

## Funicular Internship Task

Shubham Kejriwal

Github Repository: <https://github.com/shubhamkej/Web-Scraping-Model>

### **Code:**

```
#!/usr/bin/env python
```

```
# coding: utf-8
```

```
# In[8]:
```

```
import os
```

```
import selenium
```

```
from selenium import webdriver
```

```
import time
```

```
import io
```

```
import requests
```

```
import re
```

```
import mysql.connector
```

```
from webdriver_manager.chrome import ChromeDriverManager
```

```
from selenium.common.exceptions import ElementClickInterceptedException
```

```
# In[9]:
```

```
driver = webdriver.Chrome()
```

```
# Calling Web Driver
```

```
# In[10]:
```

```
def store_function(content):
```

```
    splittedcontent = content.splitlines(0)
```

```
    # Splitting into list at next line
```

```
    regex = re.compile(r'^[0-9]+.$')
```

```
    filtered = [i for i in splittedcontent if not regex.match(i)] # Delete elements that don't start  
with numbers
```

```
    while("" in filtered) :
```

```
        filtered.remove("")
```

```
    # Removing Null Elements
```

```

regex = re.compile(r'^.*Growth+.*$')
final = [i for i in filtered if regex.match(i)]          # Removing elements which does
not have growth in it

```

```

mydb = mysql.connector.connect(                          # Database Connectivity
    host="localhost",
    user="root",
    password="",
    database="finance"
)

```

```

for i in final:                                          # Splitting elements of list to store in DB
    stri = i.split(";")
    d = stri[0]
    a = stri[1]
    b = stri[2]
    n = stri[5]
    da = stri[7]

```

```

mycursor = mydb.cursor()

```

```

sql = "INSERT INTO schemevalue (SchemeCode, SchemeName, ISIN_Growth,
NetAssetValue, date) VALUES ( %s, %s, %s, %s, %s)"
val = (d, a, b, n, da)
mycursor.execute(sql, val)                             # Inserting in Database

```

```

mydb.commit()

```

```

print(mycursor.rowcount, "record inserted.")
print(mydb)

```

```

# In[22]:

```

```

list1 = ["10", "11", "12", "13", "14", "15"]           # Making list so don't have to change manually
list2 = ["Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"]
list3 = ["2016", "2017", "2018", "2019", "2020"]

```

```

sk = "http://portal.amfiindia.com/DownloadNAVHistoryReport_Po.aspx?frmdt="
for k in list3:                                         # For loop for year
    for j in list2:                                     # For loop for month
        if(k=="2016" and (j=="Jan" or j=="Feb" or j=="Mar")):
            continue;

```

```

if(k=='2020' and j=='Nov'):
    break;
for i in list1:                                # For loop for Date

driver.get('http://portal.amfiindia.com/DownloadNAVHistoryReport_Po.aspx?frmdt='+i+'-'+j+'-'+k)
    #time.sleep(5)
    content = driver.find_element_by_xpath("//pre").text        # Extracting Data
    print(content)
    store_function(content)                # Sending Data to function

```

## Algorithm:

- Call Main
- Run 3 Nested Loops for Year, Month (Not including Jan, Feb and Mar of 2016 and break after Oct 2020) and Date
- Extract Data for each date, take it as a string and send it to function
- Split data into list for every new line in string
- Delete elements that don't start with numbers
- Remove Null elements
- Remove elements which don't have "Growth" word in it.
- Database Connectivity
- Split elements of list at ";
- Store 1st, 2nd, 3rd, 6th and 8th element in the database.

Output: This is how the table looks like in MySQL. It has 6 columns :  
 Sr No(Primary Key), SchemeCode, SchemeName, INIS\_Growth,  
 NetAssetValue and date.

	Sr No	SchemeCode	SchemeName	INIS_Growth	NetAssetValue	Date
	1	109946	Edelweiss Liquid Fund - Growth Option	INF754K01BC6	1465.5773	10-Apr-2016
	2	109940	Edelweiss Liquid Fund - Retail Plan - Growth Optio	INF754K01AO3	1648.622	10-Apr-2016
	3	118636	Edelweiss Liquid Fund-Direct Plan-Growth Option	INF754K01CX0	1470.5552	10-Apr-2016
	4	120557	Religare Invesco Overnight Fund - Direct Plan - Gr	INF205K01ND2	1756.1856	10-Apr-2016
	5	106501	Religare Invesco Overnight Fund - Growth	INF205K01HR4	1746.141	10-Apr-2016
	6	120415	JM Floater Short Term Fund (Direct) - Growth Optio	INF192K01CG8	23.1748	10-Apr-2016
	7	101421	JM Floater Short Term Fund - Growth Option	INF192K01775	23.082	10-Apr-2016
	8	119905	JPMorgan India Liquid Fund - Direct Plan - Growth	INF843K01AX5	19.6315	10-Apr-2016
	9	106561	JPMorgan India Liquid Fund - Retail Plan - Growth	INF843K01195	17.8031	10-Apr-2016
	10	110150	JPMorgan India Liquid Fund - Super Institutional P	INF843K01146	19.5597	10-Apr-2016
	11	120262	Sahara Liquid Fund-Fixed Pricing - Direct -Growth	INF515L01650	2659.4602	10-Apr-2016
	12	101394	Sahara Liquid Fund-Fixed Pricing - Growth option	INF515L01056	2657.9593	10-Apr-2016
	13	101402	Sahara Liquid Fund-Variable Pricing - Growth optio	INF515L01114	2699.1201	10-Apr-2016
	14	120280	Sahara Liquid Fund-Variable Pricing -Direct - Grow	INF515L01718	2701.7903	10-Apr-2016
	15	101860	Tata Liquid High Investment Plan - Growth	INF277K01LN4	2458.4554	10-Apr-2016
	16	101858	Tata Liquid Retail Investment Plan - Growth	INF277K01LM6	3346.7993	10-Apr-2016
	17	103632	Tata Liquidity Management Fund - Regular Plan - Gr	INF277K01LW5	2010.8592	10-Apr-2016
	18	119444	Tata Liquidity Management Fund- Direct Plan-Growth	INF277K01PU0	2013.8832	10-Apr-2016
	19	102153	Tata Money Market Fund Regular (Growth)	INF277K01LY1	2349.079	10-Apr-2016
	20	101996	HDFC Overnight Fund - Growth Option	INF179KB1HS3	2358.776	10-Apr-2016

Server: 127.0.0.1 » Database: finance » Table: schemevalue

Browse		Structure		SQL	Search	Insert	Export	Import	Privileges	Operations	Tracking	Triggers
← T →		Sr No	SchemeCode	SchemeName	ISIN_Growth	NetAssetValue	Date					
<input type="checkbox"/>	Edit Copy Delete	145	130021	SBI Dual Advantage Fund - Series III - Regular Pla	INF200K012J1	0	10-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	146	139123	SBI Dual Advantage Fund - Series XV- Direct Plan -	INF200KA1IT3	0	10-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	147	139121	SBI Dual Advantage Fund - Series XV- Regular Plan	INF200KA1IR7	0	10-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	148	120704	Aditya Birla Sun Life MIP - Direct Plan - Growth	INF209K01XB7	39.944	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	149	100970	Aditya Birla Sun Life MIP - Regular Plan - Growth	INF209K01694	39.2678	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	150	101713	Aditya Birla Sun Life MIP II - Saving 5 Plan - Gro	INF209K01728	28.1114	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	151	120706	Aditya Birla Sun Life MIP II - Savings 5 Plan - Gr	INF209K01XE1	28.491	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	152	120572	Aditya Birla Sun Life Monthly Income - Growth - Di	INF209K01YH2	56.974	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	153	103185	Aditya Birla Sun Life Monthly Income - Growth - Re	INF209K01AZ4	55.7308	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	154	118302	Canara Robeco InDiGo Fund - Direct Plan - Growth	INF760K01FJ9	14.6423	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	155	113101	Canara Robeco InDiGo Fund - Regular Plan - Growth	INF760K01AJ0	14.3306	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	156	118312	Canara Robeco Short Term Fund- Direct Plan - Growt	INF760K01GH1	17.3879	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	157	111879	Canara Robeco Short Term Fund- Institutional Plan	INF760K01688	18.1709	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	158	111883	Canara Robeco Short Term Fund- Regular Plan - Grow	INF760K01720	17.1631	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	159	103277	DSP BlackRock Income Opportunities Fund - Institut	INF740K01FN3	2249.8628	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	160	105666	DSP BlackRock Strategic Bond Fund - Regular Plan -	INF740K01QU5	1952.4516	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	161	104142	DSP BlackRock Ultra Short Fund - Institutional Pla	INF740K01FS2	2118.6544	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	162	109955	Edelweiss Bond Fund - Growth Option	INF754K01AH7	1801.8543	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	163	118648	Edelweiss Bond Fund -Direct Plan- Growth Option	INF754K01CQ4	1804.6639	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	164	109949	Edelweiss Bond Fund - Retail Plan - Growth Option	INF754K01AA2	1741.7784	11-Apr-2016					
<input type="checkbox"/>	Edit Copy Delete	165	118646	Edelweiss Short Term Income Fund - Direct Plan- Gr		17.322	11-Apr-2016					