Clustering Assignment

Vaibhav Sah

Overview

HELP International has collected $10 Million for aid.

Chose top 5 countries in most need of aid from on the basis of

- Lowest Income per person
- Lowest GDP per person
- Highest Child Mortality Rate

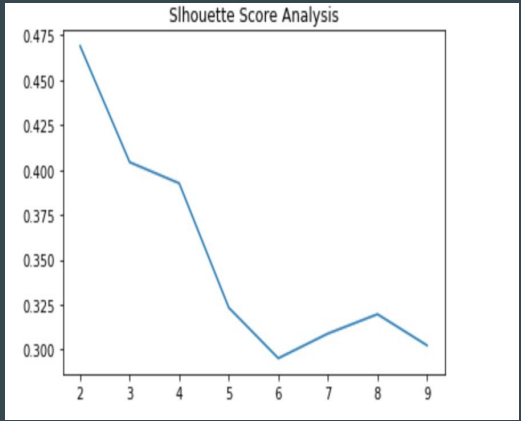
Solution Approach : Clustering

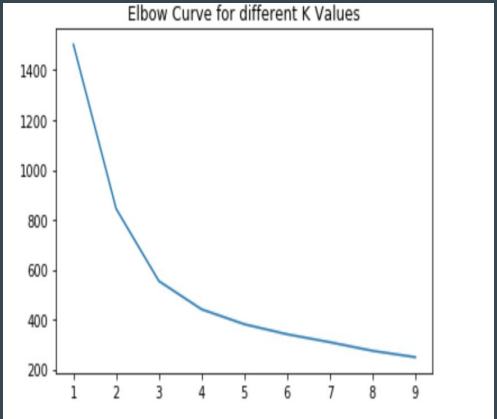
K-Means Clustering

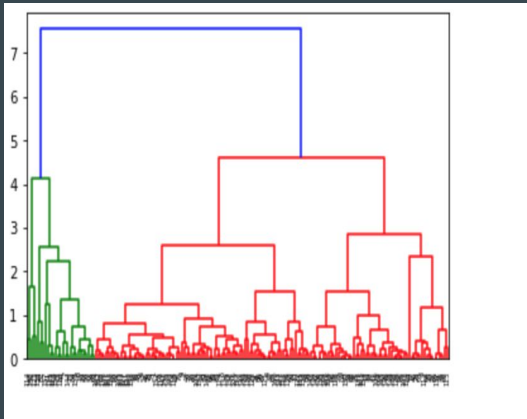
- Elbow Curve
- Silhouette Analysis
- Cluster Size = 3

Hierarchical Clustering

- Single Linkage
- Complete Linkage
- Cluster Size = 3





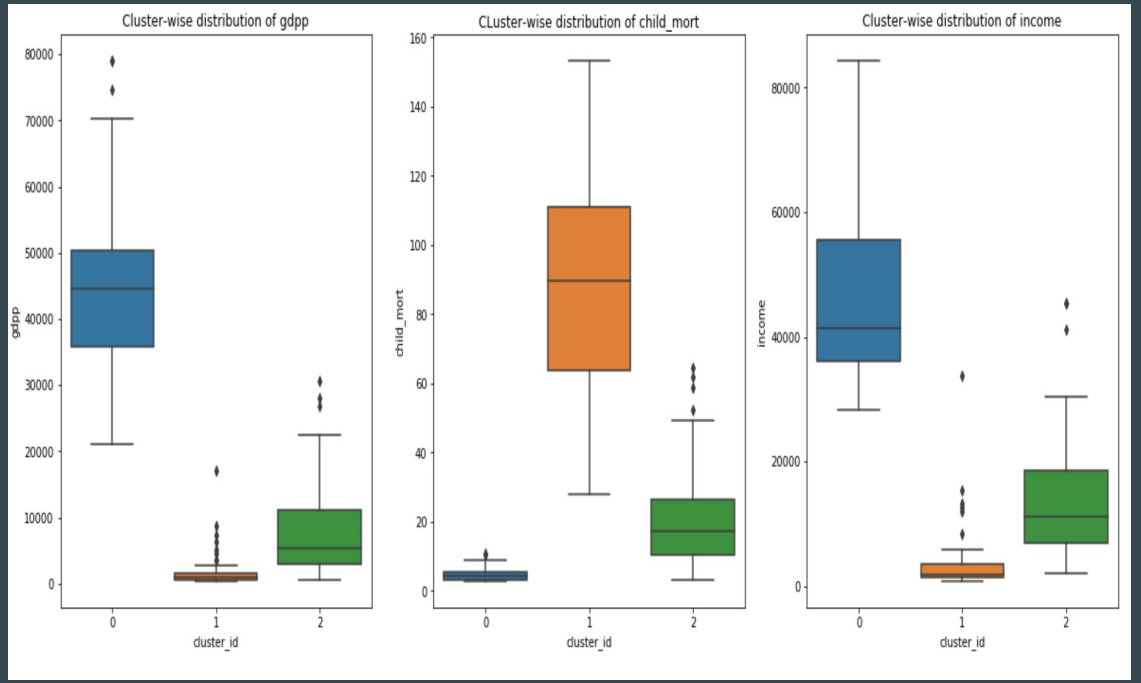


K Means Clustering

Cluster Size 3

Findings:

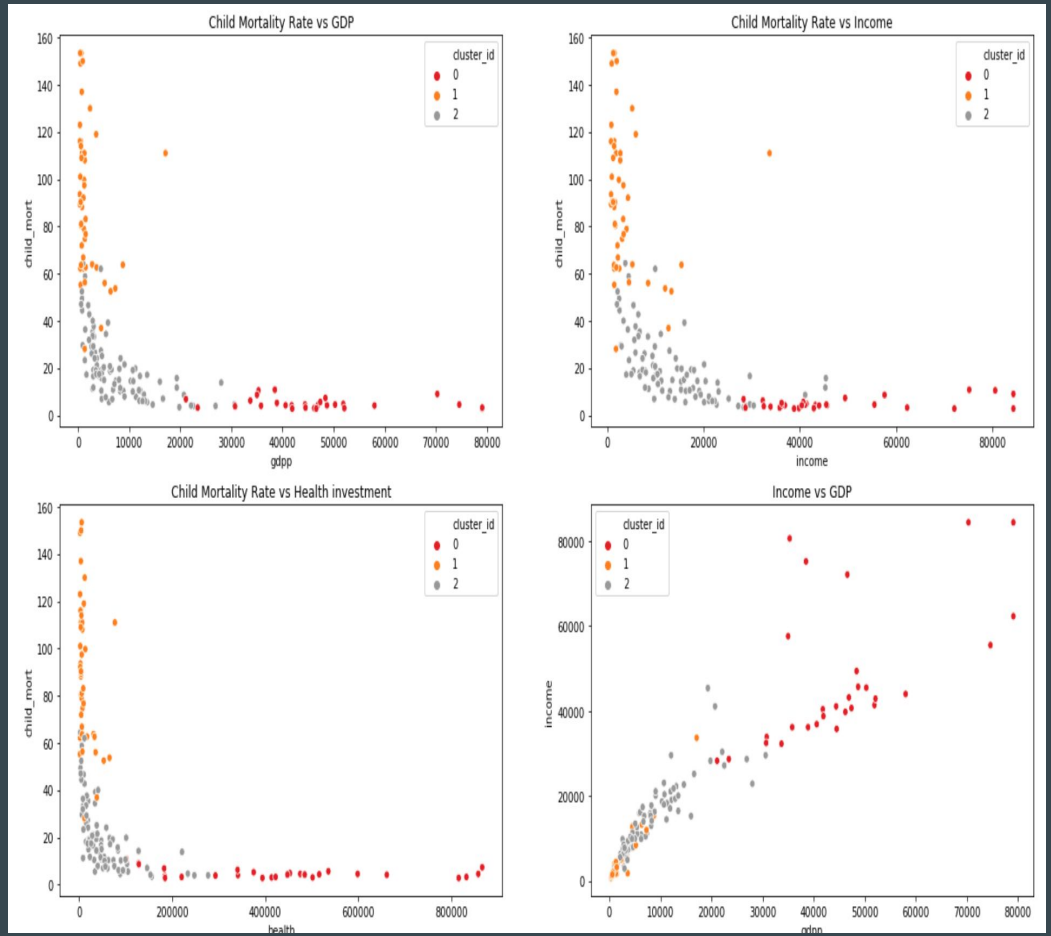
- As the GDP increases, Child Mortality Rate decreases
- As the Income per person increases, Child Mortality rate decreases
- As Investment in Health increases, Child Mortality decreases



K Means Clustering

Findings:

- 3 Clear clusters visible
- Red (0): Low Child Mortality Rate and High Income/GDP per person
- Yellow (1): Average Child Mortality Rate and Average Income/GDP per person
- Gray (2): High Child Mortality Rate and Low Income/GDP per person

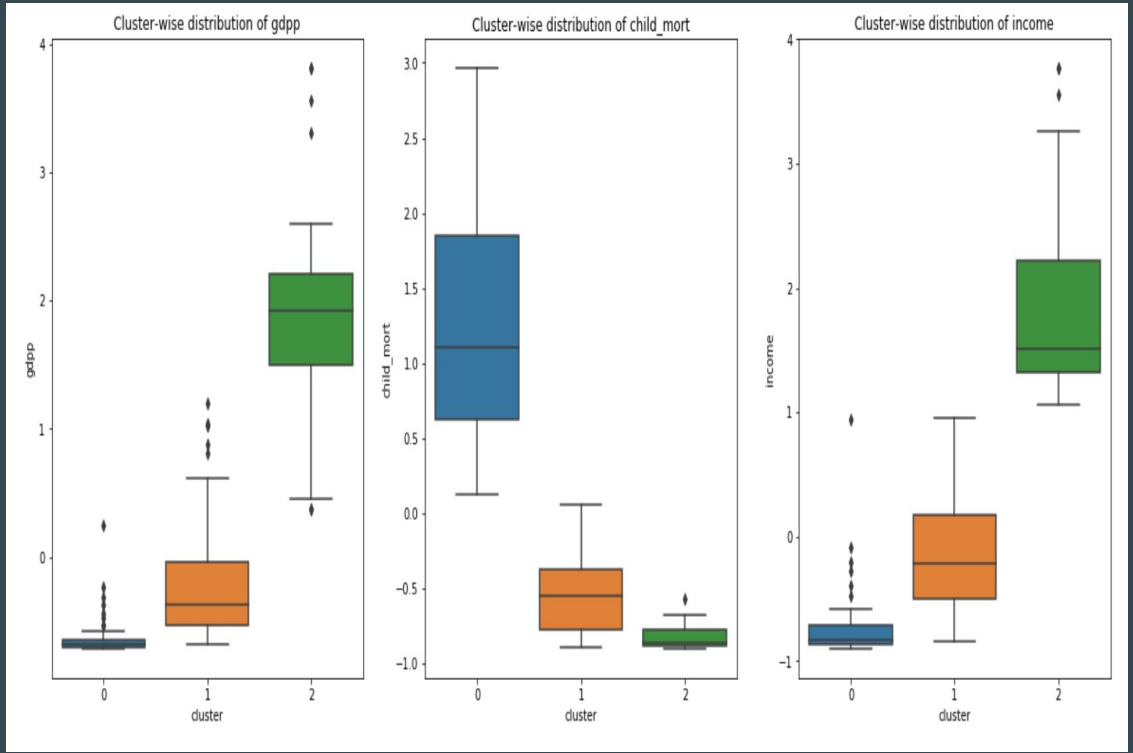


Hierarchical Clustering

Cluster Size 3

Findings:

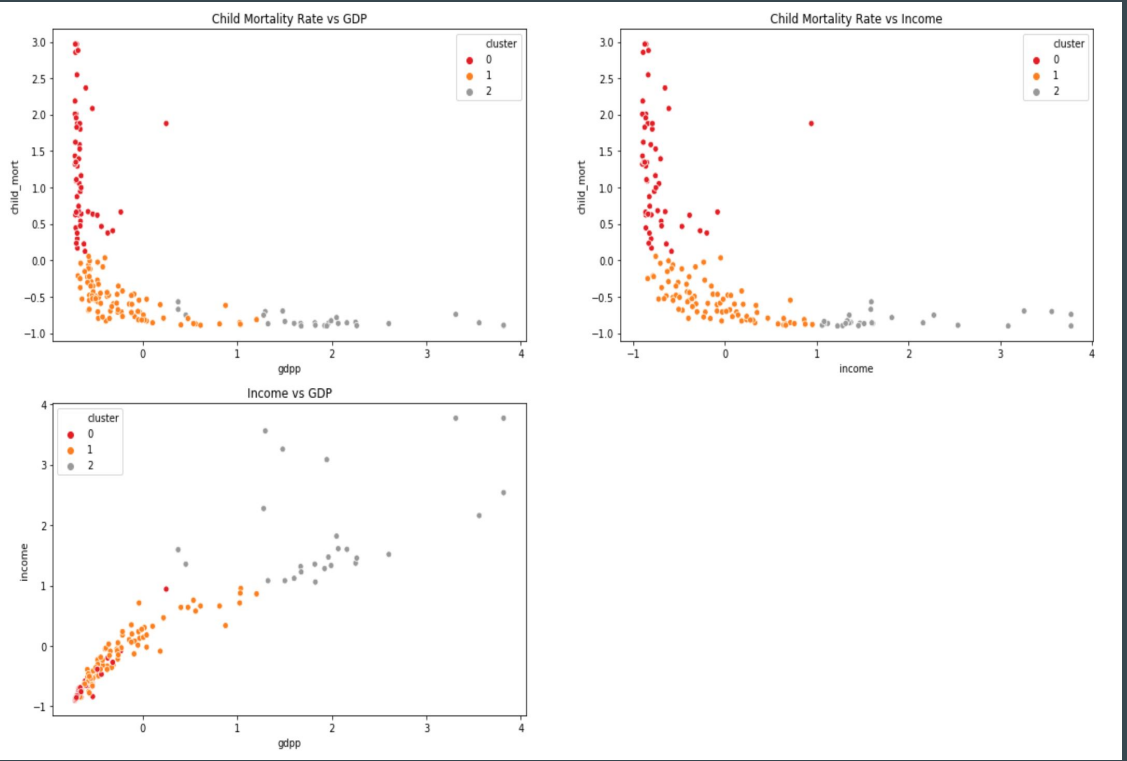
- For countries with high GDP per person, Child Mortality Rate is low
- For countries with high Income per person, Child Mortality Rate is low
- For countries with low GDP / Income per person, Child Mortality Rate is high



Hierarchical Clustering

Findings:

- 3 Clear clusters visible
- Red (0): High Child Mortality Rate and Low Income/GDP per person
- Yellow (1): Average Child Mortality Rate and Average Income/GDP per person
- Gray (2): Low Child Mortality Rate and High Income/GDP per person



Countries in Need

1. Liberia
2. Burundi
3. Congo, Dem. Rep.
4. Niger
5. Sierra Leone

Presented By:

Vaibhav Sah, Senior Engineer

Aspiring Data Scientist,
Front End Developer,
Ticketmaster