

LISTA DE EXERCÍCIO 05 – MANIPULANDO DADOS COM PANDAS

(PARTE 1)

Exercício 01 – Manipulando dataframes com Pandas.

- a) Abra o arquivo **world_happiness_report_2015.csv** e o aloque em um dataframe.
- b) Verifique o cabeçalho e o final do dataframe.
- c) Quais as colunas desse dataframe?
- d) Quais os tipos de dados temos no dataframe?
- e) Há valores faltantes ou nulos? Em quais colunas?
- f) Renomeie as variáveis como segue:

happiness rank	=>	rank_felicidade
happiness score	=>	score_felicidade
standard error	=>	stand_error
economy (GDP per Capita	=>	PIB

health (Life Expectancy)	=>	expect_vida
trust (Government Corruption)	=>	corrupcao

- g) Quais os valores médios de expect_vida? E o valor mediano? E o máximo da variável PIB?

Exercício 02 – Manipulando dados com Pandas

- a) Crie um programa em pandas para ler o arquivo em Excel imobiliária.xlsx
- b) e coloque em um dataframe.
- c) Visualize o cabeçalho dos dados;
- d) Visualize as últimas linhas;
- e) Escreva um programa para renomear o cabeçalho da tabela com os
- f) seguintes nomes:

Coluna 1	Filial
Coluna 2	Vendedor
Coluna 3	Data
Coluna 4	Tipo
Coluna 5	Área (m²)
Coluna 6	Dorms.
Coluna 7	Garag.
Coluna 8	Piscina
Coluna 8	Preço
Coluna 10	Comissão

- g) 5. Quais são os tipos de dados que temos no dataframe?
- h) 6. Há valores faltantes ou nulos?
- i) 7. Calcule a soma, média, máximo, mínimo e contagem da coluna Preço.
- j) 8. Crie um programa para criar um dataframe para mostrar as vendas por região, e salvar um arquivo em Excel para cada região.
 - Centro
 - Noroeste
 - Norte
 - Sudeste
 - Oeste
 - Leste
 - Sudoeste
 - Sul
 - Nordeste
- l) Gere um arquivo em Excel para representar as vendas efetuadas pela filial norte em 2010.
- m) Gere um arquivo em Excel para representar as vendas efetuadas pela filial Sudeste em 2010. Sendo que no relatório deve conter apenas vendas de casas sem garagem.
- n) Qual o valor máximo de comissão que a imobiliária numa venda?
- o) Quantas casas de 2 ou 3 dormitórios foram vendidas pelo Ricardo antes de 2011.

Exercício 03 – Manipulando dados com Pandas

- a) Crie um programa em pandas para ler o arquivo em Excel vendas.xlsx e
- b) coloque em um dataframe.
- c) Visualize o cabeçalho dos dados;
- d) Visualize as últimas linhas;
- e) Escreva um programa para renomear o cabeçalho da tabela com os
- f) seguintes nomes:

Coluna 1	Data
Coluna 2	Vendedor
Coluna 3	Cargo
Coluna 4	Cliente
Coluna 5	UF
Coluna 6	Vendas

- g) Quais são os tipos de dados que temos no dataframe?
- h) Há valores faltantes ou nulos?
- i) Calcule a soma, média, máximo, mínimo e contagem da coluna Preço.
- j) Crie um programa para criar um dataframe para mostrar as vendas por cargo, e salvar um arquivo em csv para com o nome de cada cargo:
 - Supervisor de Vendas PI
 - Gerente Regional Jr.
 - Supervisor de Vendas Sr
 - Coordenador de Vendas PI.
 - Vendedor Jr.
 - Vendedor Sr.
 - Vendedor PI.
 - Gerente Regional PI.
 - Coordenador de Vendas Sr.
 - Coordenador de Vendas Jr.
 - Supervisor de Vendas Jr.

9. Crie um dataframe para cada coluna, contendo dados únicos.

Exercício 04 – Criando relatórios em função do arquivo dados-vendas.xlsx.

- a) Crie um dataframe, utilizando o arquivo **dados-vendas.xlsx**. Crie data frames para dividir a planilha vendas em 4 arquivos. São eles:

Tabela Fato <ul style="list-style-type: none"> - ID-Produto; - ID-Loja; - ID-Vendedor; - Data; - ValorVenda. 	Tabela Produtos <ul style="list-style-type: none"> - ID-Produto; - Produto; - Categoria; - Segmento; - Fabricante;
Tabela Lojas <ul style="list-style-type: none"> - ID-Loja; - Cidade; - Estado 	Tabelas Vendedor <ul style="list-style-type: none"> - ID-Vendedor; - Vendedores;

Salve todos os arquivos em Excel.

b) Com base na coluna Vendedor, construa um código que permita remover as duplicadas da coluna informada, e na sequência, filtre o dataframe e gere um relatório individual para cada vendedor. Para fazer isso, siga os procedimentos a seguir:

1. Crie uma pasta chamada **relatórios-vendedores**;
2. Importe a biblioteca Pandas;
3. Declare uma variável chamada **path** para armazenar o caminho da pasta criada;
4. Declare uma variável com o nome **file** para armazenar o nome do arquivo que será manipulado;
5. Crie um dataframe chamado **dados** e utilize o comando **pd.read_excel** para acessar o arquivo.
6. Visualize o cabeçalho do dataframe;
7. Crie um dataframe chamado **df_vendedores**, para atribuir valores a ele você deve selecionar apenas a coluna Vendedores e remover as duplicadas do mesmo. Confirme se realmente o comando removeu as duplicadas.

Após isso, crie uma lista com o seguinte nome **listaVendedores**, essa lista deve receber os dados do dataframe **df_vendedores**. Para isso, você deve converter o dataframe em lista, utilize o comando **dataframe.tolist()**.

8. Construa uma estrutura de repetição (**FOR**) para que a variável **vendedor** receba os nomes que estão na lista **listaVendedores** e filtre o dataframe dados, utilizando a coluna Vendedores para que gere um novo dataframe (**df_final**) para cada vendedor. Ainda na estrutura de repetição, você deve inserir uma linha de código para gerar um arquivo em Excel para cada vendedor, utilize o comando **dataframe.to_excel**.



you should convert the dataframe to list, use the command

`dataframe.tolist()`.

9. Construct a repetition structure (**FOR**) so that the variable **vendedor** receives the names that are in the list **listaVendedores** and filter the dataframe data, using the column Vendedores so that it generates a new dataframe (**df_final**) for each seller. Still in the repetition structure, you should insert a line of code to generate an Excel file for each seller, use the command **`dataframe.to_excel`**.

Exemplo:

```
for vendedor in listaVendedores:
```

```
    df_final = dados[dados['Coluna']==variável que vai filtrar o dataframe]
```

```
    df_final.to_excel (path + variável + '.xlsx')
```

- c) Repeat this procedure for the fields Category, Segment, Manufacturer, City and State. **Reescreva todo o código para cada caso, com o objetivo de exercitar a sintaxe dos comandos.**

Exercício 05 – Construct a script in python, using the Pandas and OS (Operating System) libraries to concatenate the files that are in the subfolders.

For this, make the following steps:

- a) Import the Pandas and OS libraries;
- b) Declare a variable called **dir_path** to store the path of the directory that contains the files to be combined;
- c) Create an empty list called **res** (repository) to save the string of the path + name of the files saved in the subfolders;
- d) Create a repetition structure, using **FOR**, to capture the name of the files that are in the repository indicated in the variable **dir_path**. In this structure you should use the command **`os.listdir`**, this command lists all the files that are inside a determined directory.

Após listar estes arquivos, estes dados devem ser inseridos na variável **path** e em seguida uma estrutura de decisão deve verificar se o que está salvo neste diretório realmente é um arquivo, para isto, utilize a estrutura de decisão **if** com o comando `os.path.isfile(diretório + nome do arquivo)`, este comando verifica se o caminho informado realmente é um arquivo, caso seja ele responde verdadeiro, caso contrário, responderá false. Caso seja um arquivo, o caminho completo do arquivo deve ser inserido à lista

res.

Exemplo:

```
for path in os.listdir (informar caminho da pasta):  
    # verificar se o caminho atual é um arquivo  
    if os.path.isfile(informar o caminho completo):  
        res.append(informar o caminho do arquivo)
```

- e) Crie um dataframe vazio, e nomeie ele como `files_combined`.

`files_combined = []`

Crie uma estrutura de repetição, utilizando o comando **for**. Nesta estrutura a variável `file` deve capturar todos os caminhos dos arquivos que estão na lista **res**.

- f) Dentro da estrutura de decisão crie um dataframe e nomeie o mesmo como **df1**. Este data frame deve receber os arquivos listados na pasta, para isto utilize o comando **pd.read_excel** para ler os arquivos e atribuí-los ao data frame. Após ler o arquivo, você deve concatenar o dataframe **files_combined** com **odf1**.

```
for file in res:  
    df1 = pd.read_excel(variável com o caminho completo dos arquivos)  
    files_combined = pd.concat([dataframe 1 ,dataframe 2])
```

- g) Após agrupar os arquivos, converta o dataframe em Excel e salve o mesmo com o seguinte nome: **arquivosCombinados.xlsx**.



- h) Faça o mesmo processo com os arquivos que estão na pasta **vendas**.
Reescreva o código com o objetivo de fixar a sequência lógica e sintaxe dos comandos.