

Python: do básico à autonomia intermediária

aula 10

Bernardo Alves Furtado

March 25, 2022

Menu do dia

- ▶ Apresentar

Pandas

Pandas: crash course

Introdução

- ▶ Pandas é uma biblioteca de tabelas. Baseia-se em linhas e colunas (multidimensional).
- ▶ Quando vetor – somente uma coluna, chama Series.
- ▶ Quando mais de uma, chama-se DataFrame
- ▶ `import pandas as pd`

Lendo planilhas

- ▶ `data = pd.read_csv('nome.csv', sep=';')`
- ▶ `data.to_csv('nome.csv', sep=';', index=False)`

Básicos

- ▶ `data.head()`
- ▶ `data.columns()`
- ▶ `data.describe()`
- ▶ `data.info()`

Básicos

- ▶ `data.coluna.head()`
- ▶ `nova = data.copy()`
- ▶ Maiores 10 números
- ▶ `data.nlargest(10)`
- ▶ `data.coluna.mean()`
- ▶ `data.coluna.max()`

Modificando, assigning, renomeando colunas, frequência

- ▶ `data.loc[data.coluna1 == 'x', 'coluna2'] = 10`
- ▶ `data.loc[data.coluna1 < data.coluna2, 'coluna2'] = 20`
- ▶ `data.coluna.unique()`
- ▶ `data = data.rename(columns={'antigo': 'novo'})`
- ▶ `data.coluna.value_counts()`

Selecionando e pesquisando utilizando-se das colunas

- ▶ `data.loc[row_selction, column_selection]`
- ▶ Exemplo: seleciona todas as linhas com `age = 10` e todas as colunas
- ▶ `data.loc[data.age == 10, :]`
- ▶ Só a coluna `b`
- ▶ `data.loc[data.age == 10, 'b']`
- ▶ Para alterar esses valores
- ▶ `data.loc[data.age > 100, 'b'] = 99`

Groupby: exemplos centrais

- ▶ `df.groupby(["coluna"]).coluna.funcao()`
- ▶ `df.groupby(["coluna"]).coluna.agg([funcao1, funcao2])`
- ▶ Exercises:
https://github.com/guipsamora/pandas_exercises
- ▶ Exercises2:
<https://www.machinelearningplus.com/python/101-pandas-exercises-python/>

SQL

Para quem já sabe SQL:

[https://pandas.pydata.org/docs/getting_started/
comparison/comparison_with_sql.html](https://pandas.pydata.org/docs/getting_started/comparison/comparison_with_sql.html)