

MODALIDADE:	Aprendizagem +	Não aplicável	
CURSO:	Técnico/a Especialista em Gestão de Informação e Ciência dos Dados		
UFCD:	Limpeza e transformação de dados em Python	CÓDIGO UFCD:	10808
FORMADOR/A:	Bruno Silva	DATA:	
FORMANDO/A:		CLASSIFICAÇÃO:	

- **Exercício 1** – Crie a variável com o nome **data**, do qual vai ler o ficheiro **vgsales.csv**, indicando a localização atual do ficheiro e ler o delimitador da vírgula;
- **Exercício 2** – Mostre as informações gerais do ficheiro com a função **info()**;
- **Exercício 3** – Verifique se o ficheiro tem registos nulos com a função **isnull()** e faça a contagem com a função **sum()**;
 - **Exercício 3.1** – Se tiver registos nulos deve:
 - Criar a variável **dados_sem_nulos** e use a função **dropna()** para remover os dados nulos;
- **Exercício 4** – Verifique se o ficheiro tem registos duplicados com a função **duplicated()** e faça a contagem com a função **sum()**;
 - **Exercício 4.1** – Se tiver registos duplicados deve:
 - Criar a variável **dados_sem_nulos_duplicados** e use a função **drop_duplicates()** para remover os dados duplicados;
- **Exercício 5** – Mostre os 6 primeiros registos do dataframe com a função **head()**;
- **Exercício 6** – Mostre quantos registos existem no DataFrame com a função **len()**;
- **Exercício 7** – Mostre o nome das colunas com a propriedade **columns**;
- **Exercício 8** – Crie a variável **nova_alt** e coloque a seguinte informação: utilize a variável **data** e altere o nome de 3 colunas com a função **rename()**, mais especificamente: **Name** para Nome, **Platform** para Plataforma e **Publisher** para Editora. No final mostre a informação da nova variável;
- **Exercício 9** – Crie a variável **nova_data** e coloque a seguinte informação: utilize a variável **data** e adicione uma nova coluna (com a função **assign()**) com o nome **Total** e deve usar a função de números aleatórios para gerar números inteiros entre 1 e 5;

- **Exercício 10** – Crie a variável **dados2** e remova a coluna apenas a coluna JP_Sales, com a função **drop()**;
- **Exercício 11** – Vamos gravar a nova informação num ficheiro excel. Como tal, deve usar a funcionalidade **with pd.ExcelWriter** para gerar um ficheiro excel com várias folhas de cálculo:

```
#Guardar dados num ficheiro excel (ficheiro com várias folhas)
with pd.ExcelWriter("nome_do_ficheiro.xlsx") as ficheiro:
    # De seguida, basta indicar qual das variáveis DataFrame que deseja trabalhar
    # e de seguida invocar a função to_excel() e dentro dos parênteses indicar
    # 2 argumentos:
    #   o nome do ficheiro da variável
    #   qual o nome dessa folha de cálculo (sheet_name)

    data.to_excel(ficheiro, sheet_name="Original")
    data2.to_excel(ficheiro, sheet_name="Sem Nulos")
```

Para este caso, faça a adaptação do código mais acima e grave a informação para um ficheiro excel com o nome “TP1_primeiro_ultimo.xlsx, onde as palavras primeiro e último devem corresponder ao seu primeiro e último nome com as seguintes folhas de cálculo:

- **Criar** a folha de cálculo “**Dados Originais**” e coloque a informação da variável **quando carregou a informação do ficheiro, ou seja, a variável data do exercício 1**;
- **Criar** a folha de cálculo “**Dados sem nulos**” e coloque a informação da variável **sem dados nulos, ou seja, a variável do exercício 3**;
- **Criar** a folha de cálculo “**Nova Info**” e coloque a informação da variável **nova_data, ou seja, a variável do exercício 9**;