

DESAFIO PYTHON PANDAS

Bootcamp Data Analytics
2024

criado por:



Desafio 1 - Importando dataset pelo pydataset

Definição do problema: Dado o código do dataset retorne as seguintes informações:

1. Importe o *dataset* utilizando a seguinte função do pydataset: *data("Código")*
2. Imprimir na tela o dataset;
3. Informe o tipo de dados retornado pela função *data*;
4. Informe o número de exemplos (linhas) e características (colunas) do *dataset*.
5. Crie uma função que ao receber um DataFrame retorna o número de linhas e colunas.

Código do dataset: plantTraits

Dicas:

- plantTraits é um dataset que contém a descrição, por meio de atributos biológicos, de diferentes espécies de plantas;
- função built-in (nativa da linguagem Python): *type()*
- atributo do DataFrame: *shape*



Desafio 2: Conversão

Definição do problema: Dada a tabela com a média de altura e peso de mulheres estadunidenses, onde os nomes das colunas estão em inglês *height* (altura) e *weight* (peso) e seus respectivos valores estão nas seguintes unidades: polegada e libras.

1. Construa uma nova representação onde os nomes das colunas estão em português, e os valores de altura são convertidos para centímetros e os valores de peso convertidos para quilograma.
2. Plote o scatterplot de altura x peso da nova representação.

Código do dataset: women

Dicas:

1 polegada = 2,54 centímetros

1 libra = 0,45 quilos

use o método `.plot.scatter()`



Desafio 3: Titanic

Definição do problema: Dado que você está trabalhando com um grupo de historiadores que estão analisando o naufrágio do Titanic, informe aos historiadores as seguintes questões?

1. Quantas pessoas sobreviveram e quantas morreram?
2. Quantas pessoas eram da primeira, segunda e terceira classe?
3. Qual a porcentagem de sobreviventes na primeira classe?
4. Qual a porcentagem de mortos da terceira classe?

Código do dataset: titanic

Dicas:

- `value_counts()` (Itens 1 e 2)
- seleção de subconjuntos (Itens 3 e 4)



Desafio 4: Futebol

Definição do problema: Dado que você está trabalhando numa emissora de televisão, o narrador precisa informar o jogo com o maior número de gols na história da Bundesliga (Campeonato Alemão). Você precisa fazer uma busca em todos os jogos da história deste campeonato e verificar qual jogo é esse e passar a informação ao narrador, para posteriormente, o narrador informar aos telespectadores.

1. Verificar os jogos com maior quantidade de gols na história da Bundesliga
2. Informar a data, os times participantes e o placar do jogo

Código do dataset: Bundesliga

Dicas:

- Criar uma nova variável "TotalGols" com o total de gols da partida;
- Use o método.max()



Desafio 5: Forbes

Definição do problema: Você está trabalhando em um jornal e o editor quer que você responda as seguintes questões:

1. Qual a empresa mais valiosa da lista da Forbes 2000?
2. Construa uma tabela com as TOP 10 empresas mais lucrativas da lista.
3. Qual a média de valores de empresa das cinco categorias mais valiosas?

Código do dataset: Forbes2000

Dicas:

- Método `.max()`
- Método `nlargest()`
- Método `.groupby()`



**Juntas, somos mais
fortes!**

