Project_1_2

July 28, 2020

1 Project 1 Exercise 2. Digikala scraping

[1]:
```
import requests
from bs4 import BeautifulSoup
```

[2]:
```
url = "https://www.digikala.com/product/dkp-1891262/
↪%D8%B3%D8%A7%D8%B9%D8%AA-%D9%85%DA%86%DB%8C-%D8%B9%D9%82%D8%B1%D8%A8%D9%87-%D8%A7%DB%8C-%D9
page = requests.get(url)
```

[3]:
```
page_souped = BeautifulSoup(page.text, 'html.parser')
```

1.1 1. Product Images

[4]:
```
from IPython.display import Image
from IPython.core.display import HTML

c_gallery = page_souped.findAll('div', attrs={'class':'c-gallery'})[0]

for image in c_gallery.findAll('img'):
#     print(image['data-src'].split('?')[0])
    display(Image(url= image['data-src'].split('?')[0]))
```

```
<IPython.core.display.Image object>


<IPython.core.display.Image object>


<IPython.core.display.Image object>


<IPython.core.display.Image object>


<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

1.2 2.Product Pramiters Summary

[5]:
```
all_product_prams = page_souped.findAll('div', attrs={'class':
 ↪'c-product__params js-is-expandable'})[0]
each_product_prams = all_product_prams.findAll('li')

# product
for pram in each_product_prams:
    try:
        print(pram.findAll('span')[0].text.strip(),end=' ')
        print(pram.findAll('span')[1].text.strip())
        print('-'*50)
    except:
        pass
```

```
      :
--------------------------------------------------
    :
--------------------------------------------------
       :




--------------------------------------------------
           :
--------------------------------------------------
     :
--------------------------------------------------
          :
--------------------------------------------------
                  : 10ATM
--------------------------------------------------
     :    /      2018
--------------------------------------------------
        :
--------------------------------------------------
     :
--------------------------------------------------
    :
--------------------------------------------------
```

1.3 3.Product Summary

[6]:
```
summary = page_souped.findAll('section', attrs={'class':
↪'c-content-expert__summary'})[0].findAll('p')[0]
print(summary.text)
```

```
                  «      »    «Lamborghini TL Premio Bic 03»
                             .
 (Limited Edition)
                  .

                       .
                                   .

                    6         .
                             .

        .             TL Premio Bic 03            100
                                          .

                                  .

                          .           limited edition
      TL-Premio Bic 03
                                               .
```

1.4 4.Technical Specifications

[7]:
```
options = page_souped.findAll('article')[2] #, attrs={'class':'c-params␣
↪is-active'})
for pram in options.findAll('li'):
    print(pram.findAll('span')[0].text.strip(), end=' --> ')
    each_pram = [pram.strip() for pram in pram.findAll('span')[1].text.
↪split(',')]
    for pram in each_pram:
        print(pram, end=',')

    print('\n','-'*50)
```

```
    -->   ,
--------------------------------------------------
      -->       ,   ,   ,  ,,
--------------------------------------------------
     -->   ,
--------------------------------------------------
     -->     (   ),
```

```
------------------------------------------------
     -->    ,
------------------------------------------------
          -->                                    ,
------------------------------------------------
    -->   ,
------------------------------------------------
    -->    ,
------------------------------------------------
       -->        ,
------------------------------------------------
     -->       ,,
------------------------------------------------
      -->    ,
------------------------------------------------
           -->    ,
------------------------------------------------
     -->    ,
------------------------------------------------
          -->    ,
------------------------------------------------
           -->    ,
------------------------------------------------
                 --> 10ATM,
------------------------------------------------
     -->    /      2018,
------------------------------------------------
        -->     ,
------------------------------------------------
          -->    ,
------------------------------------------------
    --> Premio Bic 03,
------------------------------------------------
       24     -->    ,
------------------------------------------------
```

1.4.1 6.Product Price

[8]:
```
price = page_souped.findAll('div', attrs={'class':'c-product__seller-price-raw␣
↪js-price-value'})[0].text.strip()
print('Product Price: ', price)
```

```
Product Price:     ,,
```