

# PCA & CLUSTERING ASSIGNMENT

Sandeep kalyan Sunkara  
+919618923739



## PROBLEM STATEMENT

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.

After the recent funding programs, they have been able to raise around \$ 10 million.

We need to categorize the countries using some socio-economic and health factors that determine the overall development of the country.

# HOW TO PROCEED



Data cleaning



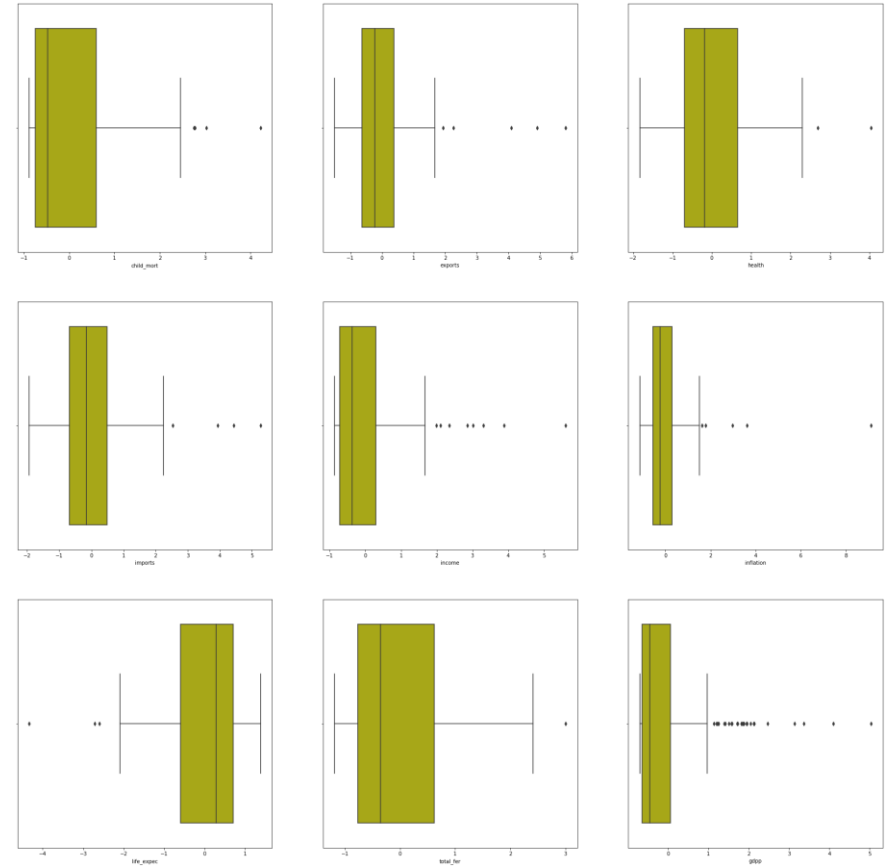
PCA & Clustering



Visualization

# DATA CLEANING

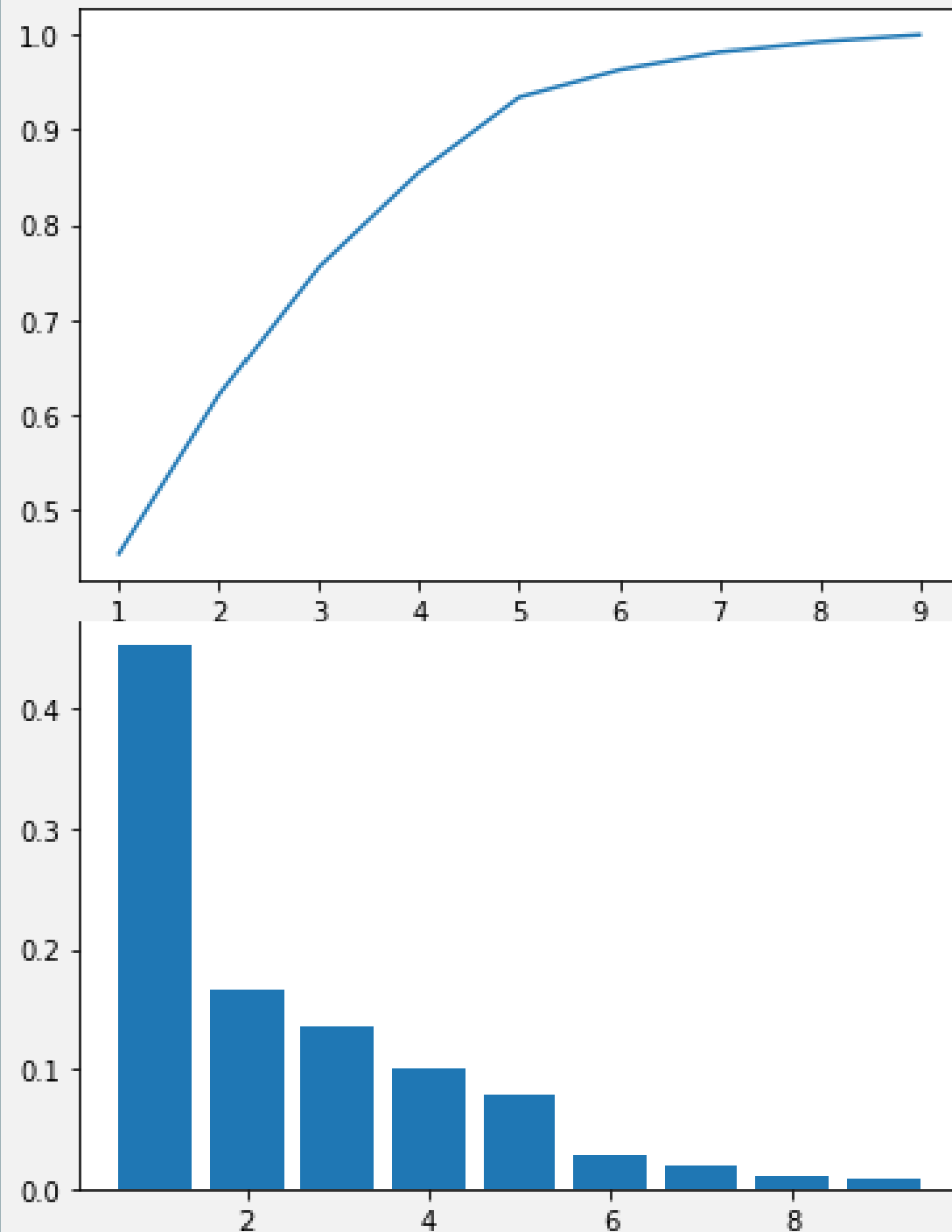
The data of 9 columns is visualized in boxplots to get the idea of the distribution involved and then scaled and vizualized again.



# PCA

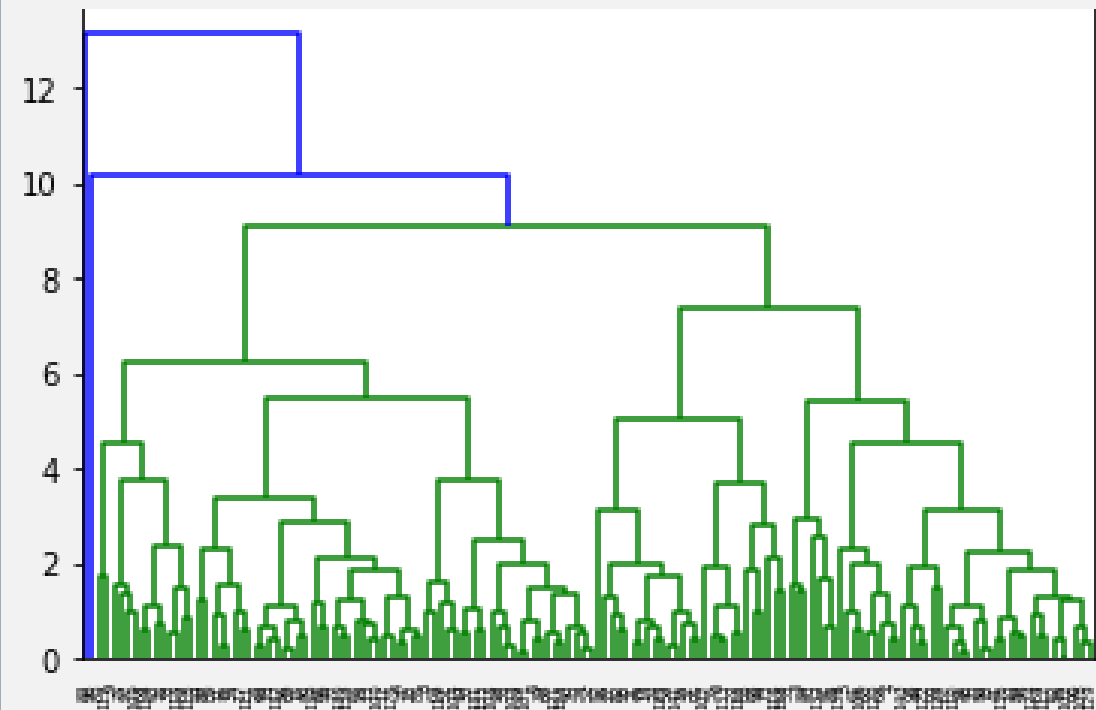
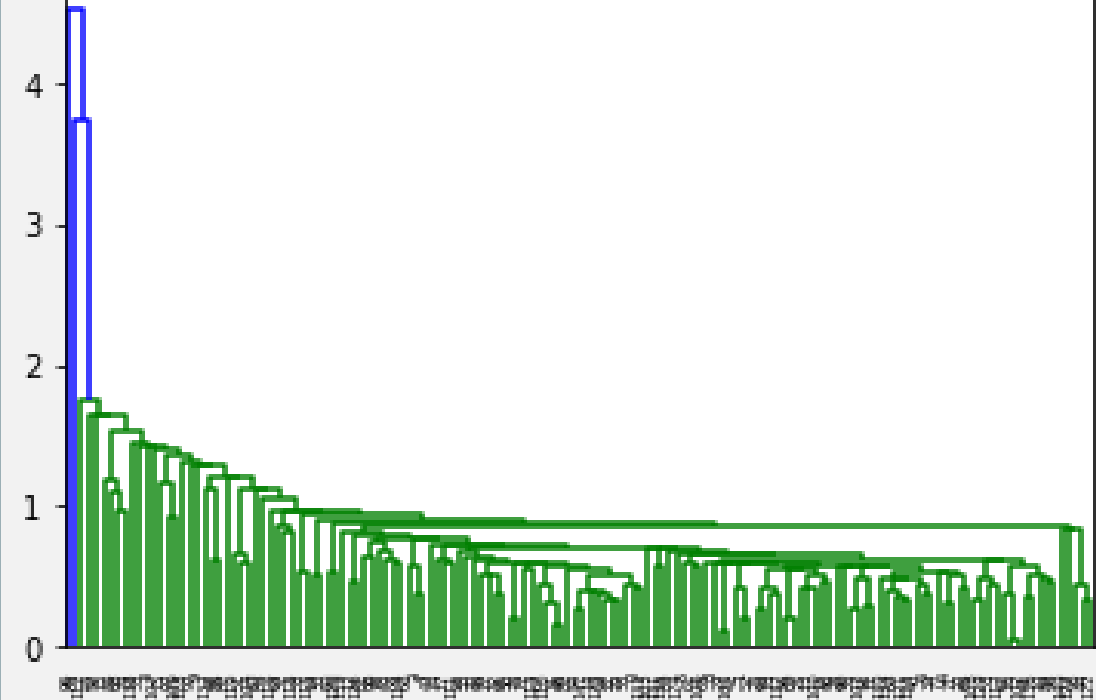
The PCA is done over a range of (1,10) to get the scree plot of the adequate number of PC's.

And we concluded that there would be 3 PC's which explain 75% of the data



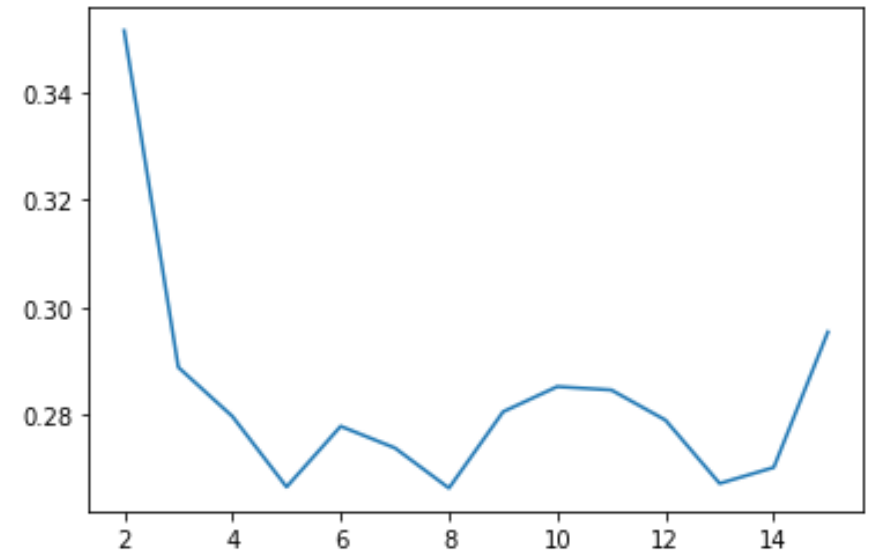
# CLUSTERING - HIERARCHY

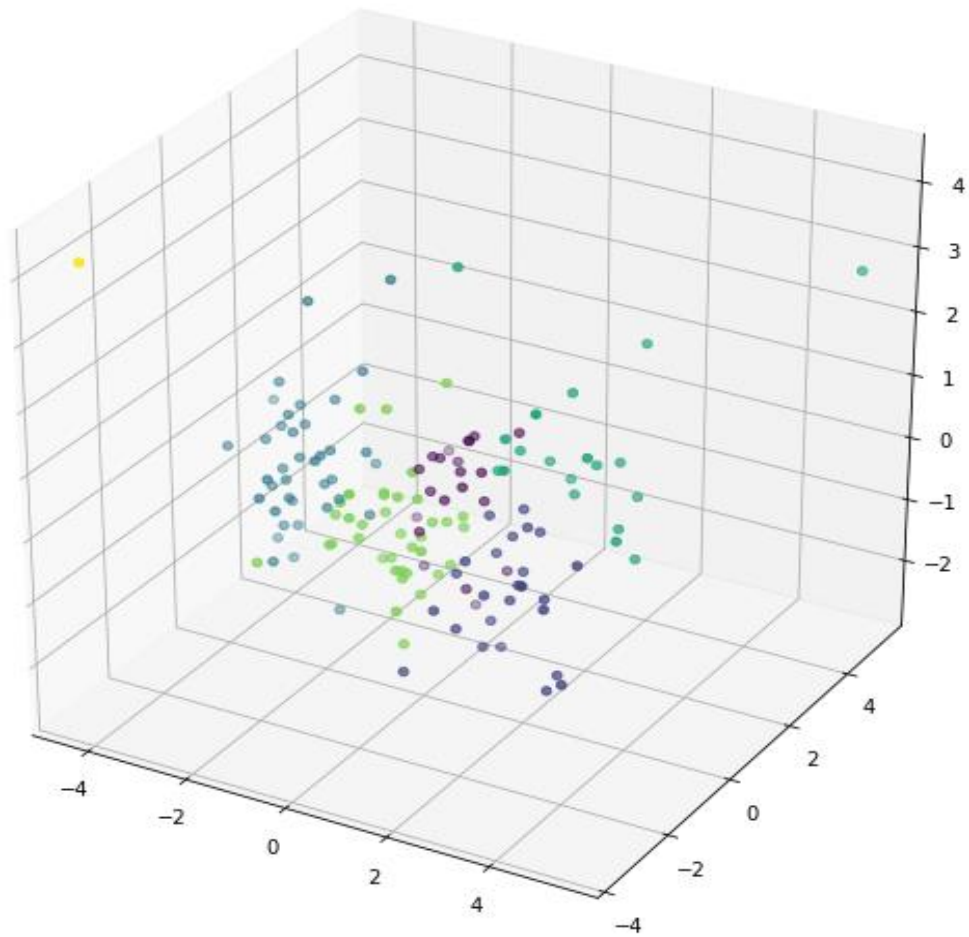
Then performed a hierarchical clustering, with both single and multiple linkage to get to know about the data distribution and data hierarchy.



## SILHOUETTE SCORE

Then performed a silhouette score analysis and concluded that the optimum number of clusters should be equal to 6.

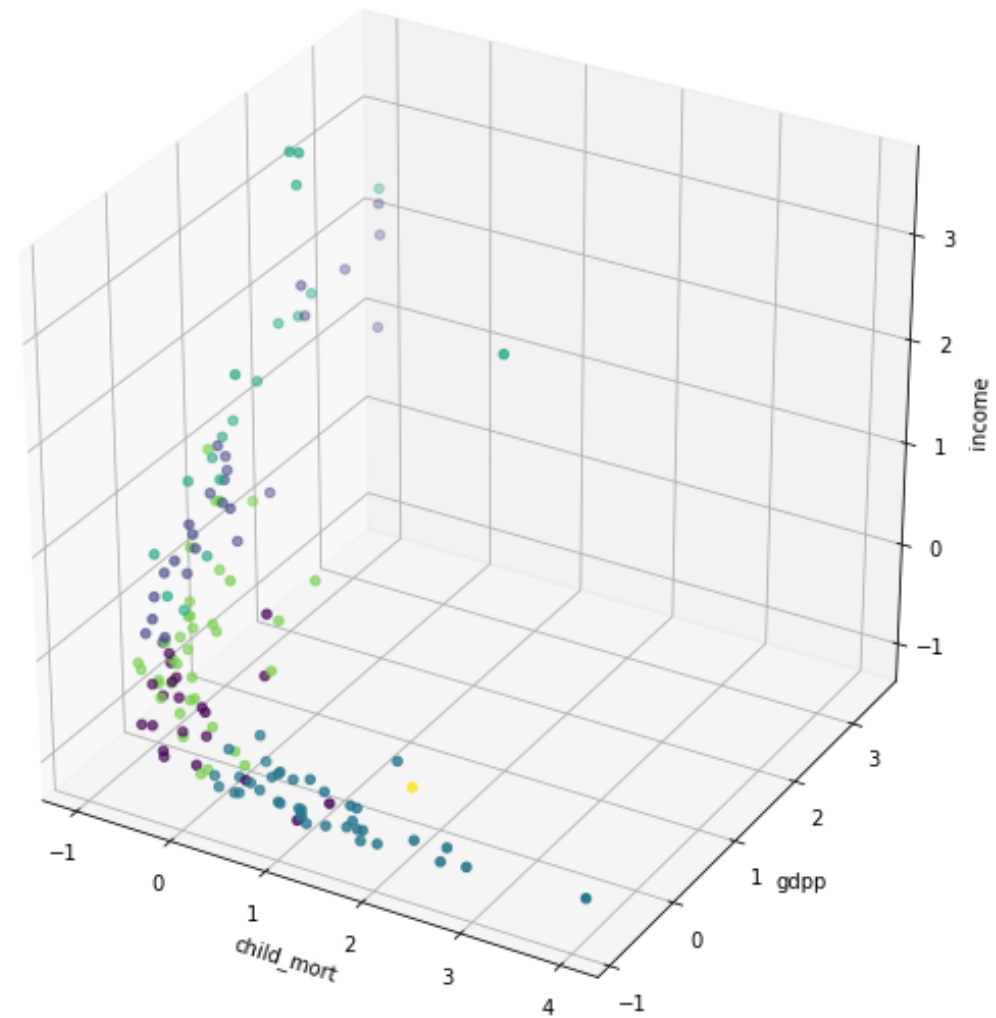




DATA ON PC-  
1, PC-2, PC-3



DATA VIZ BASED ON  
CLUSTERS FOR  
FACTORS –  
CHILD\_MORT, INCOME,  
GDPP



## FINAL LIST OF COUNTRIES FOR AID

### Code :

```
final_aid_country=list(aid_df.sort_values(by=["child_mort"],ascending=False)["country"].head(10))
```



Haiti  
Sierra Leone  
Chad  
Central African Republic  
Mali  
Niger  
Angola  
Congo, Dem. Rep.  
Burkina Faso  
Guinea-Bissau



THANK YOU