

DESAFIO PYTHON PANDAS

Bootcamp Data Analytics 2025

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: iris

Dicas:

- iris é um dataset consiste em 50 amostras de cada uma das três espécies de Iris (Iris setosa, Iris virginica e Iris versicolor).;
- função built-in (nativa da linguagem Python): *type()*
- atributo do DataFrame: *shape*



Desafio 2: PlantGrowth

Definição do problema: Dada a tabela com o peso e grupo das plantas

1. Calcule a média do peso (weight) para cada grupo (group).
2. Verifique se algum grupo tem peso médio maior que 6.
3. Crie uma tabela mostrando o peso máximo e mínimo por grupo.

Código do dataset: PlantGrowth

Dicas:

atributo do DataFrame: groupby, agg



Desafio 3: HairEyeColor

Definição do problema: Dado que você está trabalhando com um grupo de pesquisadores que estão analisando a combinação da cor e dos olhos das pessoas.

1. Calcule o número total de pessoas para cada combinação de cor de cabelo (Hair) e cor dos olhos (Eye).
2. Encontre a cor de cabelo mais comum entre pessoas com olhos castanhos (Brown).
3. Crie uma tabela com a contagem total de pessoas por cor de cabelo.

Código do dataset: HairEyeColor

Dicas:

- atributo do DataFrame: `sum`,
`reset_index`



Desafio 4: InsectSprays

Definição do problema: Dado que você está trabalhando num laboratório de inseticidas. Você precisa fazer uma pesquisa em todos os sprays e verificar qual jogo é o mais eficaz.

1. Calcule o número total de insetos mortos (count) por tipo de spray (spray).
2. Identifique qual spray foi o mais eficaz em matar insetos.
3. Calcule a média geral e filtre os dados para mostrar apenas sprays com eficácia acima da média.

Código do dataset: InsectSpray

Dicas:

- Use o método `idmax()`



Desafio 5: Chickwts

Definição do problema: Foi conduzido um experimento para medir e comparar a eficácia de vários suplementos alimentares na taxa de crescimento de galinhas.

Os pintinhos recém-nascidos foram alocados aleatoriamente em seis grupos, e cada grupo recebeu um suplemento alimentar diferente. Seus pesos em gramas após seis semanas são fornecidos junto com os tipos de ração.

1. Calcule o peso médio dos pintinhos (weight) para cada tipo de ração (feed).
2. Identifique a ração que resultou no maior peso médio.
3. Crie uma tabela que mostre o peso máximo, mínimo e médio para cada tipo de ração.

Código do dataset: chickwts

Dicas:

- Método `.reset_index()`
- Método `idxmax()`
- Método `.groupby()`



Desafio 6: Attitude

Definição do problema: De uma pesquisa com funcionários administrativos de uma grande organização financeira, os dados são agregados dos questionários de aproximadamente 35 funcionários para cada um dos 30 departamentos (selecionados aleatoriamente). Os números dão a proporção percentual de respostas favoráveis a sete perguntas em cada departamento.

Dicionário:

- rating: nota geral do departamento
- complaints: fator categórico sobre reclamações
- privileges: acesso a benefícios
- learning: fator categórico sobre aprendizado
- raises: fator categórico sobre aumentos salariais
- critical: fator categórico sobre feedback crítico
- advance: oportunidades de crescimento

1. Calcule a média de "aprendizado" (learning) por nível de motivação (rating).
2. Encontre a combinação de rating e complaints que possui o maior valor médio de aprendizado.
3. Qual é o nível médio de aprendizado para funcionários que receberam aumentos (raises) acima da média?

Código do dataset: Attitude

6

Dicas: Método `.reset_index()`
`groupby`



Desafio 7: 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. Qual a taxa de sobrevivência geral no navio?
2. Qual a taxa de sobrevivência por classe (Pclass)?
3. Qual o sexo com maior taxa de sobrevivência?
4. Baseado nos dados informados. Qual a quantidade de adultos e crianças (sobreviventes) e não sobreviventes ?

Código do dataset: titanic

Dicas:

atributo do DataFrame: shape

Método map()

Método groupby()



**Juntas, somos mais
fortes!**

