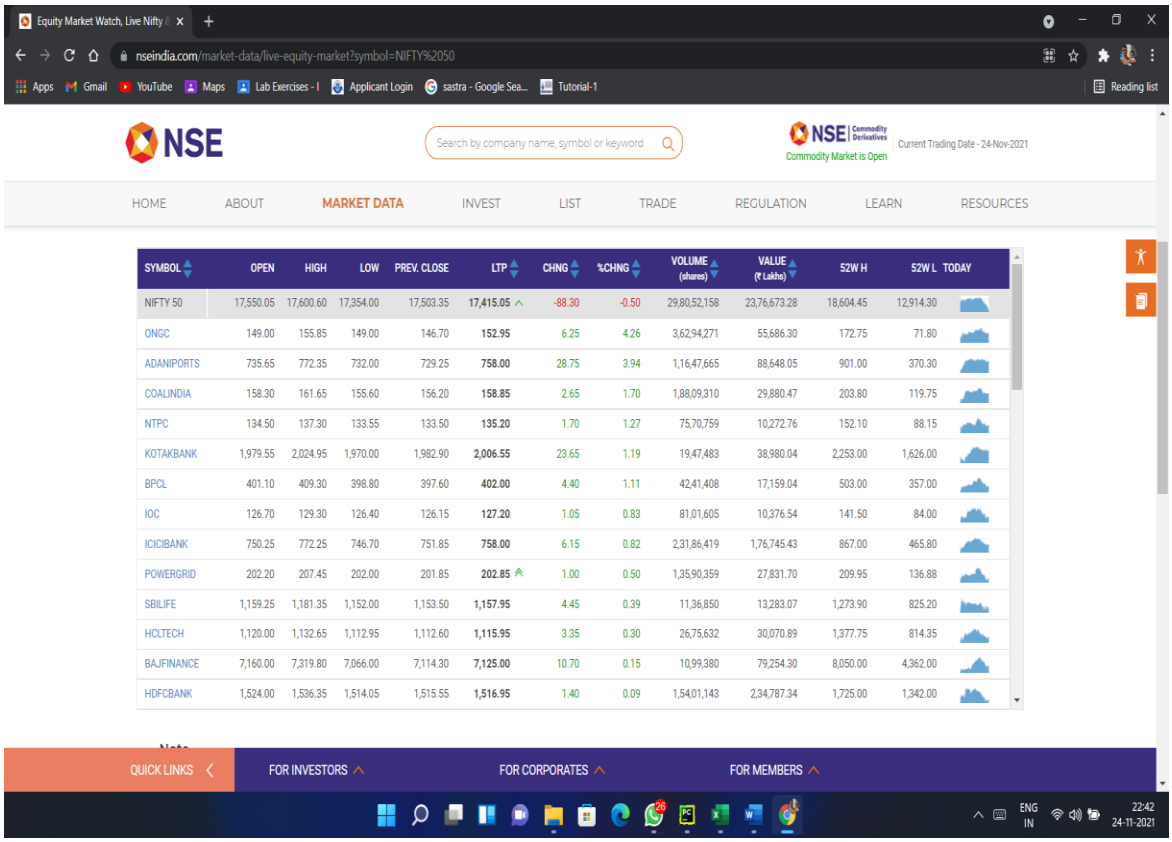
Project Title: Develop a SEO Tool to Analyse Live Web Pages

Introduction

Search Engine Optimization (SEO) is an important aspect of a Web page to gain importance for a search engine to be able to display it earlier in the search list. The optimization is based on a lot of factors such as title, description, header tags and keyword density. Different search engines will have their own mechanisms that calculate the score of a keyword on the page and thus work out its ranking in the search order.

Project Requirements

Python Version : python 3.x
Web Scraping Tool : BeatuifulSoup4
Database : SQLite
IDLE : PyCharm
Live Web Page Link : **https://www.nseindia.com/market-data/live-equity-market?symbol=NIFTY%2050**



PROGRAM

```
import bs4
from matplotlib import pyplot as plt
import requests
import xlsxwriter
```

```
import json
import xlrd
import sqlite3


class webscarp:
    final_data = []
    read_data=[]

    def data_extract(self,soup):
        data=json.loads(str(soup))
        data=data['data']
        #print(data)
        print("*************************************************")
        print("THE NUMBER OF DATA PRESENT IN THE PAGES IS",len(data))
        for dt in data:
            self.meta_data = [
            dt['symbol'],
            dt['open'],
            dt['dayHigh'],
            dt['dayLow'],
            dt['previousClose'],
            dt['lastPrice'],
            dt['change'],
            dt['pChange'],
            dt['totalTradedVolume'],
            dt['totalTradedValue'],
            dt['nearWKH'],
            dt['nearWKL']
            ]
            #print(self.meta_data)
            self.final_data.append(self.meta_data)
        #print(self.final_data)

    def excel_data(self):
        workbook=xlsxwriter.Workbook("NSE_data.xlsx")
        worksheet=workbook.add_worksheet()
        bold=workbook.add_format({'bold':True})
        worksheet.write('A1','SYMBOL',bold)
        worksheet.write('B1', 'OPEN', bold)
        worksheet.write('C1', 'HIGH', bold)
        worksheet.write('D1', 'LOW', bold)
        worksheet.write('E1', 'PREV_CLOSE', bold)
        worksheet.write('F1', 'LAST_PRICE', bold)
        worksheet.write('G1', 'CHANGE', bold)
        worksheet.write('H1','PERC_CHANGE', bold)
        worksheet.write('I1', 'VOLUME(shares)', bold)
        worksheet.write('J1','VALUE(₹lakhs)', bold)
        worksheet.write('K1','52 W_H',bold)
        worksheet.write('L1','52 W_L', bold)
        print("************************************")
        print("EXCEL SHEET CREATED SUCESSFULLY.....")
        row=1
        col=0
        for data in  self.final_data:
            worksheet.write(row,col,data[0])
            worksheet.write(row, col+1, data[1])
            worksheet.write(row, col+2, data[2])
            worksheet.write(row, col+3, data[3])
            worksheet.write(row, col+4, data[4])
            worksheet.write(row, col+5, data[5])
            worksheet.write(row, col+6, data[6])
            worksheet.write(row, col+7, data[7])
            worksheet.write(row, col+8, data[8])
            worksheet.write(row, col+9, data[9])
            worksheet.write(row, col+10, data[10])
            worksheet.write(row, col+11, data[11])
```

```
            row+=1
        print("************************************")
        print("DATA WRITTEN ON EXCEL SUCESSFULLY...")

        # chart1=workbook.add_chart({'type':'line'})
        #
chart1.add_series({'categories':'=Sheet1$B$2:$B$50','values':'Sheet1$A$2:$A$50'})
        # chart1.set_title({'name':'STOCK DATA'})
        # worksheet.insert_chart('T4',chart1)

        chart2 = workbook.add_chart({'type': 'column'})
        chart2.add_series({'categories': '=Sheet1!$A$3:$A$52', 'values':
'=Sheet1!$B$2:$B$52'})
        chart2.add_series({'categories': '=Sheet1!$A$3:$A$52', 'values':
'=Sheet1!$C$2:$C$52'})
        chart2.set_title({'name': 'STOCK DATA'})
        worksheet.insert_chart('T10',chart2)
        workbook.close()
        print("*****************************************************")
        print(" GRAPH OF THE DATA SUCESSFULLY DRAWN ON EXCEL SHEET")

    def read_excel(self):
        wb=xlrd.open_workbook("NSE_data.xlsx")
        worksheet=wb.sheet_by_name("Sheet1")
        num_rows=worksheet.nrows
        num_cols=worksheet.ncols

        for cur_row in range(0,num_rows,1):
            row_review=[]
            for cur_col in range(0,num_cols,1):
                review=worksheet.cell_value(cur_row,cur_col)
                row_review.append(review)
            self.read_data.append(row_review)
    def data_base(self):
        datas=self.read_data
        conn=sqlite3.connect("NSE_data.db")
        print("************************************")
        print("DATA BASE CONNECTED SUCESSFULLY..")
        conn.execute(''' CREATE TABLE NSE_DATA(symbol TEXT NOT NULL,
        open INTEGER NOT NULL,
        high INTEGER NOT NULL,
        low INTEGER NOT NULL,
        prev_close INTEGER NOT NULL,
        last_price INTEGER NOT NULL,
        change INTEGER NOT NULL,
        perc_change INTEGER NOT NULL,
        volume INTEGER NOT NULL,
        value INTEGER NOT NULL,
        w_h_52 INTEGER NOT NULL,
        w_l_52 INTEGER NOT NULL);''')
        cursor=conn.cursor()
        cursor.executemany("INSERT INTO
NSE_DATA(symbol,open,high,low,prev_close,last_price,change,perc_change,volume,value
,w_h_52,w_l_52) VALUES (?,?,?,?,?,?,?,?,?,?,?,?)",datas)
        conn.commit()
        print("*****************************")
        print("DATA STORED SUCESSFULLY..")
        conn.close()
    def graph(self):
        plt.bar([dt[0] for dt in self.final_data],[dt[1] for dt in self.final_data
], color='r')
        plt.legend(["STOCK HIGH", "STOCK NAME"])
        plt.show()
        print("***********************")
        print("BAR DIAGRAM DRAWN")

        plt.scatter([dt[0] for dt in self.final_data],[dt[1] for dt in
self.final_data ], label='cases', color='r')
```

```
        plt.show()
        print("*********************")
        print("SCATTER PLOT DRAWN")
try:
    print("********************************")
    print("NSE- NATIONAL STOCK EXCHANGE")
    print("********************************")
    urllink='https://www.nseindia.com/api/equity-stockIndices?index=NIFTY%2050'
    header={'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/95.0.4638.69 Safari/537.36'
    }
    response=requests.get(url=urllink,headers=header)
    soup=bs4.BeautifulSoup(response.content,'html.parser')

    #print(soup)
    w = webscarp()
    w.data_extract(soup)
    print("***************************************")
    print("WEB PAGE EXTRACTED SUCESSFULLY......")
    w.excel_data()
    w.read_excel()
    w.data_base()
    w.graph()
    print("***********************************************")
    print("DATA SCRAPPED AND COPIED TO EXCEL AND SUCESSFULLY STORED IN DATA BASE \n
GRAPHS AND BAR DIAGRAM DRAWN SUCESSFULLY...\n THANK YOU")
    print("***********************************************")
except Exception as e:
    print("Exception occurs",e,"PLEASE CHECK AND RUN AGAIN....")
```

OUTPUT:

```
C:\Users\91824\AppData\Local\Programs\Python\Python37-32\python.exe C:/Users/91824/Desktop/websracp/STOCK.py
********************************
NSE- NATIONAL STOCK EXCHANGE
********************************
**************************************************
THE NUMBER OF DATA PRESENT IN THE PAGES IS 51
***************************************
WEB PAGE EXTRACTED SUCESSFULLY......
***************************************
EXCEL SHEET CREATED SUCESSFULLY......
***********************************
DATA WRITTEN ON EXCEL SUCESSFULLY...
**************************************************
 GRAPH OF THE DATA SUCESSFULLY DRAWN ON EXCEL SHEET
***********************************
DATA BASE CONNECTED SUCESSFULLY..
*****************************
DATA STORED SUCESSFULLY..
***********************
BAR DIAGRAM DRAWN
*********************
SCATTER PLOT DRAWN
*********************************************
DATA SCRAPPED AND COPIED TO EXCEL AND SUCESSFULLY STORED IN DATA BASE
 GRAPHS AND BAR DIAGRAM DRAWN SUCESSFULLY...
 THANK YOU
*********************************************

Process finished with exit code 0
```

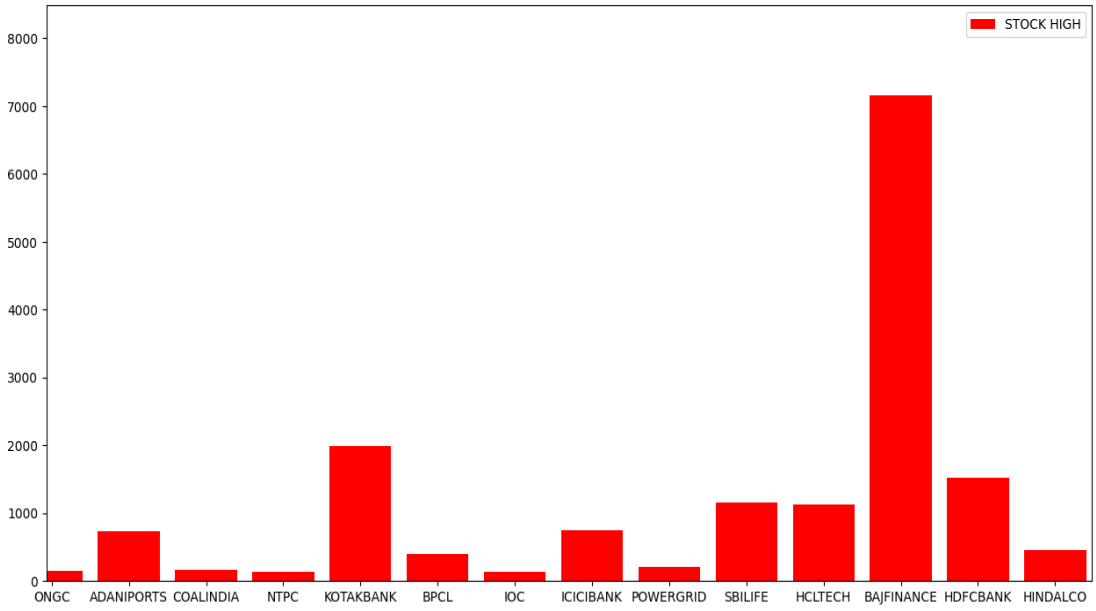
EXCEL SHEET OF ALL DATAS:

Clipboard | Font | Alignment

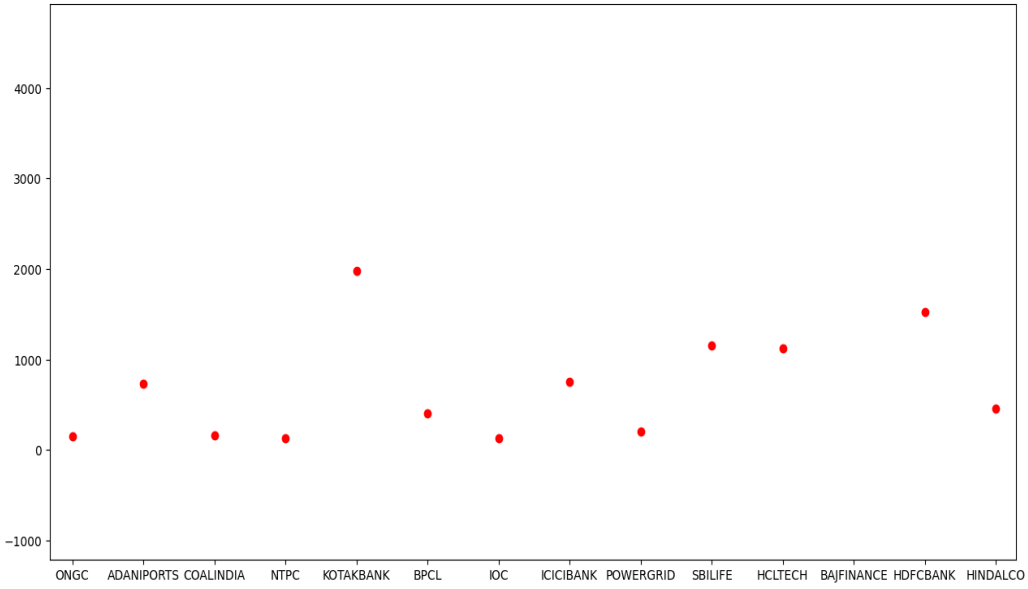
R24

	A	B	C	D	E	F	G	H	I	J	K	L
1	SYMBOL	OPEN	HIGH	LOW	PREV_CLC	LAST_PRI	CHANGE	PERC_CHA	VOLUME(	VALUE(₹	52 W_H	52 W_L
2	NIFTY 50	17550.1	17600.6	17354	17503.4	17415.1	-88.3	-0.5	3E+08	2.4E+11	6.39309	-34.8509
3	ONGC	149	155.85	149	146.7	152.95	6.25	4.26	3.6E+07	5.6E+09	11.4616	-113.022
4	ADANIPOF	735.65	772.35	732	729.25	758	28.75	3.94	1.2E+07	8.9E+09	15.8713	-104.699
5	COALINDI	158.3	161.65	155.6	156.2	158.85	2.65	1.7	1.9E+07	3E+09	22.0559	-32.6514
6	NTPC	134.5	137.3	133.55	133.5	135.2	1.7	1.27	7570759	1E+09	11.1111	-53.3749
7	KOTAKBAI	1979.55	2024.95	1970	1982.9	2006.55	23.65	1.19	1947483	3.9E+09	10.9387	-23.4041
8	BPCL	401.1	409.3	398.8	397.6	402	4.4	1.11	4241408	1.7E+09	20.0795	-12.605
9	IOC	126.7	129.3	126.4	126.15	127.2	1.05	0.83	8101605	1E+09	10.106	-51.4286
10	ICICIBANK	750.25	772.25	746.7	751.85	758	6.15	0.82	2.3E+07	1.8E+10	12.5721	-62.7308
11	POWERGF	202.2	207.45	202	201.85	202.85	1	0.5	1.4E+07	2.8E+09	3.38176	-48.1955
12	SBILIFE	1159.25	1181.35	1152	1153.5	1157.95	4.45	0.39	1136850	1.3E+09	9.10197	-40.3236
13	HCLTECH	1120	1132.65	1112.95	1112.6	1115.95	3.35	0.3	2675632	3E+09	19.002	-37.0357
14	BAJFINAN	7160	7319.8	7066	7114.3	7125	10.7	0.15	1099380	7.9E+09	11.4907	-63.3425
15	HDFCBAN	1524	1536.35	1514.05	1515.55	1516.95	1.4	0.09	1.5E+07	2.3E+10	12.0609	-13.0365
16	HINDALCC	453.9	456.9	444.2	450.2	450.2	0	0	7761523	3.5E+09	18.4199	-108.861
17	BHARTIAR	764	781.8	756.25	758.9	756.9	-2	-0.26	2.2E+07	1.7E+10	3.18496	-66.6777
18	BAJAJFINS	17432	17647	17140.4	17368.1	17320	-48.1	-0.28	274024	4.8E+09	10.3752	-109.338
19	SHREECEM	26895	27000	26665	26740.5	26665	-75.5	-0.28	29212	7.9E+08	16.7967	-18.3481
20	AXISBANK	690	695.1	682.05	685.65	683.5	-2.15	-0.31	1.1E+07	7.9E+09	21.1558	-20.2498
21	JSWSTEEL	683.3	696	670	683.3	680.95	-2.35	-0.34	4374307	3E+09	12.3052	-103.816
22	SBIN	495.8	500	489.05	493.05	490	-3.05	-0.62	1.4E+07	6.9E+09	9.64411	-105.796
23	HEROMO	2646.5	2657.9	2613	2632.2	2614.9	-17.3	-0.66	564230	1.5E+09	27.9453	-0.57308
24	NESTLEINI	19200	19290	18950	19181.6	19050	-131.55	-0.69	34355	6.6E+08	7.56533	-19.0469
25	DRREDDY	4628.9	4653.6	4586.6	4631.95	4600	-31.95	-0.69	199258	9.2E+08	18.0707	-11.2455
26	TITAN	2400	2405.25	2365	2387.7	2371.05	-16.65	-0.7	614145	1.5E+09	11.4586	-82.3394
27	UPL	726	736.1	716.15	725.25	719.95	-5.3	-0.73	1312705	9.6E+08	16.7399	-73.838
28	TCS	3472.2	3493.05	3424	3464.25	3431.95	-32.3	-0.93	2161746	7.5E+09	13.9841	-30.7684

BAR GRAPH OF STOCK HIGH IN A DAY:



SCATTER PLOT OF HIGH SOLD STOCKS IN A DAY:



EXCEL GRAPH