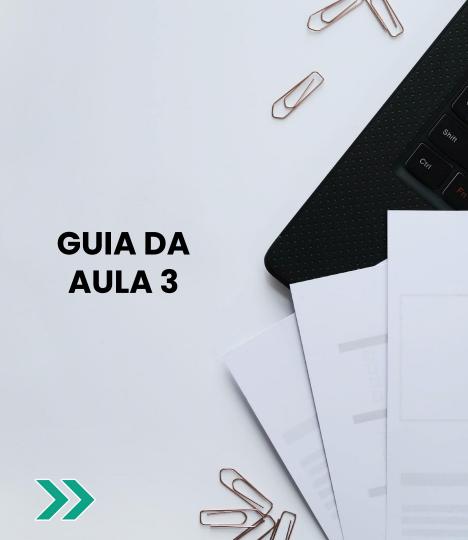


# Profissão: Analista de dados





# **COLETA DE DADOS I**







# Use arquivos de Excel

Formato

Pacote openpyxl



Acompanhe aqui os temas que serão tratados na videoaula





## 3. Arquivos Excel

#### 1. Formato

Um arquivo tabular nativo do Windows, sistema operacional da Microsoft.

Arquivo Excel: banco.xlsx

	Α	В	C	D	E	F	G	Н
1	age	job	marital	education	default	balance	housing	loan
2	30	unemployed	married	primary	no	1787	no	no
3	33	services	married	secondary	no	4789	yes	yes
4	35	management	single	tertiary	no	1350	yes	no
5	30	management	married	tertiary	no	1476	yes	yes
6	59	blue-collar	married	secondary	no	0	yes	no
7	35	management	single	tertiary	no	747	no	no
8	36	self-employed	married	tertiary	no	307	yes	no
9	39	technician	married	secondary	no	147	yes	no
10	41	entrepreneur	married	tertiary	no	221	yes	no
11	43	services	married	primary	no	-88	yes	yes
12			1-					
13								



#### Download

```
escola britânica de artes criativas & tecnologia
```

```
!wget --show-progress --continue -0 \
./banco.xlsx \
https://raw.githubusercontent.com/andre-marcos-perez/\
ebac-course-utils/main/dataset/banco.xlsx
```

### 2.2. Pacote openpyxl

Pacote Python para interagir com planilhas excel. A documentação pode ser encontrada neste <u>link</u>.

```
In []:
!pip install openpyxl
```

**Exemplo**: Média dos saldos.

```
In []:
    from openpyxl import load_workbook

    planilhas = load_workbook (filename='banco.xlsx')
    planilha = planilhas.active
```





```
In []:
    saldos = []

    cabecalho = next(planilha.values)
    indice_saldo = cabecalho.index('balance')
    saldos = [
        linha[indice_saldo] for linha in planilha.values
        if linha[indice_saldo] != 'balance'
    ]

    print(saldos)
```

```
soma_saldos = reduce(lambda saldo_a, saldo_b: saldo_a + saldo_b, saldos)
qtd_saldos = len(saldos)

media_saldos = soma_saldos / qtd_saldos
print(f"A média dos saldos é de { media_saldos}.")
```

