Name :- Rishabh Sharma

Registration Number :- 20MAI0082

Assignment Number :- 6

link for Gujarat Wikipedia is :- https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Gujarat (https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Gujarat)

link for github :- https://github.com/rishabh5197/Data-Mining/tree/main/Assignment-6 (https://github.com/rishabh5197/Data-Mining/tree/main/Assignment-6)

In [1]:
```
import numpy as np
import pandas as pd
import sklearn
import matplotlib.pyplot as plt
import seaborn as sns
from bs4 import BeautifulSoup as bs
import requests
```

In [2]:
```
link = "https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Gujarat"
print(link)
page = requests.get(link)
print(page)
```

```
https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Gujarat (https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Gujarat)
<Response [200]>
```

In [3]:
```
soup = bs(page.content,"html.parser")
```

In [4]:
```
# soup
```

There are 2 tables in Gujarat wikipedia page which are Covid-19 Cases in gujarat and Covid-

19 pandemic in Gujarat by district which are scapped as below

Covid 19 cases in gujarat

In [5]:
```
dates = soup.find_all("td",class_="bb-04em")
dates = [i.get_text() for i in dates]
date,cases,deaths=[],[],[]
for i in range(len(dates)):
    if i%3==0:
        date.append(dates[i])
    elif i%3==2:
        deaths.append(dates[i].split("(")[0])
    elif i%3==1:
        cases.append(dates[i].split("(")[0])
    else:
        print("Something went wrong at position :-",i)
```

In [6]:
```
covid_19_cases = pd.DataFrame(date,columns=['date'])
covid_19_cases["date"] = pd.to_datetime(covid_19_cases['date'])
covid_19_cases["Cases"] = cases
covid_19_cases["Deaths"] = deaths
covid_19_cases.head()
```

Out[6]:

	date	Cases	Deaths
0	2020-03-19	2	
1	2020-03-20	7	
2	2020-03-21	14	
3	2020-03-22	18	1
4	2020-03-23	30	1

In [7]:
```
covid_19_cases.shape
```

Out[7]: (418, 3)

In [8]:
```
covid_19_cases.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype         
---  ------  --------------  -----         
 0   date    418 non-null    datetime64[ns]
 1   Cases   418 non-null    object        
 2   Deaths  418 non-null    object        
dtypes: datetime64[ns](1), object(2)
memory usage: 9.9+ KB
```

In [9]:
```
covid_19_cases.to_csv("Cases.csv",index=False)
```

Covid-19 Pandemic in Gujarat by district

In [10]:
```
table = soup.find_all("td",class_=False,text="")
```

In [11]:
```
table = [i.get_text() for i in table]
```

In [12]:
```
district,total_cases,active_cases,total_recoveries,total_deaths,sample_tested,people_
for i in range(len(table[9:])):
    if i%7==0:
        district.append(table[9+i].strip("\n"))
    elif i%7==1:
        total_cases.append(table[9+i].strip("\n"))
    elif i%7==2:
        active_cases.append(table[9+i].strip("\n"))
    elif i%7==3:
        total_recoveries.append(table[9+i].strip("\n"))
    elif i%7==4:
        total_deaths.append(table[9+i].strip("\n"))
    elif i%7==5:
        sample_tested.append(table[9+i].strip("\n"))
    elif i%7==6:
        people_in_quarantine.append(table[9+i].strip("\n"))
    else:
        print("Something went wrong at position :-",i)
columns = ['district','total_cases','active_cases','total_recoveries','total_deaths','san
table = list(zip(district,total_cases,active_cases,total_recoveries,total_deaths,sample_
```

In [13]:
```
table = pd.DataFrame(table,columns=columns)
table.to_csv("Covid-19 pandemic in Gujarat.csv",index=False)
table
```

Out[13]:

	district	total_cases	active_cases	total_recoveries	total_deaths	sample_tested	peopl
0	Ahmedabad	207138	54722	149288	3128	4060936	
1	Amreli	7840	1393	6381	66	343228	
2	Anand	6577	1232	5315	30	302664	
3	Aravalli	3719	1364	2298	57	200897	
4	Banaskantha	10998	1033	9845	120	344001	
5	Bharuch	8881	1773	7023	85	276591	
6	Bhavnagar	17666	5370	12059	237	738534	
7	Botad	2000	399	1561	40	156441	
8	Chhota Udaipur	2829	630	2169	30	147640	
9	Dahod	8281	1618	6631	32	367466	
10	Dang	740	129	597	14	50661	
11	Devbhoomi Dwarka	2783	1065	1663	55	134804	
12	Gandhinagar	17911	3074	14666	171	522261	
13	Gir Somnath	5998	1861	4094	43	206204	
14	Jamnagar	29630	5809	23465	356	537645	
15	Junagadh	13815	3019	10621	175	472182	
16	Kutch	10030	2770	7143	117	529942	
17	Kheda	7931	1126	6771	34	368012	
18	Mahisagar	6355	2094	4213	48	211111	
19	Mehsana	21053	5590	15345	118	387464	
20	Morbi	6025	929	5012	84	270689	
21	Narmada	4935	1128	3798	9	150504	
22	Navsari	5489	1386	4087	16	243185	
23	Panchmahal	8810	2026	6735	49	279212	
24	Patan	9990	1534	8355	101	232034	
25	Porbandar	1815	310	1494	11	151446	
26	Rajkot	49656	4226	44825	605	1430825	
27	Sabarkantha	6887	1735	5028	124	277321	
28	Surat	129818	12661	115424	1733	4090780	
29	Surendranagar	7257	1168	5962	127	304868	
30	Tapi	3926	1346	2565	15	151998	
31	Vadodara	61156	10318	50198	640	1202040	

	district	total_cases	active_cases	total_recoveries	total_deaths	sample_tested	people
32	Valsad	4503	1320	3145	38	278500	

In [14]:
```
1 table.shape
```

Out[14]: (33, 7)