CLUSTERING PROBLEM

PROBLEM STATEMET

- HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. It runs a lot of operational projects from time to time along with advocacy drives to raise awareness as well as for funding purposes.
- After the recent funding programmers, they have been able to raise around $ 10 million. Now the CEO of the NGO needs to decide how to use this money strategically and effectively. The significant issues that come while making this decision are mostly related to choosing the countries that are in the direst need of aid.

SOLUTION TECNIQUE

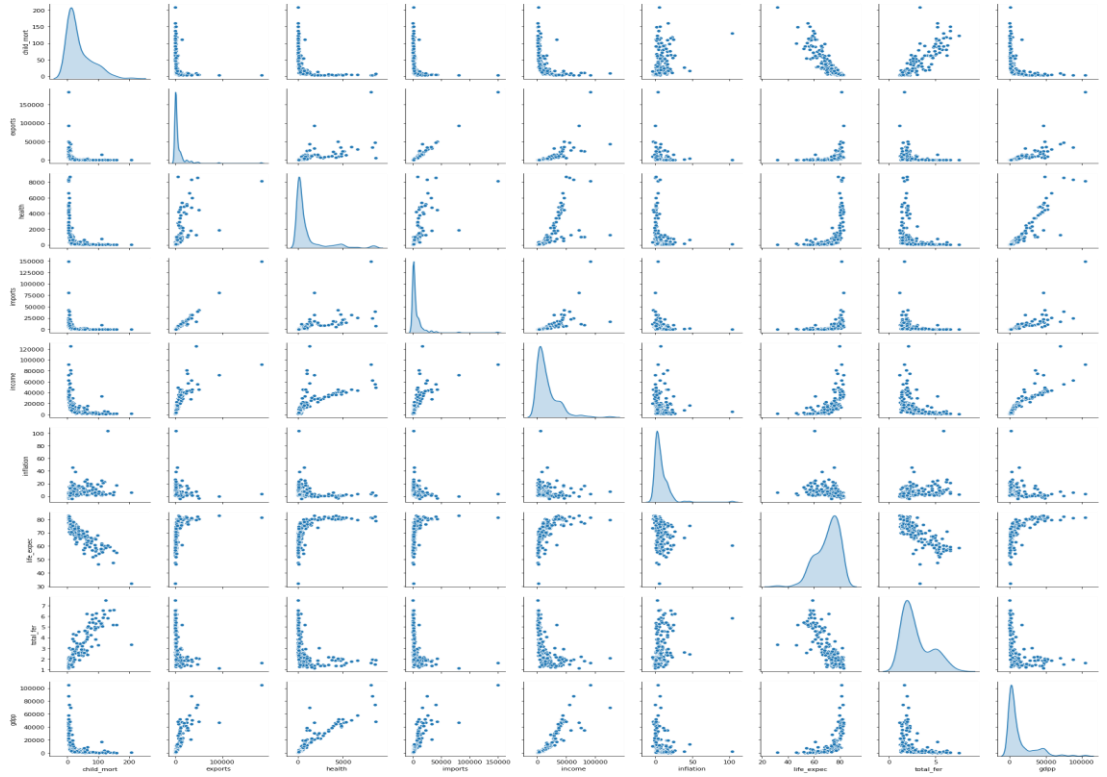
- READ THE DATA
- DO DESCRIPTIC STATICSICAL ANALYSIS
- DO DATA VISULIZATION AND CREATE INFRARENCE ABOUT THE DATA (VISUALY HOW MANY CLUSTER AVALIABLE RELATIONSHIPS BETWEEN DIFFERENT FEEATURES)
- OUTLIER CAPING
- USE KMEAN AND HIGHERARCIAL CLUSTERING TO FIND THE OPTIMAM CLUSTERS
- FIND THE CLUSTER WHICH HAS LESS INCOME AND GDP BUT HIGH CHIELD DEATH
- GET THETOP 5 COUNTRY WHICH NEED IMIDIATE HELP

INFRARENCE ABOUT THE DATA

MULTIVARIENT ANALYSIS

- from the multivariant analysis one can clearly say that there is sone relation ship present between features
- between gdpp and export,gdpp and imports there is highly positive relation ship and with health and income there is positive relation ship as well.
- between total_fer and child mortality there is positive relation ship
- between life_expct and chield_mort there is negative relation ship
- between import and export there is a high positive relation ship
- bttween income and export and health there is positive relationship
- between health and export there is high positive relation ship

VISULIZATION

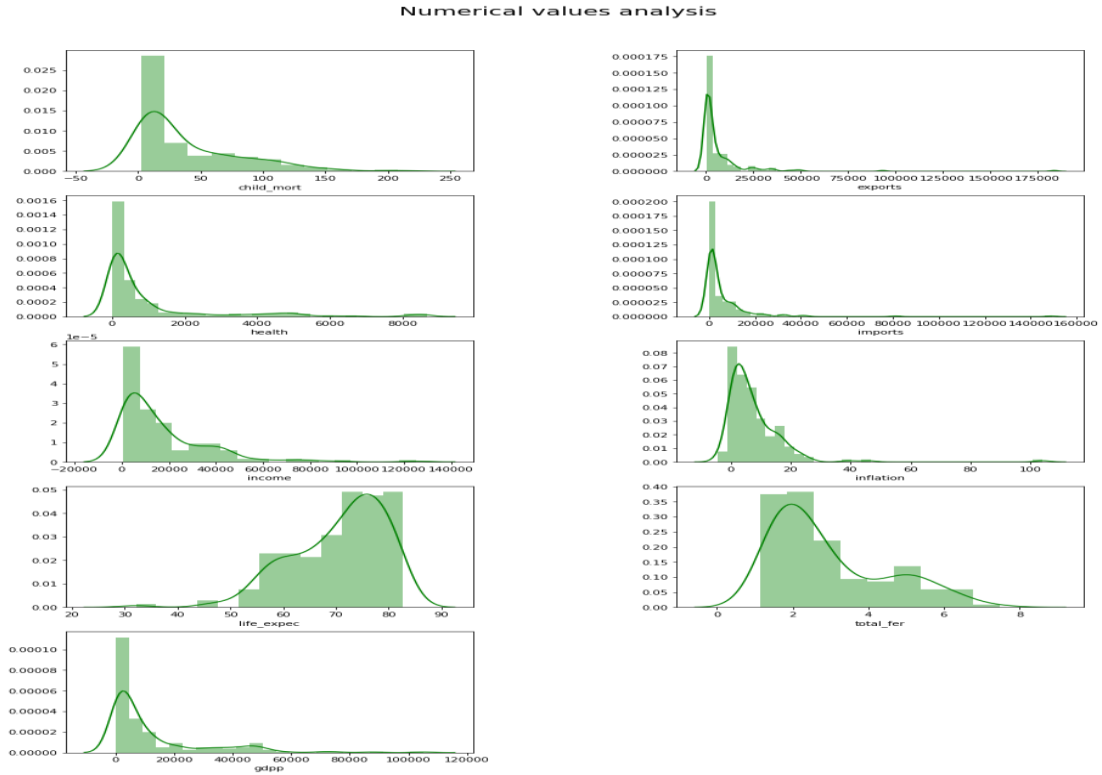


INFRARENCE ABOUT THE DATA

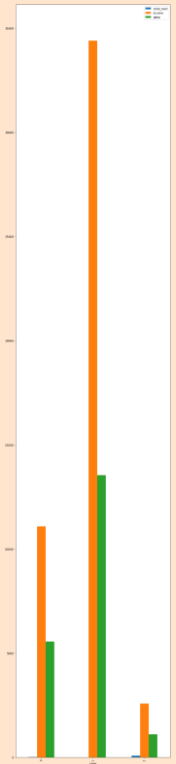
UNIVARIENT ANALYSIS

- it is clearly visible that there is two cluster from univariant analysis.
- all the feature has got rightly skewed except for the life_expec

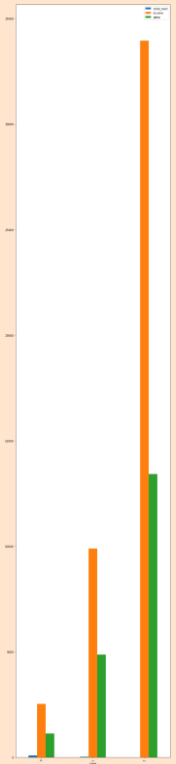
VISULIATION



SOLUTION APPROCH



KMEAN CLUSTERING



HIGHERERCIAL CLUSTERING

- DOING KMEAN AND HIGHERCIAL CLUSTERING TO GET THE CLUSTER WHICH REPRESENT HIGH CHILD DEATH LOW INCOME AND LOW GDP

RECOMANDATION

- 'Congo, Dem. Rep.',
- 'Liberia',
- 'Burundi',
- 'Niger',
- 'Central African Republic’

ARE NEED THE AID MOST

CONCLUSION

HIGHERCIAL CLUSTERING AND THE K MEAN CLUSTERING HAS BOTH PROVIDED THE TOP 5 COUNTRY WHICH MEANS THE ANALYSIS IS QUIT CONSISTANT IN TERMS 3 CLUSTER