Day 45

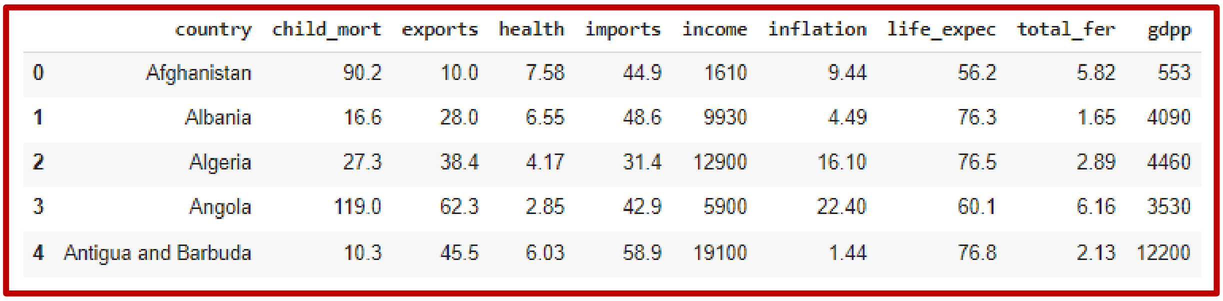
DIY

Q1. Problem Statement: Hierarchical Clustering

Load the *"Country-data.csv"* dataset into a DataFrame and perform the following tasks:

1. Create a separate DataFrame with only numeric data by removing the *"country"* column
2. Scale the data using the Standard Scaler to create a scaled DataFrame
3. Plotting dendrograms with the complete linkage method
4. Creating cluster labels using cut tree
5. Perform the 4-Component PCA on DataFrame
6. Now, from final the DataFrame, analyze how the low GDP rate corresponds to the child mortality rate around the world

Dataset:

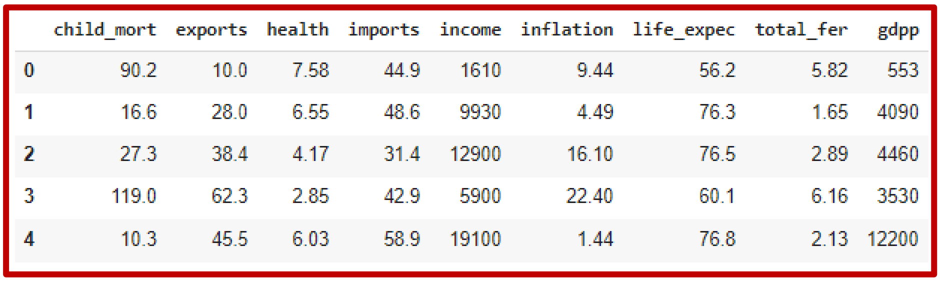


	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200

Sample Output:

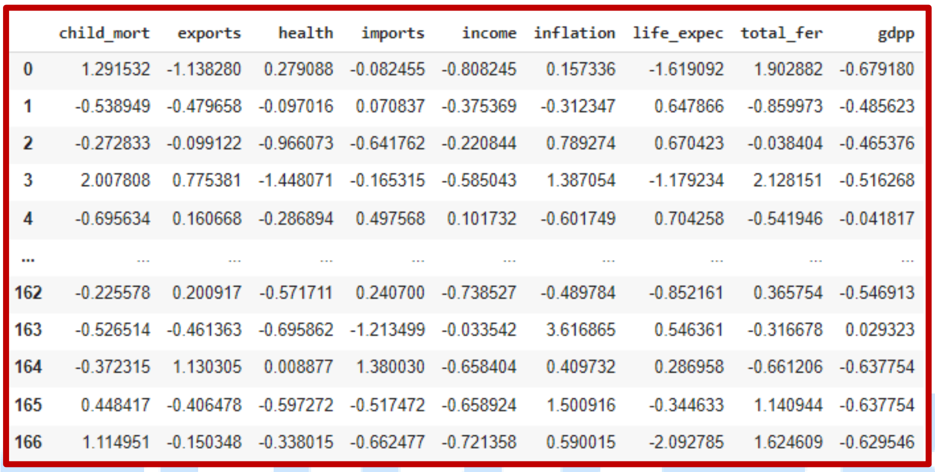
1. Create a separate DataFrame with only numeric data by removing the

"country" column



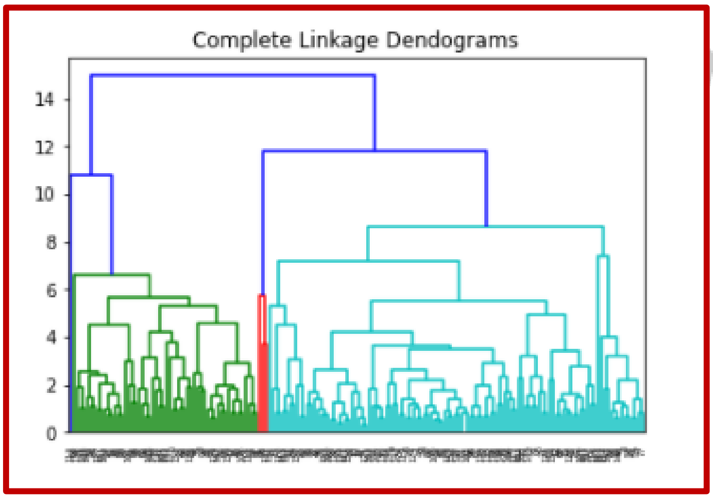
	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200

2. Scale the data using the Standard Scaler to create a scaled DataFrame

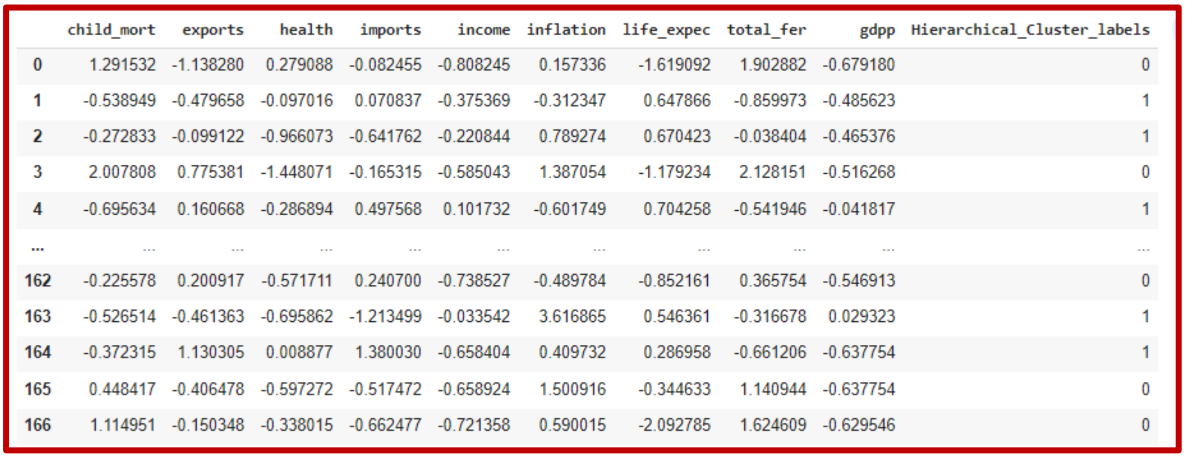


	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	1.291532	-1.138280	0.279088	-0.082455	-0.808245	0.157336	-1.619092	1.902882	-0.679180
1	-0.538949	-0.479658	-0.097016	0.070837	-0.375369	-0.312347	0.647866	-0.859973	-0.485623
2	-0.272833	-0.099122	-0.966073	-0.641762	-0.220844	0.789274	0.670423	-0.038404	-0.465376
3	2.007808	0.775381	-1.448071	-0.165315	-0.585043	1.387054	-1.179234	2.128151	-0.516268
4	-0.695634	0.160668	-0.286894	0.497568	0.101732	-0.601749	0.704258	-0.541946	-0.041817
...	...	...	...	...	...	...	...	...	...
162	-0.225578	0.200917	-0.571711	0.240700	-0.738527	-0.489784	-0.852161	0.365754	-0.546913
163	-0.526514	-0.461363	-0.695862	-1.213499	-0.033542	3.616865	0.546361	-0.316678	0.029323
164	-0.372315	1.130305	0.008877	1.380030	-0.658404	0.409732	0.286958	-0.661206	-0.637754
165	0.448417	-0.406478	-0.597272	-0.517472	-0.658924	1.500916	-0.344633	1.140944	-0.637754
166	1.114951	-0.150348	-0.338015	-0.662477	-0.721358	0.590015	-2.092785	1.624609	-0.629546

3. Plotting dendrograms with the complete linkage method

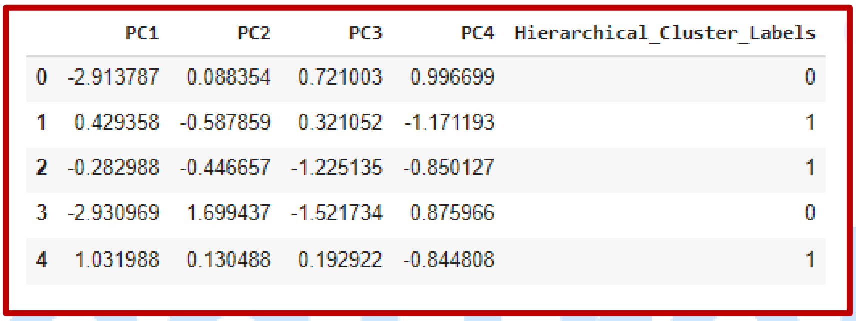


4. Creating cluster labels using cut tree



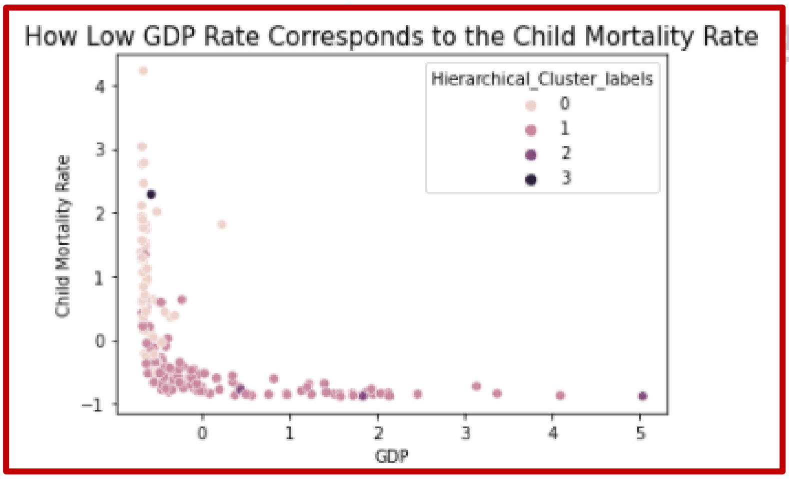
	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	Hierarchical_Cluster_labels
0	1.291532	-1.138280	0.279088	-0.082455	-0.808245	0.157336	-1.619092	1.902882	-0.679180	0
1	-0.538949	-0.479658	-0.097016	0.070837	-0.375369	-0.312347	0.647866	-0.859973	-0.485623	1
2	-0.272833	-0.099122	-0.966073	-0.641762	-0.220844	0.789274	0.670423	-0.038404	-0.465376	1
3	2.007808	0.775381	-1.448071	-0.165315	-0.585043	1.387054	-1.179234	2.128151	-0.516268	0
4	-0.695634	0.160668	-0.286894	0.497568	0.101732	-0.601749	0.704258	-0.541946	-0.041817	1
...	...	...	...	...	...	...	...	...	...	...
162	-0.225578	0.200917	-0.571711	0.240700	-0.738527	-0.489784	-0.852161	0.365754	-0.546913	0
163	-0.526514	-0.461363	-0.695862	-1.213499	-0.033542	3.616865	0.546361	-0.316678	0.029323	1
164	-0.372315	1.130305	0.008877	1.380030	-0.658404	0.409732	0.286958	-0.661206	-0.637754	1
165	0.448417	-0.406478	-0.597272	-0.517472	-0.658924	1.500916	-0.344633	1.140944	-0.637754	0
166	1.114951	-0.150348	-0.338015	-0.662477	-0.721358	0.590015	-2.092785	1.624609	-0.629546	0

5. Perform the 4-Component PCA on DataFrame



	PC1	PC2	PC3	PC4	Hierarchical_Cluster_Labels
0	-2.913787	0.088354	0.721003	0.996699	0
1	0.429358	-0.587859	0.321052	-1.171193	1
2	-0.282988	-0.446657	-1.225135	-0.850127	1
3	-2.930969	1.699437	-1.521734	0.875966	0
4	1.031988	0.130488	0.192922	-0.844808	1

6. Now, from final the DataFrame, analyze how low GDP rate corresponds to the child mortality rate around the world


How Low GDP Rate Corresponds to the Child Mortality Rate
Hierarchical_Cluster_labels
Child Mortality Rate
GDP