Problem Statement:-

Perform Clustering(Hierarchical, Kmeans & DBSCAN) for the crime data and identify the number of clusters formed and draw inferences.

Hierarchical Clustering Method

1. Import Neccesary Libraries

In [1]:
```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
```

2. Import Data

In [2]:
```
Crime_Data = pd.read_csv('crime_data.csv')
Crime_Data
```

Out[2]:

	Unnamed: 0	Murder	Assault	UrbanPop	Rape
0	Alabama	13.2	236	58	21.2
1	Alaska	10.0	263	48	44.5
2	Arizona	8.1	294	80	31.0
3	Arkansas	8.8	190	50	19.5
4	California	9.0	276	91	40.6
5	Colorado	7.9	204	78	38.7
6	Connecticut	3.3	110	77	11.1
7	Delaware	5.9	238	72	15.8
8	Florida	15.4	335	80	31.9
9	Georgia	17.4	211	60	25.8
10	Hawaii	5.3	46	83	20.2
11	Idaho	2.6	120	54	14.2
12	Illinois	10.4	249	83	24.0
13	Indiana	7.2	113	65	21.0
14	Iowa	2.2	56	57	11.3
15	Kansas	6.0	115	66	18.0
16	Kentucky	9.7	109	52	16.3
17	Louisiana	15.4	249	66	22.2
18	Maine	2.1	83	51	7.8
19	Maryland	11.3	300	67	27.8
20	Massachusetts	4.4	149	85	16.3
21	Michigan	12.1	255	74	35.1
22	Minnesota	2.7	72	66	14.9
23	Mississippi	16.1	259	44	17.1
24	Missouri	9.0	178	70	28.2
25	Montana	6.0	109	53	16.4
26	Nebraska	4.3	102	62	16.5
27	Nevada	12.2	252	81	46.0
28	New Hampshire	2.1	57	56	9.5
29	New Jersey	7.4	159	89	18.8
30	New Mexico	11.4	285	70	32.1
31	New York	11.1	254	86	26.1
32	North Carolina	13.0	337	45	16.1
33	North Dakota	0.8	45	44	7.3

	Unnamed: 0	Murder	Assault	UrbanPop	Rape
34	Ohio	7.3	120	75	21.4
35	Oklahoma	6.6	151	68	20.0
36	Oregon	4.9	159	67	29.3
37	Pennsylvania	6.3	106	72	14.9
38	Rhode Island	3.4	174	87	8.3
39	South Carolina	14.4	279	48	22.5
40	South Dakota	3.8	86	45	12.8
41	Tennessee	13.2	188	59	26.9
42	Texas	12.7	201	80	25.5
43	Utah	3.2	120	80	22.9
44	Vermont	2.2	48	32	11.2
45	Virginia	8.5	156	63	20.7
46	Washington	4.0	145	73	26.2
47	West Virginia	5.7	81	39	9.3
48	Wisconsin	2.6	53	66	10.8
49	Wyoming	6.8	161	60	15.6

3. Data Understanding

In [3]:
```
Crime_Data.shape
```

Out[3]:
```
(50, 5)
```

In [4]:
```
Crime_Data.dtypes
```

Out[4]:
```
Unnamed: 0     object
Murder        float64
Assault         int64
UrbanPop        int64
Rape          float64
dtype: object
```

In [5]:
```
Crime_Data.describe()
```

Out[5]:

	Murder	Assault	UrbanPop	Rape
count	50.00000	50.000000	50.000000	50.000000
mean	7.78800	170.760000	65.540000	21.232000
std	4.35551	83.337661	14.474763	9.366385
min	0.80000	45.000000	32.000000	7.300000
25%	4.07500	109.000000	54.500000	15.075000
50%	7.25000	159.000000	66.000000	20.100000
75%	11.25000	249.000000	77.750000	26.175000
max	17.40000	337.000000	91.000000	46.000000

In [6]:
```
Crime_Data.isna().sum()
```

Out[6]:
```
Unnamed: 0    0
Murder        0
Assault       0
UrbanPop      0
Rape          0
dtype: int64
```

4. Data Preparation

In [17]:
```
del Crime_Data['Unnamed: 0']
```

In [18]:
```
Crime_Data
```

Out[18]:

	Murder	Assault	UrbanPop	Rape
0	13.2	236	58	21.2
1	10.0	263	48	44.5
2	8.1	294	80	31.0
3	8.8	190	50	19.5
4	9.0	276	91	40.6
5	7.9	204	78	38.7
6	3.3	110	77	11.1
7	5.9	238	72	15.8
8	15.4	335	80	31.9
9	17.4	211	60	25.8
10	5.3	46	83	20.2
11	2.6	120	54	14.2
12	10.4	249	83	24.0
13	7.2	113	65	21.0
14	2.2	56	57	11.3
15	6.0	115	66	18.0
16	9.7	109	52	16.3
17	15.4	249	66	22.2
18	2.1	83	51	7.8
19	11.3	300	67	27.8
20	4.4	149	85	16.3
21	12.1	255	74	35.1
22	2.7	72	66	14.9
23	16.1	259	44	17.1
24	9.0	178	70	28.2
25	6.0	109	53	16.4
26	4.3	102	62	16.5
27	12.2	252	81	46.0
28	2.1	57	56	9.5
29	7.4	159	89	18.8
30	11.4	285	70	32.1
31	11.1	254	86	26.1
32	13.0	337	45	16.1
33	0.8	45	44	7.3
34	7.3	120	75	21.4

	Murder	Assault	UrbanPop	Rape
35	6.6	151	68	20.0
36	4.9	159	67	29.3
37	6.3	106	72	14.9
38	3.4	174	87	8.3
39	14.4	279	48	22.5
40	3.8	86	45	12.8
41	13.2	188	59	26.9
42	12.7	201	80	25.5
43	3.2	120	80	22.9
44	2.2	48	32	11.2
45	8.5	156	63	20.7
46	4.0	145	73	26.2
47	5.7	81	39	9.3
48	2.6	53	66	10.8
49	6.8	161	60	15.6

In [19]:
```
Crime_Data.isna().sum()
```

Out[19]:
```
Murder      0
Assault     0
UrbanPop    0
Rape        0
dtype: int64
```

In [20]:
```
Crime_Data.shape
```

Out[20]:
```
(50, 4)
```

Normalization

In [21]:
```
def norm_func(i):
    x = (i-i.min())/(i.max()-i.min())
    return x
```

In [22]:
```
Crime_Data=norm_func(Crime_Data)
Crime_Data
```

Out[22]:

	Murder	Assault	UrbanPop	Rape
0	0.746988	0.654110	0.440678	0.359173
1	0.554217	0.746575	0.271186	0.961240
2	0.439759	0.852740	0.813559	0.612403
3	0.481928	0.496575	0.305085	0.315245
4	0.493976	0.791096	1.000000	0.860465
5	0.427711	0.544521	0.779661	0.811370
6	0.150602	0.222603	0.762712	0.098191
7	0.307229	0.660959	0.677966	0.219638
8	0.879518	0.993151	0.813559	0.635659
9	1.000000	0.568493	0.474576	0.478036
10	0.271084	0.003425	0.864407	0.333333
11	0.108434	0.256849	0.372881	0.178295
12	0.578313	0.698630	0.864407	0.431525
13	0.385542	0.232877	0.559322	0.354005
14	0.084337	0.037671	0.423729	0.103359
15	0.313253	0.239726	0.576271	0.276486
16	0.536145	0.219178	0.338983	0.232558
17	0.879518	0.698630	0.576271	0.385013
18	0.078313	0.130137	0.322034	0.012920
19	0.632530	0.873288	0.593220	0.529716
20	0.216867	0.356164	0.898305	0.232558
21	0.680723	0.719178	0.711864	0.718346
22	0.114458	0.092466	0.576271	0.196382
23	0.921687	0.732877	0.203390	0.253230
24	0.493976	0.455479	0.644068	0.540052
25	0.313253	0.219178	0.355932	0.235142
26	0.210843	0.195205	0.508475	0.237726
27	0.686747	0.708904	0.830508	1.000000
28	0.078313	0.041096	0.406780	0.056848
29	0.397590	0.390411	0.966102	0.297158
30	0.638554	0.821918	0.644068	0.640827
31	0.620482	0.715753	0.915254	0.485788
32	0.734940	1.000000	0.220339	0.227390
33	0.000000	0.000000	0.203390	0.000000

	Murder	Assault	UrbanPop	Rape
34	0.391566	0.256849	0.728814	0.364341
35	0.349398	0.363014	0.610169	0.328165
36	0.246988	0.390411	0.593220	0.568475
37	0.331325	0.208904	0.677966	0.196382
38	0.156627	0.441781	0.932203	0.025840
39	0.819277	0.801370	0.271186	0.392765
40	0.180723	0.140411	0.220339	0.142119
41	0.746988	0.489726	0.457627	0.506460
42	0.716867	0.534247	0.813559	0.470284
43	0.144578	0.256849	0.813559	0.403101
44	0.084337	0.010274	0.000000	0.100775
45	0.463855	0.380137	0.525424	0.346253
46	0.192771	0.342466	0.694915	0.488372
47	0.295181	0.123288	0.118644	0.051680
48	0.108434	0.027397	0.576271	0.090439
49	0.361446	0.397260	0.474576	0.214470

In [23]:
```
Crime_Data.mean()
```

Out[23]:
```
Murder      0.420964
Assault     0.430685
UrbanPop    0.568475
Rape        0.360000
dtype: float64
```

In [24]:
```
Crime_Data.std()
```

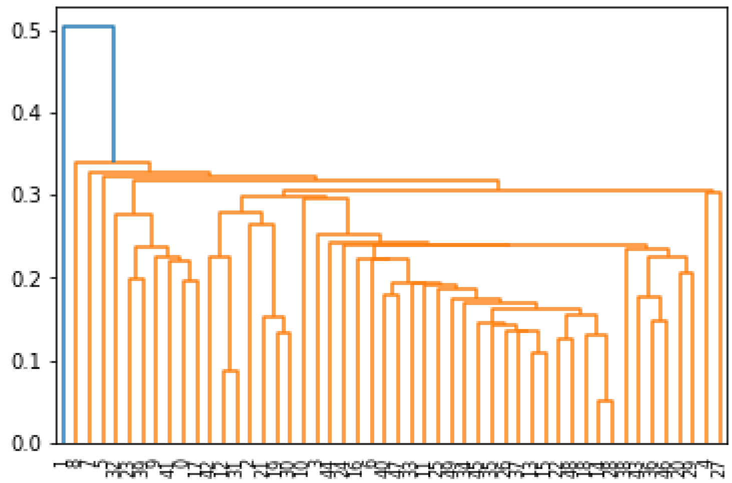
Out[24]:
```
Murder      0.262380
Assault     0.285403
UrbanPop    0.245335
Rape        0.242025
dtype: float64
```

Dendogram With All The Linkages

In [26]:
```
import scipy.cluster.hierarchy as sch
from sklearn.cluster import AgglomerativeClustering
```

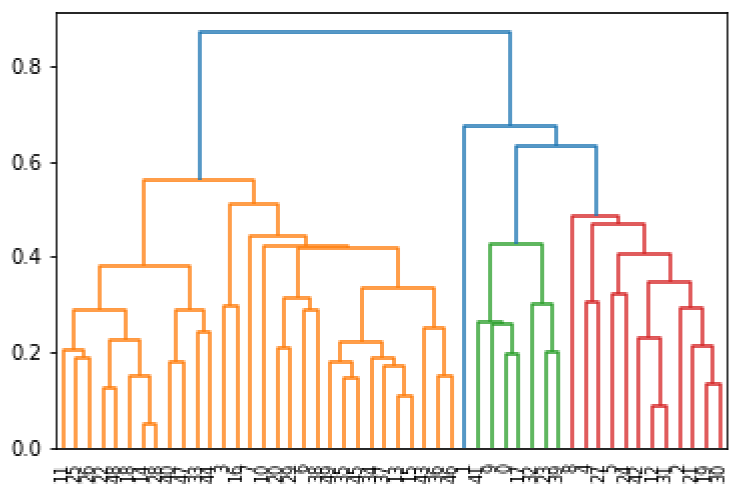
1. Single Linkage

In [27]:
```
Dendogram = sch.dendrogram(sch.linkage(Crime_Data, method = 'single'))
```



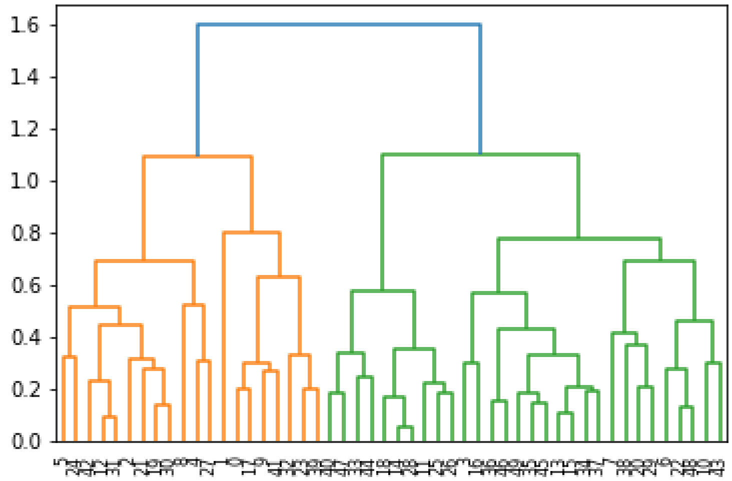
2. Average Linkage

In [28]:
```
Dendogram = sch.dendrogram(sch.linkage(Crime_Data, method = 'average'))
```



3. Complete Linkage

In [29]:
```
Dendogram = sch.dendrogram(sch.linkage(Crime_Data, method = 'complete'))
```



5. Model Building ¶

In [30]:
```
Hc = AgglomerativeClustering(n_clusters=4,  affinity='euclidean', linkage='single
Hc
```

Out[30]:
```
AgglomerativeClustering(linkage='single', n_clusters=4)
```

6. Model training

In [31]:
```
AgglomerativeClustering=Hc.fit_predict(Crime_Data)
AgglomerativeClustering
```

Out[31]:
```
array([0, 3, 0, 0, 0, 0, 0, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0], dtype=int64)
```

In [32]:
```
pd.DataFrame(AgglomerativeClustering)
```

Out[32]:

	0
0	0
1	3
2	0
3	0
4	0
5	0
6	0
7	1
8	2
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0
26	0
27	0
28	0
29	0
30	0
31	0
32	0
33	0
34	0

	0
35	0
36	0
37	0
38	0
39	0
40	0
41	0
42	0
43	0
44	0
45	0
46	0
47	0
48	0
49	0

In [33]:
```
Hc_Clusters_Data = pd.DataFrame(AgglomerativeClustering, columns = ['Clusters'])
Hc_Clusters_Data
```

Out[33]:

	Clusters
0	0
1	3
2	0
3	0
4	0
5	0
6	0
7	1
8	2
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0
26	0
27	0
28	0
29	0
30	0
31	0
32	0
33	0

	Clusters
34	0
35	0
36	0
37	0
38	0
39	0
40	0
41	0
42	0
43	0
44	0
45	0
46	0
47	0
48	0
49	0

In [34]:
```
Crime_Data['Clusters']=Hc_Clusters_Data['Clusters']
Crime_Data.head(50)
```

Out[34]:

	Murder	Assault	UrbanPop	Rape	Clusters
0	0.746988	0.654110	0.440678	0.359173	0
1	0.554217	0.746575	0.271186	0.961240	3
2	0.439759	0.852740	0.813559	0.612403	0
3	0.481928	0.496575	0.305085	0.315245	0
4	0.493976	0.791096	1.000000	0.860465	0
5	0.427711	0.544521	0.779661	0.811370	0
6	0.150602	0.222603	0.762712	0.098191	0
7	0.307229	0.660959	0.677966	0.219638	1
8	0.879518	0.993151	0.813559	0.635659	2
9	1.000000	0.568493	0.474576	0.478036	0
10	0.271084	0.003425	0.864407	0.333333	0
11	0.108434	0.256849	0.372881	0.178295	0
12	0.578313	0.698630	0.864407	0.431525	0
13	0.385542	0.232877	0.559322	0.354005	0
14	0.084337	0.037671	0.423729	0.103359	0
15	0.313253	0.239726	0.576271	0.276486	0
16	0.536145	0.219178	0.338983	0.232558	0
17	0.879518	0.698630	0.576271	0.385013	0
18	0.078313	0.130137	0.322034	0.012920	0
19	0.632530	0.873288	0.593220	0.529716	0
20	0.216867	0.356164	0.898305	0.232558	0
21	0.680723	0.719178	0.711864	0.718346	0
22	0.114458	0.092466	0.576271	0.196382	0
23	0.921687	0.732877	0.203390	0.253230	0
24	0.493976	0.455479	0.644068	0.540052	0
25	0.313253	0.219178	0.355932	0.235142	0
26	0.210843	0.195205	0.508475	0.237726	0
27	0.686747	0.708904	0.830508	1.000000	0
28	0.078313	0.041096	0.406780	0.056848	0
29	0.397590	0.390411	0.966102	0.297158	0
30	0.638554	0.821918	0.644068	0.640827	0
31	0.620482	0.715753	0.915254	0.485788	0
32	0.734940	1.000000	0.220339	0.227390	0
33	0.000000	0.000000	0.203390	0.000000	0

	Murder	Assault	UrbanPop	Rape	Clusters
34	0.391566	0.256849	0.728814	0.364341	0
35	0.349398	0.363014	0.610169	0.328165	0
36	0.246988	0.390411	0.593220	0.568475	0
37	0.331325	0.208904	0.677966	0.196382	0
38	0.156627	0.441781	0.932203	0.025840	0
39	0.819277	0.801370	0.271186	0.392765	0
40	0.180723	0.140411	0.220339	0.142119	0
41	0.746988	0.489726	0.457627	0.506460	0
42	0.716867	0.534247	0.813559	0.470284	0
43	0.144578	0.256849	0.813559	0.403101	0
44	0.084337	0.010274	0.000000	0.100775	0
45	0.463855	0.380137	0.525424	0.346253	0
46	0.192771	0.342466	0.694915	0.488372	0
47	0.295181	0.123288	0.118644	0.051680	0
48	0.108434	0.027397	0.576271	0.090439	0
49	0.361446	0.397260	0.474576	0.214470	0

1. Hierarchical Clustering with cluster '0'

In [35]:
```
Crime_Data[Crime_Data['Clusters']==0]
```

Out[35]:

	Murder	Assault	UrbanPop	Rape	Clusters
0	0.746988	0.654110	0.440678	0.359173	0
2	0.439759	0.852740	0.813559	0.612403	0
3	0.481928	0.496575	0.305085	0.315245	0
4	0.493976	0.791096	1.000000	0.860465	0
5	0.427711	0.544521	0.779661	0.811370	0
6	0.150602	0.222603	0.762712	0.098191	0
9	1.000000	0.568493	0.474576	0.478036	0
10	0.271084	0.003425	0.864407	0.333333	0
11	0.108434	0.256849	0.372881	0.178295	0
12	0.578313	0.698630	0.864407	0.431525	0
13	0.385542	0.232877	0.559322	0.354005	0
14	0.084337	0.037671	0.423729	0.103359	0
15	0.313253	0.239726	0.576271	0.276486	0
16	0.536145	0.219178	0.338983	0.232558	0
17	0.879518	0.698630	0.576271	0.385013	0
18	0.078313	0.130137	0.322034	0.012920	0
19	0.632530	0.873288	0.593220	0.529716	0
20	0.216867	0.356164	0.898305	0.232558	0
21	0.680723	0.719178	0.711864	0.718346	0
22	0.114458	0.092466	0.576271	0.196382	0
23	0.921687	0.732877	0.203390	0.253230	0
24	0.493976	0.455479	0.644068	0.540052	0
25	0.313253	0.219178	0.355932	0.235142	0
26	0.210843	0.195205	0.508475	0.237726	0
27	0.686747	0.708904	0.830508	1.000000	0
28	0.078313	0.041096	0.406780	0.056848	0
29	0.397590	0.390411	0.966102	0.297158	0
30	0.638554	0.821918	0.644068	0.640827	0
31	0.620482	0.715753	0.915254	0.485788	0
32	0.734940	1.000000	0.220339	0.227390	0
33	0.000000	0.000000	0.203390	0.000000	0
34	0.391566	0.256849	0.728814	0.364341	0
35	0.349398	0.363014	0.610169	0.328165	0
36	0.246988	0.390411	0.593220	0.568475	0
37	0.331325	0.208904	0.677966	0.196382	0

	Murder	Assault	UrbanPop	Rape	Clusters
38	0.156627	0.441781	0.932203	0.025840	0
39	0.819277	0.801370	0.271186	0.392765	0
40	0.180723	0.140411	0.220339	0.142119	0
41	0.746988	0.489726	0.457627	0.506460	0
42	0.716867	0.534247	0.813559	0.470284	0
43	0.144578	0.256849	0.813559	0.403101	0
44	0.084337	0.010274	0.000000	0.100775	0
45	0.463855	0.380137	0.525424	0.346253	0
46	0.192771	0.342466	0.694915	0.488372	0
47	0.295181	0.123288	0.118644	0.051680	0
48	0.108434	0.027397	0.576271	0.090439	0
49	0.361446	0.397260	0.474576	0.214470	0

2. Hierarchical Clustering with cluster '1'

In [36]:
```
Crime_Data[Crime_Data['Clusters']==1]
```

Out[36]:

	Murder	Assault	UrbanPop	Rape	Clusters
7	0.307229	0.660959	0.677966	0.219638	1

3. Hierarchical Clustering with cluster '2'

In [37]:
```
Crime_Data[Crime_Data['Clusters']==2]
```

Out[37]:

	Murder	Assault	UrbanPop	Rape	Clusters
8	0.879518	0.993151	0.813559	0.635659	2

4. Hierarchical Clustering with cluster '3'

In [38]:
```
Crime_Data[Crime_Data['Clusters']==3]
```

Out[38]:

	Murder	Assault	UrbanPop	Rape	Clusters
1	0.554217	0.746575	0.271186	0.96124	3

** Hierarchical Clustering Ends **

K Means Clustering Method

In [39]:
```
Crime_Data_Kmeans = pd.read_csv('crime_data.csv')
Crime_Data_Kmeans
```

Out[39]:

	Unnamed: 0	Murder	Assault	UrbanPop	Rape
0	Alabama	13.2	236	58	21.2
1	Alaska	10.0	263	48	44.5
2	Arizona	8.1	294	80	31.0
3	Arkansas	8.8	190	50	19.5
4	California	9.0	276	91	40.6
5	Colorado	7.9	204	78	38.7
6	Connecticut	3.3	110	77	11.1
7	Delaware	5.9	238	72	15.8
8	Florida	15.4	335	80	31.9
9	Georgia	17.4	211	60	25.8
10	Hawaii	5.3	46	83	20.2
11	Idaho	2.6	120	54	14.2
12	Illinois	10.4	249	83	24.0
13	Indiana	7.2	113	65	21.0
14	Iowa	2.2	56	57	11.3
15	Kansas	6.0	115	66	18.0
16	Kentucky	9.7	109	52	16.3
17	Louisiana	15.4	249	66	22.2
18	Maine	2.1	83	51	7.8
19	Maryland	11.3	300	67	27.8
20	Massachusetts	4.4	149	85	16.3
21	Michigan	12.1	255	74	35.1
22	Minnesota	2.7	72	66	14.9
23	Mississippi	16.1	259	44	17.1
24	Missouri	9.0	178	70	28.2
25	Montana	6.0	109	53	16.4
26	Nebraska	4.3	102	62	16.5
27	Nevada	12.2	252	81	46.0
28	New Hampshire	2.1	57	56	9.5
29	New Jersey	7.4	159	89	18.8
30	New Mexico	11.4	285	70	32.1
31	New York	11.1	254	86	26.1
32	North Carolina	13.0	337	45	16.1
33	North Dakota	0.8	45	44	7.3

	Unnamed: 0	Murder	Assault	UrbanPop	Rape
34	Ohio	7.3	120	75	21.4
35	Oklahoma	6.6	151	68	20.0
36	Oregon	4.9	159	67	29.3
37	Pennsylvania	6.3	106	72	14.9
38	Rhode Island	3.4	174	87	8.3
39	South Carolina	14.4	279	48	22.5
40	South Dakota	3.8	86	45	12.8
41	Tennessee	13.2	188	59	26.9
42	Texas	12.7	201	80	25.5
43	Utah	3.2	120	80	22.9
44	Vermont	2.2	48	32	11.2
45	Virginia	8.5	156	63	20.7
46	Washington	4.0	145	73	26.2
47	West Virginia	5.7	81	39	9.3
48	Wisconsin	2.6	53	66	10.8
49	Wyoming	6.8	161	60	15.6

In [40]:
```
del Crime_Data_Kmeans['Unnamed: 0']
Crime_Data_Kmeans.head()
```

Out[40]:

	Murder	Assault	UrbanPop	Rape
0	13.2	236	58	21.2
1	10.0	263	48	44.5
2	8.1	294	80	31.0
3	8.8	190	50	19.5
4	9.0	276	91	40.6

In [41]:
```
Crime_Data_Kmeans
```

Out[41]:

	Murder	Assault	UrbanPop	Rape
0	13.2	236	58	21.2
1	10.0	263	48	44.5
2	8.1	294	80	31.0
3	8.8	190	50	19.5
4	9.0	276	91	40.6
5	7.9	204	78	38.7
6	3.3	110	77	11.1
7	5.9	238	72	15.8
8	15.4	335	80	31.9
9	17.4	211	60	25.8
10	5.3	46	83	20.2
11	2.6	120	54	14.2
12	10.4	249	83	24.0
13	7.2	113	65	21.0
14	2.2	56	57	11.3
15	6.0	115	66	18.0
16	9.7	109	52	16.3
17	15.4	249	66	22.2
18	2.1	83	51	7.8
19	11.3	300	67	27.8
20	4.4	149	85	16.3
21	12.1	255	74	35.1
22	2.7	72	66	14.9
23	16.1	259	44	17.1
24	9.0	178	70	28.2
25	6.0	109	53	16.4
26	4.3	102	62	16.5
27	12.2	252	81	46.0
28	2.1	57	56	9.5
29	7.4	159	89	18.8
30	11.4	285	70	32.1
31	11.1	254	86	26.1
32	13.0	337	45	16.1
33	0.8	45	44	7.3
34	7.3	120	75	21.4

	Murder	Assault	UrbanPop	Rape
35	6.6	151	68	20.0
36	4.9	159	67	29.3
37	6.3	106	72	14.9
38	3.4	174	87	8.3
39	14.4	279	48	22.5
40	3.8	86	45	12.8
41	13.2	188	59	26.9
42	12.7	201	80	25.5
43	3.2	120	80	22.9
44	2.2	48	32	11.2
45	8.5	156	63	20.7
46	4.0	145	73	26.2
47	5.7	81	39	9.3
48	2.6	53	66	10.8
49	6.8	161	60	15.6

In [42]:
```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
Scaled_Data = scaler.fit_transform(Crime_Data_Kmeans)
Scaled_Data
```

Out[42]:
```
array([[ 1.25517927,  0.79078716, -0.52619514, -0.00345116],
       [ 0.51301858,  1.11805959, -1.22406668,  2.50942392],
       [ 0.07236067,  1.49381682,  1.00912225,  1.05346626],
       [ 0.23470832,  0.23321191, -1.08449238, -0.18679398],
       [ 0.28109336,  1.2756352 ,  1.77678094,  2.08881393],
       [ 0.02597562,  0.40290872,  0.86954794,  1.88390137],
       [-1.04088037, -0.73648418,  0.79976079, -1.09272319],
       [-0.43787481,  0.81502956,  0.45082502, -0.58583422],
       [ 1.76541475,  1.99078607,  1.00912225,  1.1505301 ],
       [ 2.22926518,  0.48775713, -0.38662083,  0.49265293],
       [-0.57702994, -1.51224105,  1.21848371, -0.11129987],
       [-1.20322802, -0.61527217, -0.80534376, -0.75839217],
       [ 0.60578867,  0.94836277,  1.21848371,  0.29852525],
       [-0.13637203, -0.70012057, -0.03768506, -0.0250209 ],
       [-1.29599811, -1.39102904, -0.5959823 , -1.07115345],
       [-0.41468229, -0.67587817,  0.03210209, -0.34856705],
       [ 0.44344101, -0.74860538, -0.94491807, -0.53190987],
       [ 1.76541475,  0.94836277,  0.03210209,  0.10439756],
       [-1.31919063, -1.06375661, -1.01470522, -1.44862395],
       [ 0.81452136,  1.56654403,  0.10188925,  0.70835037],
       [-0.78576263, -0.26375734,  1.35805802, -0.53190987],
       [ 1.00006153,  1.02108998,  0.59039932,  1.49564599],
       [-1.1800355 , -1.19708982,  0.03210209, -0.68289807],
       [ 1.9277624 ,  1.06957478, -1.5032153 , -0.44563089],
       [ 0.28109336,  0.0877575 ,  0.31125071,  0.75148985],
       [-0.41468229, -0.74860538, -0.87513091, -0.521125  ],
       [-0.80895515, -0.83345379, -0.24704653, -0.51034012],
       [ 1.02325405,  0.98472638,  1.0789094 ,  2.671197  ],
       [-1.31919063, -1.37890783, -0.66576945, -1.26528114],
       [-0.08998698, -0.14254532,  1.63720664, -0.26228808],
       [ 0.83771388,  1.38472601,  0.31125071,  1.17209984],
       [ 0.76813632,  1.00896878,  1.42784517,  0.52500755],
       [ 1.20879423,  2.01502847, -1.43342815, -0.55347961],
       [-1.62069341, -1.52436225, -1.5032153 , -1.50254831],
       [-0.11317951, -0.61527217,  0.66018648,  0.01811858],
       [-0.27552716, -0.23951493,  0.1716764 , -0.13286962],
       [-0.66980002, -0.14254532,  0.10188925,  0.87012344],
       [-0.34510472, -0.78496898,  0.45082502, -0.68289807],
       [-1.01768785,  0.03927269,  1.49763233, -1.39469959],
       [ 1.53348953,  1.3119988 , -1.22406668,  0.13675217],
       [-0.92491776, -1.027393  , -1.43342815, -0.90938037],
       [ 1.25517927,  0.20896951, -0.45640799,  0.61128652],
       [ 1.13921666,  0.36654512,  1.00912225,  0.46029832],
       [-1.06407289, -0.61527217,  1.00912225,  0.17989166],
       [-1.29599811, -1.48799864, -2.34066115, -1.08193832],
       [ 0.16513075, -0.17890893, -0.17725937, -0.05737552],
       [-0.87853272, -0.31224214,  0.52061217,  0.53579242],
       [-0.48425985, -1.08799901, -1.85215107, -1.28685088],
       [-1.20322802, -1.42739264,  0.03210209, -1.1250778 ],
       [-0.22914211, -0.11830292, -0.38662083, -0.60740397]])
```

In [54]:
```
from sklearn.cluster import KMeans
#WCSS - Within Clusters Sum Of Squares
```

In [55]:
```
import warnings
warnings.filterwarnings('ignore')
```

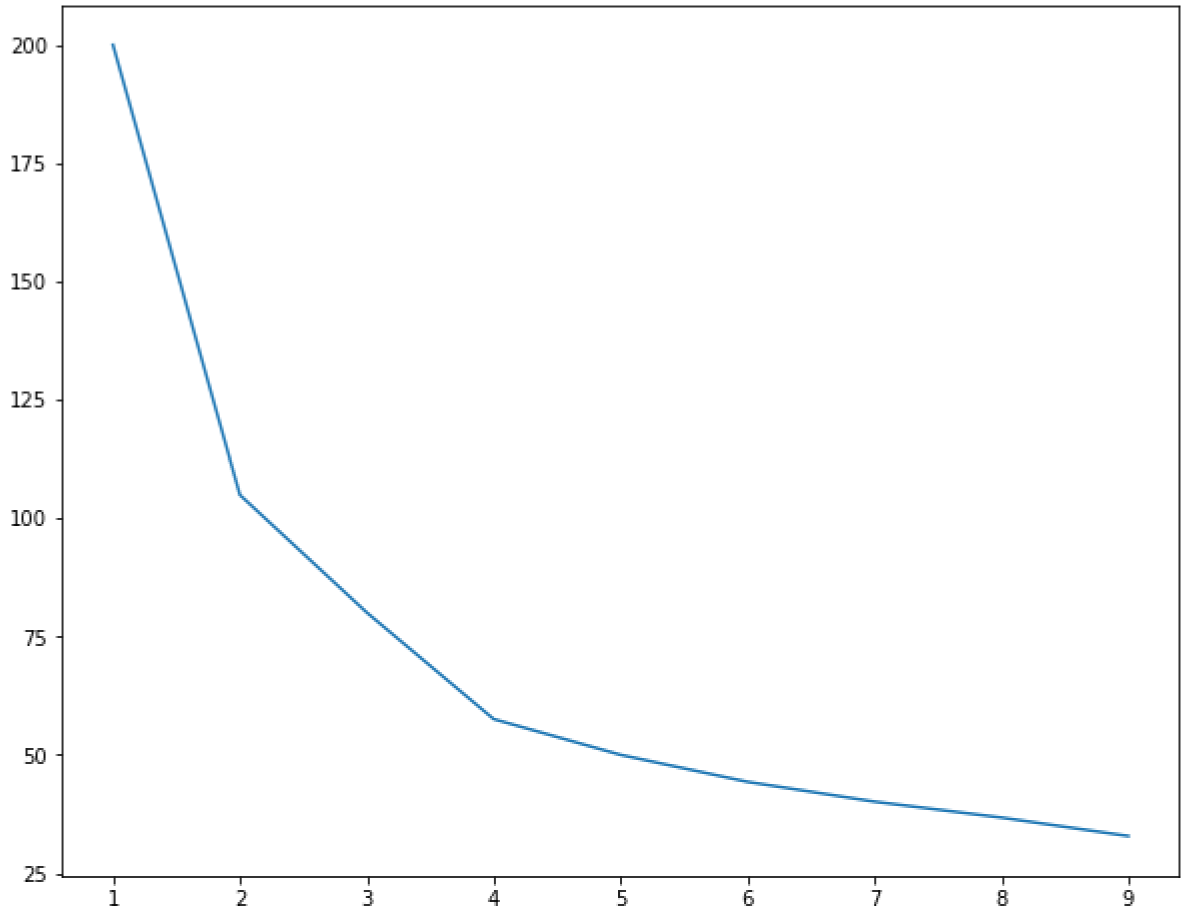
In [65]:
```
wcss = []

for i in range(1,10):
    Kmeans=KMeans(n_clusters=i)
    Kmeans.fit(Scaled_Data)
    wcss.append(Kmeans.inertia_)
```

In [66]:
```
wcss
```

Out[66]:
```
[200.0,
 104.96163315756873,
 80.08569526137276,
 57.55425863091106,
 49.993842813267484,
 44.28763884786578,
 40.099208067792866,
 36.73347182362056,
 32.827947464538525]
```

Finding iptimized value of n using Elbow method

In [70]:
```
plt.figure(figsize=(10,8))
plt.plot(range(1,10), wcss)
plt.show()
```



In [77]:
```
x=KMeans(n_clusters=4, max_iter=500, algorithm='auto')
x.fit(Crime_Data_Kmeans)
```

Out[77]:
```
KMeans(max_iter=500, n_clusters=4)
```

In [78]:
```
k_clusters = x.predict(Crime_Data_Kmeans)
```

In [83]:
```
Crime_Data_Kmeans['k_clusters']=k_clusters
```

In [84]:
```
Crime_Data_Kmeans.head(20)
```

Out[84]:

	Murder	Assault	UrbanPop	Rape	k_clusters
0	13.2	236	58	21.2	1
1	10.0	263	48	44.5	1
2	8.1	294	80	31.0	1
3	8.8	190	50	19.5	3
4	9.0	276	91	40.6	1
5	7.9	204	78	38.7	3
6	3.3	110	77	11.1	0
7	5.9	238	72	15.8	1
8	15.4	335	80	31.9	1
9	17.4	211	60	25.8	3
10	5.3	46	83	20.2	2
11	2.6	120	54	14.2	0
12	10.4	249	83	24.0	1
13	7.2	113	65	21.0	0
14	2.2	56	57	11.3	2
15	6.0	115	66	18.0	0
16	9.7	109	52	16.3	0
17	15.4	249	66	22.2	1
18	2.1	83	51	7.8	2
19	11.3	300	67	27.8	1

1. KMeans Clustering for Cluster '0'

In [87]:
```
Crime_Data_Kmeans[Crime_Data_Kmeans['k_clusters']==0]
```

Out[87]:

	Murder	Assault	UrbanPop	Rape	k_clusters
6	3.3	110	77	11.1	0
11	2.6	120	54	14.2	0
13	7.2	113	65	21.0	0
15	6.0	115	66	18.0	0
16	9.7	109	52	16.3	0
25	6.0	109	53	16.4	0
26	4.3	102	62	16.5	0
34	7.3	120	75	21.4	0
37	6.3	106	72	14.9	0
43	3.2	120	80	22.9	0

2. KMeans Clustering for Cluster '1'

In [90]:
```
Crime_Data_Kmeans[Crime_Data_Kmeans['k_clusters']==1]
```

Out[90]:

	Murder	Assault	UrbanPop	Rape	k_clusters
0	13.2	236	58	21.2	1
1	10.0	263	48	44.5	1
2	8.1	294	80	31.0	1
4	9.0	276	91	40.6	1
7	5.9	238	72	15.8	1
8	15.4	335	80	31.9	1
12	10.4	249	83	24.0	1
17	15.4	249	66	22.2	1
19	11.3	300	67	27.8	1
21	12.1	255	74	35.1	1
23	16.1	259	44	17.1	1
27	12.2	252	81	46.0	1
30	11.4	285	70	32.1	1
31	11.1	254	86	26.1	1
32	13.0	337	45	16.1	1
39	14.4	279	48	22.5	1

3. KMeans Clustering for Cluster '2'

In [91]:
```
Crime_Data_Kmeans[Crime_Data_Kmeans['k_clusters']==2]
```

Out[91]:

	Murder	Assault	UrbanPop	Rape	k_clusters
10	5.3	46	83	20.2	2
14	2.2	56	57	11.3	2
18	2.1	83	51	7.8	2
22	2.7	72	66	14.9	2
28	2.1	57	56	9.5	2
33	0.8	45	44	7.3	2
40	3.8	86	45	12.8	2
44	2.2	48	32	11.2	2
47	5.7	81	39	9.3	2
48	2.6	53	66	10.8	2

4. KMeans Clustering for Cluster '3'

In [93]:
```
Crime_Data_Kmeans[Crime_Data_Kmeans['k_clusters']==3]
```

Out[93]:

	Murder	Assault	UrbanPop	Rape	k_clusters
3	8.8	190	50	19.5	3
5	7.9	204	78	38.7	3
9	17.4	211	60	25.8	3
20	4.4	149	85	16.3	3
24	9.0	178	70	28.2	3
29	7.4	159	89	18.8	3
35	6.6	151	68	20.0	3
36	4.9	159	67	29.3	3
38	3.4	174	87	8.3	3
41	13.2	188	59	26.9	3
42	12.7	201	80	25.5	3
45	8.5	156	63	20.7	3
46	4.0	145	73	26.2	3
49	6.8	161	60	15.6	3

5. KMeans Clustering for Cluster '4'

In [94]:
```
Crime_Data_Kmeans[Crime_Data_Kmeans['k_clusters']==4]
```

Out[94]:

	Murder	Assault	UrbanPop	Rape	k_clusters

Conclusion:- KMeans clustering is more optimized has comapred to hierarchical clustering

DBSCAN

1. Import data

In [95]:
```
DB_Scan = pd.read_csv('crime_data.csv')
DB_Scan
```

Out[95]:

	Unnamed: 0	Murder	Assault	UrbanPop	Rape
0	Alabama	13.2	236	58	21.2
1	Alaska	10.0	263	48	44.5
2	Arizona	8.1	294	80	31.0
3	Arkansas	8.8	190	50	19.5
4	California	9.0	276	91	40.6
5	Colorado	7.9	204	78	38.7
6	Connecticut	3.3	110	77	11.1
7	Delaware	5.9	238	72	15.8
8	Florida	15.4	335	80	31.9
9	Georgia	17.4	211	60	25.8
10	Hawaii	5.3	46	83	20.2
11	Idaho	2.6	120	54	14.2
12	Illinois	10.4	249	83	24.0
13	Indiana	7.2	113	65	21.0
14	Iowa	2.2	56	57	11.3
15	Kansas	6.0	115	66	18.0
16	Kentucky	9.7	109	52	16.3
17	Louisiana	15.4	249	66	22.2
18	Maine	2.1	83	51	7.8
19	Maryland	11.3	300	67	27.8
20	Massachusetts	4.4	149	85	16.3
21	Michigan	12.1	255	74	35.1
22	Minnesota	2.7	72	66	14.9
23	Mississippi	16.1	259	44	17.1
24	Missouri	9.0	178	70	28.2
25	Montana	6.0	109	53	16.4
26	Nebraska	4.3	102	62	16.5
27	Nevada	12.2	252	81	46.0
28	New Hampshire	2.1	57	56	9.5
29	New Jersey	7.4	159	89	18.8
30	New Mexico	11.4	285	70	32.1
31	New York	11.1	254	86	26.1
32	North Carolina	13.0	337	45	16.1
33	North Dakota	0.8	45	44	7.3

	Unnamed: 0	Murder	Assault	UrbanPop	Rape
34	Ohio	7.3	120	75	21.4
35	Oklahoma	6.6	151	68	20.0
36	Oregon	4.9	159	67	29.3
37	Pennsylvania	6.3	106	72	14.9
38	Rhode Island	3.4	174	87	8.3
39	South Carolina	14.4	279	48	22.5
40	South Dakota	3.8	86	45	12.8
41	Tennessee	13.2	188	59	26.9
42	Texas	12.7	201	80	25.5
43	Utah	3.2	120	80	22.9
44	Vermont	2.2	48	32	11.2
45	Virginia	8.5	156	63	20.7
46	Washington	4.0	145	73	26.2
47	West Virginia	5.7	81	39	9.3
48	Wisconsin	2.6	53	66	10.8
49	Wyoming	6.8	161	60	15.6

2. Data Preparation

In [115]:

```
DB_Scan=Crime_Data
del DB_Scan['Clusters']
DB_Scan.head()
```

Out[115]:

	Murder	Assault	UrbanPop	Rape
0	0.746988	0.654110	0.440678	0.359173
1	0.554217	0.746575	0.271186	0.961240
2	0.439759	0.852740	0.813559	0.612403
3	0.481928	0.496575	0.305085	0.315245
4	0.493976	0.791096	1.000000	0.860465

In [116]:
```
DB_Scan
```

Out[116]:

	Murder	Assault	UrbanPop	Rape
0	0.746988	0.654110	0.440678	0.359173
1	0.554217	0.746575	0.271186	0.961240
2	0.439759	0.852740	0.813559	0.612403
3	0.481928	0.496575	0.305085	0.315245
4	0.493976	0.791096	1.000000	0.860465
5	0.427711	0.544521	0.779661	0.811370
6	0.150602	0.222603	0.762712	0.098191
7	0.307229	0.660959	0.677966	0.219638
8	0.879518	0.993151	0.813559	0.635659
9	1.000000	0.568493	0.474576	0.478036
10	0.271084	0.003425	0.864407	0.333333
11	0.108434	0.256849	0.372881	0.178295
12	0.578313	0.698630	0.864407	0.431525
13	0.385542	0.232877	0.559322	0.354005
14	0.084337	0.037671	0.423729	0.103359
15	0.313253	0.239726	0.576271	0.276486
16	0.536145	0.219178	0.338983	0.232558
17	0.879518	0.698630	0.576271	0.385013
18	0.078313	0.130137	0.322034	0.012920
19	0.632530	0.873288	0.593220	0.529716
20	0.216867	0.356164	0.898305	0.232558
21	0.680723	0.719178	0.711864	0.718346
22	0.114458	0.092466	0.576271	0.196382
23	0.921687	0.732877	0.203390	0.253230
24	0.493976	0.455479	0.644068	0.540052
25	0.313253	0.219178	0.355932	0.235142
26	0.210843	0.195205	0.508475	0.237726
27	0.686747	0.708904	0.830508	1.000000
28	0.078313	0.041096	0.406780	0.056848
29	0.397590	0.390411	0.966102	0.297158
30	0.638554	0.821918	0.644068	0.640827
31	0.620482	0.715753	0.915254	0.485788
32	0.734940	1.000000	0.220339	0.227390
33	0.000000	0.000000	0.203390	0.000000
34	0.391566	0.256849	0.728814	0.364341

	Murder	Assault	UrbanPop	Rape
35	0.349398	0.363014	0.610169	0.328165
36	0.246988	0.390411	0.593220	0.568475
37	0.331325	0.208904	0.677966	0.196382
38	0.156627	0.441781	0.932203	0.025840
39	0.819277	0.801370	0.271186	0.392765
40	0.180723	0.140411	0.220339	0.142119
41	0.746988	0.489726	0.457627	0.506460
42	0.716867	0.534247	0.813559	0.470284
43	0.144578	0.256849	0.813559	0.403101
44	0.084337	0.010274	0.000000	0.100775
45	0.463855	0.380137	0.525424	0.346253
46	0.192771	0.342466	0.694915	0.488372
47	0.295181	0.123288	0.118644	0.051680
48	0.108434	0.027397	0.576271	0.090439
49	0.361446	0.397260	0.474576	0.214470

3. Model Building

In [117]:
```
from sklearn.cluster import DBSCAN
```

In [118]:
```
Dbs =DBSCAN(min_samples=2,eps=0.2)
```

In [119]:
```
Dbs_Clusters = Dbs.fit_predict(DB_Scan)
Dbs_Clusters
```

Out[119]:
```
array([ 0, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,  1,  2,  1,  1,  1, -1,
        0,  1,  3, -1,  3,  1,  4, -1,  1,  1, -1,  1, -1,  3,  2, -1,  1,
        1,  1,  5,  1, -1,  4,  1, -1, -1,  5, -1,  1,  5,  1,  1,  1],
      dtype=int64)
```

In [120]:
```
DB_Scan['Dbs_Clusters']=Dbs_Clusters
DB_Scan
```

Out[120]:

	Murder	Assault	UrbanPop	Rape	Dbs_Clusters
0	0.746988	0.654110	0.440678	0.359173	0
1	0.554217	0.746575	0.271186	0.961240	-1
2	0.439759	0.852740	0.813559	0.612403	-1
3	0.481928	0.496575	0.305085	0.315245	-1
4	0.493976	0.791096	1.000000	0.860465	-1
5	0.427711	0.544521	0.779661	0.811370	-1
6	0.150602	0.222603	0.762712	0.098191	-1
7	0.307229	0.660959	0.677966	0.219638	-1
8	0.879518	0.993151	0.813559	0.635659	-1
9	1.000000	0.568493	0.474576	0.478036	-1
10	0.271084	0.003425	0.864407	0.333333	-1
11	0.108434	0.256849	0.372881	0.178295	1
12	0.578313	0.698630	0.864407	0.431525	2
13	0.385542	0.232877	0.559322	0.354005	1
14	0.084337	0.037671	0.423729	0.103359	1
15	0.313253	0.239726	0.576271	0.276486	1
16	0.536145	0.219178	0.338983	0.232558	-1
17	0.879518	0.698630	0.576271	0.385013	0
18	0.078313	0.130137	0.322034	0.012920	1
19	0.632530	0.873288	0.593220	0.529716	3
20	0.216867	0.356164	0.898305	0.232558	-1
21	0.680723	0.719178	0.711864	0.718346	3
22	0.114458	0.092466	0.576271	0.196382	1
23	0.921687	0.732877	0.203390	0.253230	4
24	0.493976	0.455479	0.644068	0.540052	-1
25	0.313253	0.219178	0.355932	0.235142	1
26	0.210843	0.195205	0.508475	0.237726	1
27	0.686747	0.708904	0.830508	1.000000	-1
28	0.078313	0.041096	0.406780	0.056848	1
29	0.397590	0.390411	0.966102	0.297158	-1
30	0.638554	0.821918	0.644068	0.640827	3
31	0.620482	0.715753	0.915254	0.485788	2
32	0.734940	1.000000	0.220339	0.227390	-1
33	0.000000	0.000000	0.203390	0.000000	1

	Murder	Assault	UrbanPop	Rape	Dbs_Clusters
34	0.391566	0.256849	0.728814	0.364341	1
35	0.349398	0.363014	0.610169	0.328165	1
36	0.246988	0.390411	0.593220	0.568475	5
37	0.331325	0.208904	0.677966	0.196382	1
38	0.156627	0.441781	0.932203	0.025840	-1
39	0.819277	0.801370	0.271186	0.392765	4
40	0.180723	0.140411	0.220339	0.142119	1
41	0.746988	0.489726	0.457627	0.506460	-1
42	0.716867	0.534247	0.813559	0.470284	-1
43	0.144578	0.256849	0.813559	0.403101	5
44	0.084337	0.010274	0.000000	0.100775	-1
45	0.463855	0.380137	0.525424	0.346253	1
46	0.192771	0.342466	0.694915	0.488372	5
47	0.295181	0.123288	0.118644	0.051680	1
48	0.108434	0.027397	0.576271	0.090439	1
49	0.361446	0.397260	0.474576	0.214470	1

1. DBSCAN Clustering with clusters '0'

In [121]:
```
DB_Scan[DB_Scan['Dbs_Clusters']==0]
```

Out[121]:

	Murder	Assault	UrbanPop	Rape	Dbs_Clusters
0	0.746988	0.65411	0.440678	0.359173	0
17	0.879518	0.69863	0.576271	0.385013	0

2. DBSCAN Clustering with clusters '1'

In [122]:
```
DB_Scan[DB_Scan['Dbs_Clusters']==1]
```

Out[122]:

	Murder	Assault	UrbanPop	Rape	Dbs_Clusters
11	0.108434	0.256849	0.372881	0.178295	1
13	0.385542	0.232877	0.559322	0.354005	1
14	0.084337	0.037671	0.423729	0.103359	1
15	0.313253	0.239726	0.576271	0.276486	1
18	0.078313	0.130137	0.322034	0.012920	1
22	0.114458	0.092466	0.576271	0.196382	1
25	0.313253	0.219178	0.355932	0.235142	1
26	0.210843	0.195205	0.508475	0.237726	1
28	0.078313	0.041096	0.406780	0.056848	1
33	0.000000	0.000000	0.203390	0.000000	1
34	0.391566	0.256849	0.728814	0.364341	1
35	0.349398	0.363014	0.610169	0.328165	1
37	0.331325	0.208904	0.677966	0.196382	1
40	0.180723	0.140411	0.220339	0.142119	1
45	0.463855	0.380137	0.525424	0.346253	1
47	0.295181	0.123288	0.118644	0.051680	1
48	0.108434	0.027397	0.576271	0.090439	1
49	0.361446	0.397260	0.474576	0.214470	1

3. DBSCAN Clustering with clusters '2'

In [123]:
```
DB_Scan[DB_Scan['Dbs_Clusters']==2]
```

Out[123]:

	Murder	Assault	UrbanPop	Rape	Dbs_Clusters
12	0.578313	0.698630	0.864407	0.431525	2
31	0.620482	0.715753	0.915254	0.485788	2

4. DBSCAN Clustering with clusters '3'

In [124]:
```
DB_Scan[DB_Scan['Dbs_Clusters']==3]
```

Out[124]:

	Murder	Assault	UrbanPop	Rape	Dbs_Clusters
19	0.632530	0.873288	0.593220	0.529716	3
21	0.680723	0.719178	0.711864	0.718346	3
30	0.638554	0.821918	0.644068	0.640827	3

5. DBSCAN Clustering with clusters '4'

In [125]:
```
DB_Scan[DB_Scan['Dbs_Clusters']==4]
```

Out[125]:

	Murder	Assault	UrbanPop	Rape	Dbs_Clusters
23	0.921687	0.732877	0.203390	0.253230	4
39	0.819277	0.801370	0.271186	0.392765	4

6. DBSCAN Clustering with clusters '5'

In [126]:
```
DB_Scan[DB_Scan['Dbs_Clusters']==5]
```

Out[126]:

	Murder	Assault	UrbanPop	Rape	Dbs_Clusters
36	0.246988	0.390411	0.593220	0.568475	5
43	0.144578	0.256849	0.813559	0.403101	5
46	0.192771	0.342466	0.694915	0.488372	5

Final Conclusion:- As compared to all the three clustering methods DBSCAN method is more optimized with more number of clusters

THE END