LAPORAN SCRAPING WEB DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

SCRAPING WEBSITE bukalapak.com

INPUT	OUTPUT
pwd	In [5]: pwd Out[5]: 'C:\\Users\\arifs.DESKTOP-EQJJN52\\Documents\\unpam\\Semester3\\Pengantar_Big_Data\\KelasBigData'
import urllib import pandas as pd from bs4 import BeautifulSoup as bs	In [2]: import urllib import pandas as pd from bs4 import BeautifulSoup as bs
keyword='pulsar' with urllib.request.urlopen("https://www.bukalapak.com/products?from=omnisearch&from_keyword_history=false&search%5Bkeywords%5D="+keyword+"&search_source=omnisearch_keyword&source=navbar") as page: html = page.read() print(html)	In [6]: keyword='pulsar' with urllib.request.urlopen("https://www.bukalapak.com/products?from=omnisearch&from_keyword_history=false&search%5Bkeywords%5D=" html = page.read() print(html) b'<!DOCTYPE html><html lang="id-ID" xmlns="http://ogp.me/ns/fb#"><head><script src="/ents-we-dested-their-had-tise-Cornesse-Exeunt-to" async></script><meta charset="utf-8"><title>Jual Produk Pulsar Termurah dan Terlengkap November 2023 | Bukalapak</title><meta property="og:url" content="https://www.bukalapak.com/products"><meta property="og:site_name" content="Bukalapak"><link rel="canonical" href="https://www.bukalapak.com/products/s/pulsar"><link rel="amphtml" href="https://m.bukalapak.com/amp/products/s/pulsar"><link href="android-app://com.bukalapak.android/https/www.bukalapak.com/products/s/pulsar" rel="alternate"><link rel="alternate" media="only screen and (max-width: 640px)" href="https://m.bukalapak.com/products/s/pulsar"><meta property="google:client_id" content="1077645477673-l06bkk6fuvf3vcblkis6d057ksuh4e39.apps.googleusercontent.com"><meta property="facebook:client_id" content="727108917352926"><meta property="fb:admins" content="784705928"><meta property="fb:app_id" content="305662322778267"><meta name="alexaVerifyID" content="8InBYg0jrSWd4S2bQaibhQFgeIQ"><meta name="google" content="notranslate"><meta http-equiv="X-UA-Compatible" content="IE=edge"><link href="https://www.bukalapak.com/web-manifest/basic" rel="manifest"><link href="https://www.bukalapak.com/opensearch.xml" rel="search" title="Bukalapak Search" type="application/opensearchdescription+xml"><meta name="viewport" content="width=1200"><script type="text/javascript">var USER=JSON.parse(atob(\'eyJpZCI6IiIsImlkZW50aXR5IjoiODU5ZjBiMGY3NzU1NmZhZjI0OTIzOWIyNWRkNjk5YWQiLCJpZDM2IjoiIiwiZW1haWwiOiIiLCJlbWFpbDI1NiI6IiIsIm5hbWUiOiIiLCJ1c2VybmFtZSI6IiIsImF2YXRhclVSTCI6IiIsImdlbmRlciI6IiIsInBob25lIjoiIiwiY29uZmlybWVkUGhvbmUiOiIiLCJhZ2VudCI6eyJpc1hQnJvd3NlciI6ZmFsc2UsImlzQXV0aG9yaXRhdGl2ZSI6ZmFsc2UsImlzTW9iaWxlIjpmYWxzZSwiaXNUYWJsZXQiOmZhbHNlLCJpc2lQaG9uZSI6ZmFsc2UsImlzQW5kcm9pZCI6ZmFsc2UsImlzQmxhY2tiZXJyeSI6ZmFsc2UsImlzT3BlcmEiOmZhbHNlLCJpc0lFIjpmYWxzZSwiaXNNFZGdlIjpmYWxzZSwiaXNNJRUNvbXBhdGliaWxpdHlNb2RlIjpmYWxzZSwiaXNNTYWZhcmkiOmZhbHNlLCJpc0ZpcmVmb3giOmZhbHNlLCJpc1dlYmtpdCI6ZmFsc2UsImlzQ2hyb21lIjpmYWxzZSwiaXNNLb25xdWVyb3IiOmZhbHNlLCJpc09tbmlXZWIiOmZhbHNlLCJpc1NlYU1vbmtleSI6ZmFsc2UsImlzRmxvY2siOmZhbHNlLCJpc0FtYXlhIjpmYWxzZSwiaXNNQaGFudG9tSlMiOmZhbHNlLCJpc0VwaXBoYW55IjpmYWlzZSwiaXNNEZXNrdG9wIjpmYWxzZSwiaXNNXaW5kb3dzIjpmYWxzZSwiaXNNMaW51e

LAPORAN SCRAPING WEB DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT	OUTPUT

soup = bs(html, 'lxml')
print(type(soup))
print(soup.prettify())
	In [7]:
	`soup = bs(html, 'lxml')`
	`print(type(soup))`
	`print(soup.prettify())`
	
	`<class 'bs4.BeautifulSoup'>`
	`<!DOCTYPE html>`
	`<html lang="id-ID" xmlns="http://ogp.me/ns/fb#">`
	`<head>`
	`<script async="" src="/ents-we-dested-their-had-tise-Cornesse-Exeunt-to">`
	`</script>`
	`<meta charset="utf-8"/>`
	`<title>`
	`Jual Produk Pulsar Termurah dan Terlengkap November 2023 | Bukalapak`
	`</title>`
	`<meta content="https://www.bukalapak.com/products" property="og:url"/>`
	`<meta content="Bukalapak" property="og:site_name"/>`
	`<link href="https://www.bukalapak.com/products/s/pulsar" rel="canonical"/>`
	`<link href="https://m.bukalapak.com/amp/products/s/pulsar" rel="amphtml"/>`
	`<link href="android-app://com.bukalapak.android/https/www.bukalapak.com/products/s/pulsar" rel="alternate"/>`
	`<link href="https://m.bukalapak.com/products/s/pulsar" media="only screen and (max-width: 640px)" rel="alternate"/>`
	`<meta content="1077645477673-l06bkk6fuvf3vcblkis6d057ksuh4e39.apps.googleusercontent.com" property="google:client_id"/>`
	`<meta content="727108917352926" property="facebook:client_id"/>`
	`<meta content="784705928" property="fb:admins"/>`

produk = soup.find_all("div", "bl-product-card__description")
# produk 	
i = 1
for i in produk :
nama = i.find('a', "bl-link").get_text().replace('\n','').replace(' ','')
harga = i.find('p', 'bl-text bl-text--subheading-20 bl-text--semi-bold bl-text--ellipsis__1').get_text().replace('Rp','').replace(' ','').replace('.','').replace('\n','')
print(nama)
print(harga)
	In [41]:
	`produk = soup.find_all("div", "bl-product-card__description")`
	`# produk`
	`i = 1`
	`for i in produk :`
	`nama = i.find('a', "bl-link").get_text().replace('\n','').replace(' ','')`
	`harga = i.find('p', 'bl-text bl-text--subheading-20 bl-text--semi-bold bl-text--ellipsis__1').get_text().replace('Rp','').re`
	`print(nama)`
	`print(harga)`
	
	`Sarung Motor PULSAR 220 Silver BIRU Body Cover Penutup Pelindung Motor Pulsar 220 PREMIUM`
	`52000`
	`Sarung Motor BAJAJ PULSAR Hitam MERAH Body Cover Penutup Pelindung Motor Bajaj PREMIUM`
	`52000`
	`SEPASANG KANAN DAN KIRI HANDFAT HANDGRIP UNIVERSAL MOTOR YAMAHA HONDA KAWASAKI SUZUKI MOCIN PULSAR HARGA MURAH`
	`15000`
	`SEPASANG BANDUL STANG SETIR STIR UNIVERSAL MOTOR DRAT BAUT 10 FOR YAMAHA HONDA SUZUKI KAWASAKI MOCIN PULSAR BAHAN BESI PADAT CROM MULUS`
	`25000`
	`GAS SPONTAN YZ THAILAND SETT SEPASANG KARET HANDGRIP DAN SEPASANG BANDUL STANG NYA FOR UNIVERSAL MOTOR YAMAHA HONDA SUZUKI KAWASAKI MOCIN PULSAR DLL`
	`75000`
	`(PAKET) RUMAH KOPLING BAJAJ PULSAR 135 PULSAR125 XCD KAMPAS KOPLING BAJAJ PULSAR 125 135`
	`125000`
	`ROCKER ARM PELATUK PLATUK KLEP BAJAJ PULSAR 180 200 220`
	`145000`
	`fullset paking packing perpak lengkap komplit bajaj pulsar 135`
	`64000`
	`Packing paking perpak gasket fullset full set kawasaki bajaj pulsar 200Ns 200 ns`

LAPORAN SCRAPING WEB DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT

```
nama_produk = []
harga_produk = []
produk = soup.find_all("div", "bl-product-card__description")
# produk
i = 1
for i in produk :
    nama = i.find('a', "bl-link").get_text().replace('\n','').replace('  ','')
    harga = i.find('p', 'bl-text bl-text--subheading-20 bl-text--semi-bold bl-text--ellipsis__1').get_text().replace('Rp','').replace('.','').replace('\n','')
#     print(nama)
#     print(harga)
    nama_produk.append(nama)
    harga_produk.append(int(harga))

data = {'nama': nama_produk, 'harga': harga_produk}
df = pd.DataFrame(data, columns = ['nama', 'harga'])

df
```

OUTPUT

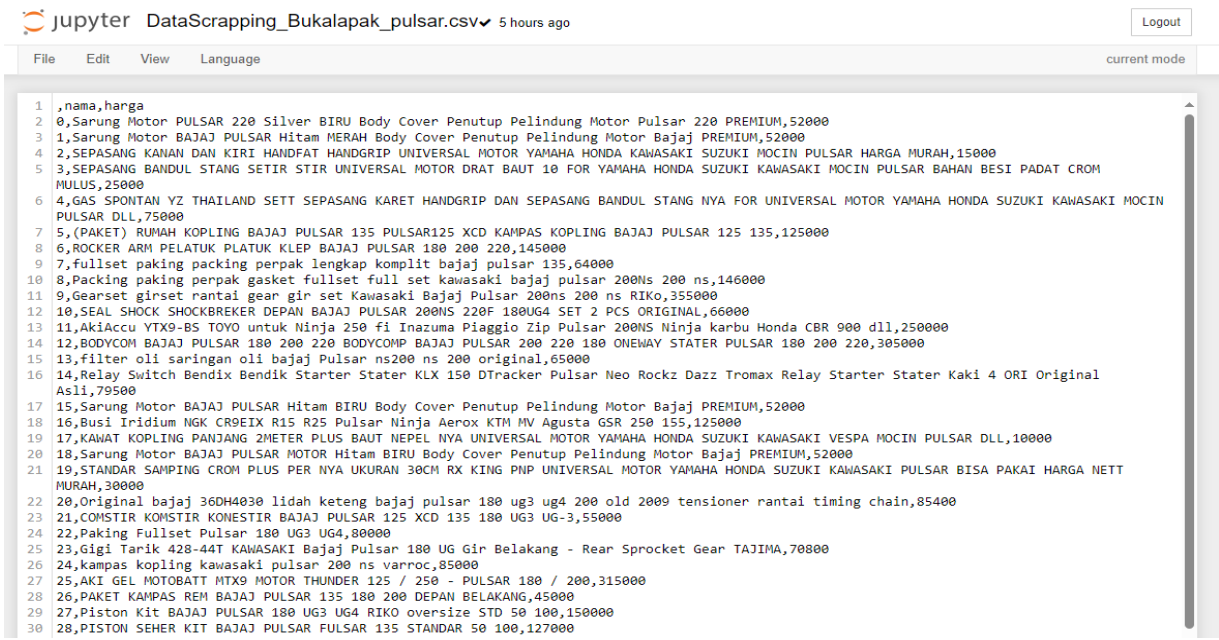
In [42]:

```
nama_produk = []
harga_produk = []
produk = soup.find_all("div", "bl-product-card__description")
# produk
i = 1
for i in produk :
    nama = i.find('a', "bl-link").get_text().replace('\n','').replace('  ','')
    harga = i.find('p', 'bl-text bl-text--subheading-20 bl-text--semi-bold bl-text--ellipsis__1').get_text().replace('Rp','').rep
#     print(nama)
#     print(harga)
    nama_produk.append(nama)
    harga_produk.append(int(harga))

data = {'nama': nama_produk, 'harga': harga_produk}
df = pd.DataFrame(data, columns = ['nama', 'harga'])

df
```

Out[42]:

	nama	harga
0	Sarung Motor PULSAR 220 Silver BIRU Body Cover...	52000
1	Sarung Motor BAJAJ PULSAR Hitam MERAH Body Cov...	52000
2	SEPASANG KANAN DAN KIRI HANDFAT HANDGRIP UNIVE...	15000
3	SEPASANG BANDUL STANG SETIR STIR UNIVERSAL MOT...	25000
4	GAS SPONTAN YZ THAILAND SETT SEPASANG KARET HA...	75000
5	(PAKET) RUMAH KOPLING BAJAJ PULSAR 135 PULSAR1...	125000
6	ROCKER ARM PELATUK PLATUK KLEP BAJAJ PULSAR 18...	145000
7	fullset paking packing perpak lengkap komplit ...	64000
8	Packing paking perpak gasket fullset full set ...	146000
9	Gearset girset rantai gear gir set Kawasaki Ba...	355000
10	SEAL SHOCK SHOCKBREKER DEPAN BAJAJ PULSAR 200N...	66000
11	AkiAccu YTX9-BS TOYO untuk Ninja 250 fi Inazum...	250000
12	BODYCOM BAJAJ PULSAR 180 200 220 BODYCOMP BAJA...	305000
13	filter oli saringan oli bajaj Pulsar ns200 ns ...	65000
14	Relay Switch Bendix Bendik Starter Stater KLX ...	79500

LAPORAN SCRAPING WEB DENGAN PYTHON

ARIF FRIMA ARI SUWADJI - 221011700443 - 03SIFE003

INPUT	OUTPUT
df.to_csv("DataScrapping_Bukalapak_pulsar.csv", sep=',')