

Data Analytics Challenge

Olá Estudante!

Seja muito bem-vindo ao seu desafio de Data Analytics!

Seu desafio está em resolver o seguinte problema:

Você enquanto analista de dados de uma empresa, recebeu a missão de fazer uma análise exploratória (EDA) para que você gerasse insights relevantes na industria automobilística.

Como treino, você recebeu o [Auto MPG Data Set](#) com informações relevantes do consumo de milhas por galão de determinados automóveis.

A descrição das colunas do dataset pode ser encontrada [aqui](#).

Através desses dados, você terá que responder 20 perguntas de múltipla escolha abaixo, que envolvem : análise exploratória, programação e estatística.

OBS: Nem todas as perguntas são sobre dataset em si, pode ocorrer questões mais teóricas envolvendo conceitos fundamentais de análise exploratória, programação e estatística

matheushpavani@gmail.com [Alternar conta](#)



Não compartilhado



* Indica uma pergunta obrigatória

Nome Completo *

Sua resposta



1. Qual a dimensão do dataset inicialmente em (linhas, colunas)? *

1 ponto

- ☐ (390,9)
- ☐ (406,9)
- ☐ (406,8)
- ☐ (405,9)
- ☐ (390,7)

2. Em uma análise de dados, é muito comum fazermos as limpezas dos valores nulos para que consigamos analisar principalmente dados numéricos, onde podemos substituir os valores nulos por algum valor que deixe os dados mais próximos da realidade, ou simplesmente excluir as linhas onde eles estão. Se optarmos excluir as linhas com valores nulos em qualquer coluna para que possamos, nosso dataset passará a ter quantas linhas?

* 1 ponto

- ☐ 401
- ☐ 400
- ☐ 392
- ☐ 350
- ☐ 399



3. Se quisermos carregar o dataset via python, precisaremos da biblioteca "pandas". Qual a sintaxe correta para essa ação? * 1 ponto

- ☐ `pd.to_excel('auto-mpg-original.csv', sep = ';')`
- ☐ `pd.to_csv('auto-mpg-original.csv', sep = ';')`
- ☐ `pd.read_csv('auto-mpg-original.csv', sep = ' ', '')`
- ☐ `pd.read_excel('auto-mpg-original.csv', sep = ' ', '')`
- ☐ `pd.read_csv('auto-mpg-original.csv', sep = ' ; ')`

4. Olhando para o ano de fabricação, qual ano corresponde a carros com maior economia de milhas por galão (mpg) ? * 1 ponto

- ☐ 78
- ☐ 75
- ☐ 72
- ☐ 80
- ☐ 79

5. No intervalo interquartil, diga o valor para o 3º quartil da coluna "horsepower" * 1 ponto

- ☐ 840
- ☐ 650
- ☐ 320
- ☐ 780
- ☐ 575



6. Após a limpeza dos dados(dropar as linhas com valores nulos), quantos * 1 ponto
tipos de carros distintos nós temos em nossa base?

- ☐ 300
- ☐ 301
- ☐ 302
- ☐ 310
- ☐ 305

7. Qual o desvio padrão aproximado da coluna "mpg"? *

1 ponto

- ☐ 27.2
- ☐ 10.2
- ☐ 7.8
- ☐ 23.4
- ☐ 25

8. Qual o valor da correlação entre as colunas "cylinders" e "weight"? Obs: * 1 ponto
Considere o coeficiente de correlação de Pearson

- ☐ 0.4
- ☐ 0.7
- ☐ 0.2
- ☐ 0.5
- ☐ 0.9



9. Qual a função nativa do pandas que modifica o tipo de dado um uma coluna do dataframe?

* 1 ponto

- ☐ dtype()
- ☐ astype()
- ☐ type()
- ☐ to_type()
- ☐ value()

10. O z-score é uma métrica estatística fundamental para análise de dados, e sua equação é dada abaixo. Qual o significado do z-score?

* 1 ponto

$$z = \frac{x - \mu}{\sigma}$$

- ☐ Determinar o tipo de distribuição do dado
- ☐ Identificar Outliers
- ☐ Calcular a discrepância estatística
- ☐ Verificar o quão distante em desvios padrões os dados estão em relação à média.
- ☐ Achar uma tendência média em torno da variância

11. Quando fazemos a correlação da variáveis, os valores negativos representam o que?

* 1 ponto

- ☐ A relação das variáveis são inversamente proporcionais
- ☐ A relação das variáveis são diretamente proporcionais
- ☐ A relação das variáveis possuem correlações fracas
- ☐ Todas as variáveis estão correlacionadas diretamente



12. Quando montamos um gráfico para comparar qual o ano dos carros apresentaram uma maior economia no mpg, qual ano mostrou o maior desempenho?

* 1 ponto

- ☐ 70
- ☐ 82
- ☐ 75
- ☐ 80

13. Os valores nulos dentro de um conjunto de dados sempre atrapalham na visualização e na compreensão das nossas análises. Qual função do pandas identifica que há dados nulos no dataset?

* 1 ponto

- ☐ .isnull()
- ☐ .not_null()
- ☐ <>
- ☐ .to_null()

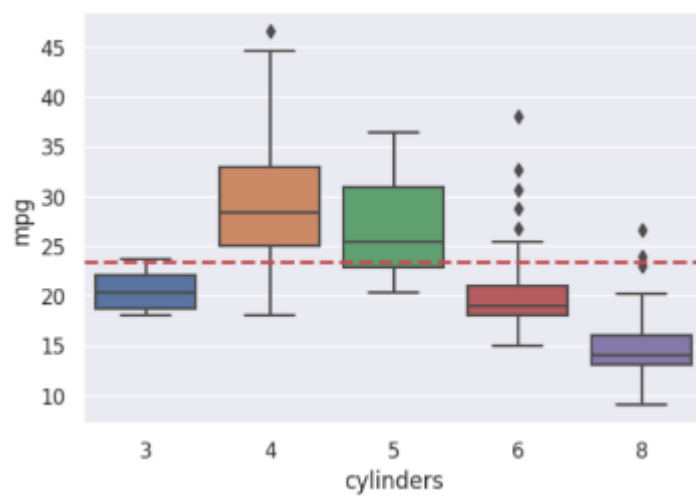
14. Muitas das vezes quando precisamos entender a tendencia de como está sendo a evolução dos carros analisados, utilizamos um determinado tipo de gráfico. Qual gráfico utilizamos?

* 1 ponto

- ☐ Dispensão
- ☐ Linha
- ☐ Barra
- ☐ Gráfico de Caixa



15. Observando o gráfico abaixo, qual das cilindradas há a maior economia de combustível? * 1 ponto



- ☐ 8
- ☐ 6
- ☐ 4
- ☐ 3



16. Olhando para o código em python, o que está sendo codado em nosso notebook? * 1 ponto

```
plt.figure(figsize=(24,20))

plt.subplot(4, 2, 1)
fig = df['mpg'].hist(bins=20)
fig.set_xlabel('mpg')
fig.set_ylabel('Número de casos')

plt.subplot(4, 2, 2)
fig = df['cylinders'].hist(bins=20)
fig.set_xlabel('cylinders')
fig.set_ylabel('Número de casos')

plt.subplot(4, 2, 3)
fig = df['displacement'].hist(bins=20)
fig.set_xlabel('displacement')
fig.set_ylabel('Número de casos')

plt.subplot(4, 2, 4)
fig = df['horsepower'].hist(bins=20)
fig.set_xlabel('horsepower')
fig.set_ylabel('Número de casos')

plt.subplot(4, 2, 5)
fig = df['weight'].hist(bins=20)
fig.set_xlabel('weight')
fig.set_ylabel('Número de casos')

plt.subplot(4, 2, 6)
fig = df['acceleration'].hist(bins=20)
fig.set_xlabel('acceleration')
fig.set_ylabel('Número de casos')

plt.subplot(4, 2, 7)
fig = df['model year'].hist(bins=20)
fig.set_xlabel('model year')
fig.set_ylabel('Número de casos')

plt.subplot(4, 2, 8)
fig = df['origin'].hist(bins=20)
fig.set_xlabel('origin')
fig.set_ylabel('Número de casos')
```

- ☐ Construção de um modelo de Machine Learning de Regressão
- ☐ Construção de gráficos de barras
- ☐ Verificando há existências de valores correlatos e separando um grande grupos





Construção de um modelo de Deep Learning - redes convolucionais

17. No python, se quiser colocar uma coluna a mais para dentro do dataframe e adicionar o nome do país nesta coluna, como fazer?

* 1 ponto

- ☐ `df['cod_pais'] = df.origin.add([1,2,3],['EUA','Europa','Japão'])`
- ☐ `df.add['cod_pais'] = df.replace([1,2,3],['EUA','Europa','Japão'])`
- ☐ `df['cod_pais'] = df.origin.replace([1,2,3],['EUA','Europa','Japão'])`
- ☐ `df['cod_pais'] = df.origin([1,2,3],['EUA','Europa','Japão'])`

18. Qual a função nativa do pandas que faz a contagem distinta de valores em uma coluna ou no dataframe todo? *

1 ponto

- ☐ `sort()`
- ☐ `agg()`
- ☐ `sum()`
- ☐ `value_counts()`
- ☐ `count()`

19. Dentro do nosso dataset, queremos saber quantos carros temos e iremos analisar. Quantos carros temos no nosso conjunto de dados?

* 1 ponto

- ☐ 312
- ☐ 83
- ☐ 406
- ☐ 13



20. Qual a função do pandas que utilizamos para evitar que a o índice de um dataframe vire uma coluna após uma modificação? * 1 ponto

- ☐ reset_index(drop=True)
- ☐ remove_index(True)
- ☐ drop_index(drop_index=True)
- ☐ sort_index(True)

Enviar

Limpar formulário

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



