

INTRODUÇÃO À CIÊNCIA DE DADOS

Biblioteca Pandas

BIBLIOTECA PANDAS



- Tem o nome originado de "panel data", e suas principais funções estão atreladas ao uso de dados multidimensionais.
- Uma das principais ferramentas para limpeza e análise de dados.
- Pandas parece muito com NumPy, entretanto o Pandas foi projetado para trabalhar com dados tabulares e heterogêneos, enquanto o NumPy é mais adequado a dados homogêneos em arrays.
- Pandas é uma das bibliotecas mais utilizadas por Cientistas de Dados ao redor do mundo.
- As estruturas fundamentais de trabalho no pandas são as "Series" e o "Dataframe".

SERIES

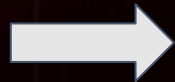
- Objeto do tipo array unidimensional contendo uma sequência de valores e um array associado de rótulos (labels) de dados chamado de índice.
- A representação em string de uma Series mostra índices à esquerda e valores à direita (vamos ver na demonstração).
- Diferente do NumPy, nas Series podemos ter rótulos nos índices.

DATAFRAME

- Representa uma tabela de dados retangular e contém uma coleção ordenada de colunas, em que cada uma pode ter um tipo de valor diferente (booleano, string, numérico, entre outros).
- O Dataframe tem índice tanto para linha quanto para coluna, podendo ser imaginado como um dicionário (dict) de Series, todos compartilhando o mesmo índice.
- A capacidade de trabalhar com vários tipos de dados, como se fosse uma tabela de banco de dados, favorece consideravelmente o trabalho de um profissional de DS.

DEMONSTRAÇÃO

- Assim como na Biblioteca NumPy, vamos fazer uma demonstração de uso da biblioteca Pandas.
- Importante ressaltar a necessidade de conhecer profundamente os recursos do Pandas, pois acelera demais a resolução de problemas do cotidiano com dados.
- Estude o livro texto, o material de apoio e abuse do uso do Pandas.
- Demonstração



FINALIZANDO

- Essencial saber explorar a Biblioteca Pandas.
- Usaremos muito essa biblioteca daqui para frente em outras tarefas que vamos realizar.
- Lembre-se sempre de explorar os recursos de Series e Dataframes do Panda, e não se esqueça que um Dataframe é formado por Series.
- Aprofunde muito seus estudos em Pandas!

INTRODUÇÃO À CIÊNCIA DE DADOS

Biblioteca Pandas