Web scraping

June 25, 2021

1 Web Automation and Information Search with Python

Challenge:

We work at an importer and the price of our products is linked to the quotation of:

- Dollar
- Euro
- Gold

We need to automatically get the quotation for these 3 items on the internet and find out how much we should charge for our products, considering a contribution margin that we have in our database.

Database: https://drive.google.com/drive/folders/1o2lpxoi9heyQV1hIlsHXWSfDkBPtze-V?usp=sharing

For this, let's create a web automation:

- We will use selenium

- Important: download the webdriver (google chrome -> chromedriver) (firefox -> geckodriver) (microsoft edge -> microsoft edge driver)

- Once it has been downloaded, I should put it at the same place than Python

```python
[2]: # Step 1 - Get the current price for each currency: Dollar, Euro and Gold
     # Step 2 - Importing the list of products
     # Step 3 - Calculate the current price for each product according to the
     ↪current currency price
```

```python
[16]: from selenium import webdriver
      from selenium.webdriver.common.keys import Keys

      # Open a new browser
      browser = webdriver.Edge("msedgedriver.exe")

      #Dollar
      browser.get('https://www.google.com/')

      # Accept the cookies with Xpath (Copy Xpath)
      browser.find_element_by_xpath('//*[@id="L2AGLb"]/div').click()
```

```python
# Xpath - Search on google
browser.find_element_by_xpath(
    '/html/body/div[1]/div[3]/form/div[1]/div[1]/div[1]/div/div[2]/input').
↪send_keys("dollar in real")

# Enter
browser.find_element_by_xpath(
    '/html/body/div[1]/div[3]/form/div[1]/div[1]/div[1]/div/div[2]/input').
↪send_keys(Keys.ENTER)

# Now we do not need write something. We need to get the current currency price.
dollar = browser.find_element_by_xpath(
    '//*[@id="knowledge-currency__updatable-data-column"]/div[1]/div[2]/
↪span[1]').get_attribute("data-value")
print(dollar)


# Open a new browser - Euro
browser.get('https://www.google.com/')


# Xpath - Search on google
browser.find_element_by_xpath(
    '/html/body/div[1]/div[3]/form/div[1]/div[1]/div[1]/div/div[2]/input').
↪send_keys("euro in real")

# Enter
browser.find_element_by_xpath(
    '/html/body/div[1]/div[3]/form/div[1]/div[1]/div[1]/div/div[2]/input').
↪send_keys(Keys.ENTER)

euro = browser.find_element_by_xpath(
    '//*[@id="knowledge-currency__updatable-data-column"]/div[1]/div[2]/
↪span[1]').get_attribute("data-value")
print(euro)


# Gold
browser.get('https://www.melhorcambio.com/ouro-hoje')

gold = browser.find_element_by_xpath(
    '//*[@id="comercial"]').get_attribute("value")
gold = gold.replace(",",".")
print(gold)
```

```
browser.quit()
```

4.921491
5.884676003
282.00

[17]:
```python
# Step 2 - Importing the list of products

import pandas as pd
table = pd.read_excel("Produtos.xlsx")
display(table)
```

	Produtos	Preço Base Original	Moeda	Cotação	Preço Base Reais \
0	Câmera Canon	999.99	Dólar	5	4999.95
1	Carro Renault	4500.00	Euro	6	27000.00
2	Notebook Dell	899.99	Dólar	5	4499.95
3	IPhone	799.00	Dólar	5	3995.00
4	Carro Fiat	3000.00	Euro	6	18000.00
5	Celular Xiaomi	480.48	Dólar	5	2402.40
6	Joia 20g	20.00	Ouro	350	7000.00

	Margem	Preço Final
0	1.40	6999.930
1	2.00	54000.000
2	1.70	7649.915
3	1.70	6791.500
4	1.90	34200.000
5	2.00	4804.800
6	1.15	8050.000

[18]:
```python
# update the currency
#table.loc[linha,coluna] = value

table.loc[table["Moeda"] == "Dólar", "Cotação"] = float(dollar)
table.loc[table["Moeda"] == "Euro", "Cotação"] = float(euro)
table.loc[table["Moeda"] == "Ouro", "Cotação"] = float(gold)

display(table)
```

	Produtos	Preço Base Original	Moeda	Cotação	Preço Base Reais \
0	Câmera Canon	999.99	Dólar	4.921491	4999.95
1	Carro Renault	4500.00	Euro	5.884676	27000.00
2	Notebook Dell	899.99	Dólar	4.921491	4499.95
3	IPhone	799.00	Dólar	4.921491	3995.00
4	Carro Fiat	3000.00	Euro	5.884676	18000.00
5	Celular Xiaomi	480.48	Dólar	4.921491	2402.40
6	Joia 20g	20.00	Ouro	282.000000	7000.00

	Margem	Preço Final

```
0    1.40     6999.930
1    2.00    54000.000
2    1.70     7649.915
3    1.70     6791.500
4    1.90    34200.000
5    2.00     4804.800
6    1.15     8050.000
```

[22]:
```python
# Step 3 - Calculate the current price for each product according to the
 ↪current currency price

#update Price Base Reais
table["Preço Base Reais"] = table["Preço Base Original"]*table["Cotação"]

#update final price
table["Preço Final"] = table["Preço Base Reais"]*table["Margem"]
display(table)
```

```
        Produtos  Preço Base Original  Moeda     Cotação  Preço Base Reais  \
0    Câmera Canon               999.99  Dólar    4.921491       4921.441785
1   Carro Renault              4500.00   Euro    5.884676      26481.042013
2   Notebook Dell               899.99  Dólar    4.921491       4429.292685
3          IPhone               799.00  Dólar    4.921491       3932.271309
4      Carro Fiat              3000.00   Euro    5.884676      17654.028009
5  Celular Xiaomi               480.48  Dólar    4.921491       2364.677996
6        Joia 20g                20.00   Ouro  282.000000       5640.000000

   Margem   Preço Final
0    1.40   6890.018499
1    2.00  52962.084027
2    1.70   7529.797565
3    1.70   6684.861225
4    1.90  33542.653217
5    2.00   4729.355991
6    1.15   6486.000000
```

[23]:
```python
# Step 4 - Save new database

table.to_excel("Produtos2.xlsx" , index=False)
```