In [6]:
```
import pandas as pd
```

In [7]:
```
data=pd.read_csv("covid_19_data.csv")
```

In [8]:
```
data
```

Out[8]:

	Date	State	Region	Confirmed	Deaths	Recovered
0	4/29/2020	NaN	Afghanistan	1939	60	252
1	4/29/2020	NaN	Albania	766	30	455
2	4/29/2020	NaN	Algeria	3848	444	1702
3	4/29/2020	NaN	Andorra	743	42	423
4	4/29/2020	NaN	Angola	27	2	7
...	...	...	...	...	...	...
316	4/29/2020	Wyoming	US	545	7	0
317	4/29/2020	Xinjiang	Mainland China	76	3	73
318	4/29/2020	Yukon	Canada	11	0	0
319	4/29/2020	Yunnan	Mainland China	185	2	181
320	4/29/2020	Zhejiang	Mainland China	1268	1	1263

321 rows × 6 columns

Explore the Data

In [9]:
```
data.count()
```

Out[9]:
```
Date         321
State        140
Region       321
Confirmed    321
Deaths       321
Recovered    321
dtype: int64
```

In [10]:
```
data.isnull().sum()
```

Out[10]:
```
Date           0
State        181
Region         0
Confirmed      0
Deaths         0
Recovered      0
dtype: int64
```

In [11]:
```
.sum()
```

Loading [MathJax]/extensions/Safe.js

Out[11]:
```
Date           0
State        181
Region         0
Confirmed      0
Deaths         0
Recovered      0
dtype: int64
```

Q.1) Show the number of comfirmed,Death and recovered cases in each Region.

In [12]:
```
data.head(2)
```

Out[12]:

	Date	State	Region	Confirmed	Deaths	Recovered
0	4/29/2020	NaN	Afghanistan	1939	60	252
1	4/29/2020	NaN	Albania	766	30	455

Q.2) Show me top 5 highest Death rates Region?

In [13]:
```
top_5 = data.nlargest(5,"Deaths")
print(top_5[['Region','Deaths']])
```

```
    Region  Deaths
80    Italy   27682
168      UK   26097
153   Spain   24275
57   France   24087
265      US   23477
```

In [14]:
```
data.groupby('Region').agg({
    'Confirmed': 'sum',
    'Deaths': 'sum',
    'Recovered': 'sum'
})
```

Loading [MathJax]/extensions/Safe.js

Out[14]:

Region	Confirmed	Deaths	Recovered
Afghanistan	1939	60	252
Albania	766	30	455
Algeria	3848	444	1702
Andorra	743	42	423
Angola	27	2	7
...	...	...	...
West Bank and Gaza	344	2	71
Western Sahara	6	0	5
Yemen	6	0	1
Zambia	97	3	54
Zimbabwe	32	4	5

187 rows × 3 columns

Q.3) Combine Region and states Column

In [15]:
```
data["State"].fillna('No State',inplace=True)
data["rs"]=data["Region"] + "_" + data["State"]
data.head(59)
```

Loading [MathJax]/extensions/Safe.js

Out[15]:

	Date	State	Region	Confirmed	Deaths	Recovered	rs
0	4/29/2020	No State	Afghanistan	1939	60	252	Afghanistan_No State
1	4/29/2020	No State	Albania	766	30	455	Albania_No State
2	4/29/2020	No State	Algeria	3848	444	1702	Algeria_No State
3	4/29/2020	No State	Andorra	743	42	423	Andorra_No State
4	4/29/2020	No State	Angola	27	2	7	Angola_No State
5	4/29/2020	No State	Antigua and Barbuda	24	3	11	Antigua and Barbuda_No State
6	4/29/2020	No State	Argentina	4285	214	1192	Argentina_No State
7	4/29/2020	No State	Armenia	1932	30	900	Armenia_No State
8	4/29/2020	No State	Austria	15402	580	12779	Austria_No State
9	4/29/2020	No State	Azerbaijan	1766	23	1267	Azerbaijan_No State
10	4/29/2020	No State	Bahamas	80	11	23	Bahamas_No State
11	4/29/2020	No State	Bahrain	2921	8	1455	Bahrain_No State
12	4/29/2020	No State	Bangladesh	7103	163	150	Bangladesh_No State
13	4/29/2020	No State	Barbados	80	7	39	Barbados_No State
14	4/29/2020	No State	Belarus	13181	84	2072	Belarus_No State
15	4/29/2020	No State	Belgium	47859	7501	11283	Belgium_No State
16	4/29/2020	No State	Belize	18	2	9	Belize_No State
17	4/29/2020	No State	Benin	64	1	33	Benin_No State
18	4/29/2020	No State	Bhutan	7	0	5	Bhutan_No State
19	4/29/2020	No State	Bolivia	1110	59	117	Bolivia_No State
20	4/29/2020	No State	Bosnia and Herzegovina	1677	65	710	Bosnia and Herzegovina_No State

Loading [MathJax]/extensions/Safe.js

	Date	State	Region	Confirmed	Deaths	Recovered	rs
21	4/29/2020	No State	Botswana	23	1	5	Botswana_No State
22	4/29/2020	No State	Brazil	79685	5513	34132	Brazil_No State
23	4/29/2020	No State	Brunei	138	1	124	Brunei_No State
24	4/29/2020	No State	Bulgaria	1447	64	243	Bulgaria_No State
25	4/29/2020	No State	Burkina Faso	641	43	498	Burkina Faso_No State
26	4/29/2020	No State	Burma	150	6	27	Burma_No State
27	4/29/2020	No State	Burundi	11	1	4	Burundi_No State
28	4/29/2020	No State	Cabo Verde	114	1	2	Cabo Verde_No State
29	4/29/2020	No State	Cambodia	122	0	119	Cambodia_No State
30	4/29/2020	No State	Cameroon	1832	61	934	Cameroon_No State
31	4/29/2020	No State	Central African Republic	50	0	10	Central African Republic_No State
32	4/29/2020	No State	Chad	52	2	19	Chad_No State
33	4/29/2020	No State	Chile	14885	216	8057	Chile_No State
34	4/29/2020	No State	Colombia	6207	278	1411	Colombia_No State
35	4/29/2020	No State	Congo (Brazzaville)	207	8	19	Congo (Brazzaville)_No State
36	4/29/2020	No State	Congo (Kinshasa)	491	30	59	Congo (Kinshasa)_No State
37	4/29/2020	No State	Costa Rica	713	6	323	Costa Rica_No State
38	4/29/2020	No State	Croatia	2062	67	1288	Croatia_No State
39	4/29/2020	No State	Cuba	1467	58	617	Cuba_No State
40	4/29/2020	No State	Cyprus	843	15	148	Cyprus_No State

Loading [MathJax]/extensions/Safe.js

	Date	State	Region	Confirmed	Deaths	Recovered	rs
41	4/29/2020	No State	Czech Republic	7579	227	3108	Czech Republic_No State
42	4/29/2020	No State	Denmark	9008	443	6366	Denmark_No State
43	4/29/2020	No State	Diamond Princess	712	13	645	Diamond Princess_No State
44	4/29/2020	No State	Djibouti	1077	2	599	Djibouti_No State
45	4/29/2020	No State	Dominica	16	0	13	Dominica_No State
46	4/29/2020	No State	Dominican Republic	6652	293	1228	Dominican Republic_No State
47	4/29/2020	No State	Ecuador	24675	883	1557	Ecuador_No State
48	4/29/2020	No State	Egypt	5268	380	1335	Egypt_No State
49	4/29/2020	No State	El Salvador	377	9	106	El Salvador_No State
50	4/29/2020	No State	Equatorial Guinea	315	1	9	Equatorial Guinea_No State
51	4/29/2020	No State	Eritrea	39	0	19	Eritrea_No State
52	4/29/2020	No State	Estonia	1666	50	236	Estonia_No State
53	4/29/2020	No State	Eswatini	91	1	10	Eswatini_No State
54	4/29/2020	No State	Ethiopia	130	3	58	Ethiopia_No State
55	4/29/2020	No State	Fiji	18	0	12	Fiji_No State
56	4/29/2020	No State	Finland	4906	206	2800	Finland_No State
57	4/29/2020	No State	France	165093	24087	48228	France_No State
58	4/29/2020	No State	Gabon	276	3	67	Gabon_No State

Q.4) Show me the countries which have zero Death rate?

In [16]:
```
data[data["Deaths"]==0][["rs","Deaths"]]
```

Loading [MathJax]/extensions/Safe.js

Out[16]:

	rs	Deaths
18	Bhutan_No State	0
29	Cambodia_No State	0
31	Central African Republic_No State	0
45	Dominica_No State	0
51	Eritrea_No State	0
55	Fiji_No State	0
64	Grenada_No State	0
70	Holy See_No State	0
90	Laos_No State	0
99	Madagascar_No State	0
110	Mongolia_No State	0
113	Mozambique_No State	0
114	Namibia_No State	0
115	Nepal_No State	0
126	Papua New Guinea_No State	0
135	Rwanda_No State	0
136	Saint Kitts and Nevis_No State	0
137	Saint Lucia_No State	0
138	Saint Vincent and the Grenadines_No State	0
140	Sao Tome and Principe_No State	0
144	Seychelles_No State	0
152	South Sudan_No State	0
163	Timor-Leste_No State	0
169	Uganda_No State	0
175	Vietnam_No State	0
177	Western Sahara_No State	0
178	Yemen_No State	0
184	UK_Anguilla	0
192	Netherlands_Bonaire, Sint Eustatius and Saba	0
204	US_Diamond Princess cruise ship	0
206	UK_Falkland Islands (Malvinas)	0
207	Denmark_Faroe Islands	0
210	France_French Polynesia	0

Loading [MathJax]/extensions/Safe.js

	rs	Deaths
214	UK_Gibraltar	0
215	Canada_Grand Princess	0
217	Denmark_Greenland	0
237	Mainland China_Jiangsu	0
244	Macau_Macau	0
259	Canada_New Brunswick	0
260	France_New Caledonia	0
267	Mainland China_Ningxia	0
271	Australia_Northern Territory	0
272	Canada_Northwest Territories	0
279	Canada_Prince Edward Island	0
281	Mainland China_Qinghai	0
284	Canada_Recovered	0
285	US_Recovered	0
286	France_Reunion	0
288	France_Saint Barthelemy	0
289	France_Saint Pierre and Miquelon	0
294	Mainland China_Shanxi	0
305	Mainland China_Tibet	0
318	Canada_Yukon	0

Q.5) How many Regions which have zero Death rate?

In [17]:
```
data.head(2)
```

Out[17]:

	Date	State	Region	Confirmed	Deaths	Recovered	rs
0	4/29/2020	No State	Afghanistan	1939	60	252	Afghanistan_No State
1	4/29/2020	No State	Albania	766	30	455	Albania_No State

In [18]:
```
death=data.loc[data["Deaths"]==0]["Region"].nunique()
print("No of regions with zero Deaths=",death)
```

```
No of regions with zero Deaths= 36
```

Loading [MathJax]/extensions/Safe.js

In [ ]:

In [19]:
```
data_Copy=data.copy()
data_Copy["Region"].replace("Afghan","Afghansitan",inplace=True)
data
```

Out[19]:

	Date	State	Region	Confirmed	Deaths	Recovered	rs
0	4/29/2020	No State	Afghanistan	1939	60	252	Afghanistan_No State
1	4/29/2020	No State	Albania	766	30	455	Albania_No State
2	4/29/2020	No State	Algeria	3848	444	1702	Algeria_No State
3	4/29/2020	No State	Andorra	743	42	423	Andorra_No State
4	4/29/2020	No State	Angola	27	2	7	Angola_No State
...	...	...	...	...	...	...	...
316	4/29/2020	Wyoming	US	545	7	0	US_Wyoming
317	4/29/2020	Xinjiang	Mainland China	76	3	73	Mainland China_Xinjiang
318	4/29/2020	Yukon	Canada	11	0	0	Canada_Yukon
319	4/29/2020	Yunnan	Mainland China	185	2	181	Mainland China_Yunnan
320	4/29/2020	Zhejiang	Mainland China	1268	1	1263	Mainland China_Zhejiang

321 rows × 7 columns

Q.6) Top 5 rs(Resion_states) which have high recovered cases?

In [27]:
```
data.head(2)
```

Out[27]:

	Date	State	Region	Confirmed	Deaths	Recovered	rs
0	4/29/2020	No State	Afghanistan	1939	60	252	Afghanistan_No State
1	4/29/2020	No State	Albania	766	30	455	Albania_No State

In [28]:
```
top5=data.nlargest(5,'Recovered')
print(top5[['rs','Recovered']])
```

Loading [MathJax]/extensions/Safe.js

```
                    rs  Recovered
153    Spain_No State     132929
285      US_Recovered     120720
61   Germany_No State     120400
76      Iran_No State      73791
80     Italy_No State      71252
```

Q.7)Regions which have 0 recovered cases?

In [38]:
```
a=data[data['Recovered']==0][['rs','Recovered']]
a
```

Out[38]:

	rs	Recovered
98	MS Zaandam_No State	0
116	Netherlands_No State	0
126	Papua New Guinea_No State	0
152	South Sudan_No State	0
168	UK_No State	0
...	...	...
312	US_Washington	0
313	US_West Virginia	0
315	US_Wisconsin	0
316	US_Wyoming	0
318	Canada_Yukon	0

78 rows × 2 columns

Q.8)total number of regions which have 0 recovered cases?

In [40]:
```
print("Total region with zero Recovered cases=",len(a))
```

```
Total region with zero Recovered cases= 78
```

In [42]:
```
data.head(2)
```

Out[42]:

	Date	State	Region	Confirmed	Deaths	Recovered	rs
0	4/29/2020	No State	Afghanistan	1939	60	252	Afghanistan_No State
1	4/29/2020	No State	Albania	766	30	455	Albania_No State

Loading [MathJax]/extensions/Safe.js

Q.9) top 5 regions which have highest confirmed cases?

In [46]:
```
t5=data.nlargest(5,"Confirmed")
print(t5[["rs","Confirmed"]])
```

```
                  rs  Confirmed
265      US_New York     299691
153   Spain_No State     236899
80    Italy_No State     203591
168      UK_No State     165221
57   France_No State     165093
```

Q.10) Which regoin have least cases top least 5 regions?

In [47]:
```
l5=data.nsmallest(5,"Confirmed")
print(l5[["rs","Confirmed"]])
```

```
                                    rs  Confirmed
203  Canada_Diamond Princess cruise ship          0
284                  Canada_Recovered          0
285                      US_Recovered          0
289    France_Saint Pierre and Miquelon          1
305                Mainland China_Tibet          1
```

Q.11) Removed all the records where Confirmed cases is less than 10?

In [50]:
```
data[~(data.Confirmed <10)]
```

Loading [MathJax]/extensions/Safe.js

Out[50]:

	Date	State	Region	Confirmed	Deaths	Recovered	rs
0	4/29/2020	No State	Afghanistan	1939	60	252	Afghanistan_No State
1	4/29/2020	No State	Albania	766	30	455	Albania_No State
2	4/29/2020	No State	Algeria	3848	444	1702	Algeria_No State
3	4/29/2020	No State	Andorra	743	42	423	Andorra_No State
4	4/29/2020	No State	Angola	27	2	7	Angola_No State
...	...	...	...	...	...	...	...
316	4/29/2020	Wyoming	US	545	7	0	US_Wyoming
317	4/29/2020	Xinjiang	Mainland China	76	3	73	Mainland China_Xinjiang
318	4/29/2020	Yukon	Canada	11	0	0	Canada_Yukon
319	4/29/2020	Yunnan	Mainland China	185	2	181	Mainland China_Yunnan
320	4/29/2020	Zhejiang	Mainland China	1268	1	1263	Mainland China_Zhejiang

304 rows × 7 columns

Q.12) How many Confirmed ,deaths and Recovered cases were reported from US till 29 April 2020?

In [58]:
```
s=data.loc[(data["Region"]=="US") & (data["Date"]<="4/29/2020")]
s
```

Loading [MathJax]/extensions/Safe.js

Out[58]:

	Date	State	Region	Confirmed	Deaths	Recovered	rs
181	4/29/2020	Alabama	US	6912	256	0	US_Alabama
182	4/29/2020	Alaska	US	355	9	0	US_Alaska
186	4/29/2020	Arizona	US	7209	308	0	US_Arizona
187	4/29/2020	Arkansas	US	3193	57	0	US_Arkansas
195	4/29/2020	California	US	48747	1946	0	US_California
199	4/29/2020	Colorado	US	14758	766	0	US_Colorado
200	4/29/2020	Connecticut	US	26767	2169	0	US_Connecticut
202	4/29/2020	Delaware	US	4655	144	0	US_Delaware
204	4/29/2020	Diamond Princess cruise ship	US	49	0	0	US_Diamond Princess cruise ship
205	4/29/2020	District of Columbia	US	4106	205	0	US_District of Columbia
208	4/29/2020	Florida	US	33193	1218	0	US_Florida
213	4/29/2020	Georgia	US	25775	1101	0	US_Georgia
216	4/29/2020	Grand Princess	US	103	3	0	US_Grand Princess
219	4/29/2020	Guam	US	141	5	0	US_Guam
224	4/29/2020	Hawaii	US	613	16	0	US_Hawaii
231	4/29/2020	Idaho	US	1952	60	0	US_Idaho
232	4/29/2020	Illinois	US	50358	2215	0	US_Illinois
233	4/29/2020	Indiana	US	17182	964	0	US_Indiana
235	4/29/2020	Iowa	US	6843	148	0	US_Iowa
240	4/29/2020	Kansas	US	3839	134	0	US_Kansas
241	4/29/2020	Kentucky	US	4537	234	0	US_Kentucky
243	4/29/2020	Louisiana	US	27660	1845	0	US_Louisiana
245	4/29/2020	Maine	US	1056	52	0	US_Maine
248	4/29/2020	Maryland	US	20849	1078	0	US_Maryland
249	4/29/2020	Massachusetts	US	60265	3405	0	US_Massachusetts
251	4/29/2020	Michigan	US	40399	3670	0	US_Michigan
252	4/29/2020	Minnesota	US	4644	319	0	US_Minnesota
253	4/29/2020	Mississippi	US	6569	250	0	US_Mississippi
254	4/29/2020	Missouri	US	7660	338	0	US_Missouri
255	4/29/2020	Montana	US	451	16	0	US_Montana
		Nebraska	US	3851	56	0	US_Nebraska

Loading [MathJax]/extensions/Safe.js

	Date	State	Region	Confirmed	Deaths	Recovered	rs
258	4/29/2020	Nevada	US	4934	230	0	US_Nevada
261	4/29/2020	New Hampshire	US	2058	60	0	US_New Hampshire
262	4/29/2020	New Jersey	US	116365	6771	0	US_New Jersey
263	4/29/2020	New Mexico	US	3213	112	0	US_New Mexico
265	4/29/2020	New York	US	299691	23477	0	US_New York
268	4/29/2020	North Carolina	US	10180	382	0	US_North Carolina
269	4/29/2020	North Dakota	US	1033	19	0	US_North Dakota
270	4/29/2020	Northern Mariana Islands	US	14	2	0	US_Northern Mariana Islands
274	4/29/2020	Ohio	US	17303	937	0	US_Ohio
275	4/29/2020	Oklahoma	US	3473	214	0	US_Oklahoma
277	4/29/2020	Oregon	US	2446	101	0	US_Oregon
278	4/29/2020	Pennsylvania	US	46327	2373	0	US_Pennsylvania
280	4/29/2020	Puerto Rico	US	1433	86	0	US_Puerto Rico
285	4/29/2020	Recovered	US	0	0	120720	US_Recovered
287	4/29/2020	Rhode Island	US	8247	251	0	US_Rhode Island
298	4/29/2020	South Carolina	US	5882	231	0	US_South Carolina
299	4/29/2020	South Dakota	US	2373	13	0	US_South Dakota
302	4/29/2020	Tennessee	US	10366	195	0	US_Tennessee
303	4/29/2020	Texas	US	27257	754	0	US_Texas
307	4/29/2020	Utah	US	4497	45	0	US_Utah
308	4/29/2020	Vermont	US	862	47	0	US_Vermont
310	4/29/2020	Virgin Islands	US	57	4	0	US_Virgin Islands
311	4/29/2020	Virginia	US	14962	522	0	US_Virginia
312	4/29/2020	Washington	US	14070	801	0	US_Washington
313	4/29/2020	West Virginia	US	1110	38	0	US_West Virginia
315	4/29/2020	Wisconsin	US	6520	308	0	US_Wisconsin
316	4/29/2020	Wyoming	US	545	7	0	US_Wyoming

In [59]:
```
len(s)
```

Out[59]: 58

Loading [MathJax]/extensions/Safe.js

Q.13) Sort ths data wrt Confirmed cases in ascending order?

In [61]:
```
data.sort_values(by=["Confirmed"],ascending=True)
```

Out[61]:

	Date	State	Region	Confirmed	Deaths	Recovered	rs
285	4/29/2020	Recovered	US	0	0	120720	US_Recovered
284	4/29/2020	Recovered	Canada	0	0	20327	Canada_Recovered
203	4/29/2020	Diamond Princess cruise ship	Canada	0	1	0	Canada_Diamond Princess cruise ship
305	4/29/2020	Tibet	Mainland China	1	0	1	Mainland China_Tibet
289	4/29/2020	Saint Pierre and Miquelon	France	1	0	0	France_Saint Pierre and Miquelon
...	...	...	...	...	...	...	...
57	4/29/2020	No State	France	165093	24087	48228	France_No State
168	4/29/2020	No State	UK	165221	26097	0	UK_No State
80	4/29/2020	No State	Italy	203591	27682	71252	Italy_No State
153	4/29/2020	No State	Spain	236899	24275	132929	Spain_No State
265	4/29/2020	New York	US	299691	23477	0	US_New York

321 rows × 7 columns

Loading [MathJax]/extensions/Safe.js