escola britânica de artes criativas & tecnologia

Profissão: Analista de dados

escola britânica de artes criativas & tecnologia

DATA WRANGLING I

GUIA DA AULA 1

escola britânica de artes criativas & tecnologia

Introdução a DataFrames Pandas

- **Pacote Pandas**
- **Anatomia de um DataFrame**

Acompanhe aqui os temas que serão tratados na videoaula

Estruturas de dados

Não estruturado: texto, imagem, áudio etc.

Semi estruturado: html, json etc.

****Estruturado****: tabelas, planilhas etc.

DataFrame Pandas

1. Pacote Pandas

Pacote Python para manipulação de dados (talvez o mais utilizado). A abstração base do pandas é o DataFrame, uma estrutura de dados Python de duas dimensões utilizado para representar tabelas. A documentação pode ser encontrada no *link* https://pandas.pydata.org/docs/

escola britânica de artes criativas & tecnologia

Exemplo: Criar um DataFrame Pandas a partir de um dicionário Python.

In [ ]:
```
transacoes = dict(
    id=[571, 572, 573],
    data=['19-01-2021', '19-01-2021', '23-01-2021'],
    valor=[371.30, 57.19, 101.21],
    categoria=['supermercado', 'farmacia', 'outros']
)
```

In [ ]:
```
import pandas as pd

transacoes_df = pd.DataFrame(transacoes)
```

In [ ]:
```
transacoes_df
```

In [ ]:
```
type(transacoes_df)
```

Atributos

In [ ]:
```
transacoes_df.columns
```

In [ ]:
```
transacoes_df.dtypes
```

In [ ]:
```
transacoes_df.index
```

In [ ]:
```
transacoes_df.shape
```

Métodos

In [ ]:
```
transacoes_df.head(n=1)
```

In [ ]:
```
transacoes_df.tail()
```

In [ ]:
```
transacoes_df.info()
```

In [ ]:
```
transacoes_df[['id', 'valor']].describe().T # colunas numéricas
```

In [ ]:
```
transacoes_df[['data', 'categoria']].describe().T # colunas categóricas
```

Exemplo: Criar um DataFrame Pandas a partir de um arquivo csv.

In [ ]:
```
%%writefile github.csv
ranking;project;language;stars;stars_today;forks
1;plow;go;1304;574;38
2;n8n;typescript;15668;280;1370
3;slides;go;3218;265;80
4;defi-developer-road-map;;636;247;49
5;pytorch-image-models;python;11065;101;1646
6;javascript-algorithms;javascript;110768;248;18331
7;paddleclas;python;1429;283;323
8;reddit_sentiment_trader;python;369;71;60
9;augly;python;2849;393;99
10;self-taught-guide-to-cloud-computing;;863;179;84
```

In [ ]:
```
import pandas as pd

github_df = pd.read_csv('github.csv', sep=';')
```

In [ ]:
```
github_df
```

In [ ]:
```
type(github_df)
```

Atributos

In [ ]:
```
github_df.columns
```

In [ ]:
```
github_df.dtypes
```

In [ ]:
```
github_df.index
```

In [ ]:
```
github_df.shape
```

Métodos

In [ ]:
```
github_df.head()
```

In [ ]:
```
github_df.tail()
```

In [ ]:
```
github_df.info()
```

In [ ]:
```
github_df[['project', 'language']].describe().T # colunas categóricas
```

In [ ]:
```
github_df[['ranking', 'stars', 'stars_today', 'forks']].describe().T
# colunas numéricas
```

2. Anatomia de um DataFrame

- **Série**: `Coluna` de um DataFrame;
- **Índice**: Indentificador de uma `linha` de um DataFrame.

In [ ]:
```
github_df
```