CRISIS IN AFRICA DATA

Mutwoki Dennis Mureti
95003

Dataset Description

Dataframe statistics

	systemic_crisis	exch_usd	gdp_weighted_default	inflation_annual_cpi	independence	currency_crises	inflation_crises
count	1059	1059	1059	1059	1059	1059	1059
mean	0.0774315	43.1408	0.00640227	20848.9	0.776204	0.1322	0.129367
std	0.267401	111.475	0.0435716	675727	0.416984	0.349847	0.335765
min	0	0	0	-28.5021	0	0	0
25%	0	0.19535	0	2.08616	1	0	0
50%	0	0.8684	0	5.76233	1	0	0
75%	0	8.46275	0	11.644	1	0	0
max	1	744.306	0.4	2.19897e+07	1	2	1

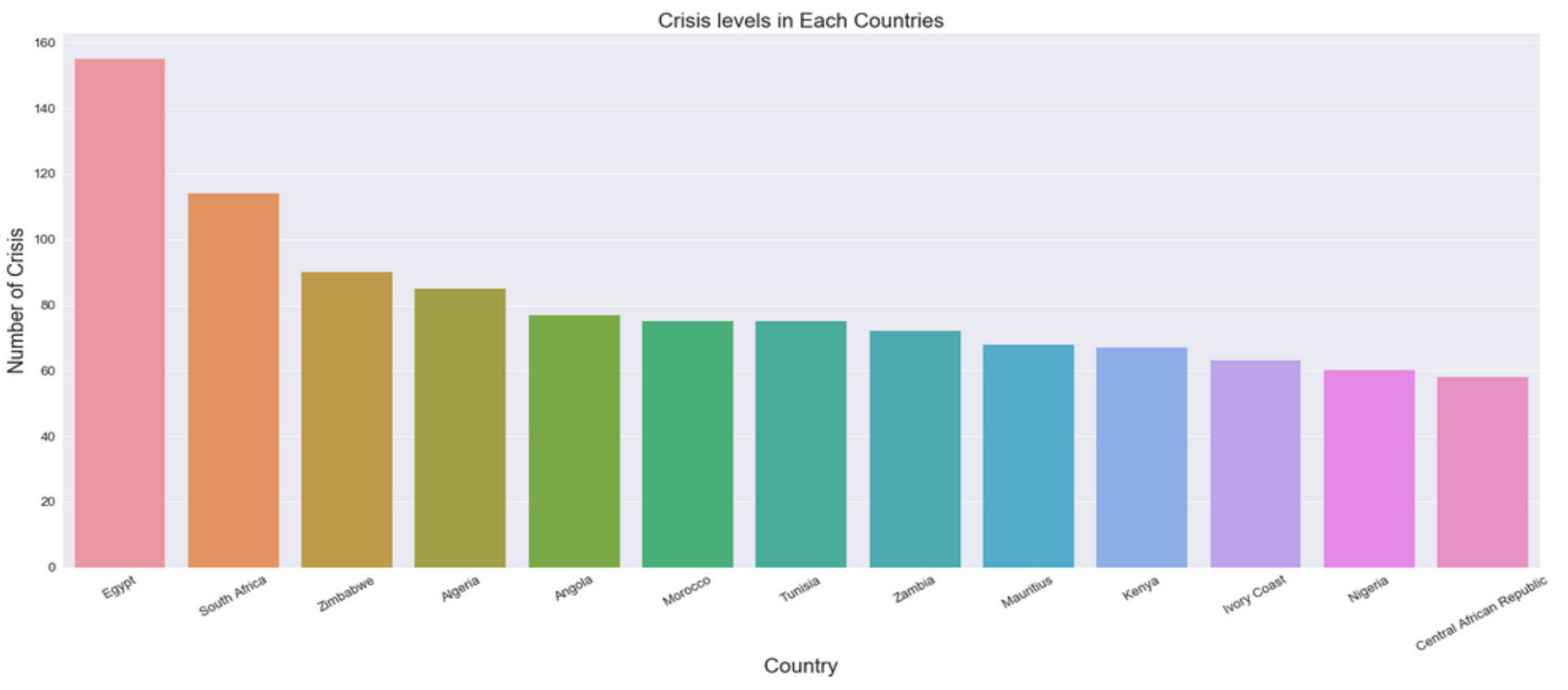
Research Questions

- Does the systemic crisis give details on the banking crisis?
- Does the crisis significantly affect the GDP of countries?
- Does having a crisis mean an effect on the exchange USD?
- Does the time period year a country got independence signify the level of crisis per country?

Methodology

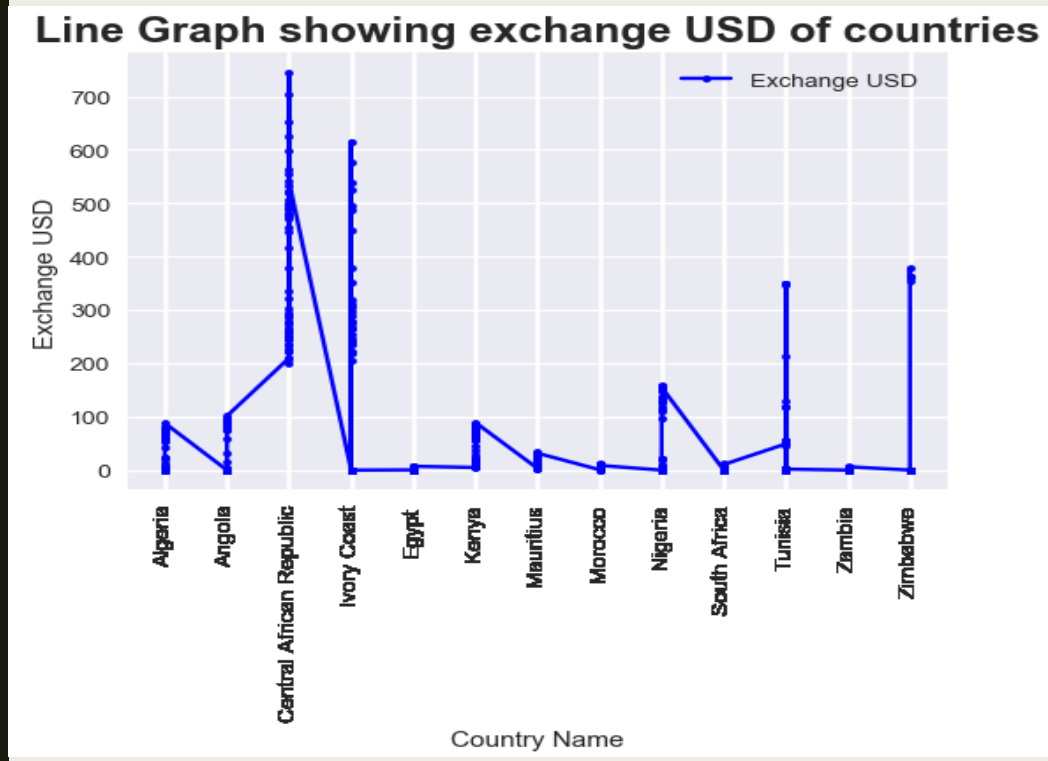
- For the purpose of my data analysis I obtained data that captured on the effect and variables that can be used to detect a crisis that spans within a period of over 130 years to capture also on the independence of the countries making the data quite huge for analysis.
- Tool of analysis that I used for data analysis was panda. I also used libraries such as pandas, numpy, seaborn, matplotlip, sklearn, statsmodels to used them for visualisation and for my regression analysis on the data.
- For feature engineering I focused on the numeric data e.g. exchange USD, inflation e.t.c. Observation of data was by the use of different graphing tool such as line graphs, histograms, bar plots to correlate the large set of data.
- For my regression analysis and machine learning I used ; logistic regression model, classification model i.e. random forest and KNN, models.

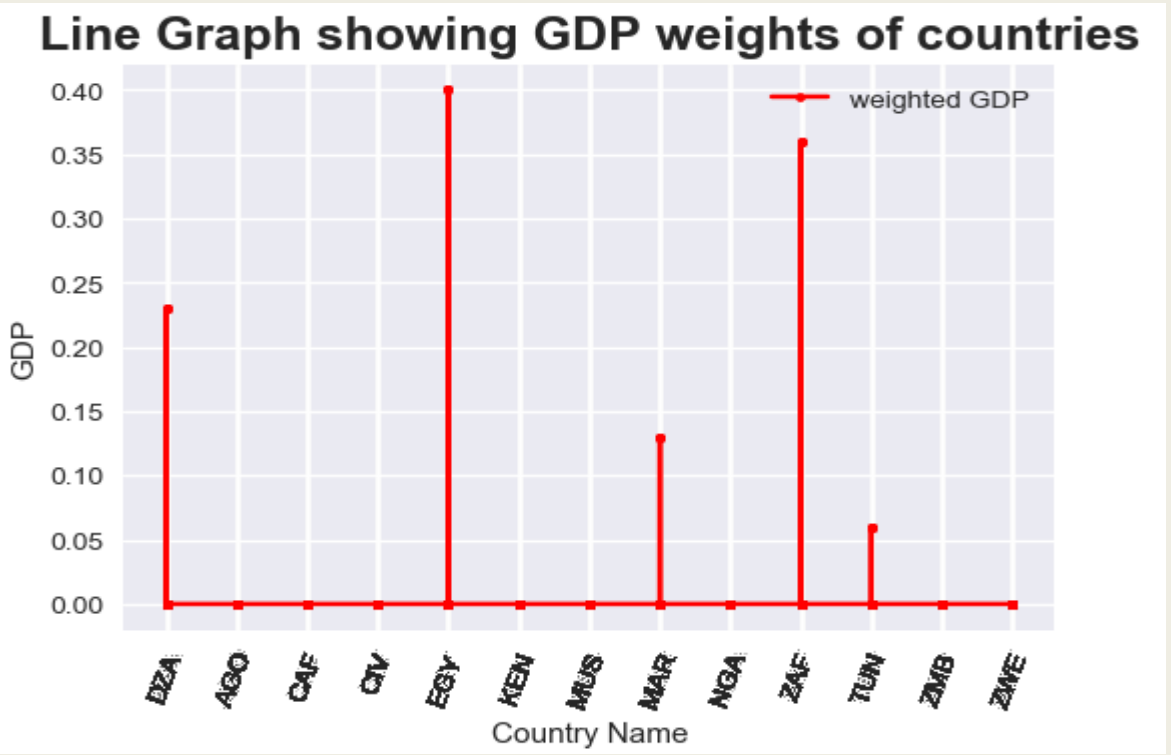
Crisis levels for each country



- A graphical representation of the countries' crisis levels showing that Egypt had the highest level of years that it had crisis and Central African republic with the lowest crisis levels. This data spans over a period of 130+ years.

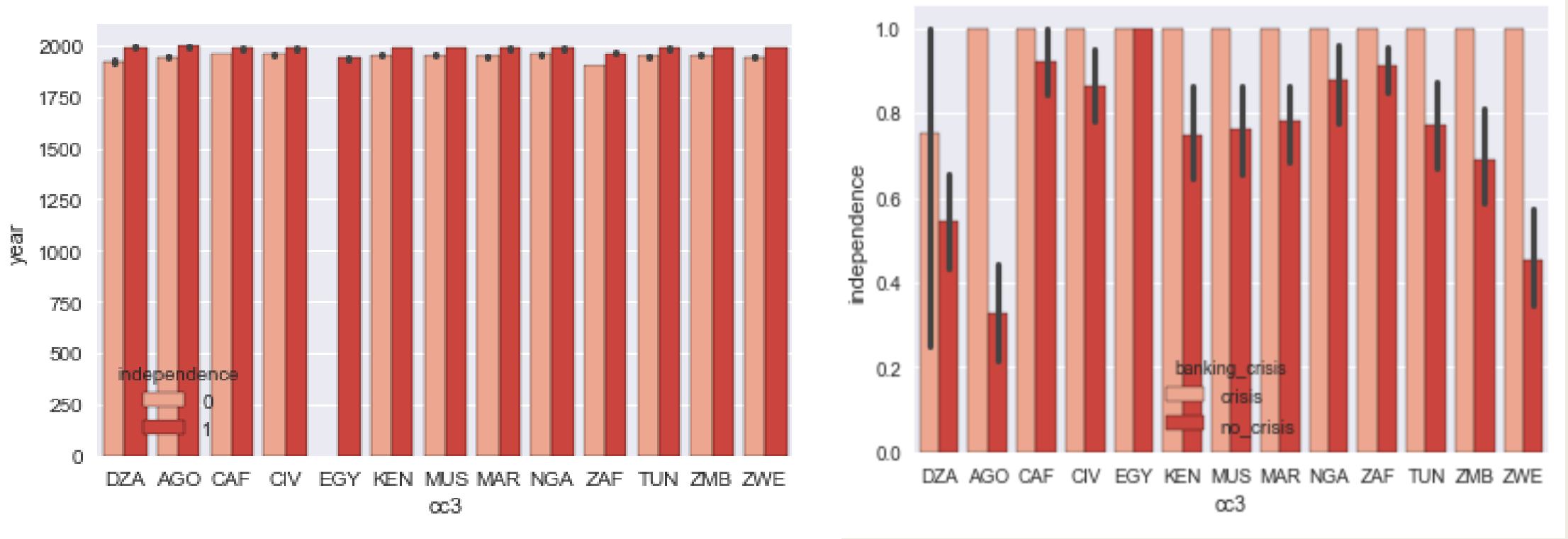
Line graph showing crisis relates statistics





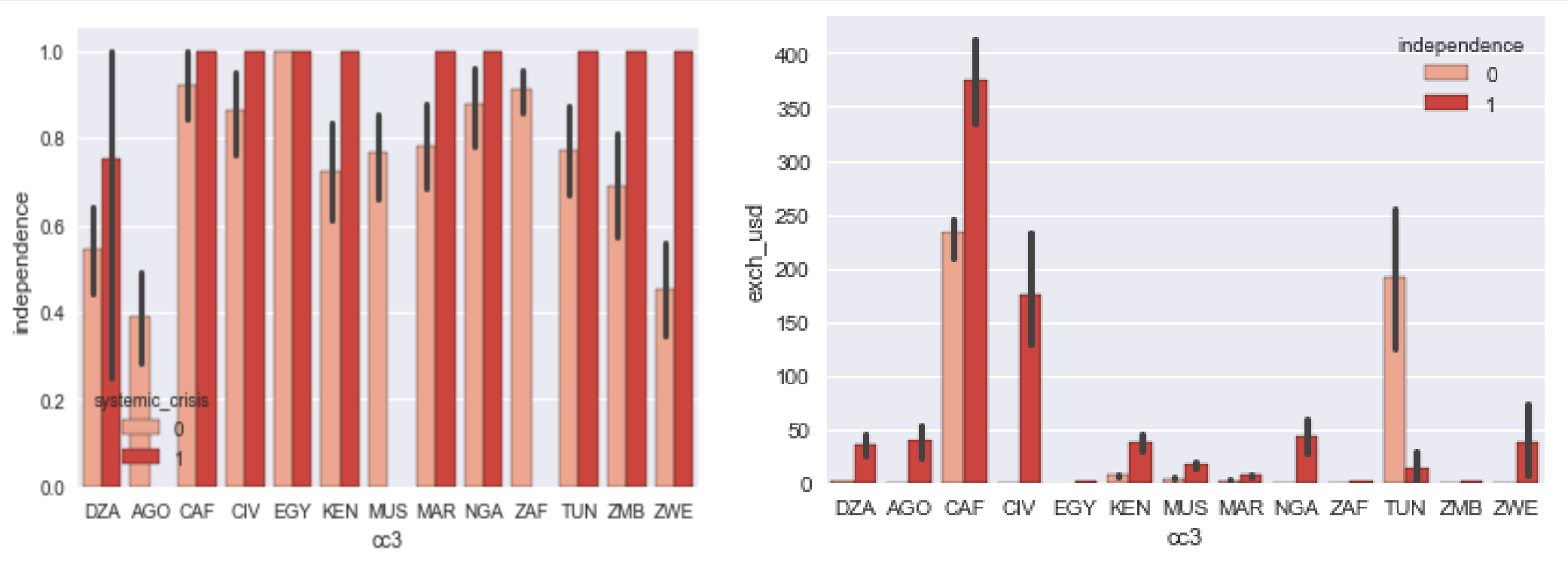
- Line graphs represents other crisis related statistics and is evident that the level of crisis per country later reflects on the GDS, exchange rate and even inflation rate. With Egypt having the highest level of crisis it is evident that it records the best exchange rate as at 2014 also being that they had an averagly longer time to act and stabilize the economy due to the crisis. It also records the highest GDP as of 2014 due to the response time they has as compared to a country like Central African republic.

Statistics corelated to independence



- First bar plot uses three variables i.e. Independence, year and the country where 1 represents the start year of independence. For the second bar plot, it corelates; banking crisis, independence and the country. From the two graphs I derived that Egypt attained its independence earliest from the mentioned countries and thus the low level of the 0 bar for independence. Having attained its independence early, thus means that most of its crisis also occurred after its attainment of independence. Most of the banking crisis having occurred mostly after independence for most of the countries.

Systemic analysis and exchange USD in relation to independence.



- I observed that most the systemic crisis occurred before independence and with Egypt having attained its earlier was inevitable but i=occurred after independence. The USD exchange rate for most of the countries went up mostly after independence except for a country like Tunisia which was better before independence.

ML (Model evaluations of the dataset)

CLASSIFICATION (RANDOM FOREST)			
Banking Crisis		**Systemic Crisis**	
Accuracy	0.9150	Accuracy	0.9119
Precision	0.67	Precision	0.89
Recall	0.92	Recall	0.91
F1-Score	0.90	F1-Score	0.90

CLASSIFICATION (KNN)			
Banking Crisis		**Systemic Crisis**	
Accuracy	0.8993	Accuracy	0.8742
Precision	0.90	Precision	0.87
Recall	0.90	Recall	0.87
F1-Score	0.90	F1-Score	0.87

ML(Logistic Regression)

Logistic Regression	
Banking crisis	
Accuracy	0.92
Precision	0.92
Recall	0.92
F1-Score	0.89

- By using the tow models we can visually conclude that with logistic regression having the highest accuracy the analysis are more accurate.
- Due to the bulkiness of the data it was safe to use random forest for the data analysis as well.

Conclusion

- For the analysis it can be said that the earlier a country got its independence then the higher the number of crisis the country will have.
- Crisis in this case will significantly affect the GDP per country also due to factors such as inflation.
- We can also say that the crisis level also reflect on the USD currency exchange. Countries like Egypt have a better exchange rate also due to the response time that they had to make the economy stable.

Limitations and challenges

- Dataset for analysis was large and spanned over a long time period making it bulky.
- Filtering of data(rows) was impossible being that a country may have occupied 100+ rows thus the use of the whole dataset.
- Larger ranges within the bar graphs was inevitable while visualizing the data.
- Was difficult in the case of visualising data in form of a pie chat due to its size.

THE END