

# Profissão: Analista de dados

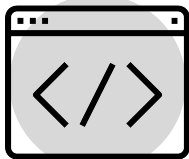


# DE OLHO NO CÓDIGO



# Projeto: Análise Exploratória de Dados em Python

- Explore dados
- Transforme e limpe dados
- Visualize dados



Confira boas práticas da linguagem Python por assunto relacionado às aulas.



# Explore dados

Wrangling de dados refere-se ao processo de limpeza, transformação e preparação de dados para a análise. Existem algumas boas práticas de Wrangling que podem ajudar a garantir que os dados estejam em um estado adequado para análise. Acompanhe algumas dessas boas práticas.



## **Identifique dados incompletos, duplicados ou inválidos**

É importante identificar dados ausentes, duplicados ou inválidos e decidir como tratá-los. Você pode preencher dados ausentes com valores médios ou remover dados duplicados ou inválidos.

## **Seja cuidadoso com a formatação**

A consistência na formatação de dados é importante para a análise. Certifique-se de que os dados estejam formatados de maneira uniforme e coerente.

# Explore dados

Wrangling de dados refere-se ao processo de limpeza, transformação e preparação de dados para a análise. Existem algumas boas práticas de Wrangling que podem ajudar a garantir que os dados estejam em um estado adequado para análise. Acompanhe algumas dessas boas práticas.



- Padronize os dados**  
 Os dados podem vir de várias fontes e, portanto, podem ter diferentes padrões de codificação. Padronize os dados de acordo com um formato comum, para que possam ser comparados e combinados facilmente.

- Automatize o processo de Wrangling**  
 Para lidar com grandes volumes de dados, pode ser necessário automatizar o processo de Wrangling. Use ferramentas como Python e R para criar rotinas automatizadas de limpeza, transformação e preparação de dados.

# Transforme e limpe dados



## NA (Not Available)

No pandas, há a convenção de se utilizar a linguagem de programação R, referenciando os dados faltantes como NA. (Not Available). Em aplicações estatísticas, dados NA podem ser inexistentes ou dados existentes que não foram observados.



## Motivos para transformar dados

Em alguns casos, é necessário transformar os dados para que eles possam ser analisados. Por exemplo, converter unidades de medida ou codificar dados categóricos em valores numéricos.



# Transforme e limpe dados

## • **Motivos para limpar dados**

Ao limpar dados, faça a análise nos dados ausentes, a fim de identificar problemas em sua coleta ou distorções provocadas por dados ausentes.

Para aprofundar seus conhecimentos sobre o assunto leia [o Capítulo 7: Limpeza e preparação dos dados](#) no livro Python para análise de dados, de Wes McKinney.

**Saiba Mais**



# Visualize dados



## Matplotlib

A Matplotlib é um pacote de plotagem. A biblioteca exporta visualizações para todos os vetores comuns e formatos de gráficos raster (PDF, SVG, JPG, PNG, BMP, GIF etc).



## Convenção de importação com a matplotlib

Com a matplotlib, usa-se a seguinte convenção de importação:

```
In [11]: import  
matplotlib.pyplot as plt
```





# Visualize dados

## Seaborn

Importe o seaborn para mudar os esquemas de cores e os estilos de plotagem padrões da matplotlib a fim de melhorar a legibilidade e a estética.

Para aprofundar seus conhecimentos sobre o assunto leia o Capítulo 9: Plotagem e visualização no livro Python para análise de dados, de Wes McKinney.

**Saiba Mais**



# Bons estudos!

