



escola
britânica de
artes criativas
& tecnologia

Profissão: Analista de dados



GLOSSÁRIO



Coleta de dados I



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

- **Manuseie arquivos CSV**

- **Manipule arquivos TXT**

- **Use arquivos de Excel**



Manuseie arquivos CSV



Manuseie arquivos CSV

Glossário



Bibliotecas

São conjuntos de módulos e funções que estendem a funcionalidade básica do Python.



Estruturas de dados

São maneiras de organizar e armazenar informações em um programa de computador.



Código

É uma sequência de instruções que pode ser executada por um computador.



If/else

É uma estrutura de controle de fluxo que permite executar diferentes blocos de código dependendo de uma condição.



Manuseie arquivos CSV

Glossário

Iterador

É um objeto que permite iterar sobre uma coleção de elementos um de cada vez. Em Python, um iterador é definido pela interface do método `iter()` e `next()`, que retornam o próximo elemento da coleção.

Next

É usada para retornar o próximo elemento de um iterador. Ela é usada em conjunto com um objeto iterável para iterar sobre seus elementos.

len

É uma função incorporada em Python que retorna o número de itens em um objeto que suporta contagem, como uma string, lista ou tupla.

Numpy

É uma biblioteca em Python utilizada para trabalhar com matrizes e operações matemáticas.



Manuseie arquivos CSV

Glossário

Try/except

É uma estrutura de controle de fluxo utilizada para capturar e tratar exceções em Python.

While/for

São estruturas de controle de fluxo utilizadas para executar blocos de código repetidamente enquanto uma condição é verdadeira.

Tuplas

São estruturas de dados em Python semelhantes a listas, mas que não podem ser modificadas após a sua criação.

With

É uma palavra-chave utilizada em Python para gerenciar o contexto de um bloco de código, garantindo que recursos como arquivos e conexões de rede sejam fechados adequadamente ao final da execução.



Manipule arquivos TXT



Manipule arquivos TXT

Glossário

Abertura de arquivo

É o processo de acesso a um arquivo para leitura ou escrita de dados. Para abrir um arquivo em um programa de computador, geralmente é necessário especificar o caminho para o arquivo e o modo de acesso desejado (leitura, escrita ou ambos).

Dados estruturados

São dados que possuem uma estrutura bem definida e organizada, com um formato que permite a sua manipulação e análise de forma sistemática. Exemplos de dados estruturados incluem bancos de dados, planilhas eletrônicas, arquivos XML, JSON, entre outros.



Manipule arquivos TXT

Glossário

• Dados não estruturados

São dados que não possuem uma estrutura bem definida, ou seja, não estão organizados de forma sistemática em um formato específico. Exemplos de dados não estruturados incluem textos livres, imagens, áudios, vídeos, entre outros.

• Linha

É uma string que representa uma linha de texto em um arquivo.

• Função join

Método de strings em Python que une uma lista de substrings em uma única string, utilizando um separador definido. No texto, o separador utilizado é um espaço vazio ou espaço em branco.

• Operador and

operador lógico que retorna True apenas se as duas expressões que o precedem forem verdadeiras. Por exemplo, True and True é True, enquanto True and False é False.



Manipule arquivos TXT

Glossário

• Operador filter

Função que recebe uma lista e uma função como argumentos e retorna uma nova lista contendo apenas os elementos da lista original para os quais a função retorna True.

• Operador for

Utilizado em loops (laços) em Python para percorrer elementos de uma sequência (como uma lista) ou executar um bloco de código várias vezes.

• Operador not

Operador lógico que inverte o valor booleano de uma expressão. Por exemplo, not True é False.

• Operador in

Utilizado para verificar se um elemento está presente em uma sequência (como uma lista).



Manipule arquivos TXT

Glossário

• Operador or

Operador lógico que retorna True se pelo menos uma das expressões que o precedem for verdadeira. Por exemplo, True or False é True.

• Processamento de texto

Refere-se ao conjunto de técnicas e ferramentas utilizadas para manipular e transformar textos, como por exemplo, remover palavras indesejadas, extrair informações relevantes, realizar traduções automáticas, entre outras.

• Regex (ou Expressões Regulares)

É uma sequência de caracteres que define um padrão de busca em um texto. É uma linguagem usada para identificar padrões em textos, e é muito útil para a busca e manipulação de dados em grande escala, especialmente em processamento de linguagem natural e em análise de dados.



Manipule arquivos TXT

Glossário

• Vazio

Uma string que representa uma linha vazia em um arquivo.

• Writerows

É uma função da biblioteca padrão do Python para escrever linhas de dados em um arquivo CSV (Comma-Separated Values), onde cada linha é uma lista de valores separados por vírgulas. A função writerows é uma forma conveniente e eficiente de escrever muitos dados em um arquivo CSV de uma só vez, sem a necessidade de iterar sobre cada linha individualmente.



Use arquivos de Excel



Use arquivos de Excel

Glossário

• Estrutura de dados tabular

É um tipo de estrutura de dados que organiza informações em forma de tabelas, onde cada coluna representa um atributo da informação e cada linha representa uma instância ou registro. É uma forma organizada e eficiente de armazenar e manipular grandes volumes de dados relacionais, permitindo a análise, visualização e processamento de dados de maneira mais fácil e intuitiva. Alguns exemplos de formatos tabulares incluem planilhas do Excel, bancos de dados relacionais e arquivos CSV (Comma Separated Values).

• Dataframe

Um DataFrame é uma estrutura de dados tabular bidimensional e mutável, organizada em linhas e colunas, semelhante a uma planilha de Excel ou uma tabela de banco de dados. É uma das principais estruturas de dados oferecidas pela biblioteca Pandas em Python, que é muito utilizada para análise de dados.



Use arquivos de Excel

Glossário

Openpyxl

É uma biblioteca em Python que permite ler e escrever arquivos Excel (XLSX/XLSM/XLSB/XLS) usando o formato de arquivo aberto OpenXML. Com o openpyxl, os usuários podem automatizar a manipulação de dados em planilhas do Excel e criar programas para executar tarefas repetitivas, como formatação de células, cálculos e filtragem de dados.



Bons estudos!

