In [2]:

```
import os,numpy as np,pandas as pd
```

In [3]:

```
os.chdir("D:\ML Project\sql\sql")
```

In [4]:

```
os.listdir()
```

Out[4]:

```
['afiedt.buf',
 'base_sqlproc_6992.pdf',
 'countries.csv',
 'merge.csv',
 'oilprod.csv',
 'oilrsrvs.csv',
 'postalcodes.csv',
 'sql class.sas',
 'Unitedstates.csv',
 'USCITYCOORDS.csv',
 'worldcitycoords.csv',
 'worldtemps.csv',
 'worldtemps.xlsx']
```

In [4]:

```
countiers = pd.read_csv("countries.csv")
```

In [5]:

```
countiers.shape
```

Out[5]:

```
(235, 6)
```

In [6]:

```
countiers.head()
```

Out[6]:

	Rank	Name	capital	Area	Population	Continent
0	1	Afghanistan	Kabul	647500	27755775	Asia
1	2	Albania	Tirana	28748	3544841	Europe
2	3	Algeria	Algiers	2381740	32277942	Africa
3	4	American Samoa	Pago Pago	199	68688	Polynesia
4	5	Andorra	Andorra la Vella	468	68403	Europe

In [7]:

```
## sorting by cloumns
countiers.sort_values(["Name","Continent"],ascending=False)
```

Out[7]:

	Rank	Name	capital	Area	Population	Continent
234	235	Zimbabwe	Harare	390580	11376676	Africa
233	234	Zambia	Lusaka	752614	9959037	Africa
232	233	Yugoslavia	Belgrade	102350	10656929	Asia
231	232	Yemen	Sanaa	527970	18701257	Asia
230	231	Western Sahara	none	266000	256177	Others
...	...	...	...	...	...	...
4	5	Andorra	Andorra la Vella	468	68403	Europe
3	4	American Samoa	Pago Pago	199	68688	Polynesia
2	3	Algeria	Algiers	2381740	32277942	Africa
1	2	Albania	Tirana	28748	3544841	Europe
0	1	Afghanistan	Kabul	647500	27755775	Asia

235 rows × 6 columns

In [8]:

```
## subseting with contition
df=countiers[(countiers["Population"]>5000000)].Name
```

In [9]:

```
df
```

Out[9]:

```
0       Afghanistan
2           Algeria
5            Angola
8         Argentina
11        Australia
            ...
227         Vietnam
231           Yemen
232      Yugoslavia
233          Zambia
234        Zimbabwe
Name: Name, Length: 111, dtype: object
```

In [10]:

```
df = countiers[["Population","Name"]]
```

In [11]:

```
df.sort_values(by=["Population"],ascending=True)
```

Out[11]:

	Population	Name
169	47	Pitcairn Islands
45	474	Christmas Island
46	632	Cocos Islands
94	900	Holy See (Vatican City)
210	1431	Tokelau
...	...	...
27	176029560	Brazil
99	231328092	Indonesia
222	280562489	United States
98	1045845226	India
42	1284303705	China

235 rows × 2 columns

In [12]:

```
## grouping and sorting
df2=countiers.groupby("Continent")["Population"].sum().sort_values(ascending=False)
```

In [13]:

```
df2
```

Out[13]:

```
Continent
Asia             4107401495
NC America        687478118
Europe            652375206
Africa            529126224
South America     146338433
Others             88228352
Australia          23454829
Polynesia             68688
Name: Population, dtype: int64
```

In [14]:

```
## grouping with particular values
df3=countiers[countiers["Continent"].isin(["Asia","Europe"])].groupby("Continent")["Populat
```

In [15]:

```
df3
```

Out[15]:

```
Continent
Asia      4107401495
Europe     652375206
Name: Population, dtype: int64
```

In [16]:

```
## sorting particluer column value with condition
df4 = countiers[(countiers["Continent"].isin(["Africa"]))&(countiers["Population"]>200000)]
```

In [17]:

```
df4
```

Out[17]:

	Rank	Name	capital	Area	Population	Continent
193	194	Solomon Islands	Honiara	28450	494786	Africa
201	202	Swaziland	Mbabane	17363	1123605	Africa
68	69	Estonia	Tallinn	45226	1415681	Africa
78	79	Gambia, The	Banjul	11300	1455842	Africa
146	147	Namibia	Windhoek	825418	1820916	Africa
117	118	Latvia	Riga	64589	2366515	Africa
161	162	Oman	Muscat	212460	2713462	Africa
39	40	Central African Republic	Bangui	622984	3642739	Africa
67	68	Eritrea	Asmara	121320	4465651	Africa
189	190	Sierra Leone	Freetown	71740	5614743	Africa
116	117	Laos	Vientiane	236800	5777180	Africa
194	195	Somalia	Mogadishu	637657	7753310	Africa
40	41	Chad	N'Djamena	1284000	8997237	Africa
233	234	Zambia	Lusaka	752614	9959037	Africa
5	6	Angola	Luanda	1246700	10593171	Africa
154	155	Niger	Niamey	1267000	10639744	Africa
234	235	Zimbabwe	Harare	390580	11376676	Africa
41	42	Chile	Santiago	756950	15498930	Africa
145	146	Mozambique	Maputo	801590	19607519	Africa
218	219	Uganda	Kampala	236040	24699073	Africa
110	111	Kenya	Nairobi	582650	31138735	Africa
2	3	Algeria	Algiers	2381740	32277942	Africa
198	199	Sudan	Khartoum	2505810	37090298	Africa
207	208	Tanzania	Dar es Salaam	945087	37187939	Africa
195	196	South Africa	Pretoria	1219912	43647658	Africa
69	70	Ethiopia	Addis Ababa	1127127	67673031	Africa
155	156	Nigeria	Abuja	923768	129934911	Africa

In [18]:

```
countiers["Density"]=countiers["Population"]/countiers["Area"].sort_values()
```

In [19]:

```
countiers
```

Out[19]:

	Rank	Name	capital	Area	Population	Continent	Density
0	1	Afghanistan	Kabul	647500	27755775	Asia	42.866062
1	2	Albania	Tirana	28748	3544841	Europe	123.307395
2	3	Algeria	Algiers	2381740	32277942	Africa	13.552253
3	4	American Samoa	Pago Pago	199	68688	Polynesia	345.165829
4	5	Andorra	Andorra la Vella	468	68403	Europe	146.160256
...	...	...	...	...	...	...	...
230	231	Western Sahara	none	266000	256177	Others	0.963071
231	232	Yemen	Sanaa	527970	18701257	Asia	35.421060
232	233	Yugoslavia	Belgrade	102350	10656929	Asia	104.122413
233	234	Zambia	Lusaka	752614	9959037	Africa	13.232596
234	235	Zimbabwe	Harare	390580	11376676	Africa	29.127646

235 rows × 7 columns

In [20]:

```
us = pd.read_csv("USCITYCOORDS.csv")
```

In [21]:

```
us.head()
```

Out[21]:

	State	city	latitude	Longitude
0	AL	Montgomery	22	-32
1	AK	Juneau	36	-65
2	AZ	Phoenix	54	-54
3	AR	LittleRock	43	-43
4	CA	Sacramento	48	-43

In [22]:

```
usc = us[["State","city"]]
```

In [23]:

```
usc
```

Out[23]:

	State	city
0	AL	Montgomery
1	AK	Juneau
2	AZ	Phoenix
3	AR	LittleRock
4	CA	Sacramento
5	CO	Denver
6	CT	Hartford
7	DE	Dover
8	FL	Tallahassee
9	GA	Atlanta
10	HI	Honolulu
11	ID	Boise
12	IL	Springfield
13	IN	Indianapolis
14	IA	Des Moines
15	KS	Topeka
16	KY	Frankfort
17	LA	Baton Rouge
18	ME	Augusta
19	MD	Annapolis
20	MA	Boston
21	MI	Lansing
22	MN	St.Paul
23	MS	Jackson
24	MO	Jefferson City
25	MT	Helena
26	NE	Lincoln
27	NV	Carson City
28	NH	Concord
29	NJ	Trenton
30	NM	Santa Fe
31	NY	Albany
32	NC	Raleigh
33	ND	Bismarck

	State	city
34	OH	Columbus
35	OK	Oklahoma City
36	OR	Salem
37	PA	Harrisburg
38	RI	Providence
39	SC	Columbia
40	SD	Pierre
41	TN	Nashville
42	TX	Austin
43	UT	Salt Lake City
44	VT	Montpelier
45	VA	Richmond
46	WA	Olympia
47	WV	Charleston
48	WI	Madison
49	WY	Cheyenne

In [24]:

```
unit = pd.read_csv("Unitedstates.csv")
```

In [25]:

```
unit.head()
```

Out[25]:

	Name	Capital	Population	Area	Continent
0	Alabama	Montgomery	4369862	50750	North America
1	Alaska	Juneau	619500	570373	North America
2	Arizona	Phoenix	4778332	113642	North America
3	Arkansas	LittleRock	2551373	52075	North America
4	California	Sacramento	33145121	155973	North America

In [26]:

```
unit.Continent.value_counts()
```

Out[26]:

```
North America    49
Oceania           1
Name: Continent, dtype: int64
```

In [27]:

```
unit.Continent.nunique()
```

Out[27]:

```
2
```

In [28]:

```
unit.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   Name        50 non-null     object
 1   Capital     50 non-null     object
 2   Population  50 non-null     int64 
 3   Area        50 non-null     int64 
 4   Continent   50 non-null     object
dtypes: int64(2), object(3)
memory usage: 2.1+ KB
```

In [29]:

```
unit.rename(columns={"Name":"State"})
```

Out[29]:

	State	Capital	Population	Area	Continent
0	Alabama	Montgomery	4369862	50750	North America
1	Alaska	Juneau	619500	570373	North America
2	Arizona	Phoenix	4778332	113642	North America
3	Arkansas	LittleRock	2551373	52075	North America
4	California	Sacramento	33145121	155973	North America
5	Colorado	Denver	4056133	103730	North America
6	Connecticut	Hartford	3282031	4845	North America
7	Delaware	Dover	753538	1955	North America
8	Florida	Tallahassee	15111244	53997	North America
9	Georgia	Atlanta	7788240	57919	North America
10	Hawaii	Honolulu	1185497	6423	Oceania
11	Idaho	Boise	1251700	82751	North America
12	Illinois	Springfield	12128370	55593	North America
13	Indiana	Indianapolis	5942901	35870	North America
14	Iowa	Des Moines	2869413	55875	North America
15	Kansas	Topeka	2654052	81823	North America
16	Kentucky	Frankfort	3960825	39732	North America
17	Louisiana	Baton Rouge	4372035	43566	North America
18	Maine	Augusta	1253040	30865	North America
19	Maryland	Annapolis	5171634	9775	North America
20	Massachusetts	Boston	6175169	7838	North America
21	Michigan	Lansing	9863775	56809	North America
22	Minnesota	St.Paul	4775508	79617	North America
23	Mississippi	Jackson	2768619	46914	North America
24	Missouri	Jefferson City	5468338	68898	North America
25	Montana	Helena	882779	145556	North America
26	Nebraska	Lincoln	1666028	76878	North America
27	Nevada	Carson City	1809253	109806	North America
28	New Hampshire	Concord	1201134	8969	North America
29	New Jersey	Trenton	8143412	7419	North America
30	New Mexico	Santa Fe	1739844	121364	North America
31	New York	Albany	18196601	47224	North America
32	North Carolina	Raleigh	7650789	48718	North America
33	North Dakota	Bismarck	633666	68994	North America

	State	Capital	Population	Area	Continent
34	Ohio	Columbus	11256654	40953	North America
35	Oklahoma	Oklahoma City	3358044	68679	North America
36	Oregon	Salem	3316154	96003	North America
37	Pennsylvania	Harrisburg	11994016	44820	North America
38	Rhode Island	Providence	990819	1045	North America
39	South Carolina	Columbia	3885736	30111	North America
40	South Dakota	Pierre	733133	75898	North America
41	Tennessee	Nashville	5483535	41220	North America
42	Texas	Austin	20044141	261914	North America
43	Utah	Salt Lake City	2129836	82168	North America
44	Vermont	Montpelier	593740	9249	North America
45	Virginia	Richmond	6872912	39598	North America
46	Washington	Olympia	5756361	66582	North America
47	West Virginia	Charleston	1806928	24087	North America
48	Wisconsin	Madison	5250446	54314	North America

In [30]:

```
post = pd.read_csv("postalcodes.csv")
```

In [31]:

```
post.head()
```

Out[31]:

	Name	Code
0	Alabama	AL
1	Alaska	AK
2	Arizona	AZ
3	Arkansas	AR
4	California	CA

In [32]:

```
world = pd.read_excel("worldtemps.xlsx")
```

In [33]:

```
world.head()
```

Out[33]:

	Country	City	AvgHigh	Avglow
0	Algeria	Algiers	19.0	7.0
1	Albania	Tirane	9.0	2.0
2	Australia	New South Wales	22.0	21.0
3	Belgium	Brussels	11.0	7.0
4	Botswana	Gaborone	30.0	22.0

In [34]:

```
world['low_celsius']=(world['Avglow']-32)*0.55
```

In [35]:

```
world['high_celsius']=(world['AvgHigh']-32)*0.55
```

In [36]:

```
world.head()
```

Out[36]:

	Country	City	AvgHigh	Avglow	low_celsius	high_celsius
0	Algeria	Algiers	19.0	7.0	-13.75	-7.15
1	Albania	Tirane	9.0	2.0	-16.50	-12.65
2	Australia	New South Wales	22.0	21.0	-6.05	-5.50
3	Belgium	Brussels	11.0	7.0	-13.75	-11.55
4	Botswana	Gaborone	30.0	22.0	-5.50	-1.10

In [37]:

```
world['Range']=world['high_celsius']-world['low_celsius']
```

In [38]:

```
world.head()
```

Out[38]:

	Country	City	AvgHigh	Avglow	low_celsius	high_celsius	Range
0	Algeria	Algiers	19.0	7.0	-13.75	-7.15	6.60
1	Albania	Tirane	9.0	2.0	-16.50	-12.65	3.85
2	Australia	New South Wales	22.0	21.0	-6.05	-5.50	0.55
3	Belgium	Brussels	11.0	7.0	-13.75	-11.55	2.20
4	Botswana	Gaborone	30.0	22.0	-5.50	-1.10	4.40

In [39]:

```
wrdtemp=world.groupby(["Country","City"]).agg({"AvgHigh":['mean'],'Avglow':['mean']}).reset
```

In [40]:

```
wrdtemp
```

Out[40]:

	Country	City	AvgHigh	Avglow
			mean	mean
0	Albania	Tirane	9.0	2.0
1	Algeria	Algiers	19.0	7.0
2	Australia	New South Wales	22.0	21.0
3	Belgium	Brussels	11.0	7.0
4	Botswana	Gaborone	30.0	22.0
5	Canada	Edmonton	-7.0	-16.0
6	Ethiopia	Addis Ababa	23.0	6.0
7	Finland	Helsinki	-3.0	-8.0
8	Finland	Oulu	-7.0	-13.0
9	France	Strasbourg	3.0	-1.0
10	France	Toulouse	8.0	1.0
11	Germany	Munich	2.0	-4.0
12	India	New Delhi	20.0	8.0
13	Indonesia	Jakarta	28.0	23.0
14	Indonesia	Medan	30.0	23.0
15	Ireland	Dublin	7.0	2.0
16	Ireland	Galway	7.0	2.0
17	Ireland	Shannon	8.0	3.0
18	Italy	Bozen/Bolzano	5.0	-5.0
19	Japan	Tokyo	8.0	1.0
20	Malta	Valletta	15.0	9.0
21	Norway	Oslo	0.0	-6.0
22	Romania	Constanta	3.0	-1.0
23	Singapore	Singapore	29.0	23.0
24	South Africa	Durban	27.0	22.0
25	South Africa	Pretoria	27.0	19.0
26	Spain	Madrid	10.0	0.0
27	Sweden	Göteborg	1.0	-1.0
28	Sweden	Jönköping	NaN	NaN
29	Uganda	Kampala	29.0	17.0
30	Zambia	Lusaka	25.0	20.0
31	Zimbabwe	Bulawayo	27.3	16.3
32	Zimbabwe	Harare	25.0	17.0

In [43]:

```
world_city=pd.read_csv("worldcitycoords.csv")
```

In [44]:

```
world_city
```

Out[44]:

	City	Country	Latitude	Longitude
0	Aberdeen	Scotland	57	2
1	Adelaide	Australia	34	138
2	Algiers	Algeria	36	3
3	Amsterdam	Netherlands	52	4
4	Ankara	Turkey	39	32
...	...	...	...	...
111	Vienna	Austria	48	16
112	Vladivostok	Russia	43	132
113	Warsaw	Poland	52	21
114	Wellington	New Zealand	41	174
115	Z�rich	Switzerland	47	8

116 rows × 4 columns

In [45]:

```
def sal_cat(row):
    if row['Latitude']>67:
        return 'North Frigid'
    elif row['Latitude']<=67 and row ['Latitude']>=23:
        return 'North Temperate'
    elif row['Latitude']<23 and row ['Latitude']>-23:
        return 'Torrid'
    else:
        return 'South Frigid'
```

In [46]:

```
world_city['Latitude2']=world_city.apply(sal_cat,axis=1)
```

In [47]:

```
world_city
```

Out[47]:

	City	Country	Latitude	Longitude	Latitude2
0	Aberdeen	Scotland	57	2	North Temperate
1	Adelaide	Australia	34	138	North Temperate
2	Algiers	Algeria	36	3	North Temperate
3	Amsterdam	Netherlands	52	4	North Temperate
4	Ankara	Turkey	39	32	North Temperate
...	...	...	...	...	...
111	Vienna	Austria	48	16	North Temperate
112	Vladivostok	Russia	43	132	North Temperate
113	Warsaw	Poland	52	21	North Temperate
114	Wellington	New Zealand	41	174	North Temperate
115	Z�rich	Switzerland	47	8	North Temperate

116 rows × 5 columns

merging

In [48]:

```
oil = pd.read_csv("oilprod.csv")
```

In [49]:

```
oil.head()
```

Out[49]:

	Country	BarrelsPerDay
0	Algeria	1400000
1	Canada	2500000
2	China	3000000
3	Egypt	900000
4	Indonesia	1500000

In [50]:

```
oilr = pd.read_csv("oilrsrvs.csv")
```

In [51]:

```
oilr.head()
```

Out[51]:

	Country	Barrels
0	Algeria	9.200000e+09
1	Canada	7.000000e+09
2	China	2.500000e+10
3	Egypt	4.000000e+09
4	Indonesia	5.000000e+09

In [52]:

```
merge = pd.merge(oil,oilr,on="Country",how="inner")
```

In [53]:

```
merge
```

Out[53]:

	Country	BarrelsPerDay	Barrels
0	Algeria	1400000	9.200000e+09
1	Canada	2500000	7.000000e+09
2	China	3000000	2.500000e+10
3	Egypt	900000	4.000000e+09
4	Indonesia	1500000	5.000000e+09
5	Iran	4000000	9.000000e+10
6	Iraq	600000	1.100000e+11
7	Kuwait	2500000	9.500000e+10
8	Libya	1500000	3.000000e+10
9	Mexico	3400000	5.000000e+10
10	Nigeria	2000000	1.600000e+10
11	Norway	3500000	1.100000e+10
12	Oman	900000	1.600000e+10
13	Saudi Arabia	9000000	2.600000e+11
14	United States of America	8000000	1.000000e+09

In [54]:

```
oilp=oil.rename(columns={'BarrelsPerDay':'Production'})
```

In [55]:

```
oilrr=oilr.rename(columns={'Barrels':'Reserves'})
```

In [56]:

```
merge1 = pd.merge(oilp,oilrr,on="Country",how="inner")
```

In [57]:

```
merge1.sort_values(by="Production",ascending=False)
```

Out[57]:

	Country	Production	Reserves
13	Saudi Arabia	9000000	2.600000e+11
14	United States of America	8000000	1.000000e+09
5	Iran	4000000	9.000000e+10
11	Norway	3500000	1.100000e+10
9	Mexico	3400000	5.000000e+10
2	China	3000000	2.500000e+10
1	Canada	2500000	7.000000e+09
7	Kuwait	2500000	9.500000e+10
10	Nigeria	2000000	1.600000e+10
4	Indonesia	1500000	5.000000e+09
8	Libya	1500000	3.000000e+10
0	Algeria	1400000	9.200000e+09
3	Egypt	900000	4.000000e+09
12	Oman	900000	1.600000e+10
6	Iraq	600000	1.100000e+11

In [58]:

```
us = pd.read_csv('USCITYCOORDS.csv')
us.head()
```

Out[58]:

	State	city	latitude	Longitude
0	AL	Montgomery	22	-32
1	AK	Juneau	36	-65
2	AZ	Phoenix	54	-54
3	AR	LittleRock	43	-43
4	CA	Sacramento	48	-43

In [59]:

```
world = pd.read_csv('worldcitycoords.csv')
world.head()
```

Out[59]:

	City	Country	Latitude	Longitude
0	Aberdeen	Scotland	57	2
1	Adelaide	Australia	34	138
2	Algiers	Algeria	36	3
3	Amsterdam	Netherlands	52	4
4	Ankara	Turkey	39	32

In [60]:

```
mergeu = pd.merge(us[["city","latitude","State"]],world[["City","Latitude"]],how='outer',le
```

In [61]:

```
mergeu
```

Out[61]:

	city	latitude	State	City	Latitude
0	Montgomery	22.0	AL	Calcutta	22.0
1	Montgomery	22.0	AL	Hong Kong	22.0
2	Montgomery	22.0	AL	Rio de Janeiro	22.0
3	Concord	22.0	NH	Calcutta	22.0
4	Concord	22.0	NH	Hong Kong	22.0
...	...	...	...	...	...
189	NaN	NaN	NaN	Stockholm	59.0
190	NaN	NaN	NaN	Panama City	8.0
191	NaN	NaN	NaN	Paramaribo	5.0
192	NaN	NaN	NaN	Reykjav�k	64.0
193	NaN	NaN	NaN	Tananarive	18.0

194 rows × 5 columns

In [62]:

```
mergeu.to_csv("merge.csv")
```

In [63]:

```
mergeed=mergeu[(mergeu["City"]=="Cairo")
               |(mergeu["latitude"]<mergeu["Latitude"])]
```

In [64]:

```
mergeed
```

Out[64]:

	city	latitude	State	City	Latitude
147	NaN	NaN	NaN	Cairo	30.0

In [65]:

```
countries = pd.read_csv('countries.csv')
```

In [66]:

```
countries.head()
```

Out[66]:

	Rank	Name	capital	Area	Population	Continent
0	1	Afghanistan	Kabul	647500	27755775	Asia
1	2	Albania	Tirana	28748	3544841	Europe
2	3	Algeria	Algiers	2381740	32277942	Africa
3	4	American Samoa	Pago Pago	199	68688	Polynesia
4	5	Andorra	Andorra la Vella	468	68403	Europe

In [67]:

```
country = pd.merge(countries[["Name","capital"]],world,how="inner",left_on="Name",right_on=
```

In [68]:

```
country
```

Out[68]:

	Name	capital	City	Country	Latitude	Longitude
0	Algeria	Algiers	Algiers	Algeria	36	3
1	Argentina	Buenos Aires	Buenos	Argentina	34	58
2	Argentina	Buenos Aires	C�rdoba	Argentina	31	64
3	Australia	Canberra	Adelaide	Australia	34	138
4	Australia	Canberra	Brisbane	Australia	27	153
...	...	...	...	...	...	...
95	Thailand	Bangkok	Bangkok	Thailand	13	100
96	Turkey	Ankara	Ankara	Turkey	39	32
97	Ukraine	Kiev	Odessa	Ukraine	46	30
98	Uruguay	Montevideo	Montevideo	Uruguay	34	56
99	Venezuela	Caracas	Caracas	Venezuela	10	67

100 rows × 6 columns

In [69]:

```
country1 = country[(country["capital"].str.startswith('L'))&(country["capital"]=='City')&(c
```

In [70]:

```
country1
```

Out[70]:

Name	capital	City	Country	Latitude	Longitude

In [71]:

```
post = pd.read_csv('postalcodes.csv')
```

In [72]:

```
post.shape
```

Out[72]:

(50, 2)

In [73]:

```
uscity = pd.read_csv('Unitedstates.csv')
```

In [74]:

```
uscity.shape
```

Out[74]:

```
(50, 5)
```

In [75]:

```
us.shape
```

Out[75]:

```
(50, 4)
```

In [76]:

```
merge3 = pd.merge(us,uscity,how='left',left_on='city',right_on="Capital").merge(post,how="l
```

In [77]:

```
merge3.shape
```

Out[77]:

```
(50, 10)
```

In [78]:

```
merge4 = pd.merge(us,uscity,how='inner',left_on='city',right_on="Capital").merge(post,how="
```

In [79]:

```
merge4.shape
```

Out[79]:

```
(50, 10)
```

In [85]:

```
world.head()
```

Out[85]:

	City	Country	Latitude	Longitude
0	Aberdeen	Scotland	57	2
1	Adelaide	Australia	34	138
2	Algiers	Algeria	36	3
3	Amsterdam	Netherlands	52	4
4	Ankara	Turkey	39	32

In [86]:

```
countries.head()
```

Out[86]:

	Rank	Name	capital	Area	Population	Continent
0	1	Afghanistan	Kabul	647500	27755775	Asia
1	2	Albania	Tirana	28748	3544841	Europe
2	3	Algeria	Algiers	2381740	32277942	Africa
3	4	American Samoa	Pago Pago	199	68688	Polynesia
4	5	Andorra	Andorra la Vella	468	68403	Europe

In [ ]:

```
wc = pd.merge()
```

In [81]:

```
temp = pd.read_csv('worldtemps.csv')
```

In [82]:

```
temp
```

Out[82]:

	Country	City	AvgHigh	Avglow
0	Algeria	Algiers	19.0	7.0
1	Albania	Tirane	9.0	2.0
2	Australia	New South Wales	22.0	21.0
3	Belgium	Brussels	11.0	7.0
4	Botswana	Gaborone	30.0	22.0
5	Canada	Edmonton	-7.0	-16.0
6	Ethiopia	Addis Ababa	23.0	6.0
7	Finland	Helsinki	-3.0	-8.0
8	Finland	Oulu	-7.0	-13.0
9	France	Strasbourg	3.0	-1.0
10	France	Toulouse	8.0	1.0
11	Germany	Munich	2.0	-4.0
12	India	New Delhi	20.0	8.0
13	Indonesia	Jakarta	28.0	23.0
14	Indonesia	Medan	30.0	23.0
15	Ireland	Dublin	7.0	2.0
16	Ireland	Galway	7.0	2.0
17	Ireland	Shannon	8.0	3.0
18	Italy	Bozen/Bolzano	5.0	-5.0
19	Japan	Tokyo	8.0	1.0
20	Malta	Valletta	15.0	9.0
21	Norway	Oslo	0.0	-6.0
22	Romania	Constanta	3.0	-1.0
23	Singapore	Singapore	29.0	23.0
24	South Africa	Durban	27.0	22.0
25	South Africa	Pretoria	27.0	19.0
26	Spain	Madrid	10.0	0.0
27	Sweden	J�nk�ping	NaN	NaN
28	Sweden	G�teborg	1.0	-1.0
29	Uganda	Kampala	29.0	17.0
30	Zambia	Lusaka	25.0	20.0
31	Zimbabwe	Bulawayo	27.3	16.3
32	Zimbabwe	Harare	25.0	17.0

In [83]:

```
temp1 = pd.read_excel('worldtemps.xlsx')
```

In [84]:

```
temp1
```

9	France	Strasbourg	3.0	-1.0
10	France	Toulouse	8.0	1.0
11	Germany	Munich	2.0	-4.0
12	India	New Delhi	20.0	8.0
13	Indonesia	Jakarta	28.0	23.0
14	Indonesia	Medan	30.0	23.0
15	Ireland	Dublin	7.0	2.0
16	Ireland	Galway	7.0	2.0
17	Ireland	Shannon	8.0	3.0
18	Italy	Bozen/Bolzano	5.0	-5.0
19	Japan	Tokyo	8.0	1.0
20	Malta	Valletta	15.0	9.0
21	Norway	Oslo	0.0	-6.0

In [5]:

```
os.chdir("D:\ML Project\data")
```

In [6]:

```
kan = pd.read_sas("kannada.sas7bdat",encoding='latin-1')
eng = pd.read_sas("english.sas7bdat",encoding='latin-1')
bio=pd.read_sas("biology.sas7bdat",encoding='latin-1')
maths=pd.read_sas("maths.sas7bdat",encoding='latin-1')
chem=pd.read_sas("chemistry.sas7bdat",encoding='latin-1')
phy=pd.read_sas("physics.sas7bdat",encoding='latin-1')
```

In [7]:

```
merg=[kan,eng,bio,maths,chem,phy]
```

In [8]:

```
from functools import reduce
```

In [9]:

```
df = reduce(lambda x,y:pd.merge(x,y,how='inner',on='student_id'),merg)
```

In [10]:

```
df
```

Out[10]:

	student_id	k_marks	E_marks	b_marks	m_marks	c_marks	p_marks
0	1.0	73.0	84.0	84.0	72.0	83.0	82.0
1	2.0	54.0	75.0	75.0	44.0	64.0	54.0
2	3.0	59.0	77.0	77.0	51.0	69.0	61.0
3	4.0	44.0	70.0	70.0	29.0	54.0	39.0
4	5.0	77.0	86.0	86.0	79.0	87.0	89.0
...	...	...	...	...	...	...	...
495	496.0	54.0	74.0	74.0	43.0	64.0	53.0
496	497.0	61.0	78.0	78.0	54.0	71.0	64.0
497	498.0	45.0	70.0	70.0	30.0	55.0	40.0
498	499.0	53.0	74.0	74.0	42.0	63.0	52.0
499	500.0	73.0	84.0	84.0	71.0	83.0	81.0

500 rows × 7 columns

In [ ]: