**Statistics Library In Python**

**Statistics Library In Python:**

* **Pandas: ;**
* **Numpy: ;**
* **Matplotlib: ;**
* **Scipy: ;**
* **Holoview: ;**

1. **Introdução à Análise Exploratória de Dados**

**Dataframe**

df = pd.read\_csv(“.......”, sep=’;’)

\*sep = leia esse arquivo mas utilize o separador ;

* **Series –** Uma matriz de apenas uma coluna;
* **DataFrame –** Várias séries interligadas em um único bloco – tipo um arquivo excel;

**Describe**

A função Describe analisa o seu Dataframe para verificar como está a distribuição dos seus dados. Ela apresenta algumas métricas estatísticas

**Uso de API**

Uma API é um programa que é escrito para integrar sistemas.

1. **Manipulando Dados com Pandas**

**Uso de Filtros**

df = data.filter(items=['Open', 'Close'])

df

# Caso não lembre o nome complete de uma coluna você pode usar o parâmetro 'like'

# e aí você vai usar o parâmetro do eixo (axis - coluna)

df = data.filter(like='gh', axis=1)

df

**Escolhendo Colunas**

Para filtrar uma coluna podemos utilizar o método filter() ou apenas:

df = data.Close

df

ou

df = data['Close']

df

Se você quiser trabalhar com mais de uma coluna:

df = data[['Open', 'Close']]

df

**Retirando Linhas**

**Retirando Colunas**

**Map e Replace**

1. **Funções no Pandas**

**Conceito de Função**

A função nada mais é que um trecho de código e esse mesmo estará contido no código no programa para ser utilizado várias vezes. Utiliza-se funções principalmente para evitar duplicidade de código.

**Funções Matemáticas no Pandas**

**Apply**

O apply é um método que espera de entrada uma função. Para usar uma função utiliza-se o apply.

def duplicaBolsa(rows):

nova\_bolsa = rows \* 2

return nova\_bolsa

df = df\_prouni[['uf\_busca', 'curso\_busca', 'mensalidade', 'bolsa\_integral\_ampla']]

df

df['nova\_bolsa'] = df['bolsa\_integral\_ampla'].apply(duplicaBolsa)

df['nova\_bolsa'] = df['nova\_bolsa'].fillna(1)

**Lambda**

A linguagem python possui uma palavra reservada ‘Lambda’, que é uma função anônima. Para funções simples e rápidas.

A linguagem Python possui a palavra reservada 'lamba' que é uma função anônima.

\* Para funções simples e rápidas;

o que seria:

def duplica\_bolsa(rows):

nova\_bolsa = rows \* 2

=

lambda x : x \* 2

-> rows : rows \* 2

1. **Índices no Pandas**

**Conceito de Indices**

frutas = [‘Maça’, ‘Banana’, ‘Abacaxi’]

frutas[0:2]:

De 0 até 2

* 0: Inclusivo
* 2: Exclusivo

[‘Maça’, ‘Banana’]