ITL LAB: EXPERIMENT 6

NAME: Shreya Shetty

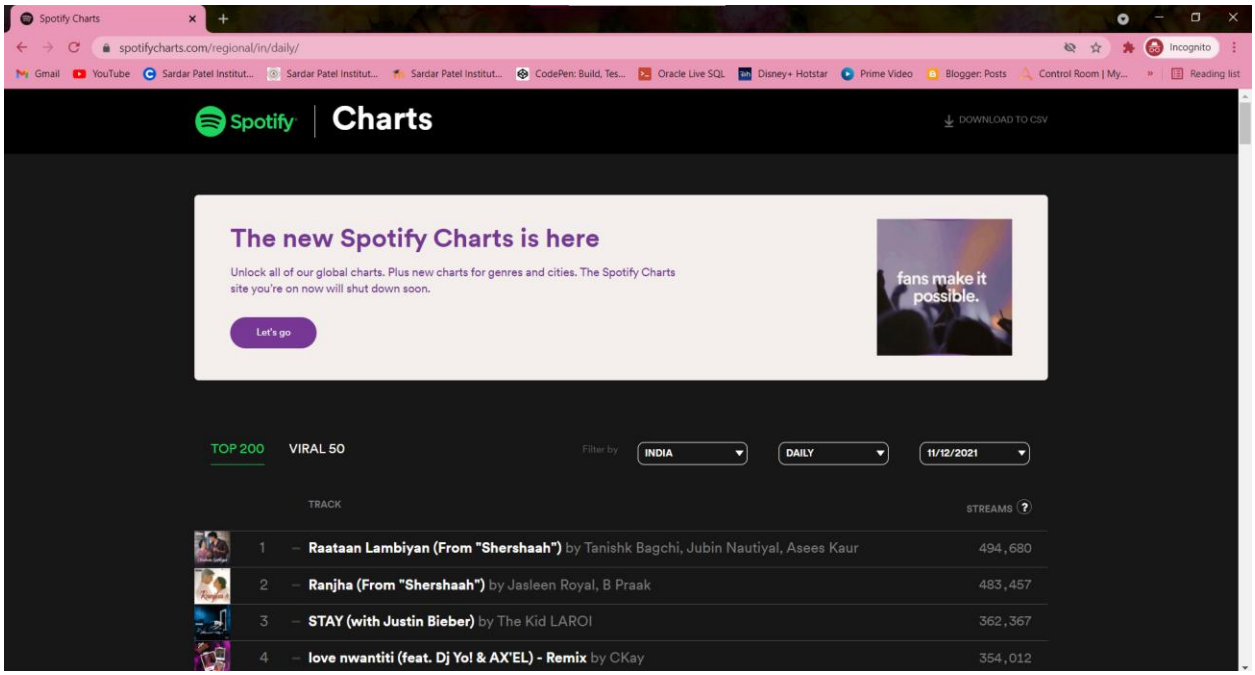
UID: 2019140059

CLASS: TE IT

BATCH: D

AIM: Write a python code to demonstrate the behaviour of Web Crawlers/ spiders (use XPATH, CSSPATH), extract information and store it in database.

WEBSITE: https://spotifycharts.com/regional/in/daily/



URLs Crawled:

https://spotifycharts.com/regional/in/daily/2021-10-1

https://spotifycharts.com/regional/in/daily/2021-10-2

.
.
https://spotifycharts.com/regional/in/daily/2021-11-1

CODE:

```
from bs4 import BeautifulSoup
import pandas as pd
import requests
from time import sleep
from datetime import date, timedelta
from lxml import etree

dir="D:\PROJECT_AND_CODES\ITL-Web-Scraper"
#create empty arrays for data we're collecting
dates=[]
url_list=[]
final = []

#map site
headers = {
    "User-Agent":
    "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/72.0.3538.102 Safari/537.36
Edge/18.19582"
}
url = "https://spotifycharts.com/regional/in/daily/"
start_date= date(2020, 11, 1)
end_date= date(2021, 11, 1)

delta= end_date-start_date
# Creating list of dates
for i in range(delta.days+1):
    day = start_date+timedelta(days=i)
    day_string= day.strftime("%Y-%m-%d")
    dates.append(day_string)

# Adding Urls to Scrape to url_list
def add_url():
    for date in dates:
        c_string = url+date
        url_list.append(c_string)

add_url()

# count stores the total number of songs scraped
count=0
# Function for going through each row in each url and finding
relevant song info
```

```
def song_scrape(x):
    global count
    song_count_date=0
    pg = x
    for tr in songs.find("tbody").findAll("tr"):
        song_count_date+=1
        # Scraping data of top 20 songs on that particular date
'x' from url 'u'
        if song_count_date>20:
            break

        # Usig css selectors to scrape data with BeautifulSou[]
        artist= tr.find("td", class_="chart-table-
track").find("span").text
        artist= artist.replace("by ","").strip()

        title= tr.find("td", class_="chart-table-
track").find("strong").text

        songId= tr.find("td", {"class": "chart-table-
image"}).find("a").get("href")
        songId= songId.split("track/")[1]

        # Using XPath to scrape data with BeautifulSoup

xpath_id='//*[@id="content"]/div/div/div/span/table/tbody/tr['+st
r(song_count_date)+']/td[1]/a/img/@src'
        songImg=dom.xpath(xpath_id)[0]


xpath_id='//*[@id="content"]/div/div/div/span/table/tbody/tr['+st
r(song_count_date)+']/td[5]/text()'
        songCount=dom.xpath(xpath_id)[0]

        url_date= x.split("daily/")[1]

        count+=1
        final.append([count,url_date,title, artist, songId,
songImg,songCount])
        print("Top Song ",song_count_date," Data added")


# Loop through urls to create an array of all of our song info
for u in url_list:
```

```
    print('URL : ',u)

    read_pg= requests.get(u, headers=headers)
    # sleep(2)

    soup= BeautifulSoup(read_pg.text, "html.parser")
    songs= soup.find(class_="chart-table")
    dom = etree.HTML(str(soup))
    song_scrape(u)

# Converting final data to data frame with pandas for easier data
manipulation
final_df = pd.DataFrame(final, columns= ["Sr. No.", "Chart Date",
"Song Title", "Artist", "Song ID", "Song Image","Streams"])
print('Total count of songs is ',count)
print('Final Dataframe is created using pandas')

# Write data to csv
with open(dir+'\song_scrape_data.csv', 'w') as f:
    final_df.to_csv(f, header= True, index=False,
line_terminator='\n')
    print('CSV file created successfully')
```

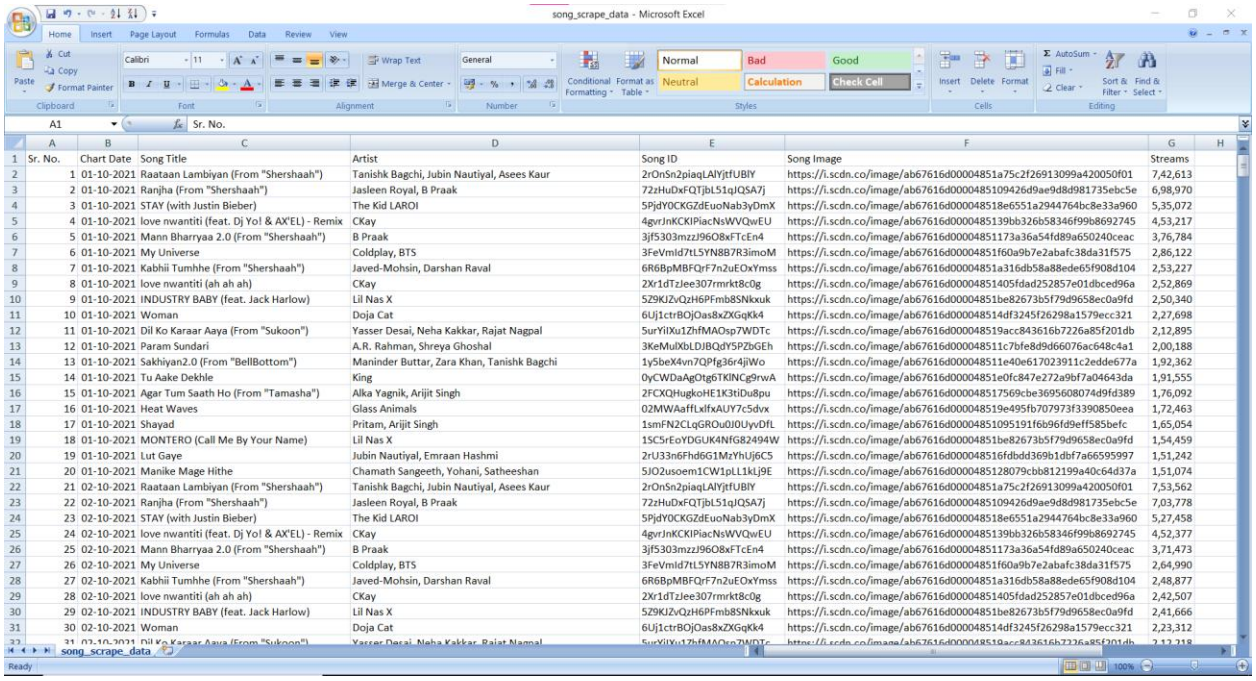
OUTPUT :

```
PS D:\PROJECT_AND_CODES> & C:/Users/AIM/AppData/Local/Programs/Python/Python38/python.exe d:/PROJECT_AND_CODES/ITL-Web-Scraper/songscrape.py
URL :  https://spotifycharts.com/regional/in/daily/2021-10-01
Top Song  1  Data added
Top Song  2  Data added
Top Song  3  Data added
Top Song  4  Data added
Top Song  5  Data added
Top Song  6  Data added
Top Song  7  Data added
Top Song  8  Data added
Top Song  9  Data added
Top Song  10  Data added
Top Song  11  Data added
Top Song  12  Data added
Top Song  13  Data added
Top Song  14  Data added
Top Song  15  Data added
Top Song  16  Data added
Top Song  17  Data added
Top Song  18  Data added
Top Song  19  Data added
Top Song  20  Data added
URL :  https://spotifycharts.com/regional/in/daily/2021-10-02
Top Song  1  Data added
Top Song  2  Data added
Top Song  3  Data added
Top Song  4  Data added
Top Song  5  Data added
Top Song  6  Data added
Top Song  7  Data added
Top Song  8  Data added
Top Song  9  Data added
Top Song  10  Data added
Top Song  11  Data added
Top Song  12  Data added
Top Song  13  Data added
Top Song  14  Data added
Top Song  15  Data added
Top Song  16  Data added
Top Song  17  Data added
Top Song  18  Data added
```

.

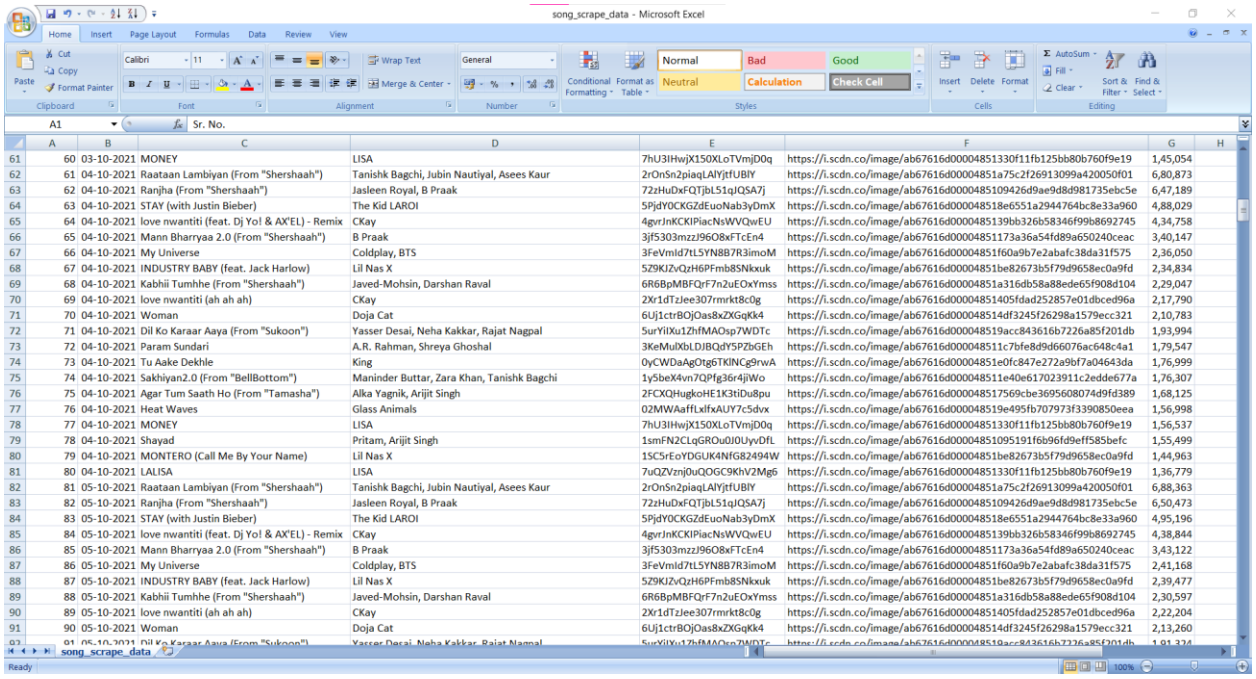
```
Top Song  5  Data added
Top Song  6  Data added
Top Song  7  Data added
Top Song  8  Data added
Top Song  9  Data added
Top Song  10  Data added
Top Song  11  Data added
Top Song  12  Data added
Top Song  13  Data added
Top Song  14  Data added
Top Song  15  Data added
Top Song  16  Data added
Top Song  17  Data added
Top Song  18  Data added
Top Song  19  Data added
Top Song  20  Data added
URL :  https://spotifycharts.com/regional/in/daily/2021-11-01
Top Song  1  Data added
Top Song  2  Data added
Top Song  3  Data added
Top Song  4  Data added
Top Song  5  Data added
Top Song  6  Data added
Top Song  7  Data added
Top Song  8  Data added
Top Song  9  Data added
Top Song  10  Data added
Top Song  11  Data added
Top Song  12  Data added
Top Song  13  Data added
Top Song  14  Data added
Top Song  15  Data added
Top Song  16  Data added
Top Song  17  Data added
Top Song  18  Data added
Top Song  19  Data added
Top Song  20  Data added
Total count of songs is  640
Final Dataframe is created using pandas
CSV file created successfully
PS D:\PROJECT AND CODES>
```

CSV FILE DATABASE :



song_scrape_data - Microsoft Excel

Sr. No.	Chart Date	Song Title	Artist	Song ID	Song Image	Streams
1	01-10-2021	Raataan Lambiyan (From "Shershaah")	Tanishk Bagchi, Jubin Nautiyal, Asees Kaur	2rOnSn2piaqLAlYjtfUBlY	https://i.scdn.co/image/ab67616d00004851a75c2f26913099a420050f01	7,42,613
2	01-10-2021	Ranjha (From "Shershaah")	Jasleen Royal, B Praak	72zHuDxFQTjbL51qJQSA7j	https://i.scdn.co/image/ab67616d00004851094260d9ae9d8d981735ebc5e	6,98,970
3	01-10-2021	STAY (with Justin Bieber)	The Kid LAROI	5PjdY0CKGZdEuoNab3yDmX	https://i.scdn.co/image/ab67616d00004851 8e6551a2944764bc8e33a960	5,35,072
4	01-10-2021	love nwantiti (feat. Dj Yo! & AX'EL) - Remix	CKay	4gvrJnKCKIPiacNsWVQwEU	https://i.scdn.co/image/ab67616d00004851 39bb326b58346f99b8692745	4,53,217
5	01-10-2021	Mann Bharryaa 2.0 (From "Shershaah")	B Praak	3jf5303mzzJ96O8xFTcEn4	https://i.scdn.co/image/ab67616d00004851 173a36a54fd89a650240ceac	3,76,784
6	01-10-2021	My Universe	Coldplay, BTS	3FeVmId7tL5YN8B7R3imoM	https://i.scdn.co/image/ab67616d00004851 f60a9b7e2abafc38da31f575	2,86,122
7	01-10-2021	Kabhii Tumhhe (From "Shershaah")	Javed-Mohsin, Darshan Raval	6R6BpMBFQrF7n2uEOxYmss	https://i.scdn.co/image/ab67616d00004851 a316db58a88ede65f908d104	2,53,227
8	01-10-2021	love nwantiti (ah ah ah)	CKay	2Xr1dTzJee307rmrkt8c0g	https://i.scdn.co/image/ab67616d00004851 405fdad252857e01dbced96a	2,52,869
9	01-10-2021	INDUSTRY BABY (feat. Jack Harlow)	Lil Nas X	5Z9KJZvQzH6PFmb8SNkxuk	https://i.scdn.co/image/ab67616d00004851 be82673b5f79d9658ec0a9fd	2,50,340
10	01-10-2021	Woman	Doja Cat	6Uj1ctrBOjOas8xZXGqKk4	https://i.scdn.co/image/ab67616d00004851 4df3245f26298a1579ecc321	2,27,698
11	01-10-2021	Dil Ko Karaar Aaya (From "Sukoon")	Yasser Desai, Neha Kakkar, Rajat Nagpal	5urYiIXu1ZhfMAOsp7WDTc	https://i.scdn.co/image/ab67616d00004851 9acc843616b7226a85f201db	2,12,895
12	01-10-2021	Param Sundari	A.R. Rahman, Shreya Ghoshal	3KeMulXbLDJBQdY5PZbGEh	https://i.scdn.co/image/ab67616d00004851 1c7bfe8d9d66076ac648c4a1	2,00,188
13	01-10-2021	Sakhiyan2.0 (From "BellBottom")	Maninder Buttar, Zara Khan, Tanishk Bagchi	1y5beX4vn7QPfg36r4jiWo	https://i.scdn.co/image/ab67616d00004851 1e40e617023911c2edde677a	1,92,362
14	01-10-2021	Tu Aake Dekhle	King	0yCWDaAgOtg6TKlNCg9rwA	https://i.scdn.co/image/ab67616d00004851 e0fc847e272a9bf7a04643da	1,91,555
15	01-10-2021	Agar Tum Saath Ho (From "Tamasha")	Alka Yagnik, Arijit Singh	2FCXQHugkoHE1K3tiDu8pu	https://i.scdn.co/image/ab67616d00004851 7569cbe3695608074d9fd389	1,76,092
16	01-10-2021	Heat Waves	Glass Animals	02MWAaffLxlfxAUY7c5dvx	https://i.scdn.co/image/ab67616d00004851 9e495fb707973f3390850eea	1,72,463
17	01-10-2021	Shayad	Pritam, Arijit Singh	1smFN2CLqGROu0J0UyvDfL	https://i.scdn.co/image/ab67616d00004851 095191f6b96fd9eff585befc	1,65,054
18	01-10-2021	MONTERO (Call Me By Your Name)	Lil Nas X	1SC5rEoYDGUK4NfG82494W	https://i.scdn.co/image/ab67616d00004851 be82673b5f79d9658ec0a9fd	1,54,459
19	01-10-2021	Lut Gaye	Jubin Nautiyal, Emraan Hashmi	2rU33n6Fhd6G1MzYhUj6C5	https://i.scdn.co/image/ab67616d00004851 6fdbdd369b1dbf7a66595997	1,51,242
20	01-10-2021	Manike Mage Hithe	Chamath Sangeeth, Yohani, Satheeshan	5JO2usoem1CW1pLL1kLj9E	https://i.scdn.co/image/ab67616d00004851 28079cbb812199a40c64d37a	1,51,074
21	02-10-2021	Raataan Lambiyan (From "Shershaah")	Tanishk Bagchi, Jubin Nautiyal, Asees Kaur	2rOnSn2piaqLAlYjtfUBlY	https://i.scdn.co/image/ab67616d00004851 a75c2f26913099a420050f01	7,53,562
22	02-10-2021	Ranjha (From "Shershaah")	Jasleen Royal, B Praak	72zHuDxFQTjbL51qJQSA7j	https://i.scdn.co/image/ab67616d00004851 094260d9ae9d8d981735ebc5e	7,03,778
23	02-10-2021	STAY (with Justin Bieber)	The Kid LAROI	5PjdY0CKGZdEuoNab3yDmX	https://i.scdn.co/image/ab67616d00004851 8e6551a2944764bc8e33a960	5,27,458
24	02-10-2021	love nwantiti (feat. Dj Yo! & AX'EL) - Remix	CKay	4gvrJnKCKIPiacNsWVQwEU	https://i.scdn.co/image/ab67616d00004851 39bb326b58346f99b8692745	4,52,377
25	02-10-2021	Mann Bharryaa 2.0 (From "Shershaah")	B Praak	3jf5303mzzJ96O8xFTcEn4	https://i.scdn.co/image/ab67616d00004851 173a36a54fd89a650240ceac	3,71,473
26	02-10-2021	My Universe	Coldplay, BTS	3FeVmId7tL5YN8B7R3imoM	https://i.scdn.co/image/ab67616d00004851 f60a9b7e2abafc38da31f575	2,64,990
27	02-10-2021	Kabhii Tumhhe (From "Shershaah")	Javed-Mohsin, Darshan Raval	6R6BpMBFQrF7n2uEOxYmss	https://i.scdn.co/image/ab67616d00004851 a316db58a88ede65f908d104	2,48,877
28	02-10-2021	love nwantiti (ah ah ah)	CKay	2Xr1dTzJee307rmrkt8c0g	https://i.scdn.co/image/ab67616d00004851 405fdad252857e01dbced96a	2,42,507
29	02-10-2021	INDUSTRY BABY (feat. Jack Harlow)	Lil Nas X	5Z9KJZvQzH6PFmb8SNkxuk	https://i.scdn.co/image/ab67616d00004851 be82673b5f79d9658ec0a9fd	2,41,666
30	02-10-2021	Woman	Doja Cat	6Uj1ctrBOjOas8xZXGqKk4	https://i.scdn.co/image/ab67616d00004851 4df3245f26298a1579ecc321	2,23,312
31	02-10-2021	Dil Ko Karaar Aaya (From "Sukoon")	Yasser Desai, Neha Kakkar, Rajat Nagpal	5urYiIXu1ZhfMAOsp7WDTc	https://i.scdn.co/image/ab67616d00004851 9acc843616b7226a85f201db	2,12,318



	A	B	C	D	E	F	G	H
61	60	03-10-2021	MONEY	LISA	7hU3IHwjX150XLoTVmjD0q	https://i.scdn.co/image/ab67616d00004851330f11fb125bb80b760f9e19	1,45,054	
62	61	04-10-2021	Raataan Lambiyan (From "Shershaah")	Tanishk Bagchi, Jubin Nautiyal, Asees Kaur	2rOnSn2piaqLAlYjtfUBlY	https://i.scdn.co/image/ab67616d00004851a75c2f26913099a420050f01	6,80,873	
63	62	04-10-2021	Ranjha (From "Shershaah")	Jasleen Royal, B Praak	72zHuDxFQTjbL51qJQSA7j	https://i.scdn.co/image/ab67616d00004851409426d9ae9d8d981735ebc5e	6,47,189	
64	63	04-10-2021	STAY (with Justin Bieber)	The Kid LAROI	5PjdY0CKGZdEuoNab3yDmX	https://i.scdn.co/image/ab67616d00004851 8e6551a2944764bc8e33a960	4,88,029	
65	64	04-10-2021	love nwantiti (feat. Dj Yo! & AX'EL) - Remix	CKay	4gvrJnKCKIPiacNsWVQwEU	https://i.scdn.co/image/ab67616d00004851 39bb326b58346f99b8692745	4,34,758	
66	65	04-10-2021	Mann Bharryaa 2.0 (From "Shershaah")	B Praak	3jf5303mzzJ96O8xFTcEn4	https://i.scdn.co/image/ab67616d00004851 173a36a54fd89a650240ceac	3,40,147	
67	66	04-10-2021	My Universe	Coldplay, BTS	3FeVmId7tL5YN8B7R3imoM	https://i.scdn.co/image/ab67616d00004851 f60a9b7e2abafc38da31f575	2,36,050	
68	67	04-10-2021	INDUSTRY BABY (feat. Jack Harlow)	Lil Nas X	5Z9KJZvQzH6PFmb8SNkxuk	https://i.scdn.co/image/ab67616d00004851 be82673b5f79d9658ec0a9fd	2,34,834	
69	68	04-10-2021	Kabhii Tumhhe (From "Shershaah")	Javed-Mohsin, Darshan Raval	6R6BpMBFQrF7n2uEOxYmss	https://i.scdn.co/image/ab67616d00004851 a316db58a88ede65f908d104	2,29,047	
70	69	04-10-2021	love nwantiti (ah ah ah)	CKay	2Xr1dTzJee307rmrkt8c0g	https://i.scdn.co/image/ab67616d00004851 405fdad252857e01dbced96a	2,17,790	
71	70	04-10-2021	Woman	Doja Cat	6Uj1ctrBOjOas8xZXGqKk4	https://i.scdn.co/image/ab67616d00004851 4df3245f26298a1579ecc321	2,10,783	
72	71	04-10-2021	Dil Ko Karaar Aaya (From "Sukoon")	Yasser Desai, Neha Kakkar, Rajat Nagpal	5urYilXu1ZhfMAOsp7WDTc	https://i.scdn.co/image/ab67616d00004851 9acc843616b7226a85f201db	1,93,994	
73	72	04-10-2021	Param Sundari	A.R. Rahman, Shreya Ghoshal	3KeMulXbLDJBQdY5PZbGEh	https://i.scdn.co/image/ab67616d00004851 1c7bfe8d9d66076ac648c4a1	1,79,547	
74	73	04-10-2021	Tu Aake Dekhle	King	0yCWDaAgOtg6TKlNCg9rwA	https://i.scdn.co/image/ab67616d00004851 e0fc847e272a9bf7a04643da	1,76,999	
75	74	04-10-2021	Sakhiyan2.0 (From "BellBottom")	Maninder Buttar, Zara Khan, Tanishk Bagchi	1y5beX4vn7QPfg36r4jiWo	https://i.scdn.co/image/ab67616d00004851 1e40e617023911c2edde677a	1,76,307	
76	75	04-10-2021	Agar Tum Saath Ho (From "Tamasha")	Alka Yagnik, Arijit Singh	2FCXQHugkoHE1K3tiDu8pu	https://i.scdn.co/image/ab67616d00004851 7569cbe3695608074d9fd389	1,68,125	
77	76	04-10-2021	Heat Waves	Glass Animals	02MWAaffLxlfxAUY7c5dvx	https://i.scdn.co/image/ab67616d00004851 9e495fb707973f3390850eea	1,56,998	
78	77	04-10-2021	MONEY	LISA	7hU3IHwjX150XLoTVmjD0q	https://i.scdn.co/image/ab67616d00004851 330f11fb125bb80b760f9e19	1,56,537	
79	78	04-10-2021	Shayad	Pritam, Arijit Singh	1smFN2CLqGROu0J0UyvDfL	https://i.scdn.co/image/ab67616d00004851 095191f6b96fd9eff585befc	1,55,499	
80	79	04-10-2021	MONTERO (Call Me By Your Name)	Lil Nas X	1SC5rEoYDGUK4NfG82494W	https://i.scdn.co/image/ab67616d00004851 be82673b5f79d9658ec0a9fd	1,44,963	
81	80	04-10-2021	LALISA	LISA	7uQZVznj0uQOGC9KhV2Mg6	https://i.scdn.co/image/ab67616d00004851 330f11fb125bb80b760f9e19	1,36,779	
82	81	05-10-2021	Raataan Lambiyan (From "Shershaah")	Tanishk Bagchi, Jubin Nautiyal, Asees Kaur	2rOnSn2piaqLAlYjtfUBlY	https://i.scdn.co/image/ab67616d00004851 a75c2f26913099a420050f01	6,88,363	
83	82	05-10-2021	Ranjha (From "Shershaah")	Jasleen Royal, B Praak	72zHuDxFQTjbL51qJQSA7j	https://i.scdn.co/image/ab67616d00004851 09426d9ae9d8d981735ebc5e	6,50,473	
84	83	05-10-2021	STAY (with Justin Bieber)	The Kid LAROI	5PjdY0CKGZdEuoNab3yDmX	https://i.scdn.co/image/ab67616d00004851 8e6551a2944764bc8e33a960	4,95,196	
85	84	05-10-2021	love nwantiti (feat. Dj Yo! & AX'EL) - Remix	CKay	4gvrJnKCKIPiacNsWVQwEU	https://i.scdn.co/image/ab67616d00004851 39bb326b58346f99b8692745	4,38,844	
86	85	05-10-2021	Mann Bharryaa 2.0 (From "Shershaah")	B Praak	3jf5303mzzJ96O8xFTcEn4	https://i.scdn.co/image/ab67616d00004851 173a36a54fd89a650240ceac	3,43,122	
87	86	05-10-2021	My Universe	Coldplay, BTS	3FeVmId7tL5YN8B7R3imoM	https://i.scdn.co/image/ab67616d00004851 f60a9b7e2abafc38da31f575	2,41,168	
88	87	05-10-2021	INDUSTRY BABY (feat. Jack Harlow)	Lil Nas X	5Z9KJZvQzH6PFmb8SNkxuk	https://i.scdn.co/image/ab67616d00004851 be82673b5f79d9658ec0a9fd	2,39,477	
89	88	05-10-2021	Kabhii Tumhhe (From "Shershaah")	Javed-Mohsin, Darshan Raval	6R6BpMBFQrF7n2uEOxYmss	https://i.scdn.co/image/ab67616d00004851 a316db58a88ede65f908d104	2,30,597	
90	89	05-10-2021	love nwantiti (ah ah ah)	CKay	2Xr1dTzJee307rmrkt8c0g	https://i.scdn.co/image/ab67616d00004851 405fdad252857e01dbced96a	2,22,204	
91	90	05-10-2021	Woman	Doja Cat	6Uj1ctrBOjOas8xZXGqKk4	https://i.scdn.co/image/ab67616d00004851 4df3245f26298a1579ecc321	2,13,260	
92	91	05-10-2021	Dil Ko Karaar Aaya (From "Sukoon")	Yasser Desai, Neha Kakkar, Rajat Nagpal	5urYilXu1ZhfMAOsp7WDTc	https://i.scdn.co/image/ab67616d00004851 9acc843616b7226a85f201db	1,81,324	

CONCLUSION:

Hence, I was successfully able to web crawl multiple URLS (https://spotifycharts.com/regional/in/daily/) related to my case study using both CSS and XPath with BeautifulSoup in Python to extract and store data in CSV file of the Top 20 songs in Spotify Charts in last 1 month.