

# Clustering Assignment

# Case Study:

HELP is an NGO committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. After raising a fund of 10 million USD, they want a **strategy in distributing funds**.

We are asked to ***analyze the socio-economic and health factors in different countries and make recommendations to the company CEO on countries that need funding/aid.***

# Analysis Approach

After identifying the problem statement, our approach is as follows:

## Technical (Stage 1)

- **Data Pre-Processing**
  - Outlier treatment
  - Scaling data
- **PCA for dimensionality reduction** (Took 5 Pc's)



## Technical (Stage 2)

- **Clustering**
  - K-Means Clustering
    - Clusters = 5
  - Hierarchical Clustering
    - Clusters = 4



## Business (Stage 3)

- **Identifying key features**
  - Child Mortality
  - Income/GDPP
- **Identifying our cluster basing on these features**
  - By Analyzing Visually



## Business (Stage 4)

- **Identifying countries in need of funding**
- **Dividing them into 3 groups basing on their immediate need**
- **Making recommendation**

# Technical Stages: Result

Let us look at the results:

## Data Pre-Processing

- Both PCA and Clustering get highly influenced by **both outliers and magnitudes**
- **1.5%** of data from either side is removed
- Data is **standardized to Z-Score**



## PCA

- **Considered 5 PC's that explained 92% of the variance**
- **Deduced the result by taking the cumulative sum of variances explained**



## KMeans Clustering:

- **Data is cluster able with Hopkin statistic of 0.68**
- **5 Clusters are considered ideal after silhouette analysis**
- **Cluster 3 is identified as our preferred**

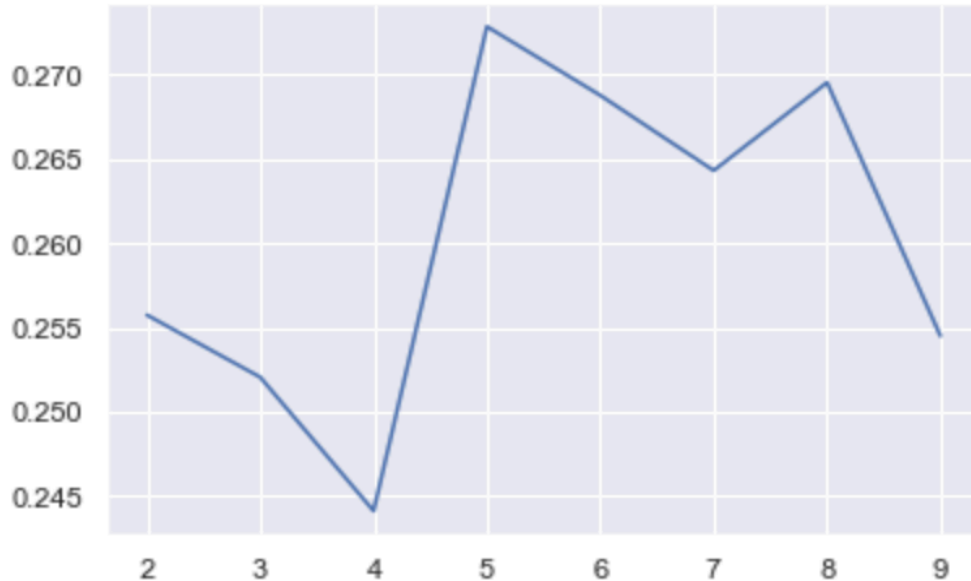


## Hierarchical Clustering

- **Deduced both Single linkage and Complete linkage dendrograms**
- **Chose "Complete Linkage dendrogram"**
- **It is evident that no of clusters is 4 from dendrogram, so  $n = 4$**

# Technical Stages: Basis for Choice of Clusters

figure 10: Silhouette Score



## K-Means Clustering

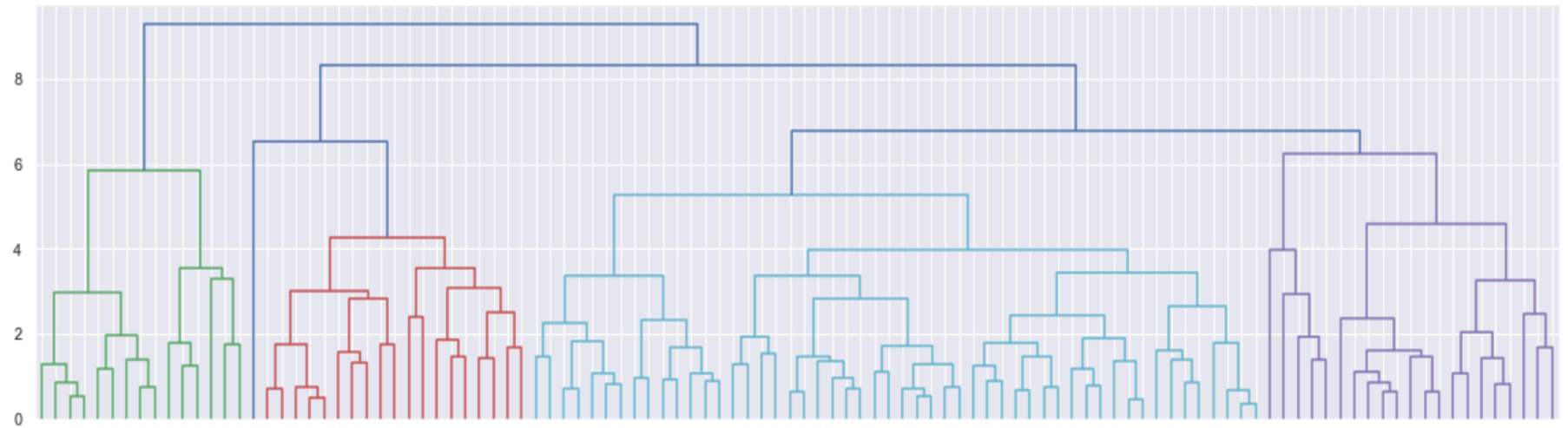
Silhouette Chart basing on which number of clusters were chosen.

**The higher the score, the better. Hence  $N = 5$**

## Hierarchical Clustering

Dendrogram clearly suggests that 4 major clusters are present

figure 14: Dendrogram for complete linkage

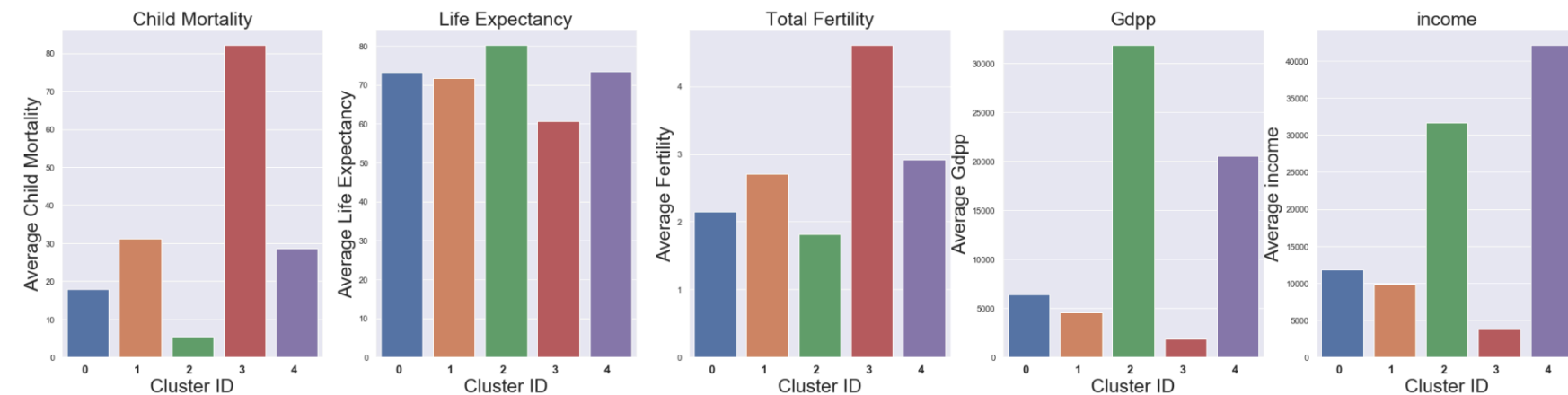
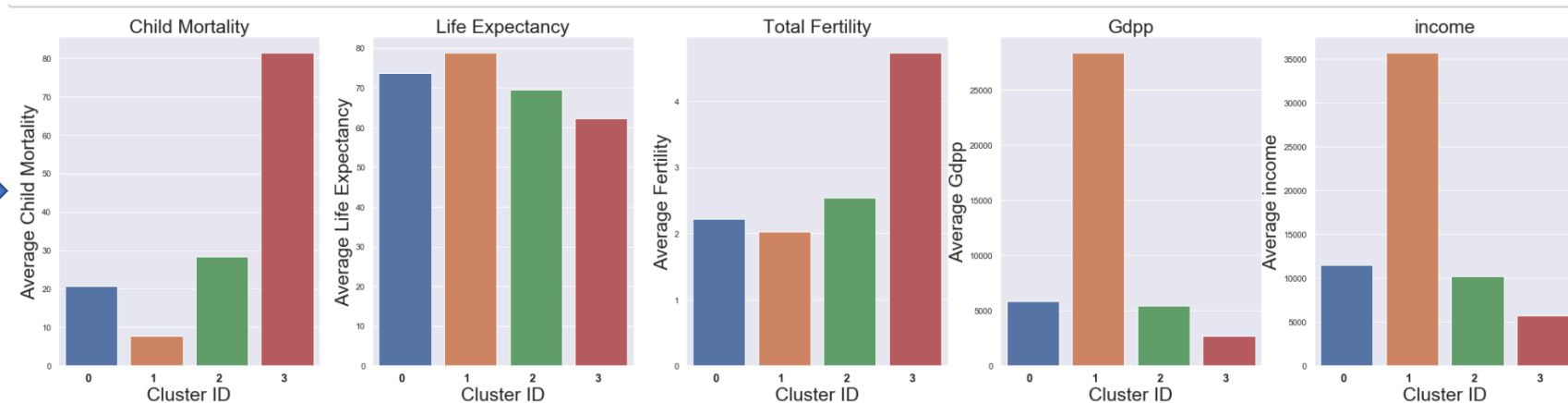


# Business Stage: Results aka Visualizations

## Hierarchical Clustering Results:

We identified that Cluster 3 had low GDPP and high Child Mortality.

Low GDPP indicates absence of sufficient economic stability to aid basic amenities or growth



K Means Clustering Results:  
We identified that Cluster 3 had low GDPP and high Child Mortality.

Similar to GDPP, high child mortality suggests there are no basic health care operations in these countries

# Final List of Countries:

In our proposal, we said, we divide the final list of countries into 3 stages.

## Countries in Stage 1

**Congo, Dem. Rep.**

**Burundi**

**Haiti**

**Liberia**

**Chad**

**Sierra Leone**

## Countries in Stage 2

**Mozambique**

**Mali**

**Niger**

**Burkina Faso**

**Togo**

**Cote d'Ivoire**

**Central African Republic**

**Afghanistan**

**Malawi**

**Sierra Leone**

**Guinea-Bissau**

**Benin**

**Nigeria**

**Equatorial Guinea**

It is advisable to give funding for stage 1 countries first since they are in dire need. Followed by Stage 2

Stage 3 Next Page

# Final List of Countries:

These are the stage 3 countries. They can be prioritized if any emergency/calamity occurs. Otherwise, they can wait a little longer.



'Iraq',  
'Tajikistan',  
'Kenya',  
'Comoros',  
'Lesotho',  
'Lao',  
'Zambia',  
'Mauritania',  
'Tanzania',  
'Sudan',  
'Guinea',  
'Yemen',  
'Botswana',  
'Uganda',  
'Senegal',  
'Kiribati',  
'Pakistan',  
'Cameroon',  
'Namibia',  
'Ghana',  
'Gabon',  
'South Africa'

## Concluding Remarks:

- Effectively 41 countries need aids.
- 31 are common while performing. Clustering with both the techniques.
- We understand, it is difficult to fund 41 countries at a time.
- We made a small outlier analysis and divided them into 3 funding stages