upGrad

#LifeKoKaroLift

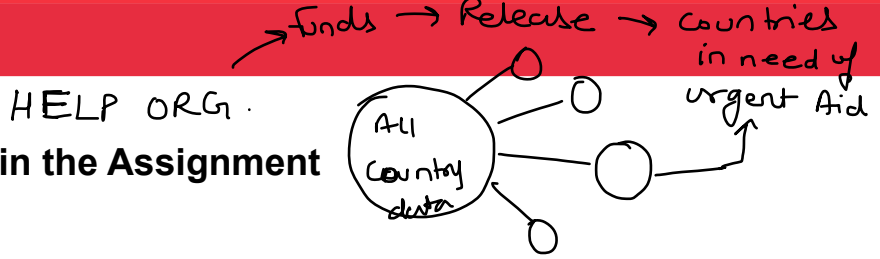
Clustering: Pre-Assignment Session

upGrad

Course : Data Science

Lecture On : Pre-Assignment

Instructor : Sumit Shukla



Step to procedure in the Assignment

Let's first understand the problem statement:

Identify top countries that are direst need of aid. Your job is to categorise the countries using some socio-economic and health factors that determine the overall development of the country. Then you need to suggest the countries which the CEO needs to focus on the most.

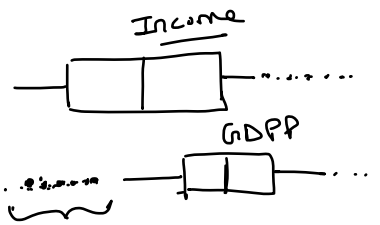
	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1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

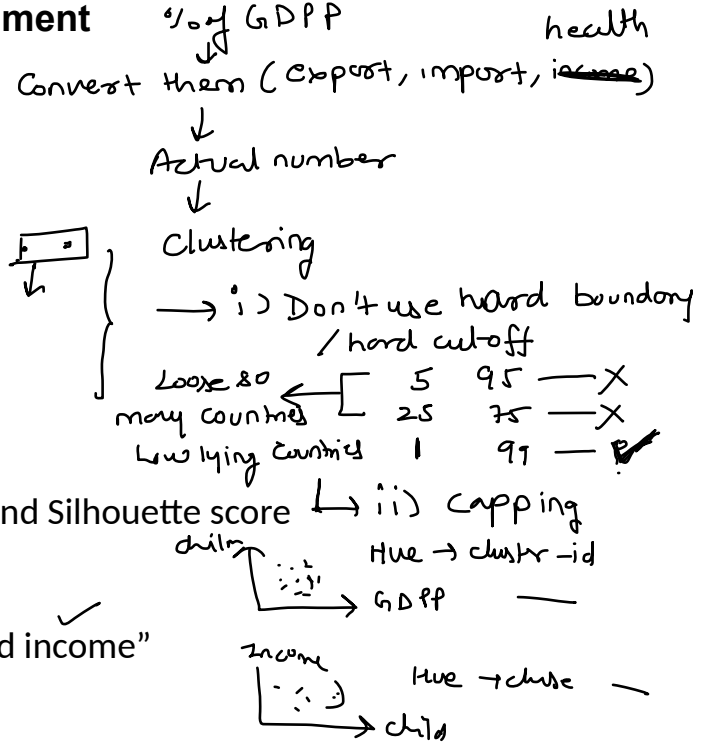
Clustering Assignment

upGrad

Step to procedure in the Assignment

1. Data Understanding.
   a. Hint: Don't forget to read the data description properly.
   b. EDA
2. Perform Clustering.
   a. Data preparation for clustering.
      i. Outlier treatment
      ii. Hopkins check
   b. Clustering
      i. K-MEANS
         1. Run K-Means and choose K using both Elbow and Silhouette score
         2. Run K-Means with the chosen K
         3. Visualise the clusters
         4. Clustering profiling using "gdpp, child_mort and income"





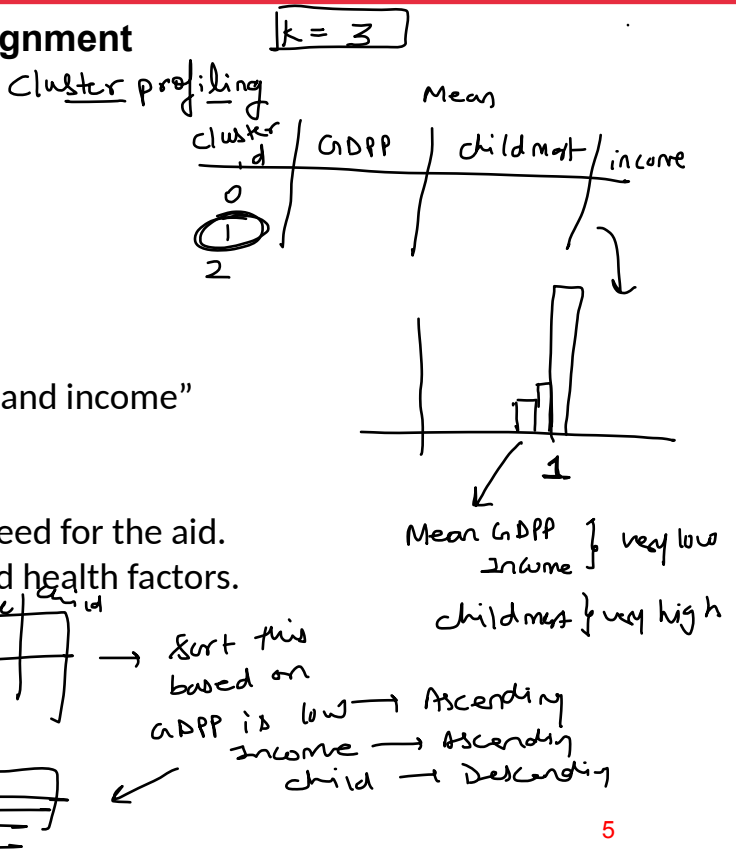
Clustering Assignment

upGrad

Step to procedure in the Assignment

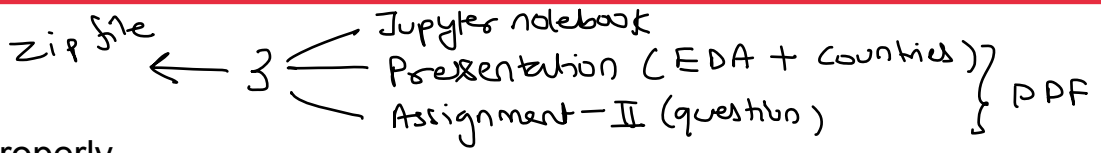
1. Perform Clustering.
   a. Clustering
      i. Hierarchical Clustering
         1. Use both Single and Complete linkage
         2. Choose one method based on the results
         3. Visualise the clusters
         4. Clustering profiling using “gdpp, child_mort and income”

2. Country Identification
   a. Based on the analysis, choose the countries that are in need for the aid.
   b. Choose the countries based on some socio-economic and health factors.



What to keep in mind

upGrad



- You need to comment your code properly.
- You need to submit 3 files zipped as one file. A python notebook with all the code, A PPT(Converted as PDF) with all the recommendation and A PDF with part-II of the assignment.
- PPT is for managers, so don't add unnecessary pages. PPT should cover main points from your analysis.
- Choose K wisely as this reflects the final solution.
- Mention your assumptions in your notebook, If taken any.
- For any help regarding assignment, use Discussion Forum.
- For any help regarding coding error, use online portals such as StackOverflow.

upGrad

#LifeKoKaroLift

Thank You!