

Exercícios de Pandas em Python

Exercício 1: Leitura e visualização de dados

1. Importe a biblioteca pandas com o alias 'pd'.
2. Carregue um arquivo CSV chamado 'dados.csv' em um DataFrame chamado 'df'.
3. Visualize as primeiras 5 linhas do DataFrame 'df'.
4. Verifique o número de linhas e colunas do DataFrame.
5. Exiba informações sobre as colunas do DataFrame, como nome, tipo e quantidade de valores não nulos.

Exercício 2: Manipulação de dados

1. Remova todas as linhas que possuem valores ausentes (NaN) no DataFrame 'df'.
2. Substitua os valores ausentes nas colunas numéricas pela média dos valores presentes.
3. Crie uma nova coluna chamada 'total_vendas', que será a soma das colunas 'vendas_janeiro', 'vendas_fevereiro' e 'vendas_março'.
4. Ordene o DataFrame em ordem decrescente com base na coluna 'total_vendas'.
5. Renomeie as colunas 'vendas_janeiro', 'vendas_fevereiro' e 'vendas_março' para 'jan', 'fev' e 'mar', respectivamente.

Exercício 3: Filtragem e seleção

1. Selecione todas as linhas do DataFrame 'df' em que a coluna 'idade' seja maior que 30.
2. Filtre o DataFrame para exibir apenas as colunas 'nome', 'cidade' e 'idade'.
3. Selecione todas as linhas em que o nome comece com a letra 'A'.
4. Filtre o DataFrame para exibir apenas as linhas em que a cidade seja 'São Paulo' e a idade esteja entre 25 e 35 anos.

Exercício 4: Agrupamento e agregação

1. Agrupe o DataFrame 'df' por cidade e calcule a média das colunas numéricas para cada grupo.
2. Encontre a idade média das pessoas em cada cidade.
3. Calcule a soma total das vendas por cidade.
4. Encontre o nome da pessoa mais jovem e da mais velha em cada cidade.

Exercício 5: Visualização de dados

1. Plote um gráfico de barras mostrando a quantidade de pessoas em cada cidade.
2. Crie um gráfico de linha para mostrar as vendas ao longo dos meses (jan, fev, mar) para uma pessoa específica.
3. Plote um gráfico de dispersão mostrando a relação entre a idade e o total de vendas.
4. Crie um gráfico de pizza para mostrar a proporção das vendas de cada mês em relação ao total.

Exercício 6: Outras operações

1. Exporte o DataFrame 'df' para um arquivo Excel chamado 'dados_processados.xlsx'.

2. Calcule a mediana das vendas para o mês de janeiro.
3. Calcule o desvio padrão da idade das pessoas no DataFrame.
4. Crie uma coluna chamada 'faixa_etaria' que categorize as pessoas em "jovem" (< 30 anos), "adulto" (30-60 anos) e "idoso" (> 60 anos).

Exercício 7: Tratamento de Dados Temporais

1. Carregue um arquivo CSV chamado 'dados_temporais.csv' em um DataFrame chamado 'df_temp'.
2. Verifique o tipo de dado da coluna que contém as datas e, se necessário, converta-o para o tipo de dado correto (datetime).
3. Crie uma nova coluna chamada 'ano' que contenha o ano de cada data.
4. Crie uma nova coluna chamada 'mes' que contenha o mês de cada data.
5. Agrupe o DataFrame 'df_temp' pelo ano e mês, e calcule a média das colunas numéricas para cada grupo.

Exercício 8: Junção de DataFrames

1. Carregue dois arquivos CSV: 'vendas_por_produto.csv' e 'estoque_por_produto.csv', em DataFrames chamados 'df_vendas' e 'df_estoque', respectivamente.
2. Ambos os DataFrames possuem uma coluna em comum chamada 'produto_id'. Faça a junção (merge) dos DataFrames com base nessa coluna.
3. Se necessário, trate valores ausentes nas colunas resultantes da junção.
4. Crie uma nova coluna chamada 'valor_total_vendido', que será o produto entre as colunas 'quantidade_vendida' e 'preco_unitario'.
5. Ordene o DataFrame resultante em ordem decrescente com base na coluna 'valor_total_vendido'.

Exercício 9: Dados Faltantes e Interpolação

1. Carregue um arquivo CSV chamado 'dados_faltantes.csv' em um DataFrame chamado 'df_faltantes'.
2. Verifique a quantidade de dados faltantes em cada coluna do DataFrame.
3. Preencha os valores faltantes nas colunas numéricas usando a técnica de interpolação linear.
4. Preencha os valores faltantes nas colunas categóricas usando o método 'forward fill' ou 'backward fill'.

Exercício 10: Análise Exploratória de Dados

1. Carregue um conjunto de dados de sua escolha (por exemplo, dados de vendas, dados de clima, dados financeiros, etc.) em um DataFrame.
2. Realize uma análise exploratória dos dados, calculando estatísticas resumidas, visualizando distribuições e plotando gráficos relevantes.
3. Identifique possíveis padrões, outliers ou relações interessantes entre as variáveis presentes nos dados.

Bônus (opcional):

1. Explore a biblioteca seaborn para criar gráficos mais elaborados e estilizados.
2. Utilize técnicas avançadas do pandas para transformar e manipular os dados de formas mais complexas.
3. Pesquise e aplique técnicas de pré-processamento de dados para lidar com situações específicas, como dados desbalanceados, escalonamento, codificação de variáveis categóricas, entre outros.
4. Resolva os exercícios acima usando a função lambda e/ou os métodos apply, map e applymap.

5. Pesquise e aplique outras operações do Pandas não abordadas nos exercícios acima.

Divirta-se praticando Pandas!