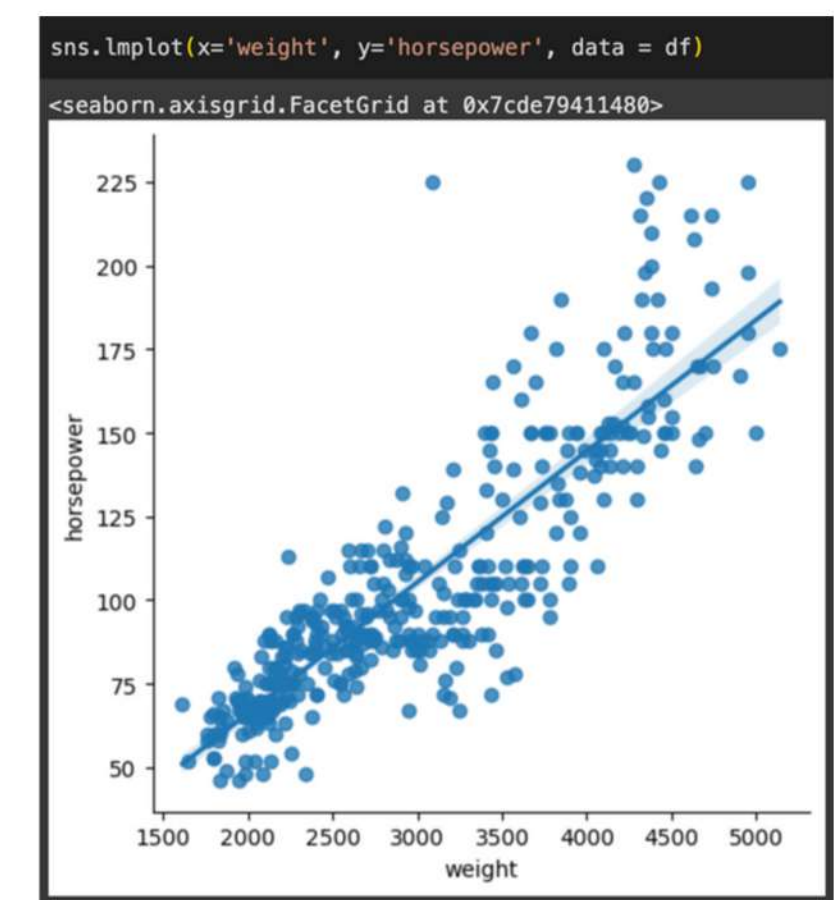
**MPG x Peso**



**MPG x Aceleração**



**Peso x Potência**





# REGRESSÃO LINEAR

Objetivo de validar o teste quais as features que mais afetam o gasto de combustível.

Para entender o quanto as variáveis (X) estão relacionadas e quanto cada uma delas contribui para o valor final de mpg (y).

O modelo foi separado em 80% de treino e 20% teste.

## Analizando os coeficientes:

Quanto mais cilindros, potência e peso, a eficiência do mpg diminui.

A cada aumento de uma unidade de volume do motor, potência, aceleração e ano, o mpg também aumenta.



```
# Separando variáveis
X = df[['cylinders', 'displacement', 'horsepower', 'weight', 'acceleration', 'model_year']]
y = df['mpg']

# Dividindo os dados em treino e teste
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

print(f"Tamanho do treino X: {X_train.shape}, Tamanho do teste X: {X_test.shape}")
```

Tamanho do treino X: (313, 6), Tamanho do teste X: (79, 6)

```
# Criação do modelo de Regressão Linear
model = LinearRegression()

# Treinando o modelo
model.fit(X_train, y_train)

# Coeficientes do modelo
coeficientes = pd.DataFrame(model.coef_, X.columns)
coeficientes.columns = ['Coeficientes']
coeficientes

coeficientes
```

### Coeficientes

	Coeficientes
cylinders	-0.116173
displacement	0.001013
horsepower	-0.002276
weight	-0.006561
acceleration	0.061736
model_year	0.760636



- Para cada aumento de 1 cilindro, o MPG diminui 0.1
- Para cada aumento de 1 volume do motor, o MPG aumenta 0.0001
- Para cada aumento de 1 potência, o MPG aumenta 0.002
- Para cada aumento de 1 peso, o MPG diminui 0.006
- Para cada aumento de 1 aceleração, o MPG aumenta 0.06
- Para cada aumento de 1 ano do modelo, o MPG aumenta 0.76

# AVALIAÇÃO DO MODELO

Utilizando as métricas RMSE e R squared para verificar se o modelo foi bem treinado.

- O modelo consegue prever o consumo de combustível com um desvio médio de aproximadamente 3 milhas por galão (RMSE)
- 79% da variabilidade do mpg é explicada pelo modelo ( $R^2$ )

O modelo poderia ter resultados melhores de  $R^2$ , mas é possível ter uma boa visão dos resultados deste modelo.

```
# Previsões
y_pred = model.predict(X_test)

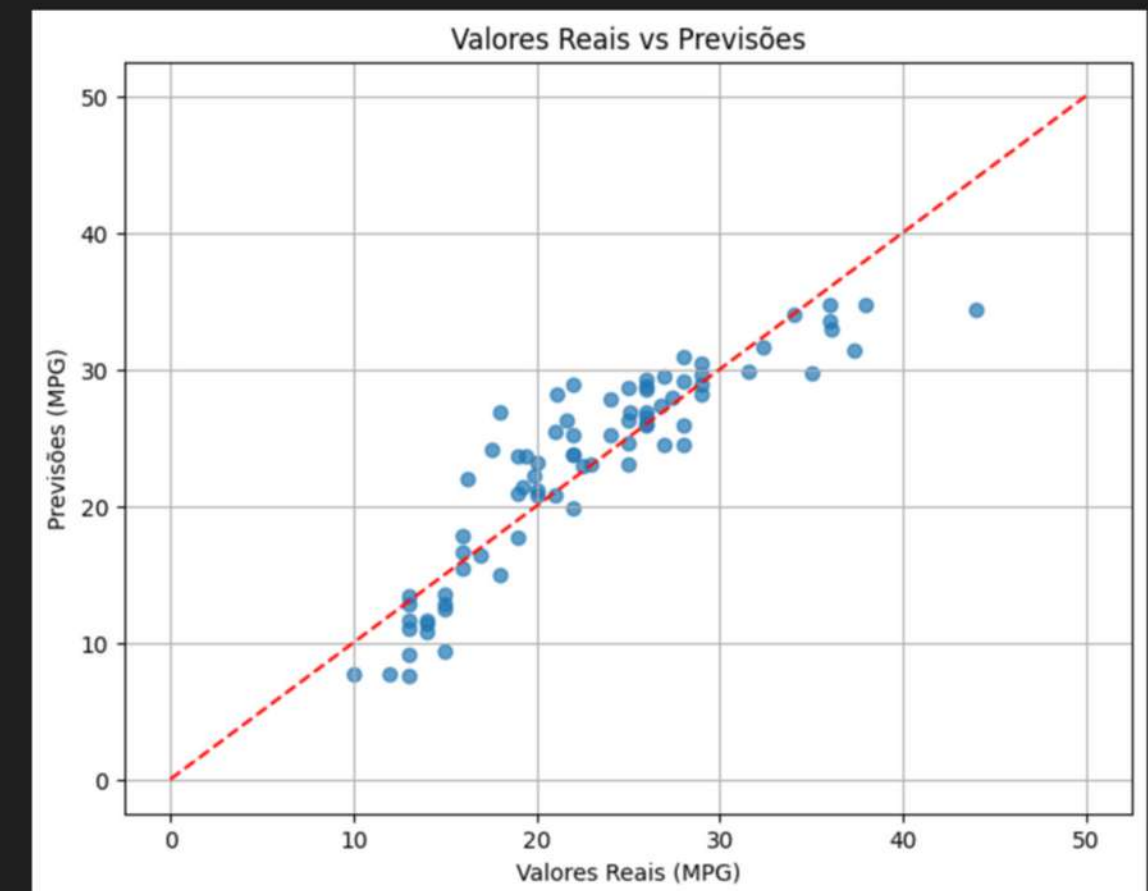
# Avaliação do modelo
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)

print(f"RMSE: {rmse:.2f}")
print(f"R²: {r2:.2f}")
```

RMSE: 3.24  
R²: 0.79

## PREVISÕES x VALORES REAIS

```
#Gráfico de Dispersão: Previsões vs Valores Reais
plt.figure(figsize=(8, 6))
plt.scatter(y_test, y_pred, alpha=0.7)
plt.plot([0, 50], [0, 50], color='red', linestyle='--') # Linha ideal
plt.xlabel("Valores Reais (MPG)")
plt.ylabel("Previsões (MPG)")
plt.title("Valores Reais vs Previsões")
plt.grid(True)
plt.show()
```





# RESULTADO

- O ano do modelo é o principal fator que impacta o consumo de combustível.
- Tamanho e peso do carro reduzem a eficiência do combustível. Quanto mais cilindros, peso e potência, mais potentes e mais pesados, consumindo mais combustível.
- Na aceleração, o coeficiente apresentou um impacto positivo, ou seja, velocidades maiores reduzem o consumo do combustível.

## Conclusão de negócio:

Para os consumidores o ideal é investir em carros mais leves, compactos e com poucos cilindros que consomem menos combustível mas não prejudiquem sua eficiência. **Como por exemplo (Ver se consigo pegar quais carros tem melhores resultados no dataset)**

Para os fabricantes, ao apostar em carros mais sofisticados, o ideal é usar tecnologias mais avançadas com motores potentes que possam otimizar a eficiência do combustível de carros maiores. **Como por exemplo (Ver se consigo pegar quais carros tem melhores resultados no dataset)**

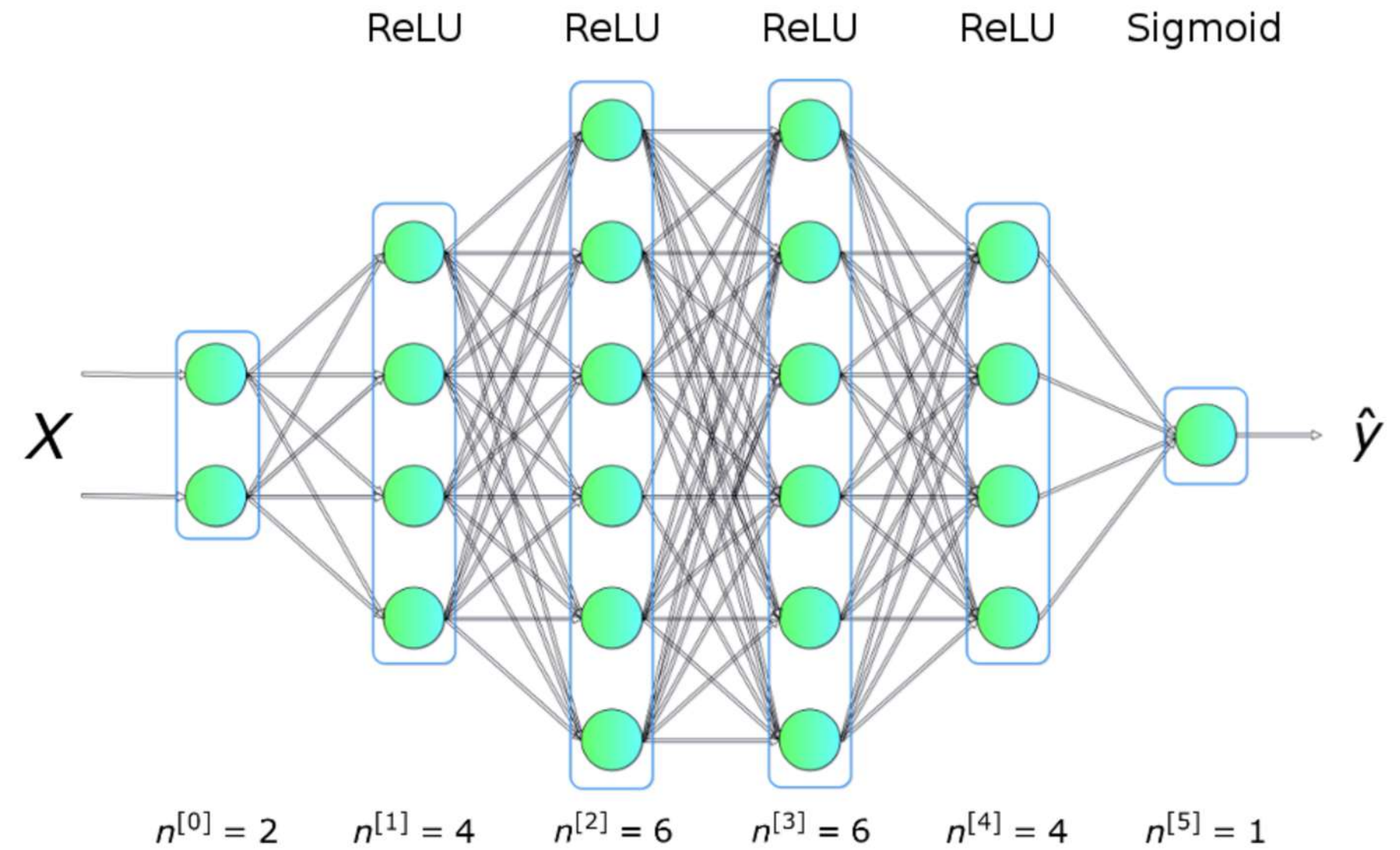




# FUTURO MODELO

Como o  $R^2$  teve um resultado bom mas que poderia ser melhor (explicou 79% da variabilidade do mpg).

Para aprofundar no assunto, seria possível usar **redes neurais** criando um modelo com camadas densas, ativação ReLU nas camadas ocultas e ajustando os hiperparâmetros, com o objetivo de classificar carros como econômicos ou não.





# Obrigada!