

ASSIGNMENT

Clustering & PCA

BY:

1. PRAJWAL RAO

Business Objective

Business Objective

- ❑ The aim of analysis is to identify and group countries by socio-economic and health factors to choose the countries that are in the direst need of aid.

Approach

- ❑ problems are systematically solved using EDA ,data visualization and clustering.



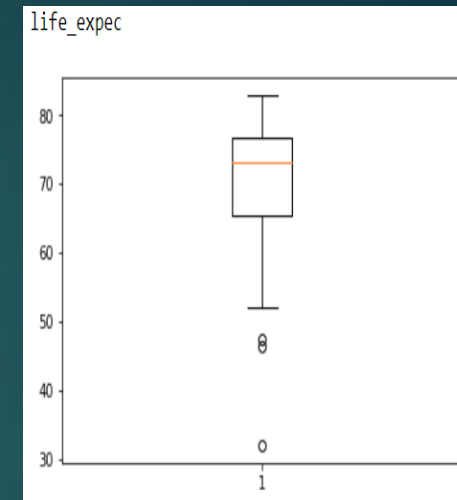
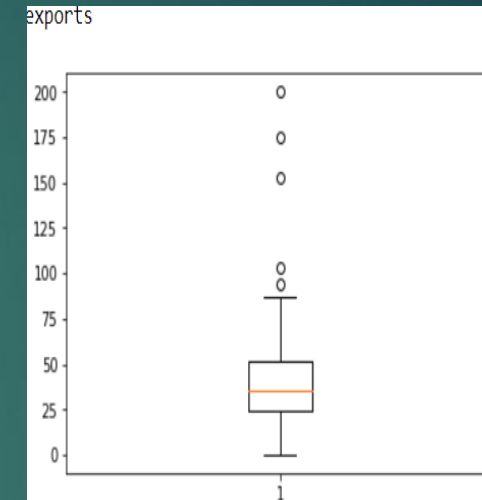
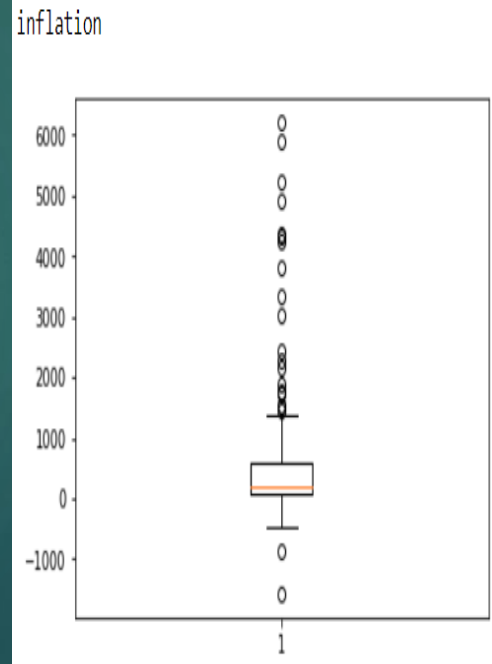
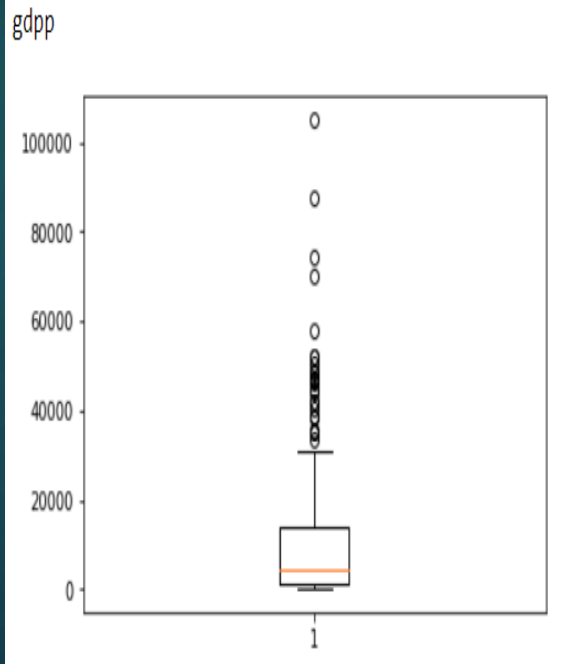
Data exploration

- ❑ **167** Countries
- ❑ **9** metrics (columns) available for socio-economic and health .
- ❑ No null values in the dataset.
- ❑ Convert inflation into absolute terms.

	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
count	167.0000	167.0000	167.0000	167.0000	167.0000	167.0000	167.0000	167.0000	167.0000
mean	38.2701	41.1090	6.8157	46.8902	17,144.6886	7.7818	70.5557	2.9480	12,964.1557
std	40.3289	27.4120	2.7468	24.2096	19,278.0677	10.5707	8.8932	1.5138	18,328.7048
min	2.6000	0.1090	1.8100	0.0659	609.0000	-4.2100	32.1000	1.1500	231.0000
25%	8.2500	23.8000	4.9200	30.2000	3,355.0000	1.8100	65.3000	1.7950	1,330.0000
50%	19.3000	35.0000	6.3200	43.3000	9,960.0000	5.3900	73.1000	2.4100	4,660.0000
75%	62.1000	51.3500	8.6000	58.7500	22,800.0000	10.7500	76.8000	3.8800	14,050.0000
max	208.0000	200.0000	17.9000	174.0000	125,000.0000	104.0000	82.8000	7.4900	105,000.0000

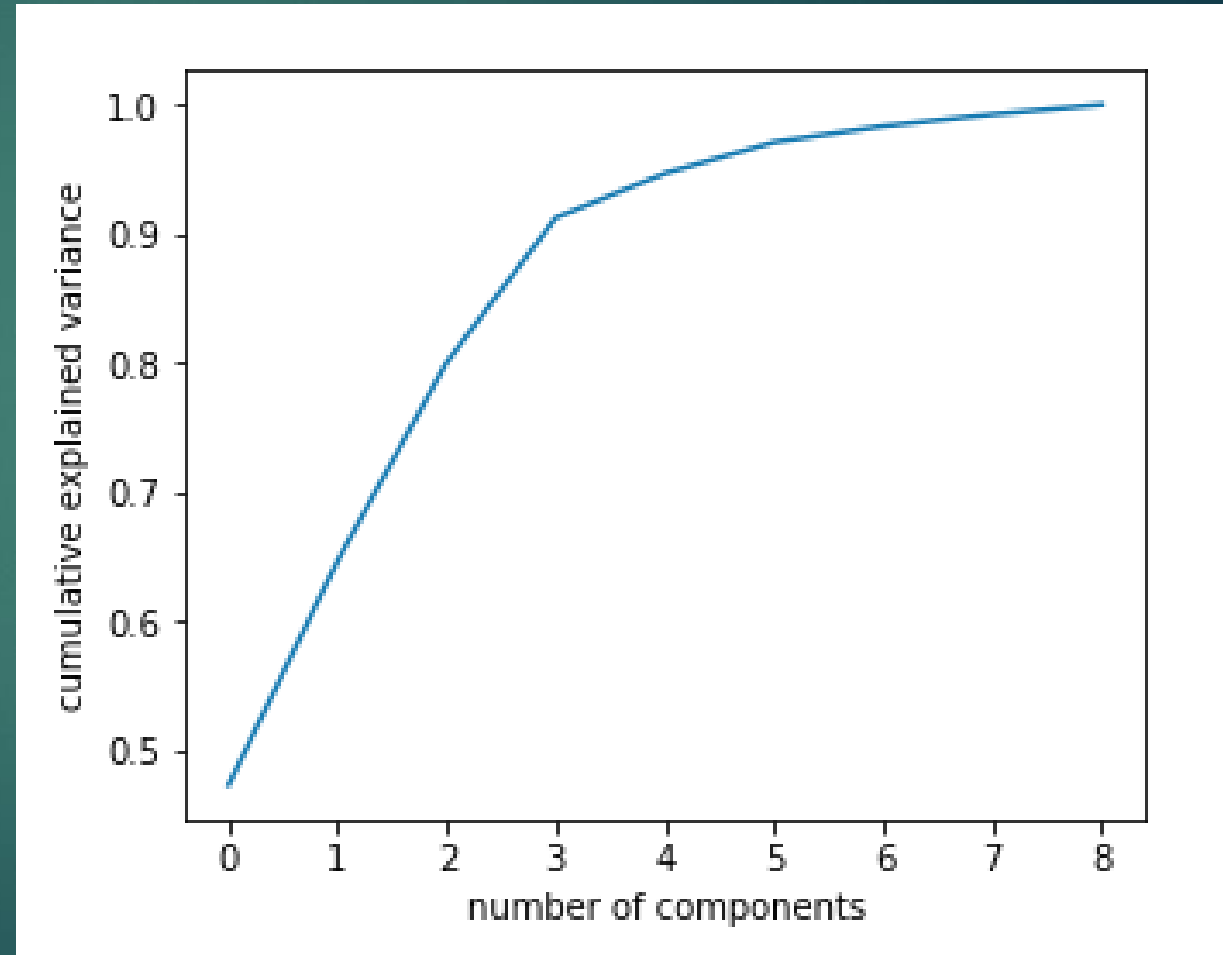
Data Analysis

- Outlier Countries detected in the data set and must be removed before clustering
- Few correlated columns detected and PCA analysis is don't to create independent columns

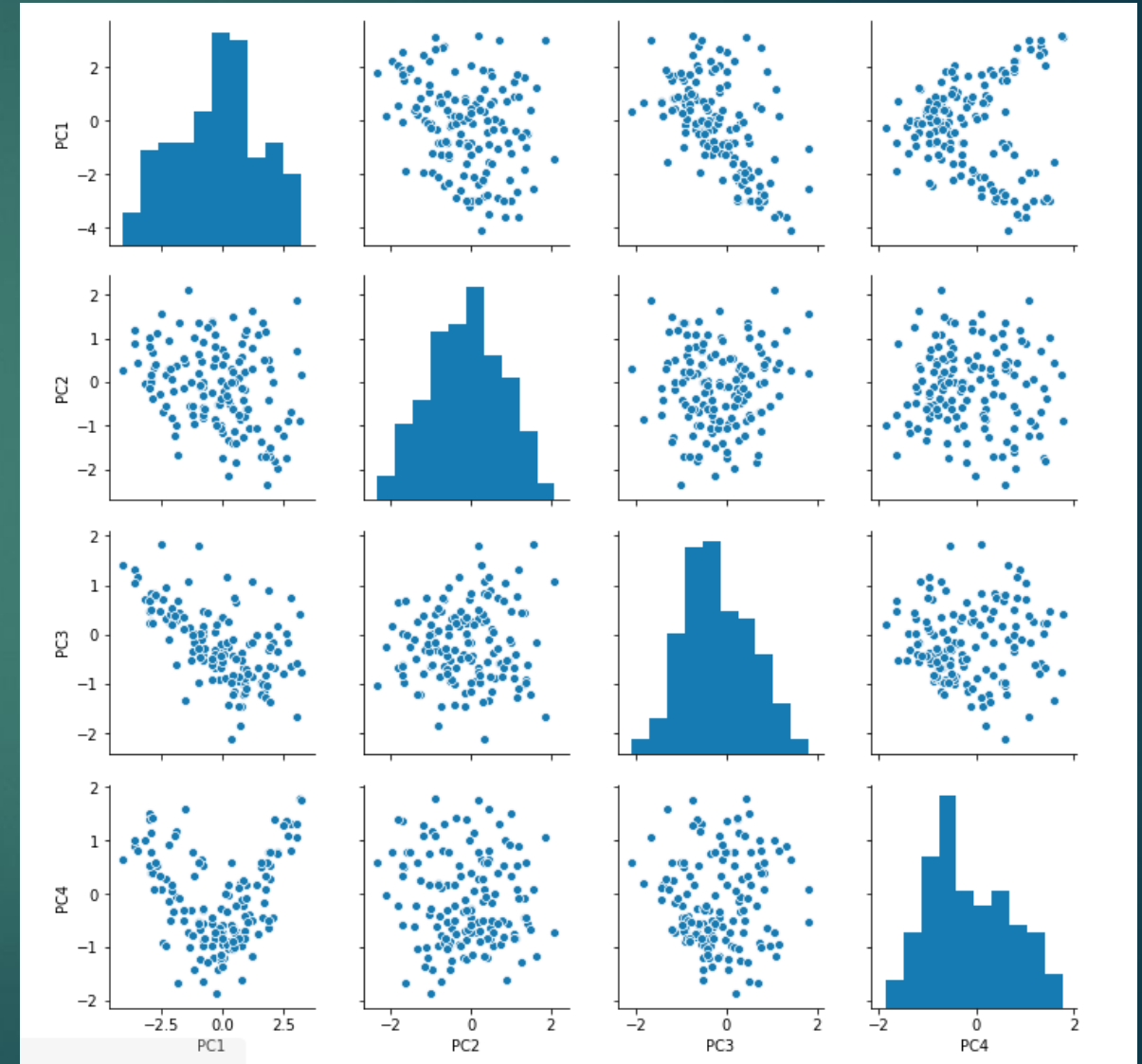
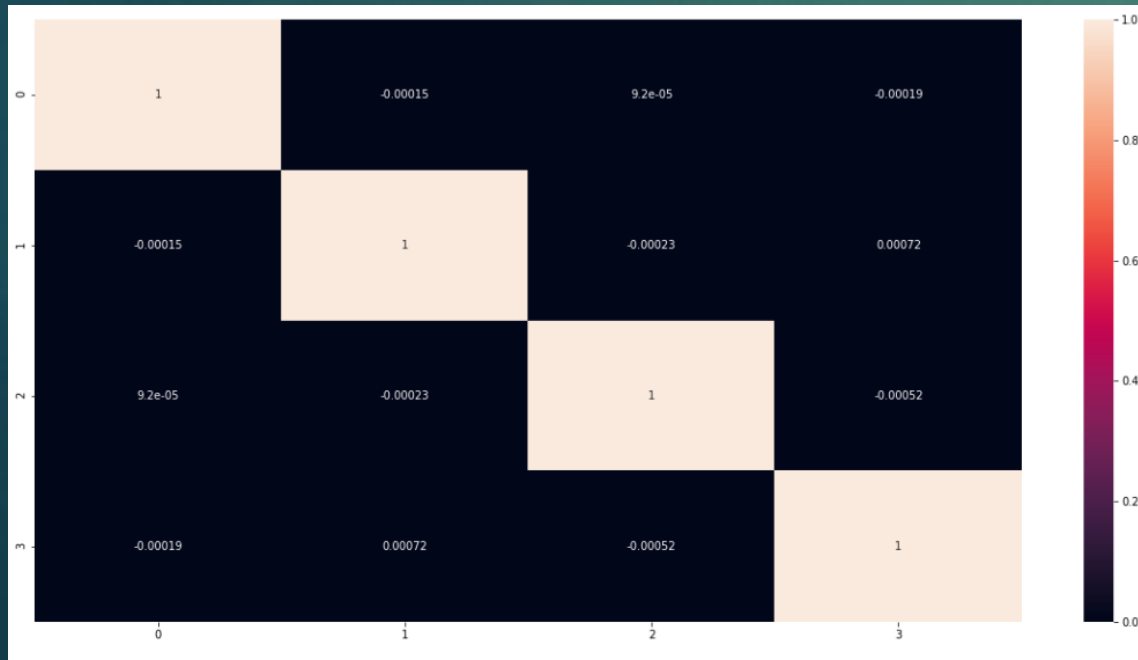


	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
child_mort	1	-0.318093	-0.200402	-0.127211	-0.524315	-0.173375	-0.886676	0.848478	-0.483032
exports	-0.318093	1	-0.114408	0.737381	0.516784	0.227577	0.316313	-0.320011	0.418725
health	-0.200402	-0.114408	1	0.0957167	0.129579	-0.176408	0.210692	-0.196674	0.345966
imports	-0.127211	0.737381	0.0957167	1	0.122406	-0.13675	0.0543905	-0.159048	0.115498
income	-0.524315	0.516784	0.129579	0.122406	1	0.60806	0.611962	-0.50184	0.895571
inflation	-0.173375	0.227577	-0.176408	-0.13675	0.60806	1	0.219201	-0.123937	0.441413
life_expec	-0.886676	0.316313	0.210692	0.0543905	0.611962	0.219201	1	-0.760875	0.600089
total_fer	0.848478	-0.320011	-0.196674	-0.159048	-0.50184	-0.123937	-0.760875	1	-0.45491
gdpp	-0.483032	0.418725	0.345966	0.115498	0.895571	0.441413	0.600089	-0.45491	1

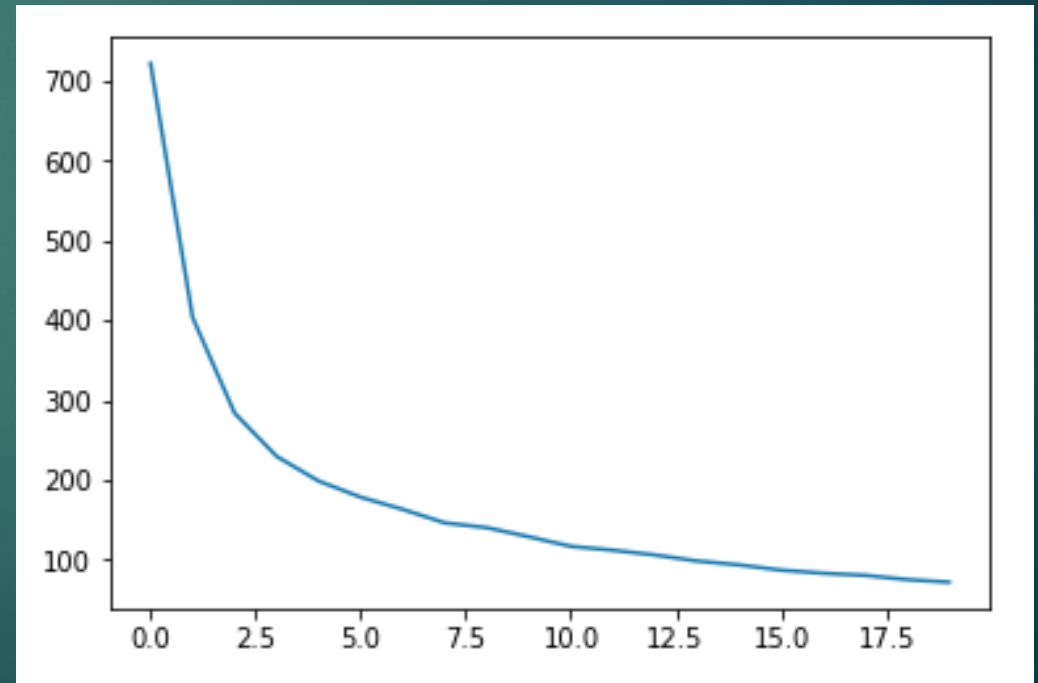
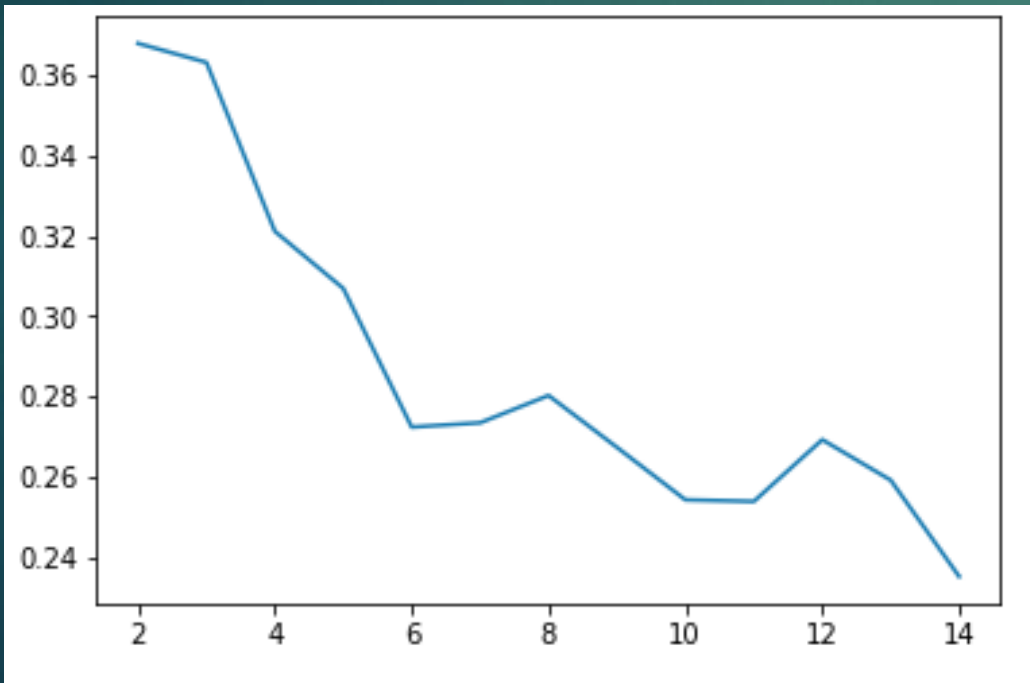
- Performed standardization of data to perform PCA
- Scree analysis shows 4 components are enough to describe 95% of the dataset and its variance.



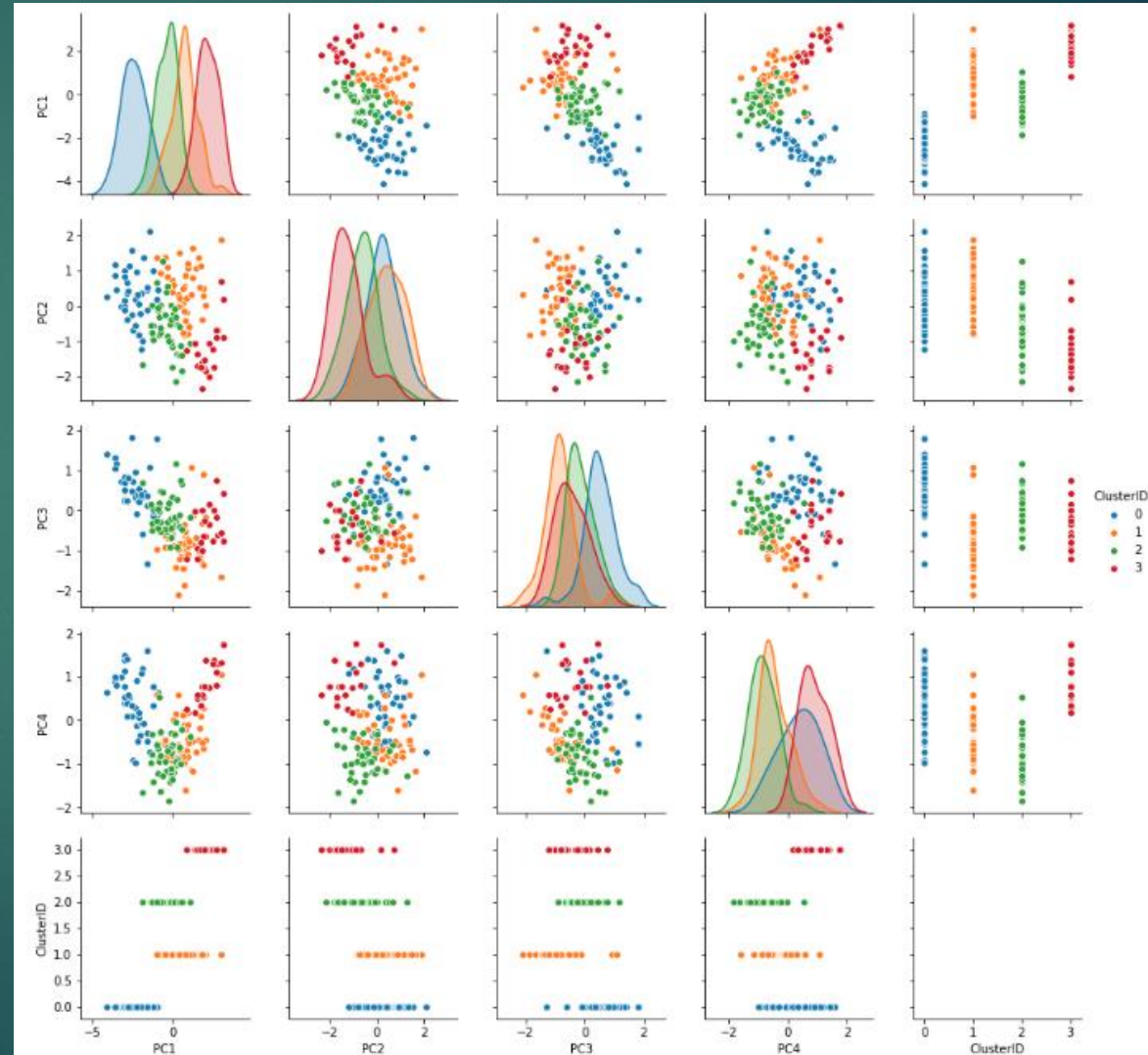
- PCA scatter plots
- After removing outlier
- PCA analysis results in independent columns



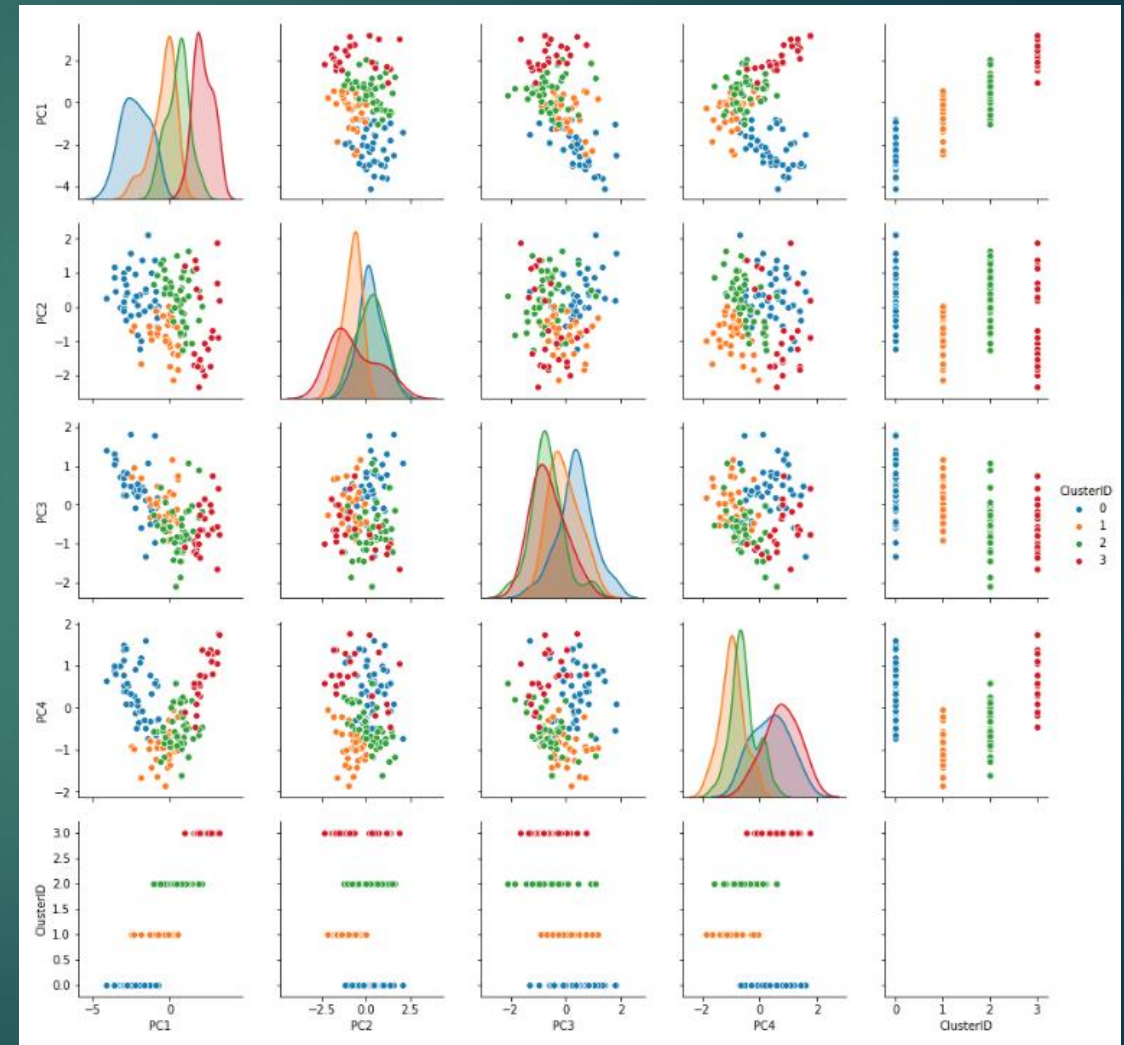
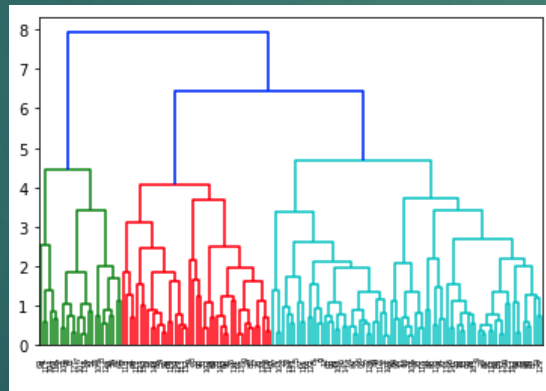
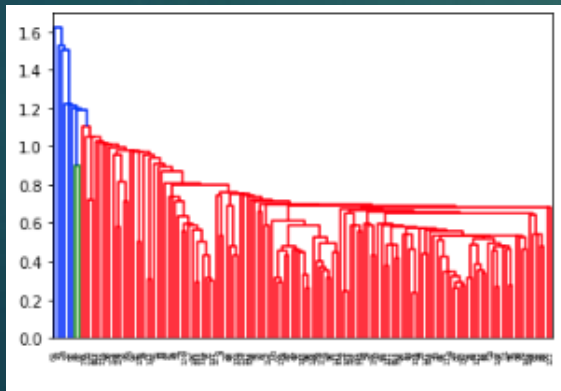
- **Optimum number** of cluster is found to be **4** by **Silhouette Analysis** and **Elbow curve**
- **3 cluster results in good clustering but 4 is chosen to target the countries with the most dire need of help**



- Optimum number of cluster is found to be **4** by **Silhouette Analysis** and **Elbow curve**



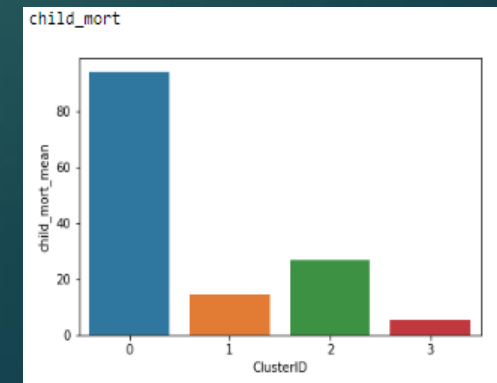
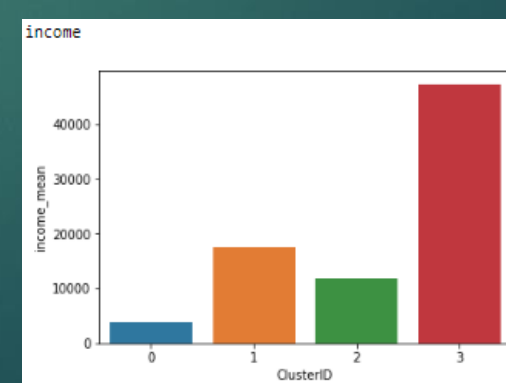
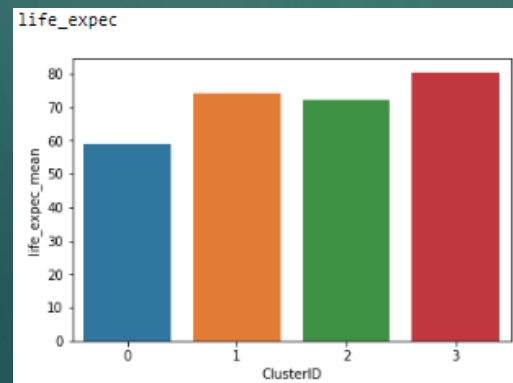
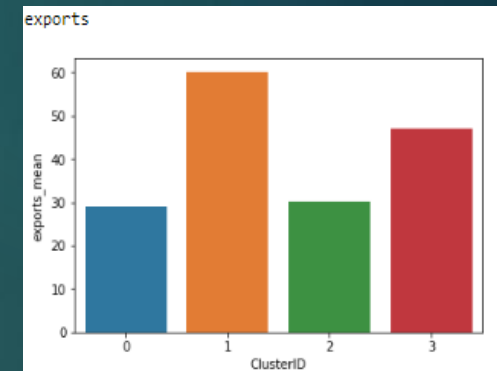
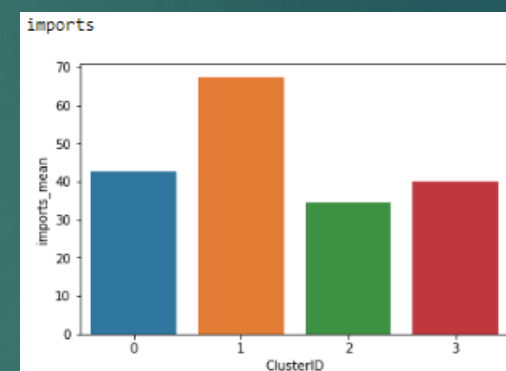
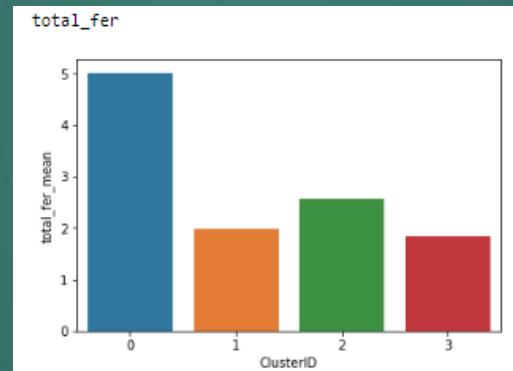
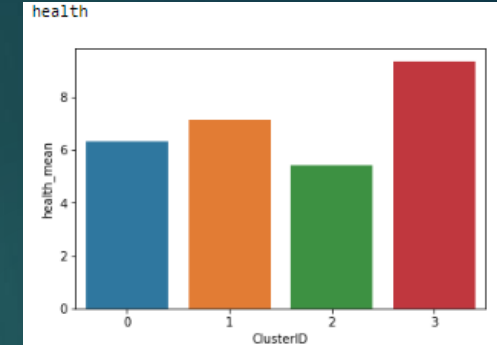
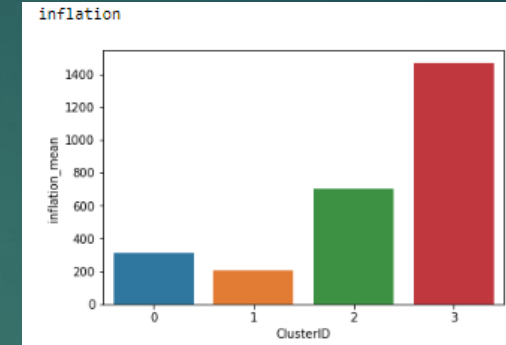
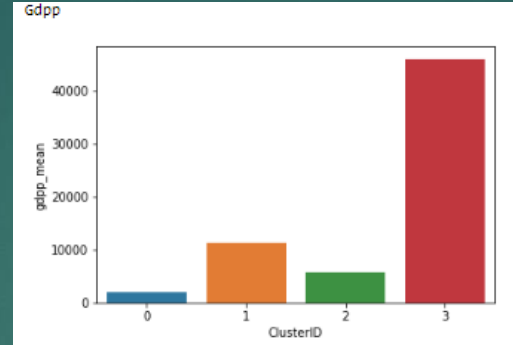
- **Optimum number of cluster** is found to be **4** by **Silhouette Analysis and Elbow curve**
- K means and hierarchical clustering give same results



*actual cluster number will change on model rerun

Key Characteristics:

1. Cluster 0:
 - highest fertility ,child mortality.
 - Lowest income, gdp, exports, life expectancy.
 - Very low inflation, health, imports.
 - **The countries in this cluster are in dire need.**
2. Cluster 1:
 - High health , imports , exports , income
 - Low inflation child mortalities
3. Cluster 2:
 - Countries rising from poverty
 - **Mainly constitute the developing countries with India and china.**
4. Cluster 3:
 - Highest health , and socio-economic factors
 - **Developed** countries



- HELP International must invest in the following countries
- countries that are in the direst need of aid are as follows:

Afghanistan	Lao	Gambia
Angola	Lesotho	Tanzania
Benin	Liberia	Timor-Leste
Botswana	Madagascar	Togo
BurkinaFaso	Malawi	Uganda
Burundi	Mali	Yemen
Cameroon	Mauritania	Zambia
CentralAfricanRepublic	Mozambique	Ghana
Chad	Namibia	Guinea
Comoros	Niger	Guinea-Bissau
Congo,Dem.Rep.	Nigeria	Haiti
Congo,Rep.	Pakistan	Kenya
Coted'Ivoire	Rwanda	Kiribati
EquatorialGuinea	Senegal	SouthAfrica
Eritrea	SierraLeone	Sudan
Gabon		