

Project Report of Data Analytics

IBRD Loan Analysis

Muhammad Ali

[19pwcse1801@uetpes
hawar.edu.pk](mailto:19pwcse1801@uetpes.hawar.edu.pk)

Abstract

The main idea of this project is to have a simple yet powerful analysis on the data of loans taken by different countries throughout the world, from World Bank. This analysis, is for laymen, in the sense, that almost all of the constraints are clearly stated, and simple to understand. A lot of data cleaning was needed to be done, and thus performed, as the dataset used is a real time dataset.

Introduction

Through this project, I have done, Analysis of Historical data of Loans, taken by countries, from IBRD, which is a child organization, of the World Bank. The project includes, analysis of the data at multiple stages, and also, prediction of 'Interest Rate', and approval of loans, based on other characteristics. As the dataset is an open source publicly available dataset, many rounds of data pre-processing were done to rectify the data and also none of the records were deleted, so that a large amount of data is available for analysis.

Background Study

A large amount of data is available for analysis regarding loans. But it requires similar number of pre-processing techniques, and this has been achieved through this project. The data, which was available, consisted of a lot of data columns, many of which were irrelevant for analysis purposes, and thus were removed. The Historical Data about Loans, provided by World Bank, upon good analysis techniques, has provided wonderful insights about the data of countries.

System Architecture

Firstly, processing was done on raw data. All the Nan values were replaced with relevant values, and extra columns, which were not at all needed for data analysis were removed. Then, Data Visualization was done on Processed Data. Then, machine learning models were applied on processed data, and the results were matched with errors. Then, some in-depth visualizations were done. The whole flowgraph of the project is shown in figure 1.

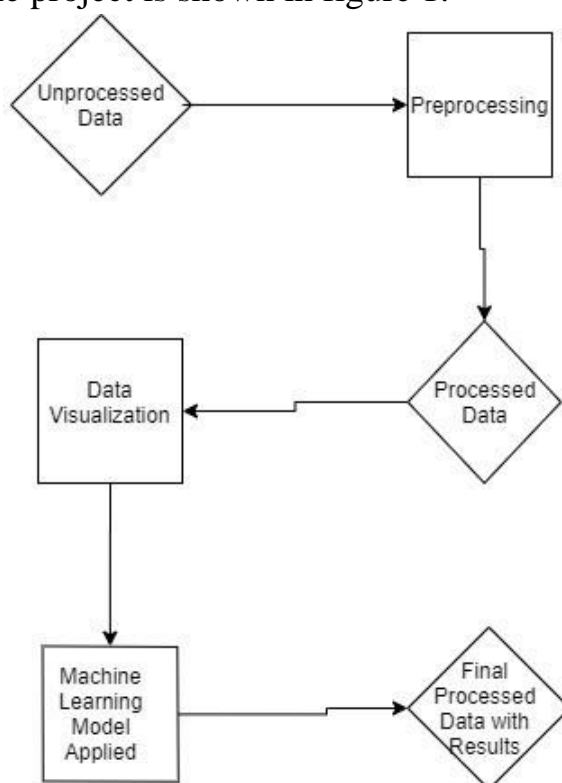


Figure 1

The data which I use, is available on World Bank website, and the source is mentioned in the references. It consisted a total of 969,426 rows and 33 columns. Many columns were removed and some were added at many stages of project, and the final data-frame was of size of 969,426 rows and 21 columns. None of the rows were deleted, and as all the analysis was performed on real time data, therefore, errors in predicted models were very less.

Novelty

The analysis which has been performed in this project, is based on actual real-world dataset, which has been discussed above. The analysis is error free, in terms of Domain analysis, and has very low calculation error, indicating that the analysis is correct, and in accordance with the actual real-world situation.

Results

The first analysis, was of all countries, a comparison of 'Loans Held' versus 'Interest Rate' at which that particular loan was taken, which was shown through a heat map. A snapshot of that graph is shown in Figure 2.

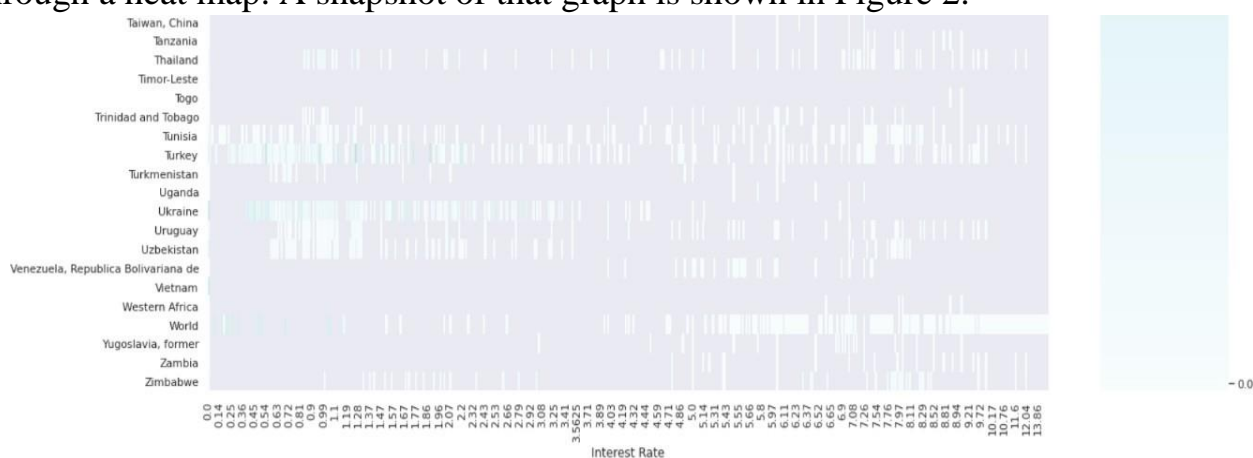


Figure 2

The second analysis is of different loan types, and their count, which was shown through histogram, in Figure 3.

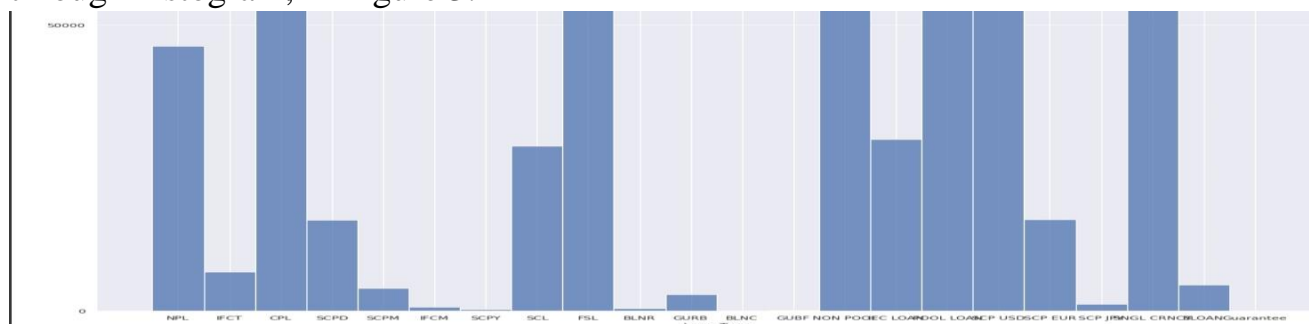


Figure 3

The third analysis was of the number of days it took a country to become debt-free, which is calculated by difference between Last Repayment Date and Last Disbursement Date, versus all countries, which is also shown through a heatmap, in Figure 4.

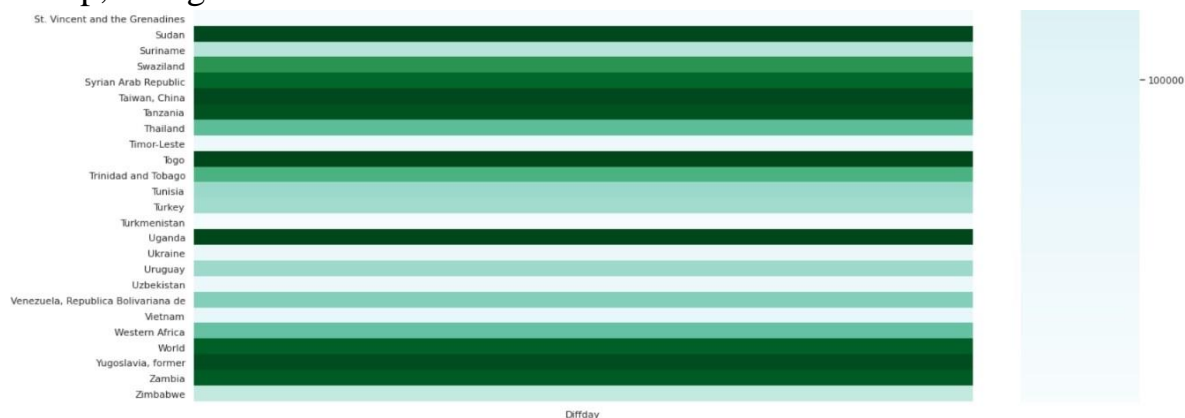


Figure 4

The fourth data visualization is showing the comparison of ‘Original Principal Amount’ of loans taken by countries, versus all countries, which is shown through a heatmap, in Figure 5.

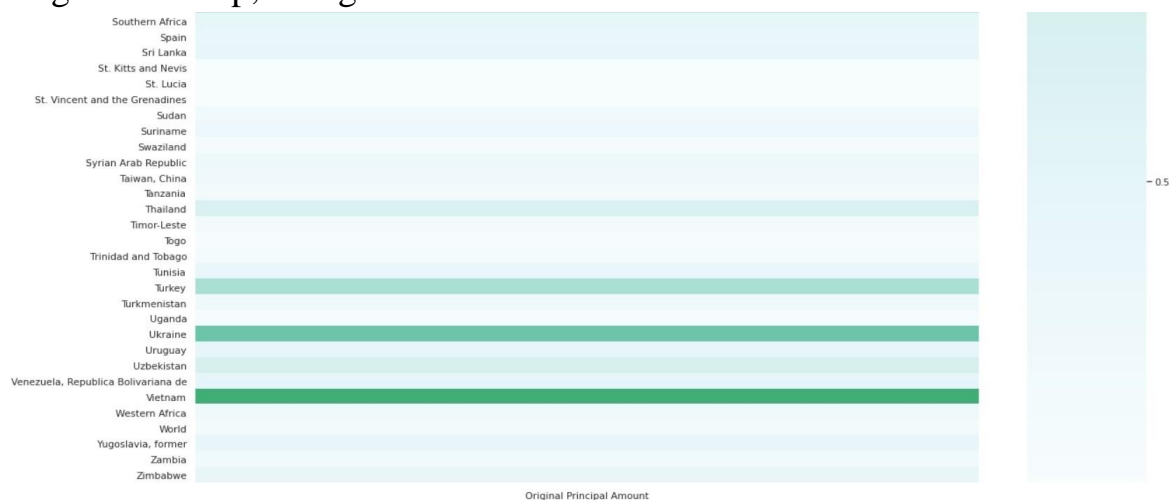


Figure 5

The fifth analysis is comparison between Original Principal Amount and Interest Rate for a particular country. Here is a snapshot for India, in Figure 6.

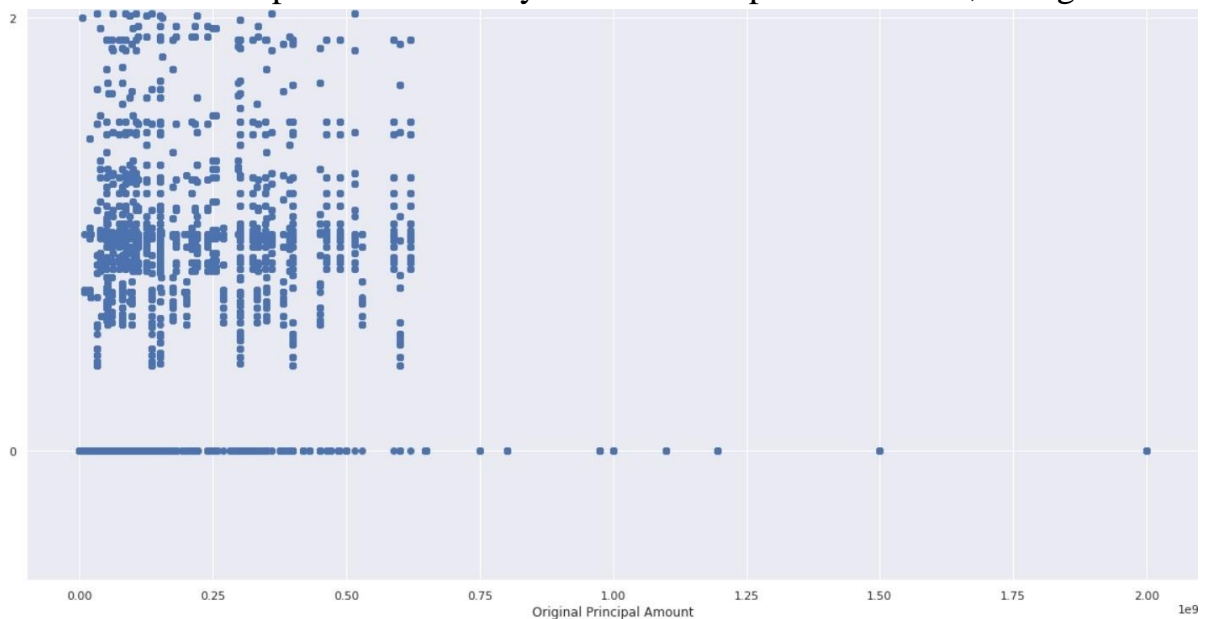


Figure 6

The sixth analysis is of Interest Rate for top 10 economies of world (whose data is present in dataset), which is shown through a 5-number-summary, using a boxplot, in Figure 7.

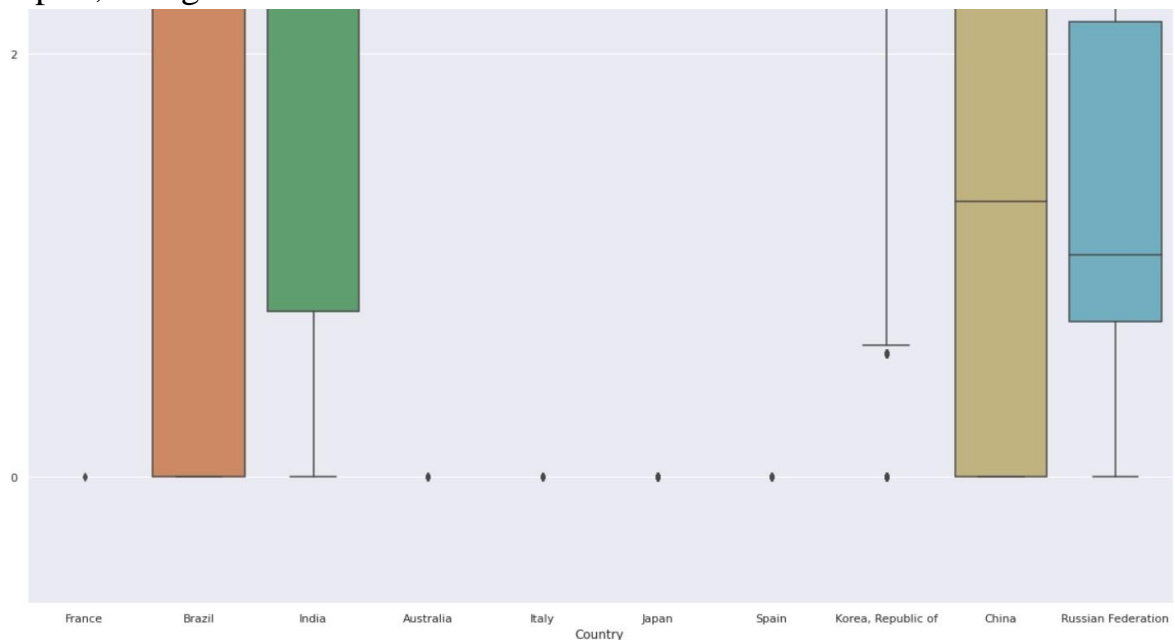
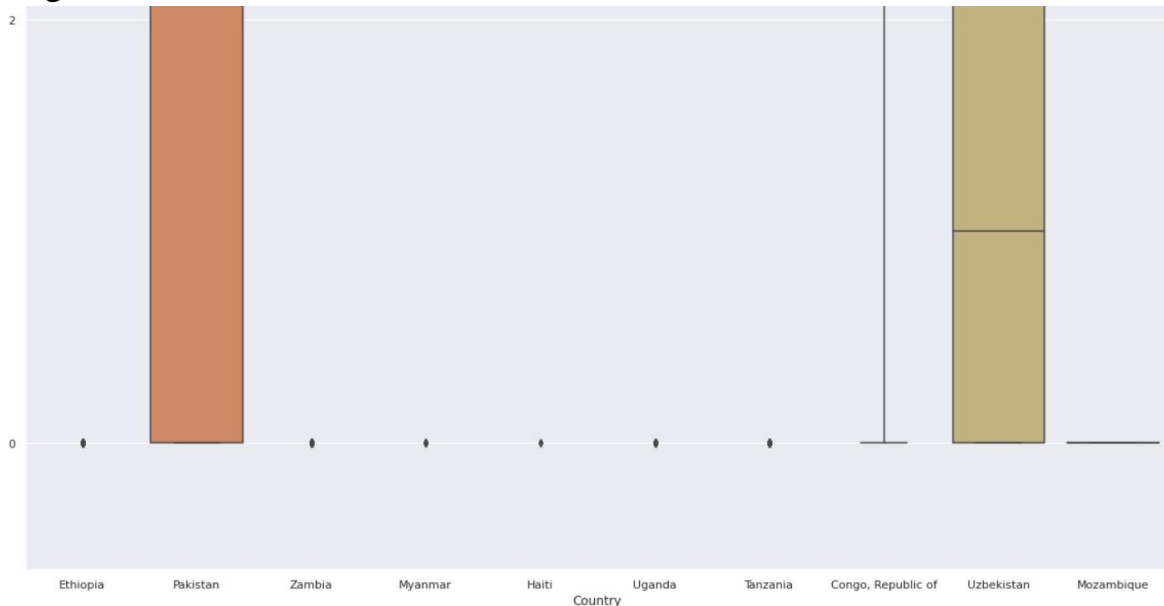


Figure 7

The seventh visualization is similar to sixth, that is comparison of interest rates for bottom 10 economies of the world, through a boxplot of 5-number-summary, in Figure 8.

Figure 8

The eight analysis is showing utilization of loans by top 10 and bottom 10

countries, by comparing percentages of undisbursed and disbursed amounts of respective loans. This is shown through pie charts.

The ninth analysis is showing comparison of Amounts repaid to IBRD and Amount due to IBRD, of top 10 and bottom 10 countries, which is shown through pie chart.

Conclusion

Through all the analysis done in the project and described above, an in-depth knowledge of different countries, and their relationship with the loans they have taken from World Bank, is analysed.

Two prediction models were used in the project, one for Classifying the approval of loan, that is a Yes or No, based on all other characteristics of loan. This was done through Random Forest Classification, and the accuracy score for many countries came to be 1.0, as the data is real world dataset, hence, no anomalies were involved in data.

Another prediction model implemented in the project is predicting Interest Rate of a particular loan, given that all other characteristics. This was done through Multivariate Linear Regression. As the data was very diverse, therefore the Mean Absolute Error was 0.8608615826006345, Mean Squared Error was 2.376010722567383 and Root Mean Squared Error was 1.5414313875639691, for India.

References

Data Source -

<https://finances.worldbank.org/Loans-and-Credits/IBRD-Statement-Of-Loans-Historical-Data/zucq-nrc3> - World Bank Group Finances – IBRD Statement of Loans – Historical Data

Top 20 Economies in the World -

<https://www.investopedia.com/insights/worlds-top-economies/>

The Poorest Countries in the World <https://www.focus-economics.com/blog/the-poorest-countries-in-the-world>

Pandas Documentation - <https://pandas.pydata.org/docs/>

NumPy Documentation - <https://numpy.org/doc/>

Scikit-learn Website - <https://scikit-learn.org/stable/>

Matplotlib Documentation - <https://matplotlib.org/3.3.3/contents.html>

Seaborn Website - <https://seaborn.pydata.org/>

Figure 1– [google.com](https://www.google.com)

Programming hints (python) - https://www.udemy.com/course/python-for-data-analysisvisualization/?utm_source=adwords&utm_medium=udemyads&utm_campaign=Python_v.PROF_la.EN_cc.ROW_ti.7380&utm_content=deal4584&utm_term=.ag_85724077624_.ad_535397245836_.kw_.de_c_.dm_.pl_.ti_dsa774930046209_.li_1011086_.pd_.&matchtype=&gclid=CjwKCAiArY2fBhB9EiwAWqHK6kjc45nP7aob8_ocuYbG7buR3LBxALQzm43kj-Auanas-63Q9k1_mhoC3BsQAvD_BwE