

Disciplina:

Linguagem de Programação

Arquivo Excel e CSV em Python

Prof^a. Dr^a. Giovana Angélica Ros Miola
giovana.miola@fatec.sp.gov.br

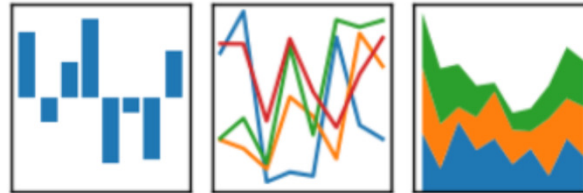


Biblioteca Pandas

- É uma biblioteca **Python**.
- Fornece ferramentas de análise de dados, estruturas de dados de alta performance e ***fáceis de usar***.

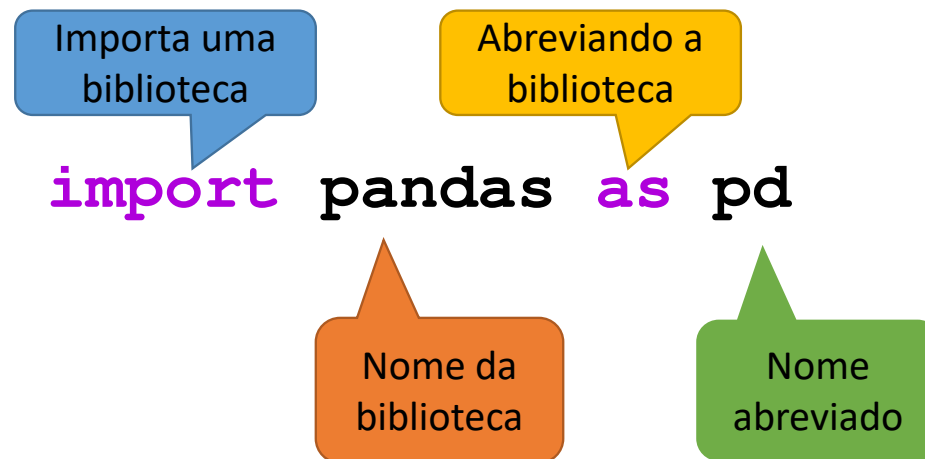
pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Biblioteca Pandas

- Exemplo, para acessar um arquivo/planilha do Excel, execute:



Biblioteca Pandas

- O Colab já tem instalado a biblioteca Pandas
- Caso queira usar em outro ambiente de programação necessitará instalar, no prompt de comandos, estas bibliotecas:
 - `pip install pandas`
 - `pip install numpy`
 - `pip install openpyxl`

Biblioteca Pandas

- DataFrame (df) é uma estrutura bidimensional de dados, como uma planilha.

- Lendo o arquivo:

```
import pandas as pd
```

```
df1 = pd.read_excel('Vendas.xlsx')
```

```
print(df1)
```



	Código Venda	Data	ID Loja	Produto
0	23014	2021-12-01	Shopping Centro Sul	Shorts praia Listrado
1	23014	2021-12-01	Shopping Centro Sul	Casaco Listrado
2	23016	2021-12-01	Shopping Bairro	Sapato Listrado
3	23016	2021-12-01	Shopping Bairro	Casaco
4	23017	2021-12-01	Shopping Horizontal	Gorro Liso
...
7084	79295	2021-12-10	Shopping Baixo	Camisa Gola V Listrado
7085	79296	2021-12-10	Shopping Residencial	Polo
7086	79296	2021-12-10	Shopping Residencial	Gorro Xadrez
7087	79297	2021-12-10	Shopping Direito	Pulseira Listrado
7088	79298	2021-12-10	Shopping Bairro	Cinto Liso

	Quantidade	Valor Unitário	Valor Final
0	5	114	570
1	1	269	269
2	2	363	726
3	1	250	250
4	3	92	276
...
7084	1	116	116
7085	2	130	260
7086	1	85	85
7087	2	79	158
7088	2	209	418

[7089 rows x 7 columns]

Biblioteca Pandas

- Arquivos com extensão .csv, é um formato comum de armazenar dados de tabelas
- CSV (Comma-Separated Values) é um formato que sugere o uso de , para separar os dados armazenados, apesar de o caracter separador poder ser o ponto-e-vírgula ou outro.

```
df2 = pd.read_csv('Propaganda.csv')
```

```
print(df2)
```



	TV	Radio	Jornal	Vendas
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	12.0
3	151.5	41.3	58.5	16.5
4	180.8	10.8	58.4	17.9
..
195	38.2	3.7	13.8	7.6
196	94.2	4.9	8.1	14.0
197	177.0	9.3	6.4	14.8
198	283.6	42.0	66.2	25.5
199	232.1	8.6	8.7	18.4

[200 rows x 4 columns]

Biblioteca Pandas

- Função `info()` para apresentar o resumo completo do dataframe. É muito útil ao fazer análises exploratórias dos dados.

```
print(df.info())
```

- Apresenta quantidade de linhas, colunas, tipos de dados, as colunas do arquivo.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7089 entries, 0 to 7088
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Código Venda          7089 non-null   int64
1   Data                  7089 non-null   datetime64[ns]
2   ID Loja               7089 non-null   object
3   Produto               7089 non-null   object
4   Quantidade            7089 non-null   int64
5   Valor Unitário        7089 non-null   int64
6   Valor Final           7089 non-null   int64
dtypes: datetime64[ns](1), int64(4), object(2)
memory usage: 387.8+ KB
None
```

Biblioteca Pandas

- A função `ravel()` apresenta a lista de cabeçalhos das colunas do arquivo Excel ou Csv, usando a propriedade `columns` do objeto `dataframe (df)`.

```
print (df.columns.ravel())
```



```
['TV' 'Radio' 'Jornal' 'Vendas']
```


Biblioteca Pandas

- A função `value_counts()`, conta a quantidade de ocorrências da coluna que quer analisar os dados.
- Na figura é apresentado que, por exemplo, Shopping Bairro, tem 315 como resultado, que significa que existe 315 produtos nesta loja.

```
import pandas as pd
df = pd.read_excel('Vendas.xlsx')
print(df['ID Loja'].value_counts())
```

Shopping Bairro	315
Shopping Centro Sul	306
Shopping Vertical	304
Shopping Direito	302
Shopping Avenida	300
Shopping Sudoeste	300
Shopping Vila	299
Shopping Esquerdo	294
Shopping Viela	290

Cálculo e apresentação do faturamento

- Faturamento: corresponde a soma de todas as vendas, utilizando a coluna ['Valor Final']

```
import pandas as pd
df = pd.read_excel('Vendas.xlsx')
faturamento = df['Valor Final'].sum()
print(f'Faturamento obtido R$ {faturamento:,.2f}')
```

- Resultado:

Faturamento obtido R\$ 2,892,968.00

Formatação das casas decimais

- Atribui a str com o valor formatado, na casa do milhar, com qualquer caracter, por exemplo, o **_**

```
f_sem_formatar = f'{faturamento:_.2f}'
```

```
print('Faturamento obtido', f_sem_formatar)
```

- Primeira função replace(), substitui o ponto, por vírgula, correspondente as casas decimais
- Segunda função replace(), troca o _ por ponto, como separador de milhar

```
f_formatado = f_sem_formatar.replace('.', ',').replace('_', '.')
```

```
print('Faturamento obtido', f_formatado)
```

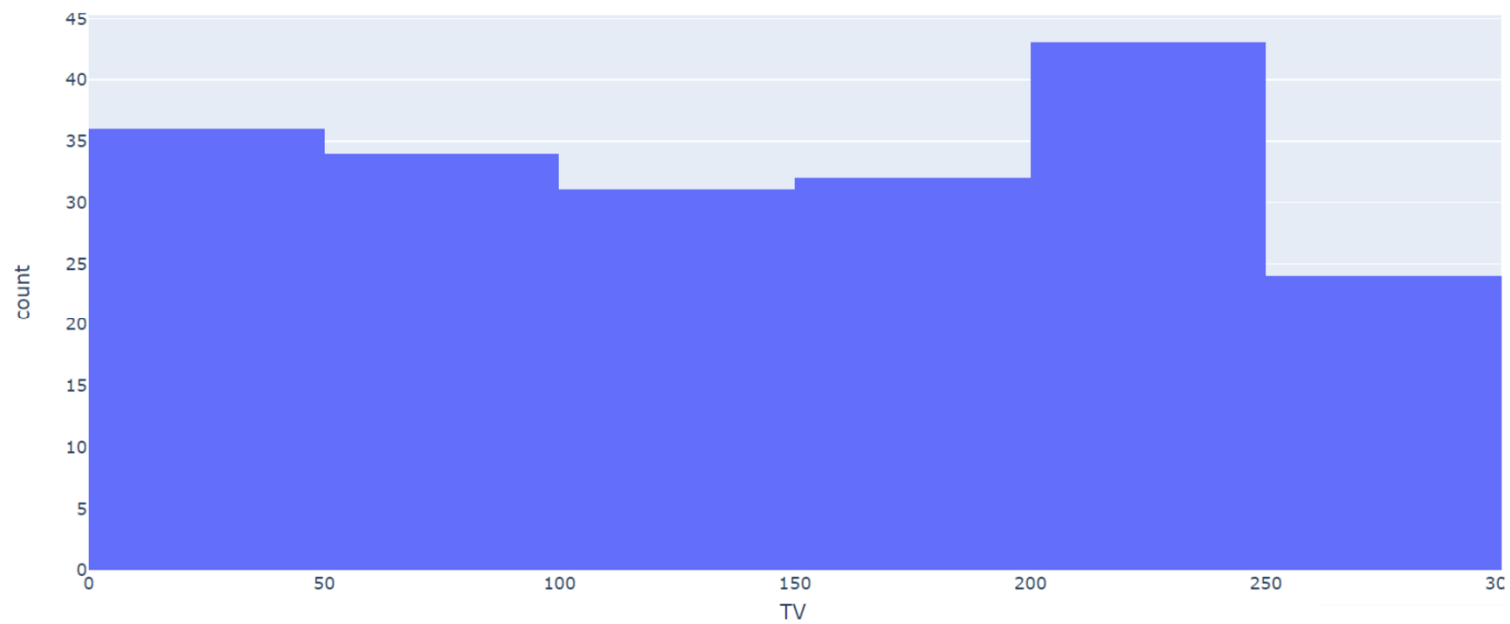
- Faturamento obtido R\$ **2_892_968.00**
- Faturamento obtido R\$ **2.892.968,00**

Biblioteca Plotly – Histograma

- Para avaliar resultados das empresas, existem várias formas para demonstrar a análise de dados e contribuir para a tomada de decisão, como por exemplo os gráficos, neste caso, histograma

```
import pandas as pd
df = pd.read_csv('Propaganda.csv')
import plotly.express as px
fig = px.histogram(df, x="TV")
fig.show()
```

Biblioteca Plotly – Histograma



Biblioteca Matplotlib – Histograma

- Outra biblioteca que oferece construção de gráficos
- rwidth indica a separação entre as barras
- 5 é a quantidade de barras

```
import matplotlib.pyplot as plt
df = pd.read_csv('Propaganda.csv')
plt.hist(df['Radio'], 5, rwidth=0.8, color = 'red')
plt.title('Propaganda Via Rádio')
plt.xlabel('Custo')
plt.ylabel('Frequência Absoluta')
plt.show()
```

