



escola
britânica de
artes criativas
& tecnologia

Módulo 07 | Módulos e Pacotes

Caderno de **Exercícios**

Professor [André Perez](#)

▼ Tópicos

1. from / import / as;
 2. Módulo;
 3. Pacote;
 4. Baixando pacotes.
-

▼ Exercícios

▼ 0. Preparação do ambiente

Neste exercício vamos utilizar a base de dados de ações da bolsa de valores dos EUA, a Dow Jones. Os dados estão disponíveis para *download* neste [link](#). Vamos utilizar o pacote `wget` para fazer o *download* dos dados.

- Instalando o pacote `wget` na versão 3.2.

```
!pip install wget==3.2
```

- Fazendo o download dos dados no arquivo compactado `dados.zip`

```
import wget
```

```
wget.download(url='https://archive.ics.uci.edu/ml/machine-learning-databases/00312/dow_
```

- Descompactando os dados na pasta dados com o pacote nativo zipfile

```
import zipfile
```

```
with zipfile.ZipFile('./dados.zip', 'r') as fp:
    fp.extractall('./dados')
```

Verifique a pasta dados criada, ela deve conter dois arquivos:

- **dow_jones_index.data**: um arquivo com os dados;
- **dow_jones_index.names**: um arquivo com a descrição completa dos dados.

É possível observar que o arquivo de dados é um arquivo separado por vírgulas, o famoso `csv`. Vamos renomear o arquivo de dados para que ele tenha a extensão `csv` com o pacote nativo `os`

- Renomeando o arquivo com o pacote nativo `os`

```
import os
```

```
os.rename('./dados/dow_jones_index.data', './dados/dow_jones_index.csv')
```

Pronto! Abra o arquivo e o Google Colab irá apresentar uma visualização bem legal dos dados.

▼ 1. Pandas

Para processar os dados, vamos utilizar o pacote `pandas` na versão `1.1.5`. A documentação completa por ser encontrada neste [link](#)

```
!pip install pandas==1.1.5
```

Vamos importar o pacote com o apelido (alias) `pd`

```
import pandas as pd
```

Estamos prontos para ler o arquivo.

```
df = pd.read_csv('./dados/dow_jones_index.csv')
```

O pandas trabalha com o conceito de dataframe, uma estrutura de dados com muitos métodos e atributos que aceleram o processamento de dados. Alguns exemplos:

- Visualizando as `n` primeiras linhas:

```
df.head(n=10)
```

- Visualizando o nome das colunas:

```
df.columns.to_list()
```

- Verificando o número de linhas e colunas.

```
linhas, colunas = df.shape
print(f'Número de linhas: {linhas}')
print(f'Número de colunas: {colunas}')
```

Vamos selecionar os valores de abertura, fechamento, máximo e mínimo das ações do McDonalds, listado na Dow Jones como MCD:

- Selecionando as linha do dataframe original `df` em que a coluna `stock` é igual a `MCD`

```
df_mcd = df[df['stock'] == 'MCD']
```

- Selecionando apenas as colunas de data e valores de ações.

```
df_mcd = df_mcd[['date', 'open', 'high', 'low', 'close']]
```

Excelente, o problema é que as colunas com os valores possuem o carater `$` e são do tipo `texto` (object no pandas).

```
df_mcd.head(n=10)
```

```
df_mcd.dtypes
```

Vamos limpar as colunas com o método `apply`, que permite a aplicação de uma função anônima (`lambda`) qualquer. A função `lambda` remove o caracter `$` e faz a conversão do tipo de `str` para `float`

```
for col in ['open', 'high', 'low', 'close']:
    df_mcd[col] = df_mcd[col].apply(lambda value: float(value.split(sep='$')[-1]))
```

Verifique novamente os dados e seus tipos.

```
df_mcd.head(n=10)
```

```
df_mcd.dtypes
```

Excelente, agora podemos explorar os dados visualmente.

Agora é a sua vez! Conduza o mesmo processo para extrair e tratar os dados da empresa Coca-Cola (stock column igual a `KO`).

- Selecionando as linha do dataframe original `df` em que a coluna `stock` é igual a `KO`

```
# extração e tratamento dos dados da empresa Coca-Cola.
```

Vamos selecionar os valores de abertura, fechamento, máximo e mínimo das ações da empresa Coca-Cola, listado na Dow Jones como `KO`:

- Selecionando apenas as colunas de data e valores de ações.

Excelente, o problema é que as colunas com os valores possuem o carater `$` e são do tipo texto (`object` no `pandas`).

```
# Visualize os dados do dataframe
```

```
# Verifique o tipo dos dados
```

Vamos limpar as colunas com o método `apply`, que permite a aplicação de uma função

anonima (lambda) qualquer. A função lambda remove o caractere \$ e faz a conversão do tipo de str para float

Verifique novamente os dados e seus tipos.

```
# Visualize novamente os dados do dataframe
```

```
# Verifique novamente o tipo dos dados
```

Excelente, agora podemos explorar os dados visualmente.

2. Seaborn

Para visualizar os dados, vamos utilizar o pacote seaborn na versão 0.11.1. A documentação completa por ser encontrada neste [link](#)

```
!pip install seaborn==0.11.1
```

Vamos importar o pacote com o apelido (alias) sns

```
import seaborn as sns
```

Vamos visualizar os valores de abertura das ações ao longo do tempo.

```
plot = sns.lineplot(x="date", y="open", data=df_mcd)
_ = plot.set_xticklabels(labels=df_mcd['date'], rotation=90)
```

Vamos também visualizar os valores de fechamento das ações ao longo do tempo.

```
plot = sns.lineplot(x="date", y="close", data=df_mcd)
_ = plot.set_xticklabels(labels=df_mcd['date'], rotation=90)
```

Para facilitar a comparação, vamos visualizar os quatro valores no mesmo gráfico.

```
plot = sns.lineplot(x="date", y="value", hue="variable", data=pd.melt(df_mcd, id_vars="date", value_vars=["open", "close", "high", "low"]))
```

```
_ = plot.set_xticklabels(labels=df_mcd['date'], rotation=90)
```

Para finalizar, vamos salvar o gráfico numa figura.

```
plot.figure.savefig("./mcd.png")
```

Agora é a sua vez, faça o gráfico acima para a empresa Coca-Cola e salve a imagem com o nome `ko.png`

```
# visualização dos dados da Coca-Cola.
```

Vamos visualizar os valores de abertura das ações ao longo do tempo.

Vamos também visualizar os valores de fechamento das ações ao longo do tempo.

Para facilitar a comparação, vamos visualizar os quatro valores no mesmo gráfico.

Para finalizar, vamos salvar o gráfico numa figura.

Análise as duas imagens e escreva pelo menos um *insight* que você consegue extrair dos dados. Fique a vontade para escrever quantos *insights* você quiser.

Obs: *Insights* são observações sobre o que você percebe/entende/interpreta em suas análises. No caso deste exercício, você vai analisar os dados dos gráficos da empresa McDonalds e da empresa Cola-Cola e notar o que os dados gerados podem ser interessante, que tipo de interpretação o comportamento dos dados estão te trazendo.

Insight #1: ...
