GDP and Economic groups

Presented by: Sahana Muralidaran (21076516)

Course: MSc. Data Science (SW)

GitHub Link: https://github.com/sahana248/ads1_assignment3.git

Introduction

Country's development is affected by 10 main factors: Agriculture, International Trade, Culture, Economy, Education, Employment, Industrialization, Infrastructures, Politics, Health. Based on this we have taken a few factors for analysis like, Production of CO2 per capita, forest area, GDP, Percentage of renewable energy out of total energy usage, Mortality rate of children under 5 years of age, unemployment, urban population and access to electricity.

A countries economic growth is estimated using its GDP (Gross Domestic Product). Based on this, the countries are divided into 4 main categories: High income, High middle income, Low middle income and Low income. Based on these classifications, the data has been segregated.

Factors affecting development

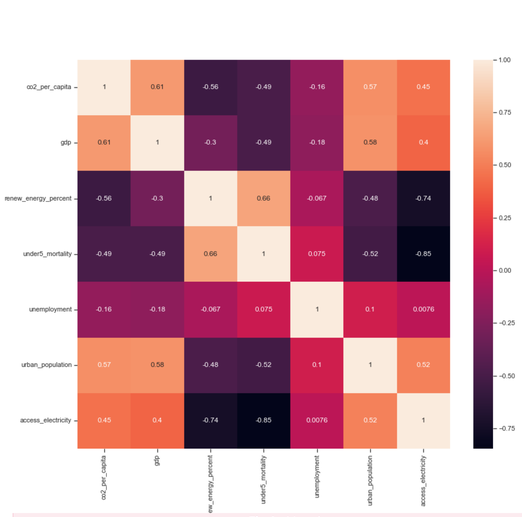

Fig1: Heatmap

Correlation Coefficient Value (r)	Direction and Strength of Correlation
-1	Perfectly negative
-0.8	Strongly negative
-0.5	Moderately negative
-0.2	Weakly negative
0	No association
0.2	Weakly positive
0.5	Moderately positive
0.8	Strongly positive
1	Perfectly positive

Fig2: Pearson Coefficient and its meaning

Data Distribution

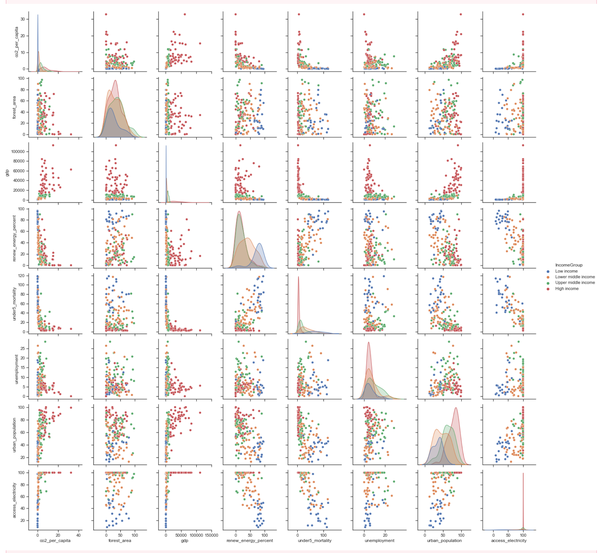

Fig3: Scatter matrix

The scatter matrix shows the correlation between various factors and they have been grouped into various clusters based on the income group.

Income group	Country Name
Low Income	Sudan
	Madagascar
Lower Middle Income	India
	Bangladesh
Upper Middle Income	Brazil
	Mexico
	United Arab Emirates
High Income	Switzerland

Below graphs shows the change in few of the above mentioned factors over the years for countries from different economic groups (as given in table).

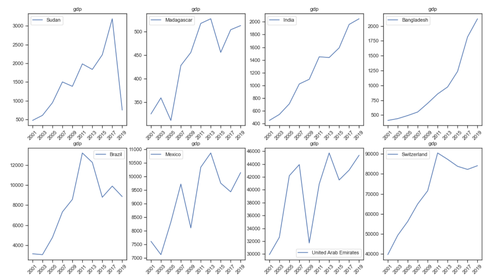

Fig4: Subplots for GDP

GDP of Mexico and UAE dropped from 2008 to 2009 due to the great recession and it recovered after 2009. GDP for Low middle income countries were not affected much and the continued to grow steadily over the years.

Data Distribution

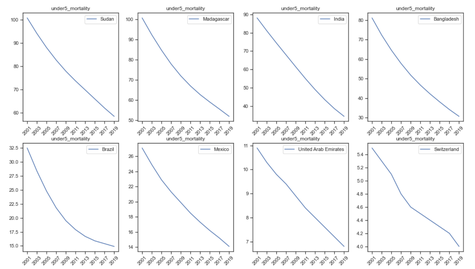

Fig5: Subplots for mortality rate of children under 5 years

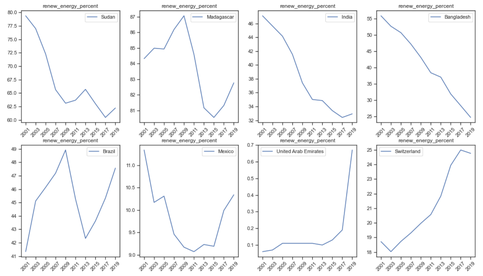

Fig6: Subplots for percentage of renewable energy use out of total energy use.

Clustering

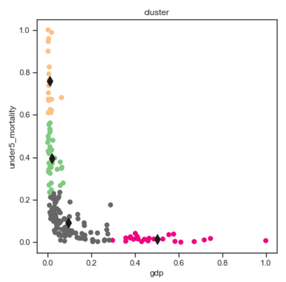

Fig7a: Cluster analysis, GDP vs mortality rate of children under 5

GDP vs mortality rate of children under 5		
	x	y
Centroid1	0.0191015	0.3937852
Centroid2	0.0102058	0.7596303
Centroid3	0.5034275	0.0156564
Centroid4	0.093203	0.0922623
GDP vs percent of renewable energy out of total energy		
	x	y
Centroid1	0.4987669	0.2045746
Centroid2	0.0504789	0.3853222
Centroid3	0.0993008	0.1002977
Centroid4	0.0314933	0.8030636
percent of renewable energy out of total energy vs mortality rate of children under 5		
	x	y
Centroid1	0.110041	0.085688
Centroid2	0.8231347	0.5785977
Centroid3	0.402792	0.4975469
Centroid4	0.4280993	0.0921706

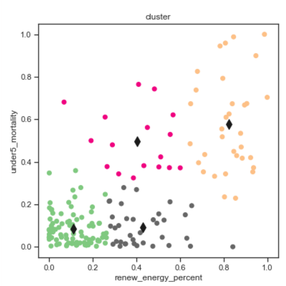

Fig7b: Cluster analysis, percent of renewable energy out of total energy vs mortality rate of children under 5

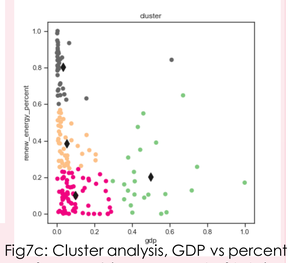

Fig7c: Cluster analysis, GDP vs percent of renewable energy out of total energy

Curve fitting

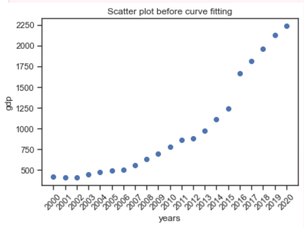

Fig8a: Scatter plot for GDP of Bangladesh between 2000 and 2020 before curve fitting

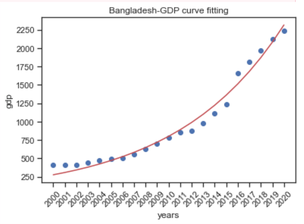

Fig8b: Scatter plot for GDP of Bangladesh between 2000 and 2020 after curve fitting

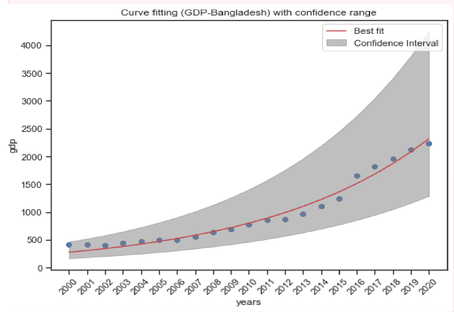

Fig8c (shown in left): Scatter plot for GDP of Bangladesh between 2000 and 2020 with curve fitting and confidence range

Year	Predicted GDP	Lower error limit	Upper error limit	Error range
2025	3940.27043	2146.886	7389.70389	5242.81789
2030	6705.77144	3577.44771	12843.0951	9265.64735
2035	11412.2549	5961.2537	22320.9337	16359.68
2040	19422.0104	9933.49129	38793.1475	28859.6562

Results and Conclusion

- Scatter pair plots for various factors which affect economic growth of a countries were plotted and it was found that, they formed 4 different clusters based on the income group of the country (low, high, low middle, upper middle)
- Cluster analysis was performed for 3 factors: GDP, mortality rate of children under 5 years and percent of renewable energy out of total energy was performed against each other and similar pattern of grouping was identified (based on income group)
- Cluster fitting was done for GDP of Bangladesh(lower middle income) between 2000 and 2020, and the GDP value for future years were also predicted based on the fit model. The graph showed a logistic growth.

References

https://www.researchgate.net/figure/Meaning-of-Pearson-correlation-coefficient-value-r_tbl1_299402589