



escola
britânica de
artes criativas
& tecnologia

Profissão: Analista de dados



GLOSSÁRIO



Data Wrangling I



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

- **Inicie sobre a dataframes Pandas**
- **Selecione e filtre dados**
- **Insira, atualize e exclua dados**



Inicie sobre a dataframes Pandas



Introdução a dataframes Pandas

Glossário

• Dados nulos ou faltantes

São valores ausentes em um conjunto de dados, que não foram registrados ou não estão disponíveis.

• Identificador de linha

É um número ou texto que identifica uma linha específica em um conjunto de dados.

• Identificador de coluna

É um nome que identifica uma coluna específica em um conjunto de dados.

• Identificador de limite

É um valor que define o limite máximo ou mínimo para um determinado conjunto de dados.



Introdução a dataframes Pandas

Glossário

• **Método head**

É um método em Python que retorna as primeiras linhas de um conjunto de dados.

• **Método tail**

É um método em Python que retorna as últimas linhas de um conjunto de dados.

• **Valores categóricos**

São valores que representam categorias, como por exemplo, cores, gêneros, tipos de veículos, dentre outros.

• **Valor de coluna**

É o valor de uma célula em uma coluna específica.



Introdução a dataframes Pandas

Glossário

● Valor de linha

É o valor de uma célula em uma linha específica.

● Valores numéricos

São valores que representam números, como por exemplo, valores monetários, idade, altura, peso, dentre outros.

● Range

É uma função em Python que retorna uma sequência de números em um determinado intervalo, com um passo especificado.



Seleção e filtre dados



Selecione e filtre dados

Glossário

● Chave dupla

É uma combinação de dois ou mais identificadores de linha em uma tabela de dados.

● Chave simples

É um único identificador de linha em uma tabela de dados.

● Estrutura de duas dimensões

É uma estrutura de duas dimensões é uma estrutura de dados que contém informações organizadas em linhas e colunas, como uma matriz, planilha ou DataFrame.

● Estrutura de uma dimensão

É uma estrutura de dados que contém informações organizadas em uma única dimensão, como uma lista, matriz unidimensional ou série.



Selecione e filtre dados

Glossário

● **Fatiamento**

O fatiamento em um DataFrame refere-se à seleção de um subconjunto dos dados em um DataFrame com base em condições especificadas.

● **Intervalo**

Um intervalo em um DataFrame é uma seleção de um conjunto de linhas ou colunas contíguas em um DataFrame.

● **Manipulação de dados**

É o processo de transformar, limpar, filtrar ou agregar dados para torná-los mais úteis para análise.

● **Mapeamento**

É o processo de associar um conjunto de valores a outro conjunto de valores. Em Pandas, isso pode ser feito usando um dicionário.



Selecione e filtre dados

Glossário

● Método loc

É um método em Pandas que permite a seleção de um conjunto de linhas e colunas de um dataframe usando rótulos de índice.

● Seleção de dados

É o processo de escolha e extração de informações específicas a partir de uma fonte de dados, como um banco de dados ou um conjunto de dados. Geralmente é feita com base em critérios predefinidos.

● Séries

Em Pandas, uma série é uma estrutura de dados unidimensional que contém um array de valores e um array de rótulos (índice). Ela pode ser vista como uma coluna em uma tabela de dados.

● Transposição

A transposição de um DataFrame é uma operação que reorganiza as linhas e colunas de um DataFrame, trocando as linhas pelas colunas.



Insira, atualize e exclua dados



Insira, atualize e exclua dados

Glossário

● Atualização

É o processo de alterar o valor de um elemento em uma estrutura de dados. Em pandas, a atualização geralmente é realizada em um DataFrame ou uma Série.

● Filtro funcional

É uma operação em que uma função é aplicada a cada elemento em uma estrutura de dados e somente os elementos que atendem a uma condição específica são mantidos.

● Inserção

É a ação de adicionar um novo elemento a uma estrutura de dados existente.

● Operação de seleção

É uma operação em que um subconjunto de elementos é selecionado de uma estrutura de dados com base em critérios específicos.



Bons estudos!

