Clustering Assignment:
HELP International NGO

By: Rohan Bhosle

Agenda:

- Scenario and Problem Statement
- Solution / Approach
- Analysis in brief
- Result

Scenario:

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

Problem Statement:

- CEO of this organization needs to decide how to spend this money strategically and effectively, and they want to make a decision on choosing the countries that are in dire need of this money. Countries are need to be categorized into developed and under-developed based on socio-economic and health factors.

Solution / Approach:

- Decide algorithm
- Data cleaning and EDA
- Check cluster Tendency
- Identify value of 'k'
- Forming clusters
- Cluster profiling

Solution / Approach:

Decide algorithm

- In order to identify the countries which are in dire need of the money, we have used clustering algorithm to make a groups and then profile them based on our requirement.

Data Cleaning, EDA & Check Cluster Tendency

- Before we start building our model we did necessary data cleaning, EDA and checked the clustering tendency using Hopkin's statistics which found to be more than 94%.

Identify value of 'k'

- We choose 3 important features 'gdpp', 'child_mort' and 'income' for our model building and we first scaled them to optimize the performance of our algorithm.

Forming Clusters

- During model building, for clustering we used the Kmeans algorithm and using SSD (elbow method) we found the best value for 'k' as 3 which we used in Kmeans algorithm.
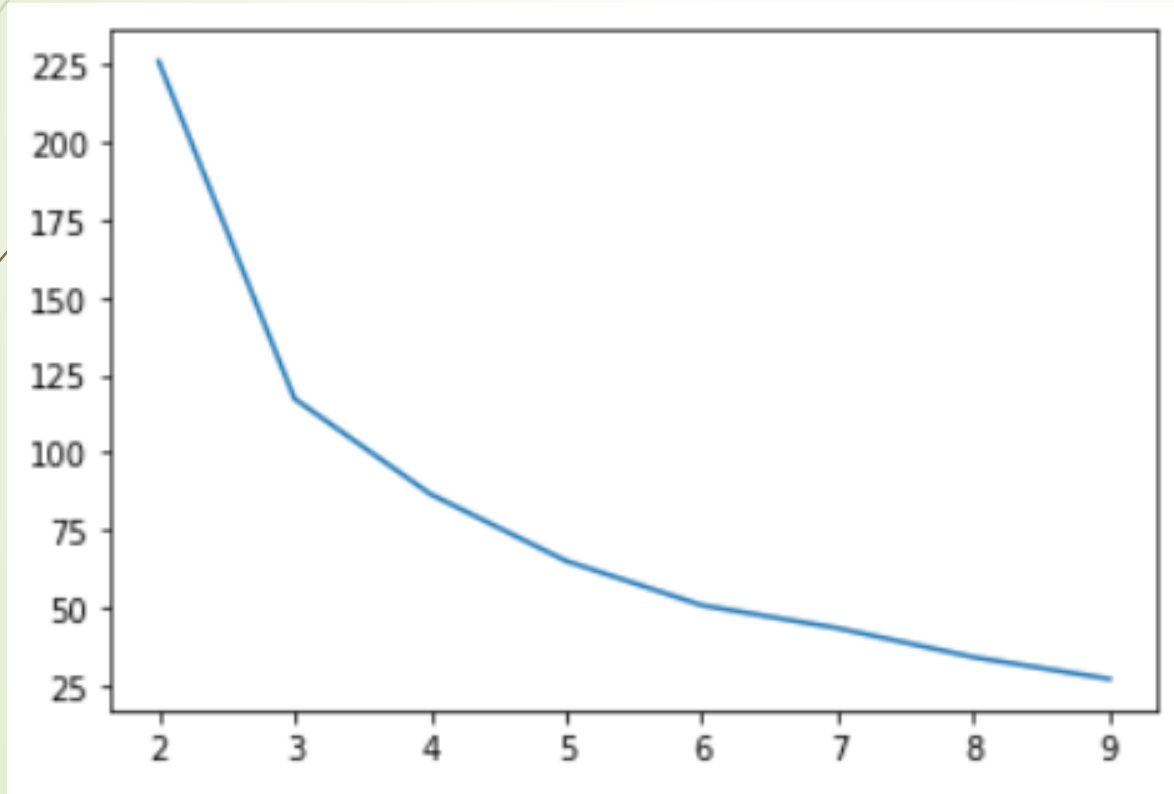
Cluster Profiling

- As we formed the 3 clusters, we now did cluster profiling in order to identify the right cluster which we need to target. We found cluster with label '0' and then sorted its data (gdpp: low to high, child_mort: high to low, income: low to high) in order to get top 5 countries in priority order which are in dire need of the aid.

Solution / Approach:

Few important visualizations we got during model building

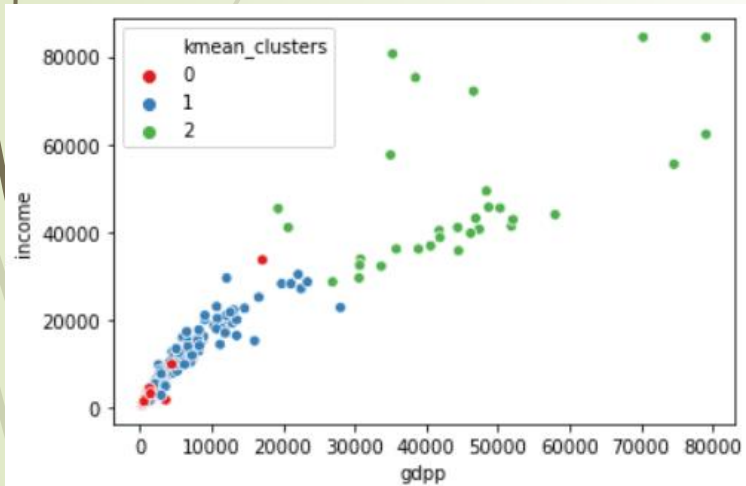■ **Finding the best value of 'k' using SSD (Elbow Method):**

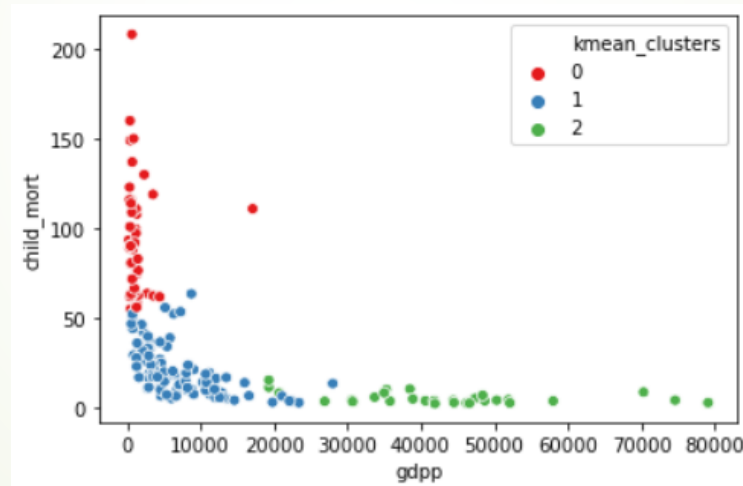Note: We can see from this plot that best value of 'k' should be 3

Solution / Approach:

Few important visualizations we got during model building

- **Visualizing cluster:**
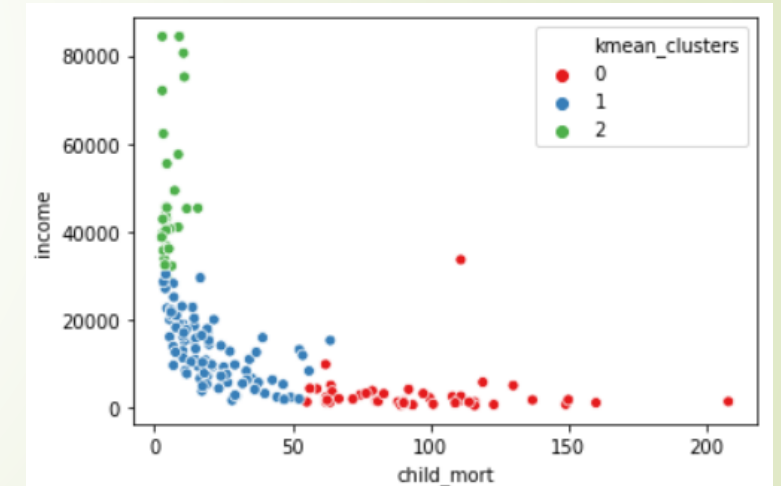
gdpp v/s income   gdpp v/s child_mort   child_mort v/s income
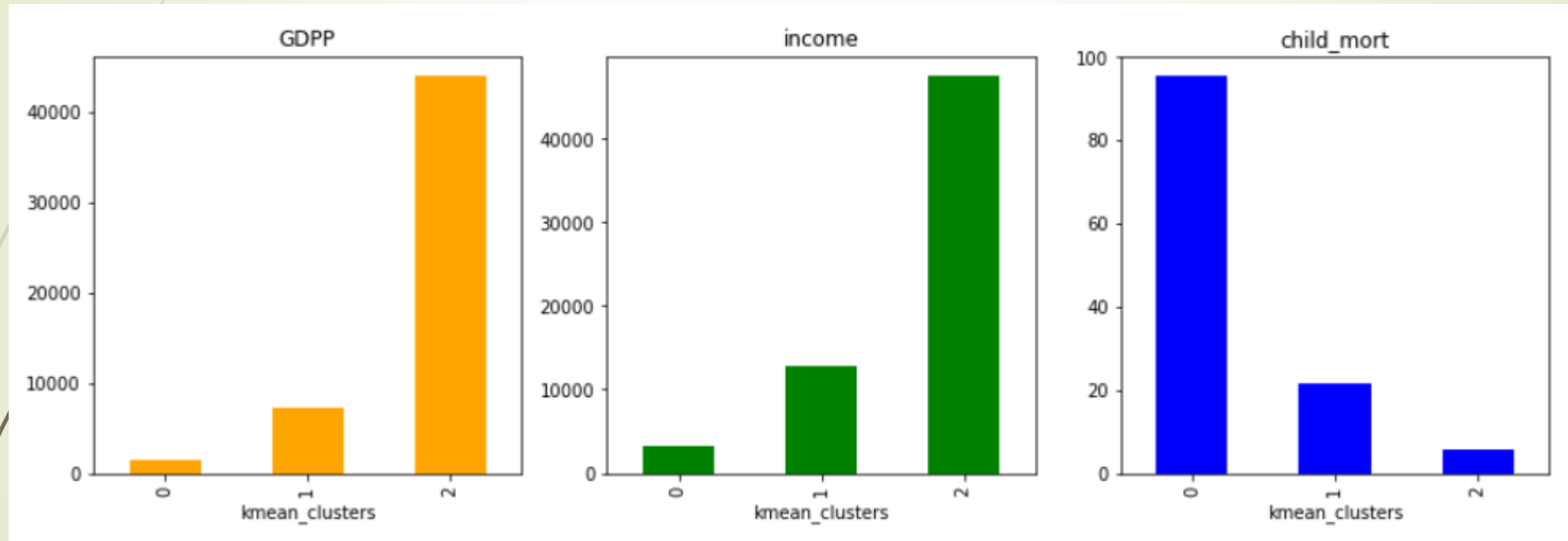

Note: We can clearly see that, 3 clusters are formed w.r.t. 3 variables i.e. gdpp, child_mort and income

Solution / Approach:

Few important visualizations we got during model building

▸ **Cluster Profiling:**

Note:
- *We can clearly see that we should be looking at the cluster label '0'*
- *As we need to identify those countries which are in dire need of the aid, we need to choose a cluster which has low gdpp, low income per person and high child mortality rate. So, we can conclude from the above bar chart that, cluster with 0 label is a clear winner here and we need to identify top 5 countries which are in a direst need of the aid.*

Result:

- After doing the cluster profiling, we sorted the data as gdpp: low to High, child_mort: High to Low, income: Low to High and we can see that below 5 countries are in the direst need of aid when we took the top 5 countries. We can report these countries to CEO of an NGO, so he can take appropriate actions.

- **Burundi**
- **Liberia**
- **Congo, Dem. Rep.**
- **Niger**
- **Sierra Leone**

Thank You