

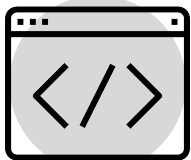
# Profissão: Cientista de Dados



# BOAS PRÁTICAS



# Agregações e operações em grupos



- **Conheça as agregações em Python**
- **Conheça o método aggregate**
- **Conheça o método apply do groupby**
- **Conheça o método mapping**
- **Utilize tabela pivô**



# Conheça as agregações em Python

- A visualização de dados é uma parte crucial da análise de dados. A biblioteca Matplotlib permite criar uma variedade de gráficos e gráficos para ajudar a entender os dados. Use-a para visualizar dados agrupados e identificar tendências ao longo do tempo.
- Além das médias, experimente outras métricas e análises relevantes para obter uma visão mais completa dos dados, como análise de frequência e comparação de categorias.
- Utilize gráficos temporais para visualizar tendências ao longo do tempo, como a variação da média de idade das mães em diferentes datas de nascimento.



# Conheça o método aggregate

- Ao usar o método aggregate em Python, é importante lembrar que ele pode ser usado de duas maneiras: escrevendo 'aggregate' por extenso ou usando a abreviação 'AGG'. Ambas as formas são aceitáveis, mas é importante ser consistente em seu uso para evitar confusão.
- Você pode passar uma função definida pelo usuário para o método aggregate. Isso permite que você personalize o processamento dos seus dados de acordo com suas necessidades específicas.
- É possível passar uma lista de funções para o método aggregate e obter o resultado para cada uma dessas funções. Isso pode ser útil para realizar várias operações ao mesmo tempo.



# Conheça o método apply do groupby

- Lembre-se de que você pode passar argumentos adicionais para as funções que você aplica aos seus dados. Isso pode ser útil para alterar o comportamento da função dependendo do contexto.
- Ao usar o método 'apply', esteja ciente de que ele retorna um novo objeto DataFrame ou Series. Se você quiser manter as alterações em seu DataFrame original, você precisará atribuir o resultado de volta a ele.



# Conheça o método mapping

- Ao criar novos grupos com base em uma variável específica, como a escolaridade da mãe, certifique-se de definir claramente as categorias e usar um dicionário para mapear as novas categorias.
- Ao agrupar dados com base em uma variável geográfica, como municípios, crie um dicionário para definir quais unidades pertencem a cada grupo.
- Lembre-se de que o mapeamento é uma ferramenta poderosa para transformar e categorizar dados, tornando-os mais úteis para análises subsequentes.



# Utilize tabela pivô

- Ao criar variáveis categóricas a partir de variáveis numéricas, considere o uso da função `cut` do pandas. Esta função permite criar faixas de valores que podem ser úteis para agrupar os dados de maneira mais significativa.
- A função `qcut` é uma alternativa à função `cut` quando você deseja criar faixas baseadas em quantis, em vez de valores fixos. Isso pode ser útil quando você deseja dividir seus dados em grupos de tamanhos iguais.
- Você também pode aplicar diferentes funções de agregação a diferentes colunas, passando um dicionário para `'aggfunc'`. Isso permite que você personalize ainda mais a maneira como seus dados são resumidos.
- Ao renomear as colunas resultantes em uma tabela pivô, você pode passar uma tupla para `'aggfunc'` em vez de uma única função ou lista de funções. Isso pode tornar seus resultados mais fáceis de entender.





# Bons estudos!

