



---

# DIGITAL ASSIGNMENT - 6

---



MAY 20, 2021

RISHABH SHARMA

20MAI0082

Github Link :- <https://github.com/rishabh5197/Data-Mining/tree/main/Assignment-6>

Wikipedia link for Gujarat:- [https://en.wikipedia.org/wiki/COVID-19\\_pandemic\\_in\\_Gujarat](https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Gujarat)

**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)  
([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]: 1 import numpy as np
        2 import pandas as pd
        3 import sklearn
        4 import matplotlib.pyplot as plt
        5 import seaborn as sns
        6 from bs4 import BeautifulSoup as bs
        7 import requests
```

```
In [2]: 1 link = "https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Gujarat"
        2 print(link)
        3 page = requests.get(link)
        4 print(page)
```

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

```
In [3]: 1 soup = bs(page.content,"html.parser")
```

```
In [4]: 1 # soup
```

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

## Covid 19 cases in gujarat

```
In [6]: 1 covid_19_cases = pd.DataFrame(date,columns=['date'])
2 covid_19_cases["date"] = pd.to_datetime(covid_19_cases['date'])
3 covid_19_cases["Cases"] = cases
4 covid_19_cases["Deaths"] = deaths
5 covid_19_cases.head()
```

	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

Out[7]: (418, 3)

localhost:8888/notebooks/Assignment.ipynb#Assignment-Number:-6

```
In [9]: 1 covid_19_cases.to_csv("Cases.csv",index=False)
```

## Covid-19 Pandemic in Gujarat by district

```
In [10]: 1 table = soup.find_all("td",class_=False,text="")
```

```
In [11]: 1 table = [i.get_text() for i in table]
```

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

In [13]:

```
1 table = pd.DataFrame(table,columns=columns)
2 table.to_csv("Covid-19 pandemic in Gujarat.csv",index=False)
3 table
```

Out[13]:

	district	total_cases	active_cases	total_recoveries	total_deaths	sample_tested	people
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)

Excel file output of  
Cases

date	Cases	Deaths
19-03-2020	2	
20-03-2020	7	
21-03-2020	14	
22-03-2020	18	1
23-03-2020	30	1
24-03-2020	36	1
25-03-2020	38	2
26-03-2020	44	3
27-03-2020	47	4
28-03-2020	55	4
29-03-2020	63	5
30-03-2020	70	6
31-03-2020	74	6
04-01-2020	87	6
04-02-2020	88	7
04-03-2020	95	9
04-04-2020	108	10
04-05-2020	128	11
04-06-2020	146	12
04-07-2020	175	14
04-08-2020	186	16
04-09-2020	262	17
04-10-2020	378	19
04-11-2020	468	22
04-12-2020	516	24
13-04-2020	572	26
14-04-2020	650	28
15-04-2020	766	33
16-04-2020	929	36
17-04-2020	1,099	41
18-04-2020	1,376	53
19-04-2020	1,743	63
20-04-2020	1,939	71
21-04-2020	2,178	90
22-04-2020	2,407	103
23-04-2020	2,624	112
24-04-2020	2,815	127
25-04-2020	3,071	133
26-04-2020	3,301	151
27-04-2020	3,548	162
28-04-2020	3,774	181
29-04-2020	4,082	197
30-04-2020	4,395	214
05-01-2020	4,721	236
05-02-2020	5,054	262
05-03-2020	5,428	290



05-04-2020	5,804	319
05-05-2020	6,245	368
05-06-2020	6,625	396
05-07-2020	7,012	425
05-08-2020	7,402	449
05-09-2020	7,796	472
05-10-2020	8,194	493
05-11-2020	8,541	513
05-12-2020	8,903	537
13-05-2020	9,267	566
14-05-2020	9,591	586
15-05-2020	9,931	606
16-05-2020	10,988	625
17-05-2020	11,379	659
18-05-2020	11,745	694
19-05-2020	12,140	719
20-05-2020	12,537	749
21-05-2020	12,905	773
22-05-2020	13,268	802
23-05-2020	13,664	829
24-05-2020	14,056	858
25-05-2020	14,460	888
26-05-2020	14,821	915
27-05-2020	15,195	938
28-05-2020	15,562	960
29-05-2020	15,933	980
30-05-2020	16,343	1,007
31-05-2020	16,779	1,038
06-01-2020	17,200	1,063
06-02-2020	17,615	1,092
06-03-2020	18,100	1,122
06-04-2020	18,584	1,155
06-05-2020	19,094	1,190
06-06-2020	19,592	1,219
06-07-2020	20,070	1,249
06-08-2020	20,545	1,280
06-09-2020	21,014	1,313
06-10-2020	21,521	1,347
06-11-2020	22,032	1,385
06-12-2020	22,527	1,415
13-06-2020	23,038	1,449
14-06-2020	23,590	1,478
15-06-2020	24,104	1,506
16-06-2020	24,628	1,534
17-06-2020	25,148	1,561
18-06-2020	25,658	1,592
19-06-2020	26,198	1,619

20-06-2020	26,737	1,639
21-06-2020	27,317	1,664
22-06-2020	27,880	1,685
23-06-2020	28,429	1,711
24-06-2020	29,001	1,739
25-06-2020	29,578	1,754
26-06-2020	30,158	1,772
27-06-2020	30,773	1,789
28-06-2020	31,397	1,809
29-06-2020	32,023	1,828
30-06-2020	32,558	1,847
07-01-2020	33,318	1,869
07-02-2020	33,923	1,888
07-03-2020	34,686	1,906
07-04-2020	35,398	1,927
07-05-2020	36,123	1,945
07-06-2020	36,858	1,962
07-07-2020	37,636	1,979
07-08-2020	38,419	1,995
07-09-2020	39,280	2,010
07-10-2020	40,155	2,024
07-11-2020	41,026	2,034
07-12-2020	41,906	2,046
13-07-2020	42,808	2,057
14-07-2020	43,723	2,071
15-07-2020	44,648	2,081
16-07-2020	45,567	2,091
17-07-2020	46,516	2,108
18-07-2020	47,476	2,127
19-07-2020	48,441	2,147
20-07-2020	49,439	2,167
21-07-2020	50,465	2,201
22-07-2020	51,485	2,229
23-07-2020	52,563	2,257
24-07-2020	53,631	2,283
25-07-2020	54,712	2,305
26-07-2020	55,822	2,326
27-07-2020	56,874	2,348
28-07-2020	57,982	2,372
29-07-2020	59,126	2,396
30-07-2020	60,285	2,418
31-07-2020	61,438	2,441
08-01-2020	62,574	2,465
08-02-2020	63,575	2,487
08-03-2020	64,684	2,509
08-04-2020	65,704	2,534
08-05-2020	66,777	2,557

08-06-2020	67,811	2,584
08-07-2020	68,885	2,606
08-08-2020	69,986	2,629
08-09-2020	71,064	2,654
08-10-2020	72,120	2,674
08-11-2020	73,163	2,695
08-12-2020	74,318	2,713
13-08-2020	75,408	2,731
14-08-2020	76,480	2,746
15-08-2020	77,559	2,765
16-08-2020	78,680	2,785
17-08-2020	79,710	2,800
18-08-2020	80,802	2,820
19-08-2020	81,942	2,837
20-08-2020	83,107	2,853
21-08-2020	84,311	2,867
22-08-2020	85,523	2,881
23-08-2020	86,624	2,895
24-08-2020	87,691	2,908
25-08-2020	88,805	2,928
26-08-2020	89,994	2,945
27-08-2020	91,179	2,962
28-08-2020	92,452	2,976
29-08-2020	93,734	2,989
30-08-2020	95,009	3,006
31-08-2020	96,300	3,020
09-01-2020	97,629	3,034
09-02-2020	98,888	3,046
09-03-2020	1,00,213	3,062
09-04-2020	1,01,533	3,076
09-05-2020	1,02,844	3,091
09-06-2020	1,04,179	3,105
09-07-2020	1,05,509	3,120
09-08-2020	1,06,804	3,133
09-09-2020	1,08,133	3,149
09-10-2020	1,09,465	3,164
09-11-2020	1,10,809	3,180
09-12-2020	1,12,174	3,195
13-09-2020	1,13,500	3,210
14-09-2020	1,14,834	3,227
15-09-2020	1,16,183	3,244
16-09-2020	1,17,547	3,256
17-09-2020	1,18,926	3,270
18-09-2020	1,20,336	3,286
19-09-2020	1,21,768	3,302
20-09-2020	1,23,175	3,319
21-09-2020	1,24,605	3,336

22-09-2020	1,26,007	3,352
23-09-2020	1,27,379	3,367
24-09-2020	1,28,787	3,381
25-09-2020	1,30,229	3,393
26-09-2020	1,31,646	3,406
27-09-2020	1,33,057	3,416
28-09-2020	1,34,461	3,428
29-09-2020	1,35,842	3,439
30-09-2020	1,37,232	3,450
10-01-2020	1,38,583	3,460
10-02-2020	1,39,893	3,475
10-03-2020	1,41,236	3,487
10-04-2020	1,42,538	3,496
10-05-2020	1,43,865	3,509
10-06-2020	1,45,200	3,519
10-07-2020	1,46,511	3,528
10-08-2020	1,47,789	3,538
10-09-2020	1,49,032	3,547
10-10-2020	1,50,253	3,557
10-11-2020	1,51,434	3,566
10-12-2020	1,52,603	3,574
13-10-2020	1,53,761	3,584
14-10-2020	1,54,936	3,595
15-10-2020	1,56,121	3,606
16-10-2020	1,57,312	3,617
17-10-2020	1,58,473	3,626
18-10-2020	1,59,564	3,635
19-10-2020	1,60,560	3,643
20-10-2020	1,61,686	3,651
21-10-2020	1,62,823	3,660
22-10-2020	1,63,959	3,667
23-10-2020	1,65,071	3,673
24-10-2020	1,66,092	3,679
25-10-2020	1,67,011	3,686
26-10-2020	1,67,919	3,690
27-10-2020	1,68,911	3,695
28-10-2020	1,69,891	3,701
29-10-2020	1,70,878	3,705
30-10-2020	1,71,847	3,711
31-10-2020	1,72,782	3,716
11-01-2020	1,73,642	3,721
11-02-2020	1,74,517	3,725
11-03-2020	1,75,471	3,731
11-04-2020	1,76,446	3,737
11-05-2020	1,77,436	3,744
11-06-2020	1,78,471	3,748
11-07-2020	1,79,517	3,753

11-08-2020	1,80,537	3,760
11-09-2020	1,81,508	3,765
11-10-2020	1,82,557	3,770
11-11-2020	1,83,682	3,776
11-12-2020	1,84,802	3,782
13-11-2020	1,85,954	3,788
14-11-2020	1,87,078	3,794
15-11-2020	1,88,148	3,800
16-11-2020	1,89,074	3,805
17-11-2020	1,90,199	3,812
18-11-2020	1,91,480	3,820
19-11-2020	1,92,820	3,827
20-11-2020	1,94,240	3,834
21-11-2020	1,95,755	3,843
22-11-2020	1,97,250	3,856
23-11-2020	1,98,737	3,873
24-11-2020	2,00,247	3,889
25-11-2020	2,01,787	3,903
26-11-2020	2,03,347	3,919
27-11-2020	2,04,954	3,935
28-11-2020	2,06,552	3,950
29-11-2020	2,08,116	3,966
30-11-2020	2,09,618	3,986
12-01-2020	2,11,095	4,001
12-02-2020	2,12,607	4,015
12-03-2020	2,14,147	4,028
12-04-2020	2,15,657	4,046
12-05-2020	2,17,171	4,061
12-06-2020	2,18,626	4,078
12-07-2020	2,20,006	4,092
12-08-2020	2,21,331	4,107
12-09-2020	2,22,649	4,120
12-10-2020	2,23,919	4,132
12-11-2020	2,25,142	4,145
12-12-2020	2,26,346	4,157
13-12-2020	2,27,521	4,168
14-12-2020	2,28,641	4,179
15-12-2020	2,29,751	4,190
16-12-2020	2,30,911	4,200
17-12-2020	2,32,026	4,208
18-12-2020	2,33,101	4,217
19-12-2020	2,34,127	4,224
20-12-2020	2,35,137	4,231
21-12-2020	2,36,097	4,238
22-12-2020	2,37,085	4,245
23-12-2020	2,38,043	4,251
24-12-2020	2,39,033	4,259

25-12-2020	2,39,943	4,265
26-12-2020	2,40,833	4,272
27-12-2020	2,41,683	4,279
28-12-2020	2,42,493	4,285
29-12-2020	2,43,297	4,292
30-12-2020	2,44,096	4,299
31-12-2020	2,44,876	4,303
01-01-2021	2,45,610	4,306
01-02-2021	2,46,351	4,311
01-03-2021	2,47,066	4,315
01-04-2021	2,47,764	4,318
01-05-2021	2,48,419	4,322
01-06-2021	2,49,084	4,326
01-07-2021	2,49,751	4,329
01-08-2021	2,50,436	4,332
01-09-2021	2,51,111	4,337
01-10-2021	2,51,782	4,341
01-11-2021	2,52,397	4,344
01-12-2021	2,52,999	4,347
13-01-2021	2,53,582	4,351
14-01-2021	2,54,152	4,354
15-01-2021	2,54,687	4,357
16-01-2021	2,55,192	4,360
17-01-2021	2,55,710	4,362
18-01-2021	2,56,205	4,364
19-01-2021	2,56,690	4,366
20-01-2021	2,57,180	4,368
21-01-2021	2,57,651	4,369
22-01-2021	2,58,102	4,371
23-01-2021	2,58,525	4,372
24-01-2021	2,58,935	4,373
25-01-2021	2,59,325	4,376
26-01-2021	2,59,705	4,378
27-01-2021	2,60,058	4,379
28-01-2021	2,60,404	4,381
29-01-2021	2,60,739	4,382
30-01-2021	2,61,062	4,384
31-01-2021	2,61,378	4,384
02-01-2021	2,61,676	4,385
02-02-2021	2,61,961	4,386
02-03-2021	2,62,244	4,388
02-04-2021	2,62,519	4,389
02-05-2021	2,62,786	4,390
02-06-2021	2,63,038	4,391
02-07-2021	2,63,282	4,392
02-08-2021	2,63,514	4,393
02-09-2021	2,63,748	4,394

02-10-2021	2,64,003	4,394
02-11-2021	2,64,288	4,396
02-12-2021	2,64,556	4,397
13-02-2021	2,64,835	4,397
14-02-2021	2,65,082	4,398
15-02-2021	2,65,331	4,398
16-02-2021	2,65,594	4,399
17-02-2021	2,65,872	4,400
18-02-2021	2,66,135	4,400
19-02-2021	2,66,401	4,401
20-02-2021	2,66,659	4,401
21-02-2021	2,66,942	4,402
22-02-2021	2,67,257	4,403
23-02-2021	2,67,605	4,403
24-02-2021	2,67,985	4,404
25-02-2021	2,68,409	4,405
26-02-2021	2,68,869	4,405
27-02-2021	2,69,320	4,406
28-02-2021	2,69,727	4,407
03-01-2021	2,70,154	4,408
03-02-2021	2,70,608	4,408
03-03-2021	2,71,083	4,409
03-04-2021	2,71,563	4,409
03-05-2021	2,72,078	4,410
03-06-2021	2,72,649	4,411
03-07-2021	2,73,224	4,412
03-08-2021	2,73,779	4,413
03-09-2021	2,74,360	4,415
03-10-2021	2,75,035	4,415
03-11-2021	2,75,745	4,415
03-12-2021	2,76,460	4,417
13-03-2021	2,77,235	4,419
14-03-2021	2,78,045	4,421
15-03-2021	2,78,935	4,422
16-03-2021	2,79,889	4,424
17-03-2021	2,81,011	4,427
18-03-2021	2,82,287	4,430
19-03-2021	2,83,702	4,434
20-03-2021	2,85,267	4,440
21-03-2021	2,86,847	4,447
22-03-2021	2,88,487	4,451
23-03-2021	2,90,217	4,455
24-03-2021	2,92,007	4,463
25-03-2021	2,93,968	4,470
26-03-2021	2,96,158	4,476
27-03-2021	2,98,434	4,481
28-03-2021	3,00,704	4,489

29-03-2021	3,02,956	4,497
30-03-2021	3,05,176	4,507
31-03-2021	3,07,536	4,516
04-01-2021	3,09,946	4,525
04-02-2021	3,12,586	4,536
04-03-2021	3,15,401	4,549
04-04-2021	3,18,276	4,563
04-05-2021	3,21,436	4,578
04-06-2021	3,24,716	4,595
04-07-2021	3,28,291	4,617
04-08-2021	3,32,312	4,652
04-09-2021	3,36,853	4,694
04-10-2021	3,41,864	4,743
04-11-2021	3,47,333	4,797
04-12-2021	3,53,354	4,852
13-04-2021	3,60,044	4,919
14-04-2021	3,67,454	4,992
15-04-2021	3,75,606	5,073
16-04-2021	3,84,526	5,167
17-04-2021	3,94,067	5,264
18-04-2021	4,04,407	5,374
19-04-2021	4,15,810	5,491
20-04-2021	4,28,016	5,612
21-04-2021	4,40,569	5,737
22-04-2021	4,53,674	5,874
23-04-2021	4,67,478	6,016
24-04-2021	4,81,575	6,168
25-04-2021	4,95,871	6,325
26-04-2021	5,10,211	6,483
27-04-2021	5,24,563	6,653
28-04-2021	5,38,683	6,827
29-04-2021	5,53,010	7,007
30-04-2021	5,67,615	7,180
05-01-2021	5,81,462	7,352
05-02-2021	5,94,440	7,505
05-03-2021	6,07,260	7,645
05-04-2021	6,20,310	7,776
05-05-2021	6,33,265	7,909
05-06-2021	6,45,810	8,032
05-07-2021	6,57,874	8,151
05-08-2021	6,69,766	8,270
05-09-2021	6,80,850	8,391
05-10-2021	6,92,442	8,508



# Excel file output of Covid-19 pandemic in Gujarat

district	total_cases	active_cases	total_recoveries	total_deaths	sample_tested	people_in_quarantine
Ahmedabad	207138	54722	149288	3128	4060936	62604
Amreli	7840	1393	6381	66	343228	105610
Anand	6577	1232	5315	30	302664	2736
Aravalli	3719	1364	2298	57	200897	1227
Banaskantha	10998	1033	9845	120	344001	7452
Bharuch	8881	1773	7023	85	276591	9123
Bhavnagar	17666	5370	12059	237	738534	66739
Botad	2000	399	1561	40	156441	981
Chhota Udaipur	2829	630	2169	30	147640	980
Dahod	8281	1618	6631	32	367466	14712
Dang	740	129	597	14	50661	901
Devbhoomi Dwa	2783	1065	1663	55	134804	34
Gandhinagar	17911	3074	14666	171	522261	13085
Gir Somnath	5998	1861	4094	43	206204	806
Jamnagar	29630	5809	23465	356	537645	28294
Junagadh	13815	3019	10621	175	472182	4292
Kutch	10030	2770	7143	117	529942	2854
Kheda	7931	1126	6771	34	368012	8042
Mahisagar	6355	2094	4213	48	211111	978
Mehsana	21053	5590	15345	118	387464	1436
Morbi	6025	929	5012	84	270689	838
Narmada	4935	1128	3798	9	150504	933
Navsari	5489	1386	4087	16	243185	8563
Panchmahal	8810	2026	6735	49	279212	4409
Patan	9990	1534	8355	101	232034	1437
Porbandar	1815	310	1494	11	151446	1310
Rajkot	49656	4226	44825	605	1430825	37384
Sabarkantha	6887	1735	5028	124	277321	3137
Surat	129818	12661	115424	1733	4090780	31145
Surendranagar	7257	1168	5962	127	304868	16872
Tapi	3926	1346	2565	15	151998	724
Vadodara	61156	10318	50198	640	1202040	10039
Valsad	4503	1320	3145	38	278500	14025