



RPA, **na** PRATICA

Aula – Pandas

Explicação

Pandas é uma biblioteca do Python que permite extrair dados de diversos arquivos, como Excel, Json, HTML, SQL, etc...

A mesma possui funcionalidades voltadas para análises de dados e ferramentas de manipulação de tabelas.

Glossário

DataFrame

Estrutura para dados que é composta de duas dimensões, estas sendo linhas e colunas.

Assemelha-se à uma tabela de Excel.

	CPF	Nome_Completo	Data_Nascimento	Telefone	Email	CEP
0	116415509	Analú Julia da Paz	1974-04-05	79982386771	aanalujuliadapaz@alwan.com.br	49026130
1	4847726278	Kamilly Clara Caldeira	1982-04-21	27998331799	kamillyclaracaldeira__kamillyclaracaldeira@caf...	29194248
2	7287125803	Eduarda Emily Almada	1941-04-20	32985692709	eduardaemilyalmada_@peopleside.com.br	36200142
3	17924324534	Miguel Matheus Nogueira	1998-09-07	85998428735	mmiguelmatheusnogueira@br.rhodia.com	61654090
4	18356547113	Lívia Andrea Porto	1979-01-02	61991134556	lliviaandreaporto@gmx.de	70645240

Series

Estrutura de dados de uma dimensão, sendo basicamente composta dos valores de uma linha ou coluna que representa apenas uma fatia do DataFrame.

```
CPF                116415509
Nome_Completo      Analú Julia da Paz
Data_Nascimento    1974-04-05 00:00:00
Telefone           79982386771
Email              aanalujuliadapaz@alwan.com.br
CEP                49026130
Name: 0, dtype: object
```



Comandos

- Instalando a biblioteca

```
pip install pandas
```

- Importando a biblioteca

```
import pandas as pd
```

- Lendo um arquivo Excel

```
df = pd.read_excel("tabela_clientes.xlsx")
```

- Resumo das dimensões do DataFrame

```
df.shape
```

- Ver nome das colunas

```
df.columns
```

- Ver os primeiros resultados Como padrão a quantidade de linhas à ser trazido são cinco, porém esse valor pode ser alterado se preenchido os parênteses de acordo com o desejado

```
df.head()
```

ou

```
df.head(20)
```

- Ver os últimos resultados

```
df.tail()
```



- Selecionar uma coluna específica - pelo nome O segundo método apresentado no exemplo somente é permitido de usar em nomes de colunas sem espaço entre palavras.

```
df["Nome_Completo"]
```

```
# ou
```

```
df.Nome_Completo
```

- Selecionar o primeiro valor de uma coluna específica

```
df["Nome_Completo"].values[0]
```

- Selecionar uma linha através do index

```
df.iloc[0]
```

- Selecionar uma linha e uma coluna através do index

```
df.iloc[0,0]
```

- Criando um DataFrame

```
data = {  
    "Nome": ["Harry", "Hermione", "Rony", "Luna", "Neville"],  
    "Herbologia": [8,5,6,9,7],  
    "Alquimia": [7,9,5,4,7],  
    "Adivinhacao": [7,4,7,6,8]  
}  
  
df_hp = pd.DataFrame(data)
```



- Estatísticas descritivas das colunas - media, desvio padrão, mínimo, máximo, etc ...

```
df.describe()
```

- Extraíndo o DataFrame em um arquivo Excel

```
df_hp.to_excel("Boletins_Alunos.xlsx", index=False)
```





RPA, **na** PRAÁTICA