

Profissão: Cientista de Dados



GLOSSÁRIO



Pandas e Carga de Dados



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

- **Conheça o pacote pandas**
- **Objetos Index**
- **Gere dados**
- **Construa funções**



Conheça o pacote Pandas



Conheça o pacote Pandas

• **Array Numpy**

Uma estrutura de dados em Python que armazena valores do mesmo tipo em múltiplas dimensões. É usado para realizar operações matemáticas e estatísticas de maneira eficiente.

• **CSV**

Sigla para Comma Separated Values, é um formato de arquivo que armazena dados tabulares como texto simples, onde cada linha é um registro de dados e cada registro consiste em um ou mais campos separados por vírgulas.



Conheça o pacote Pandas

• DataFrame

Uma estrutura de dados bidimensional, semelhante a uma tabela, com linhas e colunas, no pacote pandas. É uma das principais estruturas de dados em pandas e é usada para manipulação de dados.

• Índice

Em pandas, um índice é como uma 'etiqueta' para cada linha ou coluna. Pode ser um número (como o padrão 0, 1, 2, 3...) ou um nome ou data ou qualquer valor único.



Conheça o pacote Pandas

Series

Uma estrutura de dados unidimensional no pacote pandas que pode armazenar qualquer tipo de dados. É semelhante a um array do numpy, mas com rótulos adicionais, permitindo um índice nas linhas.

Valor Escalar

Um valor único ou um elemento individual. Em Pandas, uma Serie pode ser criada a partir de um valor escalar.



Objetos Index



Objetos Index

● **Condição booleana**

Uma expressão que retorna um valor verdadeiro ou falso. No contexto de filtragem de dados, uma condição booleana é usada para selecionar linhas que atendam a essa condição.

● **Método .loc**

Método do Pandas usado para acessar um grupo de linhas e colunas por rótulo(s) ou uma matriz booleana.

● **Filtragem de dados**

Processo de selecionar subconjuntos de dados com base em critérios específicos. No Pandas, isso pode ser feito criando uma condição booleana e aplicando-a ao DataFrame.

● **Método .iloc**

Método do Pandas usado para acessar um grupo de linhas e colunas por índice(s) inteiro(s).



Objetos Index

• Método `.values`

Método do Pandas que converte um DataFrame em um array NumPy, permitindo que você acesse e modifique valores diretamente.

• Sintaxe de colchetes

Método para acessar colunas em um DataFrame do Pandas. O nome da coluna é passado como uma string dentro de colchetes.

• `pd.options.display.float_format`

Método do Pandas usado para alterar o formato de exibição dos números em um DataFrame.

• Transposição de DataFrame

Processo de trocar linhas e colunas em um DataFrame. No Pandas, isso pode ser feito usando o atributo `.T`.



Opere DataFrames e Series



Opere DataFrames e Series

• Função exponencial

A função exponencial é uma função matemática que pode ser usada para transformar os dados. Em Python, a função exponencial pode ser aplicada a um DataFrame usando a biblioteca Numpy.

• NaN

NaN (Not a Number) é um valor especial definido em Numpy, usado para representar valores ausentes ou indefinidos em um DataFrame.

• Função random

A função random é uma função da biblioteca Numpy que gera números aleatórios. Nesta aula, é usada para criar DataFrames.

• Transposição

A transposição é uma operação que transforma as linhas de um DataFrame em colunas e vice-versa.



Encontre valores únicos



Encontre valores únicos

• Método 'nunique'

É um método em Python que retorna o número de elementos únicos em um objeto.

• Método 'unique'

É um método em Python que retorna elementos únicos de um objeto.



Interaja com APIs da Web e Bancos de dados



Interaja com APIs da Web e Bancos de dados

● API

Interface de Programação de Aplicações, é um conjunto de rotinas e padrões estabelecidos por um software para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do software, mas apenas usar seus serviços.

● Método 'get'

Método da biblioteca 'requests' usado para consultar um servidor web. Retorna uma resposta do servidor.

● Requests

Biblioteca Python usada para fazer solicitações HTTP. É usada para consultar um servidor web.



Salve Dataframes em CSV, Excel e JSON



Salve Dataframes em CSV, Excel e JSON

• 'to_csv'

Método utilizado para salvar um dataframe em um arquivo CSV. Permite especificar o caminho do arquivo, o separador de dados e se queremos ou não incluir o índice do dataframe no arquivo salvo.

• 'to_excel'

Método utilizado para salvar um dataframe em um arquivo Excel. Permite incluir ou não o índice do dataframe e especificar o nome da planilha. Também é possível pular a primeira linha do arquivo Excel ao salvá-lo.

• 'to_json'

Método utilizado para salvar um dataframe em um arquivo JSON.



Bons estudos!

