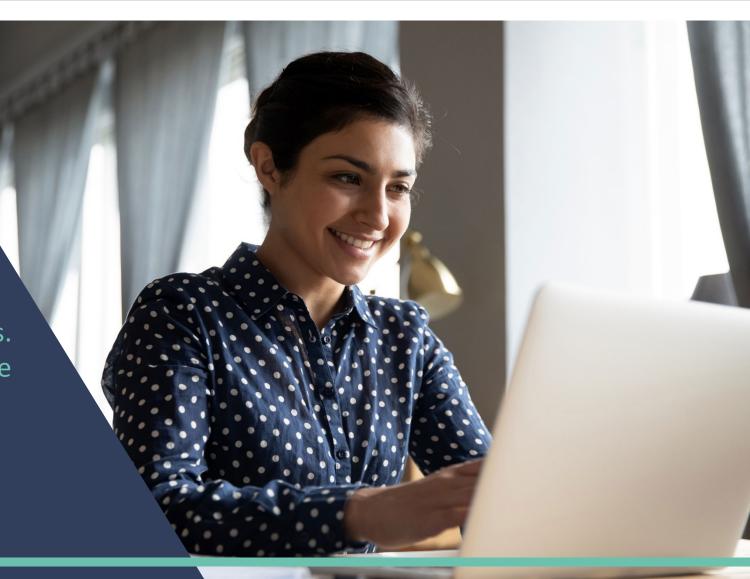




Manipulação de planilhas Excel no Pandas. Criação de gráficos customizados. Tipos de visualização.

Bloco 1





Organização e visualização de dados

A visualização de dados consiste na criação de dispositivos visuais que ajudam a compreender e interpretar dados, segundo Milovanovic (2013).

Dispositivos visuais comuns:

- Gráficos.
- Diagramas.
- Histogramas.
- Animações.



Matplotlib - Instalação

Dependências:

- Python (>= 3.6).
- *FreeType* (>= 2.3).
- *libpng* (>= 1.2).
- *NumPy* (>= 1.11).
- Setuptools.
- *cycler* (>= 0.10.0).
- dateutil (>= 2.1).

(pip install matplotlib)



Matplotlib – Exemplo de utilização

To-do:

- Gráficos de linhas com números aleatórios
 - Recursos do gráfico (grid, eixos, labels, títulos e legendas).
- Cores.
- Linhas marcadores.



Organização e visualização de dados - ETL

ETL (Extract, Transform, Load) - Extrair, Transformar, Carregar: são ferramentas cuja função é a extração e organização de dados para carregamento e utilização, segundo Denney (2013).

Características:

- Acesso aos dados em estado bruto.
- Transformações em características.
- Organização e agrupamento.



Pandas - Instalação

Dependências:

- numexpr.
- bottleneck.

(pip install pandas)



Pandas – Exemplo de utilização

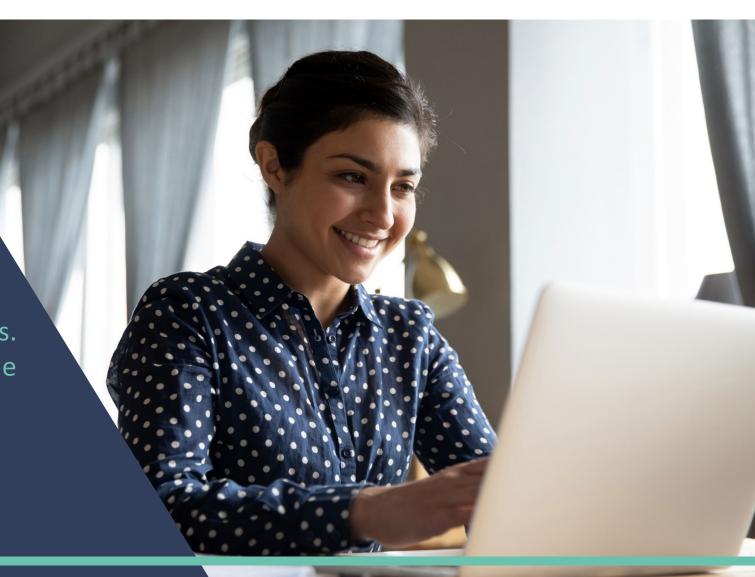
To-Do:

- Gerar séries e DataFrames com dados aleatórios.
- Criar:
 - Histograma.
 - Gráfico de linhas.
 - Gráfico de barras.
 - Gráfico de área.
 - Gráfico de pizza.
- Manipulação de arquivos:
 - Texto: CSV, JSON, HTML.
 - Binário: Excel, HDF5.



Manipulação de planilhas Excel no Pandas. Criação de gráficos customizados. Tipos de visualização.

Bloco 2





Um caso com carga, manipulação e visualização de dados

Caso:

Carregar dados de Excel de utilização de Internet, no Brasil (IBGE, 2021).

To-do:

- Seleção e limpeza das colunas.
- Verificação de totais e métricas básicas.
- Geração de histogramas.
- Geração de gráficos.



Teoria em Prática

Bloco 3



Reflita sobre a seguinte situação

- Utilização do Excel em larga escala nas empresas.
- Fazer importação de dados de *Excel* e integrar em um *ERP*.



Norte para a resolução...

- Verifique os dados de entrada e de saída.
- Limpe os dados o máximo possível.
- Verifique a integridade dos dados.
- Quanto mais fácil a integração no ERP, melhor será a solução.



Dica do (a) Professor (a)

Bloco 4



Visualize e prepare os dados até a exaustão

 Manipule os dados de forma ostensiva. É grátis e traz ideias e dicas do que fazer com os dados.

 Visualização ajuda demais no agrupamento e seleção dos dados.

 Compare os gráficos das duas bibliotecas (Pandas e Matplotlib).



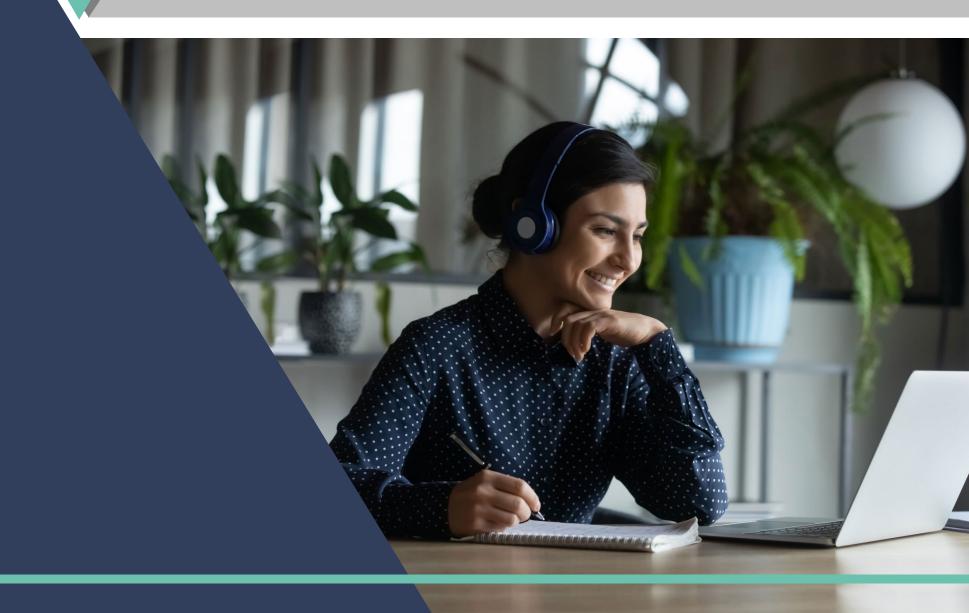
Referências

IBGE. **Utilização da Internet**. Disponível em:

https://ftp.ibge.gov.br/Acesso_a_internet_e_posse_celular/201 5/Tabelas_de_Resultados/xlsx/01_Pessoas_de_10_Anos_ou_M ais_de_Idade/01_Utilizacao_da_Internet.xlsx. Acesso em: 29 mar. 2021.

DENNEY, M. J. *et al.* Validating the extract, transform, load process used to populate a large clinical research database. **International journal of medical informatics**, v. 94, p. 271-274, 2016.

MILOVANOVIĆ, I. **Python data visualization cookbook**. Packt Publishing Ltd, 2013.



Bons estudos!