

Clustering and PCA Assignment

Submitted by: Satprem Rath

Problem Statement

- ▶ The problem statement was to find the list of countries which are in dire need of aid so that Help International can help these countries with its funding programs

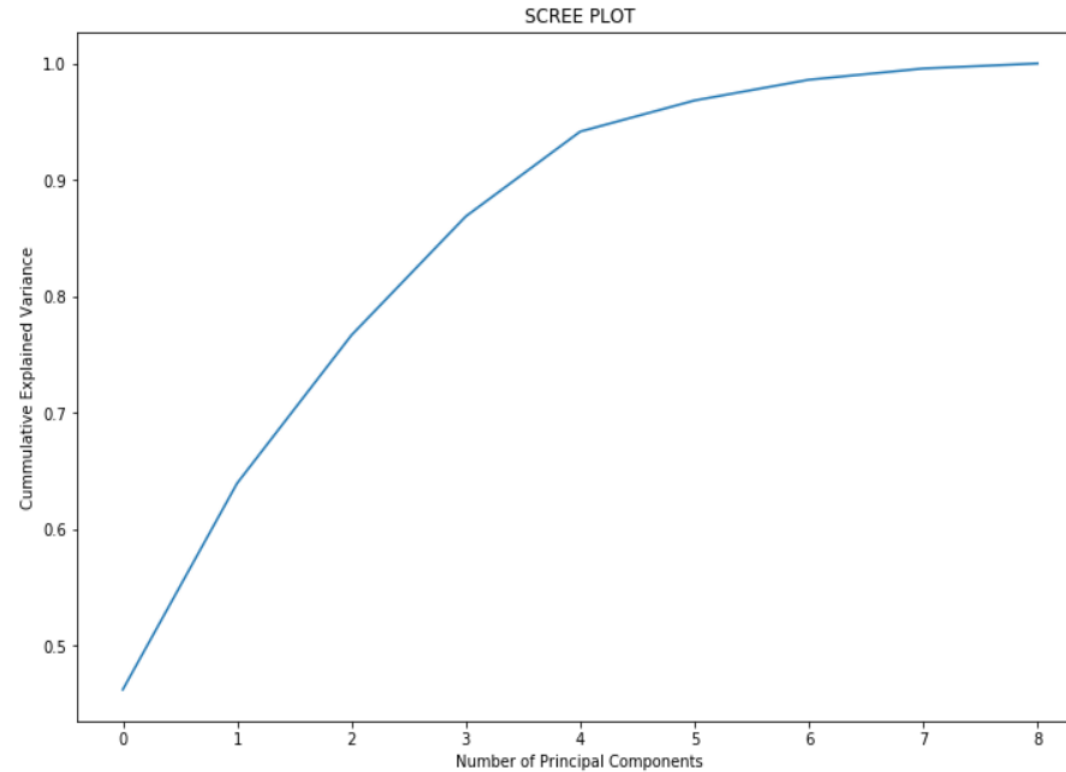
Outlier checks

Performed the outlier checks and removed some of the data which are at higher cut offs ie. Gdpp and income are higher than 95 percentile. These data are of no significance to our analysis as they belong to very highly developed countries having too high GDP and Net income per person.

After outliers removal we are left with 150/167 rows with 9 features for further analysis. It is around 90% of the original data suitable for analysis.

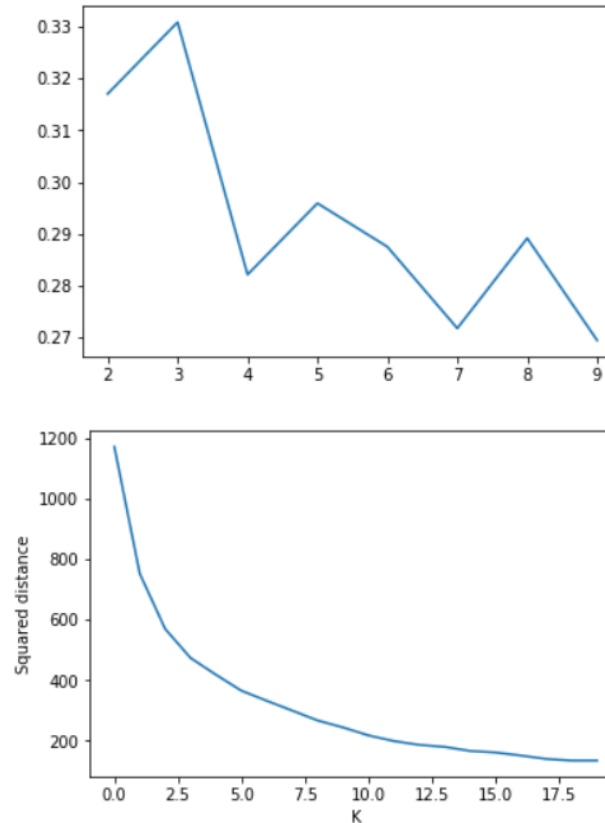
Feature Selection

- ▶ We can see that with 3 to 4 features we can explain 90-95% of the data.



Cluster classification

- ▶ We determined the number of clusters which are required to explain the data using Silhouette score and Sum of squares.
- ▶ We concluded that 3 clusters would be good to explain the dataset.



Key Features

► These are key features to find the countries with dire need of aid:

1. GDP
2. Inflation
3. Health
4. Child mortality
5. income

List of Countries

- ▶ The list of countries which have dire need of aid are:
Albania, Bosnia and Herzegovina, China, Macedonia, FYR, Tunisia.

Solutions recommended

- ▶ Arrangement of good medical and health camps in order to check the child mortality and increase the fertility rate
- ▶ Special skill jobs to improve the human resource in these countries to raise the income per person and improve the GDP.
- ▶ Special fundings to below poverty line people for basic life amenities.
- ▶ Education to poor families.
- ▶ Provide basic health facilities.