

**Fig. 1**

Bem-vindo à Análise de dados com Pandas. Nesta formação cobrimos os tópicos de Pandas e explicamos todos os conceitos com notebooks jupyter para que possa observar o funcionamento de todos os métodos em Pandas, testar, simular e rever opções de análise sobre os dados.

Primeiro de tudo, vamos entender "o que é análise de dados e por que devemos usar o Pandas para análise?".

O que é análise de dados?

Suponha que esta a trabalhar numa empresa que diariamente gera muitos dados de clientes e que lhe é atribuída uma tarefa para extrair algumas informações úteis com determinado prazo. O que fará se tiver um tempo muito limitado, não pode extrair informações apenas olhando para o conjunto de dados porque o tamanho dos dados é enorme. Estude o Pandas e descubra que o Pandas torna sua vida mais fácil do que apenas olhar para o conjunto de dados e encontrar informações úteis.

O que o Pandas realmente faz?

A documentação oficial diz:

O Pandas é um pacote Python que fornece estruturas de dados rápidas, flexíveis e expressivas projetadas para tornar o trabalho com dados "relacionais" ou "rotulados" fácil e intuitivo. Ele visa ser o bloco de construção fundamental de alto nível para fazer análises de dados práticas e do mundo real em Python. Além disso, tem o objetivo mais amplo de se tornar a ferramenta de análise / manipulação de dados de código aberto mais

poderosa e flexível disponível em qualquer idioma. Já está bem encaminhado para esse objetivo.

O Pandas é adequado para muitos tipos diferentes de dados:

- \* Dados tabulares com colunas de tipo heterogêneo, como em uma tabela SQL ou folhas de cálculo Excel.
- \* Dados de séries temporais ordenados e não ordenados (não necessariamente de frequência fixa).
- \* Dados de matriz arbitrária (tipo homogêneo ou heterogêneo) com rótulos de linha e coluna.
- \* Qualquer outra forma de conjuntos de dados observacionais/estatísticos. Os dados realmente não precisam ser rotulados para serem colocados em uma estrutura de dados Pandas.

Dependências:

```
pip install pandas
pip install numpy
```

Índice:

- [Introdução aos pandas](#)
- [Objeto Dataframe](#)
- [Lendo, escrevendo arquivo CSV e EXCEL](#)
- [Manipulação de dados ausentes parte-1](#)
- [Manipulação de dados ausentes parte-2](#)
- [Agrupar: Dividir, Combinar e Aplicar](#)
- [Concat Dataframes](#)
- [Combinar Dataframes](#)
- [Pivots e Tabelas Dinâmicas](#)
- [Remodelar o Dataframe](#)

Introdução ao Pandas:

## Fig.2

O Pandas é usado como ferramenta de limpeza de dados no campo da ciência de dados. Pode fazer qualquer operação que desejar no conjunto de dados com esta ferramenta. Agora surge a pergunta, podemos limpar ou alterar o valor no conjunto de dados manualmente? A resposta é sim, podemos se o tamanho do conjunto de dados for pequeno. E se tivermos um conjunto de dados grande, não podemos fazê-lo manualmente, levará muito tempo. O Pandas torna a ciência de dados muito fácil e eficaz.

Para usar o Pandas precisa primeiro importar o módulo pandas no seu programa

```
import pandas as pd
```

**Ler folhas de cálculo CSV e Excel:**

**d=pd.read\_csv("caminho"):** \* `pd.read_csv()` é a função para ler o ficheiro CSV (valores separados por vírgula) do seu computador. \* Na função tem que passar "caminho" do ficheiro CSV com aspas. \* Armazene o dataframe em qualquer variável, aqui eu armazenei na variável "d". \* A função `read_csv()` transforma o arquivo CSV em dataframe para que possa acedê-lo como um dicionário.

**d=pd.read\_excel("caminho") :** \* É o mesmo que `read_csv()`, mas lê folhas de cálculo ou ficheiro do Excel. Aqui usando o conjunto de dados meteorológicos que possui todos os dados do clima. No arquivo `weather.csv` está no meu diretório atual e é por isso que o caminho do arquivo é o próprio nome do arquivo.

```
d=pd.read_csv('datasets/weather.csv')
print(d)
```

## Referências:

- [Documentação Oficial Pandas](#)
- [Tutoriais](#)