

CURSO	CIENCIAS DE DADOS
COMPONENTE CURRICULAR:	PROJETO APLICADO I
TRABALHO REALIZADO:	ANÁLISE DE DADOS DA NETFLIX.py

RELATÓRIO DE ANÁLISE DE DADOS DA NETFLIX:

INTRODUÇÃO:

O código implementado realiza a leitura e análise de uma base de dados contendo informações sobre assinaturas da Netflix. Ele organiza as informações por país e permite ao usuário selecionar qual país deseja analisar. O objetivo é fornecer uma análise estruturada dos dados para facilitar a compreensão do comportamento dos usuários de acordo com diferentes regiões.

Estrutura do Código

1. Bibliotecas Utilizadas

- pandas: Usada para manipulação e análise de dados. O Pandas permite o carregamento da planilha Excel e a estruturação dos dados em um DataFrame.
- openpyxl: Necessária para trabalhar com arquivos Excel (.xlsx) no Python.
- numpy: Biblioteca usada para manipulações matemáticas e computacionais, como cálculo de médias e somas.

2. Leitura da Planilha

- O código utiliza `pd.read_excel()` para carregar o arquivo Excel. Os dados são lidos e armazenados em um DataFrame, que é uma estrutura de dados tabular eficiente para análise e manipulação.

3. Listagem dos Países

- Após a leitura do arquivo, o código extrai a lista única de países disponíveis na base de dados. Isso é feito com `df['País'].unique()`, que identifica os países distintos presentes na coluna "País".

4. Interação com o Usuário

- O programa lista os países disponíveis e solicita ao usuário a escolha de um país para realizar a análise.
- A seleção é feita pelo número correspondente ao país desejado. Se a escolha for inválida (um número fora da lista ou um nome incorreto), o programa pede uma nova entrada até que uma seleção válida seja fornecida.

5. Estrutura de Loop

- O código está em um loop `while True`, garantindo que o programa continuará executando até que o usuário forneça uma entrada válida. Após a escolha do país, o programa exibe os dados relevantes e pergunta se o usuário deseja realizar outra consulta. Se a resposta for negativa, o loop é interrompido e o programa termina.

6. Filtragem e Apresentação dos Dados

- Após a seleção de um país, os dados são filtrados usando `df[df['País'] == pais_selecionado]`. Este processo seleciona apenas as informações referentes ao país escolhido.

- Os resultados são então agrupados por categoria de assinatura (por exemplo, Basic, Premium, Standard), gênero (Homem ou Mulher) e dispositivo utilizado (por exemplo, Laptop, Smartphone). Para cada categoria, o código calcula a soma do valor das assinaturas mensais e exibe esses valores de forma organizada, imitando a aparência de uma tabela dinâmica.

7. Saída Estruturada

- O programa apresenta os resultados de forma hierárquica, com o país selecionado no topo, seguido pelas categorias de assinatura e, dentro de cada categoria, o gênero e o dispositivo utilizado, juntamente com a soma do valor das assinaturas mensais para cada combinação.

8. Exemplo de Saída

- A saída para um país, por exemplo, "França", segue este formato:

```
País: França
- Tipo de Assinatura: Basic
  - Homem:
    - Laptop: R$ 200,00
    - Smartphone: R$ 150,00
  - Mulher:
    - Laptop: R$ 180,00
    - Tablet: R$ 170,00
- Tipo de Assinatura: Premium
```

9. Estruturação do Código:

```
PR PROJETO I Version control Current File ANÁLISE DE DADOS DA NETFLIX.py
1 import pandas as pd
2
3 # Carregar a planilha Excel
4 file_path = r'C:\FACULDADE\PROJETO APLICADO\PROJETO NETFLIX\BASE - NETFLIX.xlsx'
5 df = pd.read_excel(file_path, sheet_name='BASE - NETFLIX')
6
7
8 # Função para gerar a tabela hierarquizada por país
9 def gerar_tabela_pais(pais_escolhido):
10     df_pais = df[df['País'].str.lower() == pais_escolhido.lower()] # Faz a busca sem case sensitive
11
12     if df_pais.empty:
13         print(f'Não há dados para o país: {pais_escolhido}')
14         return
15
16     # Agrupar por gênero, Tipo de assinatura, Dispositivo e somar os valores da assinatura
17     pivot_table = df_pais.groupby(['gênero', 'Tipo de assinatura', 'Dispositivo Utilizado'])['Valor da assinatura/mês'].sum()
18
19
20     # Converter para DataFrame e formatar valores de forma monetária
21     pivot_table_df = pivot_table.reset_index()
22     pivot_table_df['Valor da assinatura/mês'] = pivot_table_df['Valor da assinatura/mês'].apply(
23         lambda x: f'R$ {x:,.2f}')
24
25     # Exibir a tabela gerada
26     print(f'\nTabela de análise para {pais_escolhido}:\n')
27     print(pivot_table_df.to_string(index=False))
28
29
30 # Função para listar países e pedir a escolha
31 def listar_países_e_perguntar():
32     # Listar os países únicos
33     países_unicos = df['País'].unique()
34
35     while True: # Loop contínuo até o usuário escolher sair
36         print('\nPaíses disponíveis para análise:\n')
37
38         for i, pais in enumerate(países_unicos, 1):
39             print(f'{i}. {pais}')
40
41         print('\nDigite o nome ou número do país que você deseja consultar (ou "sair" para encerrar): ')
42         pais_escolhido = input().strip()
43
44         # Verificar se o usuário quer sair do loop
45         if pais_escolhido.lower() == 'sair':
46             print('Encerrando o programa...')
47             break
48
49         # Verificar se o usuário digitou o número ou o nome do país
50         if pais_escolhido.isdigit():
51             try:
52                 pais_escolhido = países_unicos[int(pais_escolhido) - 1]
53             except IndexError:
54                 print('Número inválido. Tente novamente.')
55                 continue # Pede ao usuário para tentar novamente
56
57         # Exibir a tabela do país escolhido
58         gerar_tabela_pais(pais_escolhido)
59
60
61 # Executar o processo
62 listar_países_e_perguntar()
63
```

10. Resultados:

Pesquisa 1 Realizada:

```
C:\Users\Gustavo\anaconda3\python.exe "C:\FACULDADE\PROJETO APLICADO\PROJETO NETFLIX\PROJETO I\ANÁLISE DE DADOS DA NETFLIX\ANÁLISE DE DADOS DA NETFLIX.py"

Países disponíveis para análise:

1. Estados Unidos
2. Canada
3. Reino Unido
4. Australia
5. Alemanha
6. França
7. Brasil
8. Mexico
9. Espanha
10. Italia

Digite o nome ou número do país que você deseja consultar (ou 'sair' para encerrar):
Estados Unidos

Tabela de análise para Estados Unidos:

Gênero Tipo de assinatura Dispositivo Utilizado Valor da assinatura/mês
Homem Basic Laptop R$ 758.10
Homem Basic Smart TV R$ 957.60
Homem Basic Smartphone R$ 1,037.40
Homem Basic Tablet R$ 1,077.30
Homem Premium Laptop R$ 1,229.80
Homem Premium Smart TV R$ 1,006.20
Homem Premium Smartphone R$ 894.40
Homem Premium Tablet R$ 950.30
Homem Standard Laptop R$ 245.70
Homem Standard Smart TV R$ 283.50
Homem Standard Smartphone R$ 151.20
Homem Standard Tablet R$ 396.90
Mulher Basic Laptop R$ 1,157.10
Mulher Basic Smart TV R$ 1,117.20
Mulher Basic Smartphone R$ 997.50
Mulher Basic Tablet R$ 837.90
Mulher Premium Laptop R$ 1,509.30
Mulher Premium Smart TV R$ 894.40
Mulher Premium Smartphone R$ 670.80
Mulher Premium Tablet R$ 950.30
Mulher Standard Laptop R$ 207.90
Mulher Standard Smart TV R$ 283.50
Mulher Standard Smartphone R$ 226.80
Mulher Standard Tablet R$ 226.80

Países disponíveis para análise:

1. Estados Unidos
2. Canada
3. Reino Unido
4. Australia
5. Alemanha
6. França
7. Brasil
8. Mexico
9. Espanha
10. Italia

Digite o nome ou número do país que você deseja consultar (ou 'sair' para encerrar):
```

Pesquisa 2 Realizada:

Países disponíveis para análise:

1. Estados Unidos
2. Canada
3. Reino Unido
4. Australia
5. Alemanha
6. França
7. Brasil
8. Mexico
9. Espanha
10. Italia

Digite o nome ou número do país que você deseja consultar (ou 'sair' para encerrar):

canada

Tabela de análise para canada:

Gênero	Tipo de assinatura	Dispositivo Utilizado	Valor da assinatura/mês
Homem	Basic	Laptop	R\$ 518.70
Homem	Basic	Smart TV	R\$ 957.60
Homem	Basic	Smartphone	R\$ 837.90
Homem	Basic	Tablet	R\$ 758.10
Homem	Premium	Laptop	R\$ 559.00
Homem	Premium	Smart TV	R\$ 614.90
Homem	Premium	Smartphone	R\$ 614.90
Homem	Premium	Tablet	R\$ 894.40
Homem	Standard	Laptop	R\$ 113.40
Homem	Standard	Smart TV	R\$ 151.20
Homem	Standard	Smartphone	R\$ 226.80
Homem	Standard	Tablet	R\$ 170.10
Mulher	Basic	Laptop	R\$ 678.30
Mulher	Basic	Smart TV	R\$ 638.40
Mulher	Basic	Smartphone	R\$ 518.70
Mulher	Basic	Tablet	R\$ 877.80
Mulher	Premium	Laptop	R\$ 391.30
Mulher	Premium	Smart TV	R\$ 503.10
Mulher	Premium	Smartphone	R\$ 447.20
Mulher	Premium	Tablet	R\$ 894.40
Mulher	Standard	Laptop	R\$ 207.90
Mulher	Standard	Smart TV	R\$ 189.00
Mulher	Standard	Smartphone	R\$ 283.50
Mulher	Standard	Tablet	R\$ 245.70

Países disponíveis para análise:

1. Estados Unidos
2. Canada
3. Reino Unido
4. Australia
5. Alemanha
6. França
7. Brasil
8. Mexico
9. Espanha
10. Italia

Digite o nome ou número do país que você deseja consultar (ou 'sair' para encerrar):

sair

Encerrando o programa...

Process finished with exit code 0

Conclusão

O código desenvolvido para a análise de dados da Netflix fornece uma abordagem prática e interativa para explorar a base de informações de assinantes por país, gênero, tipo de assinatura e dispositivo utilizado. A estrutura do programa foi desenhada para ser intuitiva, permitindo que o usuário selecione o país desejado e visualize uma análise detalhada dos valores de assinatura com base nesses critérios.

A aplicação foi implementada com o uso de bibliotecas robustas de manipulação de dados, como o Pandas, garantindo um processamento eficiente dos dados provenientes de um arquivo Excel. A apresentação dos resultados de forma hierárquica e organizada permite uma rápida interpretação, tornando a análise mais acessível mesmo para usuários sem grande experiência em manipulação de dados.

Além de ser uma ferramenta útil para visualizar o comportamento de assinantes da Netflix em diferentes regiões, este código pode ser expandido para análises mais avançadas, como geração de gráficos e relatórios comparativos. A modularidade do código também permite fácil adaptação para outras bases de dados ou tipos de análise.

No geral, este projeto demonstra o poder das bibliotecas de dados do Python e destaca a importância de estruturar os dados de forma clara para facilitar a tomada de decisões estratégicas. A interação direta com o usuário e a possibilidade de realizar múltiplas consultas tornam a ferramenta dinâmica e eficiente para análises contínuas.