PCA and Clustering Assignment

By:

Balakrishna Gadiyar

Abstract :

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. It runs a lot of operational projects from time to time along with advocacy drives to raise awareness as well as for funding purposes.

After the recent project that included a lot of awareness drives and funding programmes, they have been able to raise around $ 10 million

Business Objectives :

The **Objective** of analysis is to ,
- ✓ Categorize the countries based on socio economic and health factors on its development
- ✓ Suggest countries which CEO needs to focus on the most for his investment.

Approach

- Understand the raw data and business need
- Import and perform EDA
- Visualize the correlations and get further understanding on the data
- Perform SVD on the Dataframe and view the outcomes
- Conduct PCA - Scale and transform, conduct outlier analysis, remove outliers
- Review variance ratio, see how many components cover 80-85%.
- Using hopkins statistics see if the clustering needed
- Considering the components derived perform silhouette score and SSD analysis
- Analyze the clusters formed using k means and hierarchical approach
- Review/Visualize Principal components and Origial features to understand the key components that influence.
- Using Kmeans+ predict the segmentations using identified clusters
- Analyze segments on derived features by taking means on the segments and see if they really fall in correct groups
- Visualize the segmentation to arrive at conclusion
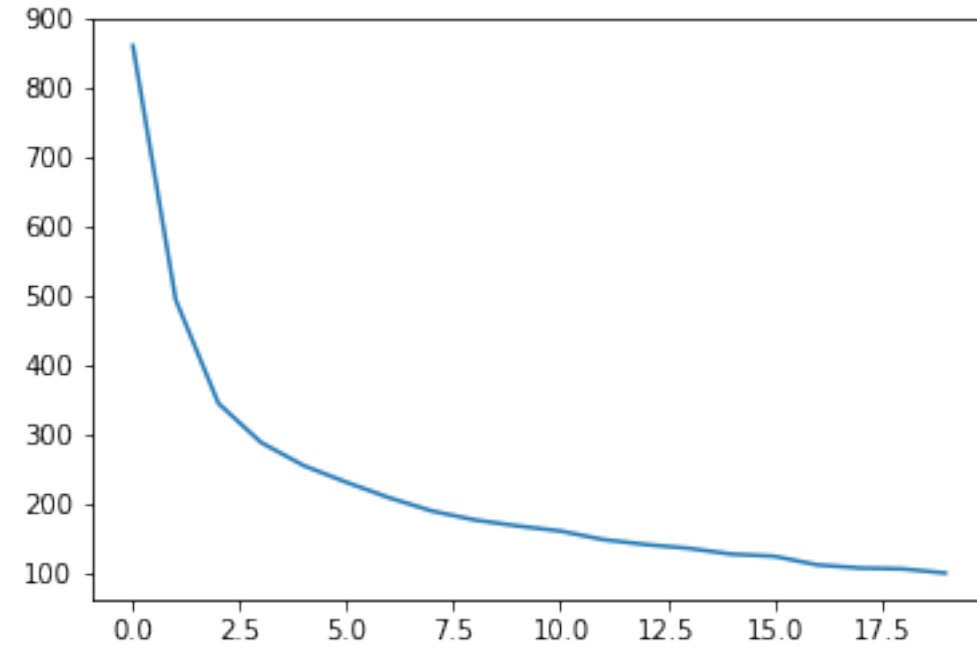- Filter countries from original dataset based on key features and suggest

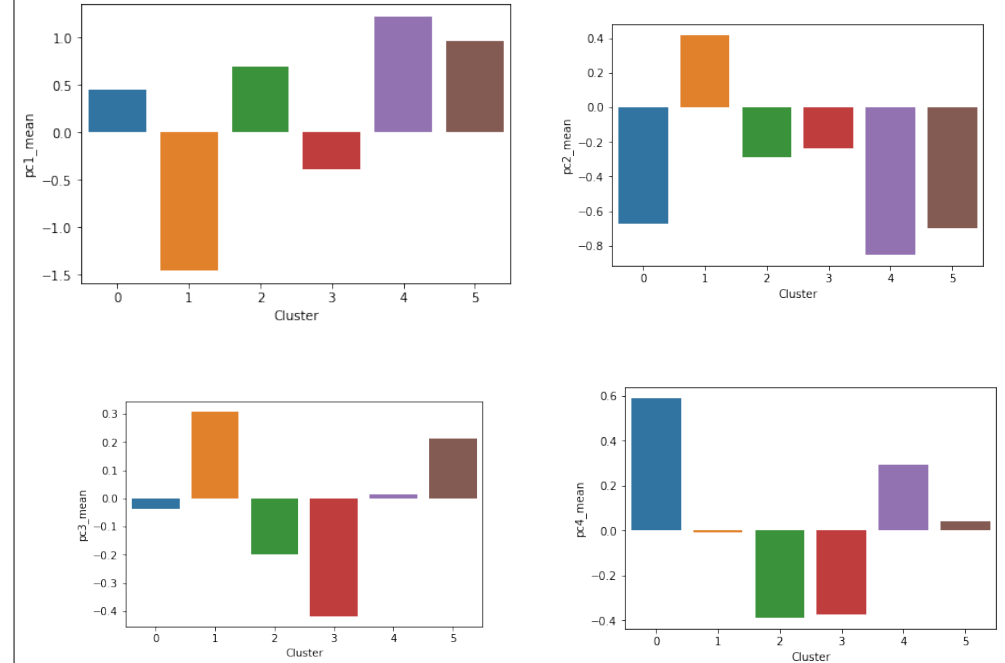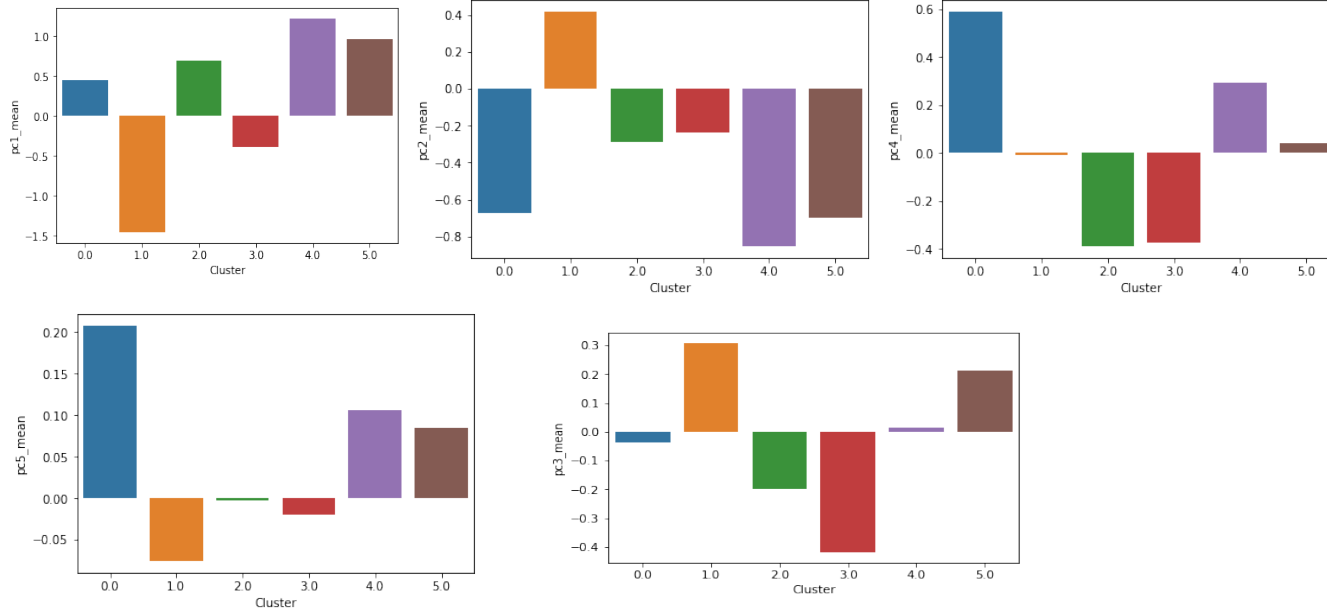# Correlations among origial features
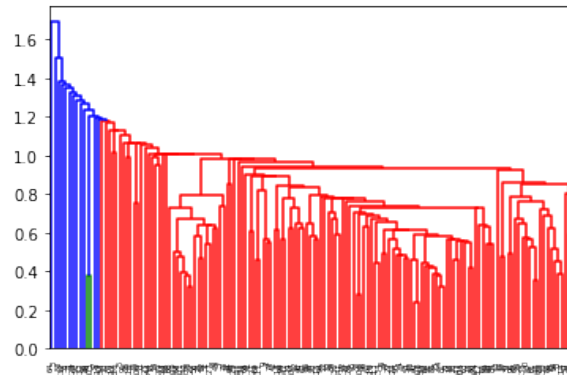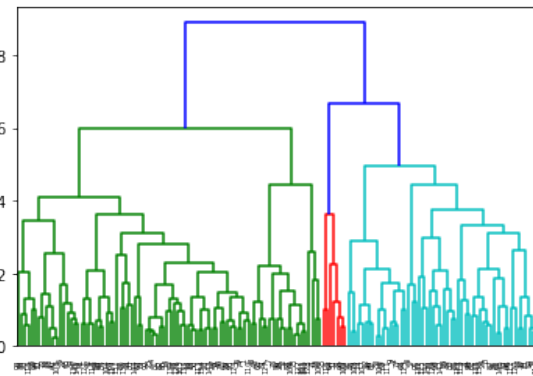
PCA – SSD/ Silhoute analysis

Silhoute

SSD

K Means/Hierarchical Analysis

K Means

Hierarchical Analysis

PC Vs Original Features
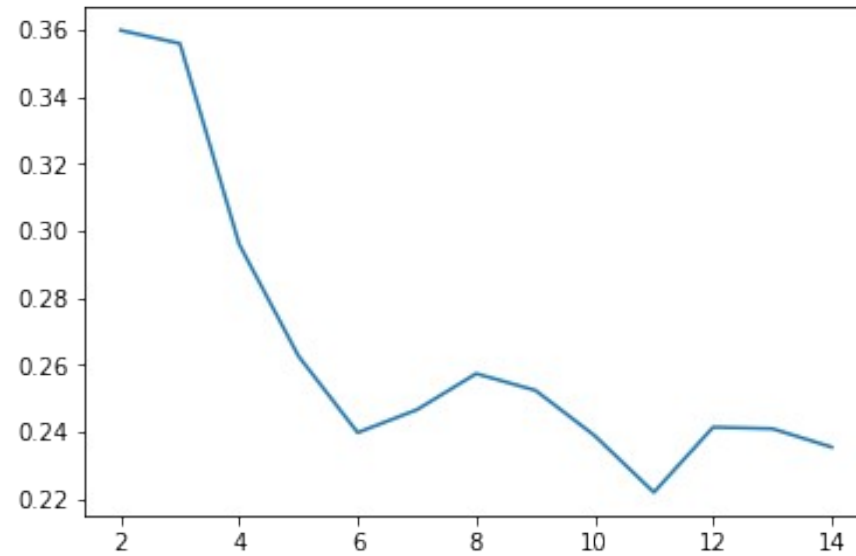
Heatmap
Original Features
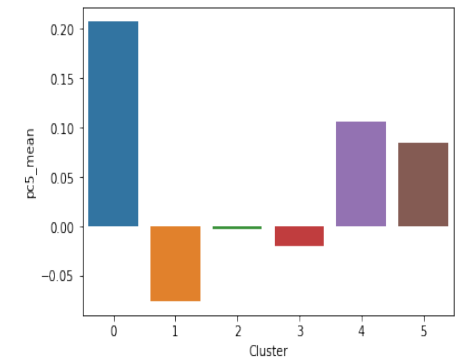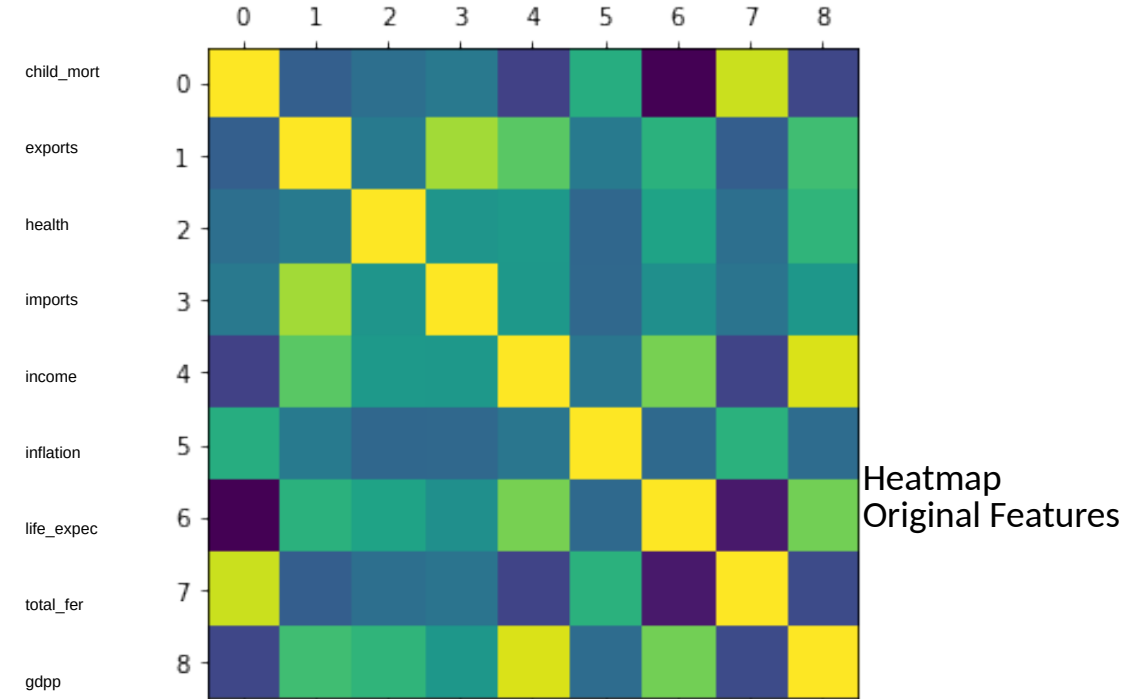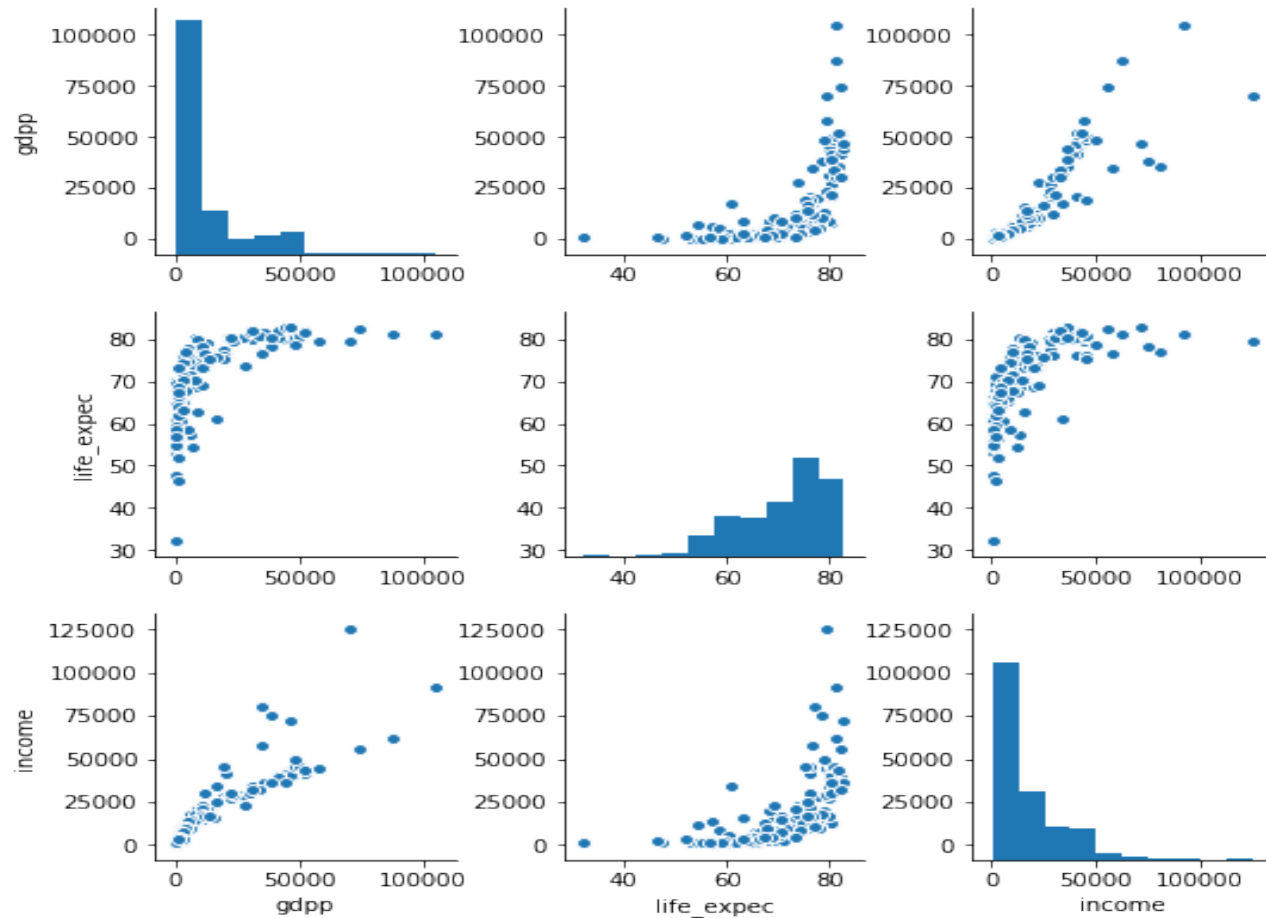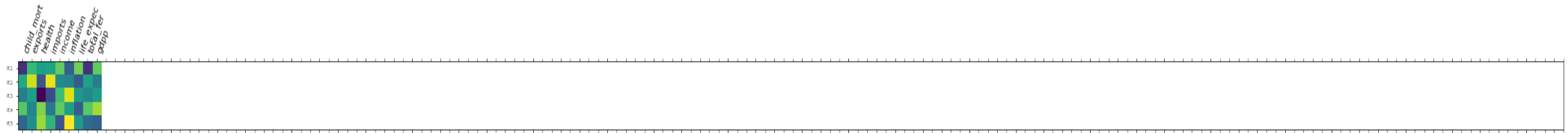
Feature Analysis

Conclusion

- Countries are divided into 3 categories
  - Poor
  - Developing
    Rich
  Primary features that are considered for decision making are
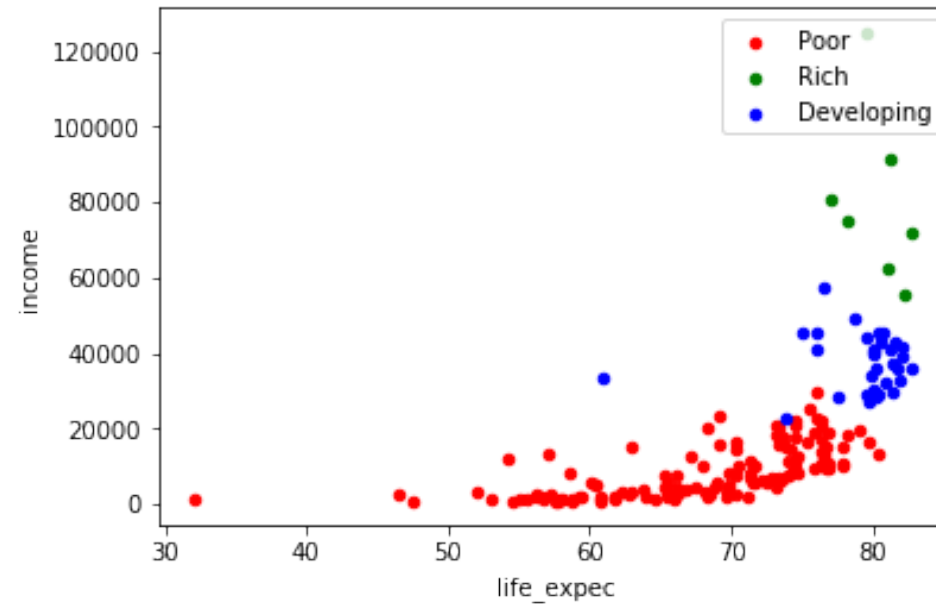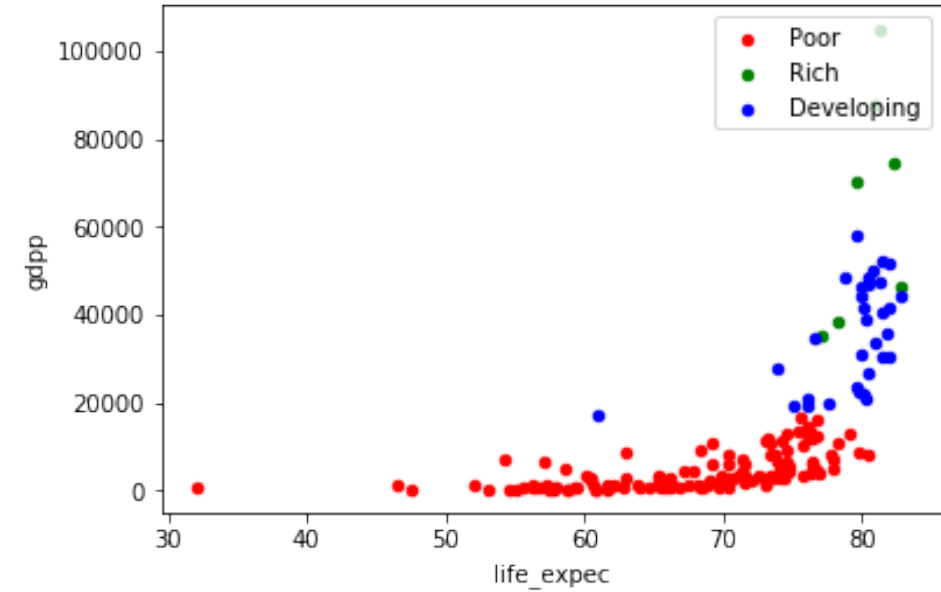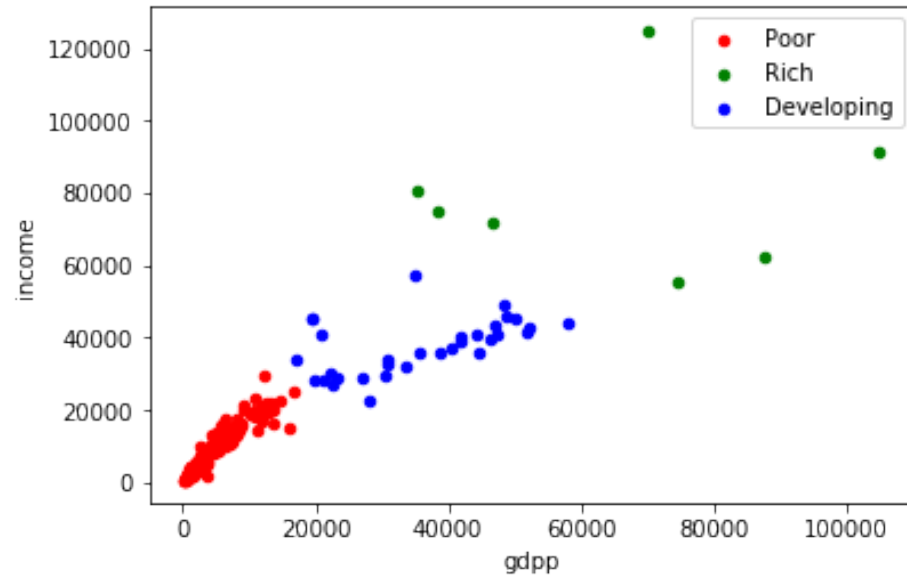 - Income
  - Gdp
  - life expectancy
- Primary investing segment should be poor counties whose gdp, income and life expectancy are low
- Child mortality feautre is also directly relates to segrement, specially poor countries.

Segment	country	gdpp	life_expec	income	child_mort
Poor	Burundi	231	57.7	764	93.6
Poor	Liberia	327	60.8	700	89.3
Poor	Congo, Dem. Rep.	334	57.5	609	116
Poor	Niger	348	58.8	814	123
Poor	Sierra Leone	399	55	1220	160
Poor	Madagascar	413	60.8	1390	62.2
Poor	Mozambique	419	54.5	918	101
Poor	Central African Republic	446	47.5	888	149
Poor	Malawi	459	53.1	1030	90.5
Poor	Eritrea	482	61.7	1420	55.2
Poor	Togo	488	58.7	1210	90.3
Poor	Guinea-Bissau	547	55.6	1390	114
Poor	Afghanistan	553	56.2	1610	90.2
Poor	Gambia	562	65.5	1660	80.3
Poor	Rwanda	563	64.6	1350	63.6
Poor	Burkina Faso	575	57.9	1430	116
Poor	Nepal	592	68.3	1990	47
Poor	Uganda	595	56.8	1540	81
Poor	Guinea	648	58	1190	109
Poor	Haiti	662	32.1	1500	208