**Final Project**

Group One

Department of Data Analytics for Business, St. Clair College

DAB103-23W-003 Analytic Tools and Decision Making

Tommaso Iacobelli

March 23, 2023

**Group members**

Aleczia Habash ID:  0508976

Noushin Asadsamani ID: 0829532

Sudharshni Radhakrishnan ID: 0833471

Sameera Sajeevan- 0826439

Iyanthika Basnayake ID: 0824786

**Background/Motivation**

Everyday hundreds of billions of dollars are traded on the bond market. Lower and middle-income countries must secure external debt to finance their operations. Countries must find a balance between debt and equity investments while still recognizing the amount of risk that they are willing to take. Bonds from foreign countries can be a more secure method of financing in comparison to stocks in the marketplace, however, having too much debt with an unbalanced portfolio can be detrimental.

**Problem Statement**

Middle-income and lower-income countries need to determine whether they can afford to take on more debt or not. To financially support an economy, it is essential to purchase foreign debt, however, purchasing too much foreign debt may lead to defaults on payments to its creditors, thereby resulting in the potential to disrupt a country’s economy in ways such as increased inflation rates, increased unemployment, and a decrease in the value of that country’s dollar.

**Project Proposal**

Our team will be creating a product to determine past trends in foreign debt statistics using descriptive analytics. This will help middle-income and lower-income countries’ governments determine how their debt levels are impacting their citizens and make predictions that determine if they have the capacity to take on additional debt.

**Analysis Questions**

1. Which 10 countries have the highest levels of foreign debt?
2. Do middle and lower income countries follow the same overall trends in debt levels that their regions do?
3. What is the overall current state of each region’s economy in terms of debt levels?
4. What countries (if any) are at risk of taking on too much foreign debt?
5. What countries, (if any) have relied the most on the International Monetary Fund to assist with their levels of debt in the past ten years?
6. Which countries have the highest levels of reserves set aside for debt protection, if needed?

**Dataset Description**

The dataset that we’ve obtained is the International Debt Statistics (IDS) dataset, from the World Bank. The dataset includes debt statistics between the years of 1970-2029 for lower income and middle income countries and has been gathered from multiple sources from international government institutions. The dataset contains over 1,500 indicators of external debt and related financial flows, such as debt stocks and flows, creditor composition, and currency composition. The data is organized by country and includes aggregates for regions and income groups. It is available in both Excel and CSV formats and includes data on external debt for 121 countries and territories. The additional workbooks in the dataset include metadata, a detailed description of each ‘series name’ variable, and the sources that the data was retrieved from. The International debt statistics dataset contains 66 variables. 6 of these variables are objects, the remaining 60 variables are floats.

**Variable Description**:

Country Name: The name of the country (str)

Country Code: The unique code of each country (str)

Counterpart-Area Name: Represents the World Bank (str)

Counterpart-Area Code: Represents the code that is assigned to the World Bank (str)

Series Name: Indicates the name of the economic debt indicator/variable that has been measured (str)

Series Code: The code of the series name (str)

1970, 1971, 1972… 2029: Represents the year that the economic debt indicators are reported for a given country (float)

**Problems with the Data**

The first issue that we came across with this data set is its size. With 76,820 rows of data, it was a challenge determining which data points would be useful in our analysis. Another issue that we noticed is that there is no primary key value that can be used that would allow the reader to determine a specific row of data. For example, simply searching by country would give hundreds of rows explaining the different types of debt (Series Name), however, searching by debt type (Series Code) would produce over 100 different data points, each pertaining to a specific country. To remedy this issue, we will need to create a primary key that combines the country code and the series code to give each row a unique value. The third issue that we came across is that there are a lot of blank values in the data set between the years 1970-2006. These statistics were not recorded in very many lower income and middle income countries and have therefore narrowed our search to the years 2006 and beyond.

**Data Collection Process**

The data has been combined from a collection of different online databases that the World Bank had amalgamated and is listed on the data sheet labelled ‘FootNote’ in the xlxs file. There are different segments/groups within this dataset that we had to take into consideration. The data was not only sorted by country, but also by region/continent. This will give us the opportunity to determine whether countries are a product of their environment and follow similar trends to their overall regions. Additionally, the data is sorted by income level – the countries in this dataset are also categorized based on their income group. The data for income groups is a cumulative value of all the countries’ values that are listed in their respective groups. It was very difficult to determine which variables we need to include and which variables we need to exclude, therefore we determined what our inclusion criteria would be based on the series name (and their respective codes) for each country. We ultimately decided that the following variables would be the best fit for our analysis:

|  |  |
| --- | --- |
| **Series Name** | **Series Code** |
| Total amount of debt rescheduled (current US$) | DT.TXR.DPPG.CD |
| Total change in external debt stocks (current US$) | DT.DOD.DECT.CD.CG |
| Total debt service (% of exports of goods, services and primary income) | DT.TDS.DECT.EX.ZS |
| Total reserves (% of total external debt) | FI.RES.TOTL.DT.ZS |
| Total reserves (includes gold, current US$) | FI.RES.TOTL.CD |
| Total reserves in months of imports | FI.RES.TOTL.MO |
| Undisbursed external debt, official creditors (UND, current US$) | DT.UND.OFFT.CD |
| Undisbursed external debt, private creditors (UND, current US$) | DT.UND.PRVT.CD |
| Undisbursed external debt, total (UND, current US$) | DT.UND.DPPG.CD |
| Use of IMF credit (DOD, current US$) | DT.DOD.DIMF.CD |
| Use of IMF credit SDR allocations (DOD, current US$) | DT.DOD.DSDR.CD |

We chose these values because they represent the total values for their respective countries.

**Data Cleaning**

The steps that we are expecting to take are to remove the column years 1970-2005 and their respective values because of the inconsistency of their statistics.

Remove the years 2023-2029 because there is not enough data to extrapolate any information from. We removed most of the NA values from the dataset by removing these columns. By removing these years mentioned we are working on this dataset in years from 2006 to 2021.

Removed all the rows for each country except the rows with the value of 11 Series Names listed above.

Remove the columns labeled “Counterpart-Area Name” and “Counterpart-Area Code” because whole observations have the same value in these two columns and don’t give us   
useful information.

In the last part of data cleaning, we checked the data for duplicate values and null values. There was no duplicate value in the dataset to be removed. But still, there were some null values in the column years, but we didn’t remove them because removing these null values the whole dataframe would be removed. So, we should work on the data with these null values.

**Expected Transformation Steps**

We need to add the following columns into the dataset. A column labelled ‘Income\_Group’ and another column labelled ‘Region’. These will denote each country’s respective income level and region that they have been categorized into.

Since the dataset contains no primary key column, we can combine the columns ‘Country code’ and ‘Series code’ to give each row a unique identification. For example, if we are looking for the total amount of debt that has been rescheduled in Afghanistan, we can type in the following code: AFG.DX.TXR.DPPG.CD.

Transforming the variable names by replacing “\_” with a space between variable names, because we faced some issues in working on the dataset with this kind of labeling.

**Data Analysis**

1. Which 10 countries have the highest levels of foreign debt?

The first visualization that we completed compares the top 10 countries with the highest levels of foreign debt. This graph plots debt that is owed to public creditors (government bodies) over time from the respective countries highlighted in the legend.

Our analysis includes changes in debt levels owed to public creditors between the years of 2011 and 2021 and illustrates that many of the listed countries have seen an increase in debt levels over time. Most follow a similar pattern, however there are a few countries that stand out. India has seen a relatively consistent increase in debts owed over the past 10 years – although their debt levels started much higher than all the other countries, they have increased the spread between themselves and many other countries. Bangladesh has followed a similar trend; however, it experienced a concerning spike in 2015 that would need further investigation.

Although Egypt experienced a similar spike to Bangladesh in 2014, its debt levels have been steadily declining since then, signaling to its creditors that it is meeting its guarantee to pay off its debts, and that investing in that country was a good investment opportunity. Vietnam has been consistent since 2011 to lower its debt levels, again, proving to its creditors that the country is a safer investment.

1. What are the topmost countries in each region that have the highest total change in external debt stocks in 2021 and how was the trend for this amount in these countries in the past 10 years?

The total change in debt stocks shows the variation in debt stock between two consecutive years. Data are in current U.S. dollars. By doing this analysis we want to know how the trend of the total change in external debt stocks has been changed for each country chosen in the region that has the highest amount of debt in 2021. To reach that we created a facet line graph for each country that has been determined earlier. In the output we have six different line graphs for each country showing the trend between 2011 and 2021 According to the visualization output, China has the highest amount of variation in debt stock in 2021, and Egypt, Arab Rep has the lowest amount of variation in debt stock in 2021. All these countries from different six regions experienced minus the total change in debt stock between 2014 and 2018, but it happened in different years for each country. The amount of total change in debt stock has fluctuated for China, Brazil and India during the past 10 years and this amount hasn’t changed a lot in 2021 in comparison to 2011. But we can say that the variation in debt stock has decreased for Russian Federation and increased for Egypt, Arab Rep in 2021 in comparison to 2011.

The analysis concludes that overall, the Russian Federation has a good policy implemented in debt stock reduction.

1. What is the overall current state of each region’s economy in terms of debt levels?

To determine what the state of each region’s economy was, we first looked at each country’s overall debt levels with the help of a geographical map. From this map, we were able to determine that the countries in the region of South Asia looked to have the highest debt levels. Although this is a good point of reference, to break our analysis down further determine which countries’ income groups contribute to the region’s overall performance we compared the income levels in each region.

From this analysis, we can determine that the Sub-Saharan Africa region has the largest income disparity between its countries. 34% of its countries are listed as lower income countries, 62.4% are listed as lower-middle income countries and only 3.53% of the countries within the region are listed as upper-middle income countries. In contrast, the region of Latin America and Caribbean are in the best position in terms of income levels, with 78.8% of the countries labelled as upper-middle income and 21.2% of the countries labelled as lower-middle income in the region. This analysis is useful as it extrapolates the different income levels from a larger perspective. For investment purposes, private and public creditors not only want to know which countries are the best to invest in, but that those countries are surrounded by other countries with similar levels of income and financial success.

1. What countries (if any) are at risk of taking on too much foreign debt?

When attempting to complete this visualization, we noticed that many of the values were filled with NAs. When we started the data cleaning process, we chose not to remove the rows with NA values because of the type of dataset that we are working with. Since we are looking at various trends in different countries over multiple years, and each year is a variable type, we did not want to make the mistake of removing a whole row if the data was missing from only one year.

This, however, proved to be a grievance for us. Since the type of debt that we were going to use to analyze whether a country would be at risk of taking on too much foreign debt could not be used due to the lack of data, we decided to use a different variable and amend the type of debt to analyze a similar metric.

The original type of debt was rescheduled debt, meaning that the terms of the contract were restructured to assist that country. We switched to a different mindset and chose to analyze which countries had the largest IMF credits. International Monetary Fund credits can either be given (in borrowing terms) to a country that needs additional monetary assistance. For instance, if a crisis occurs in a country, the IMF will loan that country funds over a specified period of time.

1. What countries, (if any) have relied the most on the International Monetary Fund to assist with their levels of debt in the past 10 years?

The countries that have relied the most on the International Monetary Fund (IMF) are China, India, and the Russian Federation. This analysis determines that these countries have relied on this fund the most to assist them and can be in some of the worst conditions. However, what this analysis does not consider is that China has used this fund to assist in expanding its infrastructure and strengthening its economy. Additionally, China was not labelled as one of the top ten countries with the highest levels of external debt owed to public creditors and has the highest levels of reserves set aside.

India is also listed as one of the top three countries that are listed as borrowers for the IMF, however, India is listed as the top country for external debt funds owed to public creditors. This insinuates that India is not financially stable enough to support itself solely through private and public creditors and needs additional financial assistance. This money could be used to support economic growth in the economy, or to support itself in terms of financial downturns.

Lastly, we consider the Russian Federation and its economy. The Russian Federation has an abundance of tariffs placed on them from other countries and is very independent in comparison to other countries in the world. This explains its need for the international Monetary Fund, and its low levels of external debt from public creditors.

1. Which countries have the highest levels of reserves set aside for debt protection, if needed?

The final analysis that was conducted compares the top 5 countries with the highest levels of reserves (including gold). This analysis determines which countries have the most cash and liquid assets set aside in case of an emergency or a unique investment opportunity.

From the visualizations, we can determine that China has the highest level of reserves set aside and assists in the claim that its use of the funds from the IMF are for economic expansion and infrastructure. This is beneficial because it highlights that China (who has been deemed an upper-middle income country) has made great accomplishments in keeping its debt levels at an average level while benefitting its members of society between the years 2011 and 2021. There is, however, a large imbalance between China and the remaining four countries that are ranked at the top.

The remaining upper-middle income countries (the Russian Federation and Brazil) have much smaller reserves set aside. It could be argued that the size of the economy is a factor that contributes to reserve levels, however, the difference between China, Brazil, and the Russian Federation are substantial.

From this analysis, we can determine that the overall reserve levels align with each country’s economic performance and follow a similar pattern that their overall IMF levels do.

**Conclusion**

There are several conclusions that can be drawn from our analysis and visualizations. From the first visualization, we determined that India had the highest levels of external debt, and Bangladesh was following closely behind. India’s increasing levels of debt hold significance for one of the following reasons, either its debt levels are so high that it is beginning to default on its payments, or the additional investment in foreign expenditure is to increase economic growth in the economy.

In 2020, many middle-income and lower-income countries experienced a decline in GDP growth. As seen in the visualization from analysis question two, all but one country revealed that their debt stocks increased in 2020, presenting the notion that the global economy was weakening. The pandemic had a significant effect on the economy, and several nations not only saw a decline in their GDP, but a rise in their debt levels. Higher income countries were able to adapt better economically because of the infrastructure that they have in place and their resources, however, the gap between higher income-earning countries and lower/middle income countries post-pandemic continued to increase economically. China was one of the few countries to achieve positive GDP growth in 2020 due to its significant contributions in global exports and manufacturing abilities.

Our third analysis question examined the regions that had the highest external debt levels as well as the income levels of those regions. From our analysis, we determined that the Sub-Saharan African region had the highest level of income disparity, with 34% of its countries deemed “lower income countries” and over 62% of its countries deemed “lower-middle income countries”. This allowed us to see which regions receive the least amount of support and which regions had the highest income levels in comparison.

Over the past 10 years, China has kept a relatively steady borrowing rate with the IMF, however, China uses the IMF for economic growth and stability. Its levels of borrowing from the IMF surpass all other countries by almost three times the amount. Although China has been continually borrowing with the IMF, its levels do not change drastically over the 10 year period, which is a sign of continued economic growth.

In an ideal world, every country would follow the same data trends and make the same investments to minimize risk and maximize return, however, there are always anomalies in the data. For example, the Russian Federation has a large number of reserves set aside. Initially, this may seem odd to an individual, however, if real-world indicators are taken into consideration, the anomalies make sense. The Russian Federation has several sanctions placed on their economy from many other countries; holding more reserves is a means of ensuring that they have money set aside in case of an emergency.

In general, the countries with higher GDP growth are likely to have lower foreign debt levels and higher levels of reserves whereas countries with low GDP growth tend to have higher foreign debt levels and lower reserves. These lower GDP countries use foreign investments to stimulate the economy. For example, a loan is received, and the borrowing country uses this loan to invest in new infrastructure. They plan on building a hotel near a beach. This has multiple long-run benefits, such as an increase in employment and an increase in tourism (once the hotel is ready) which will stimulate growth in the economy.

Overall, building a strong economy is essential for proper debt and reserve management in the long run. A strong economy ensures that a country can maintain sustainable debt levels that have a guaranteed payback period from the borrower. Additionally, having reserves on hand ensures that countries have liquid assets in case of an emergency or a quick, but profitable investment.

**Recommendations**

Based on the results of our analysis and visualizations, we believe that the following recommendations are beneficial to lower-income and middle-income countries:

The first recommendation that we have is regarding the improvement of robust infrastructure and resources in lower-income and middle-income countries. Investments in infrastructure will increase economic growth, and consequently GDP. We believe that this is something that India has already implemented, however, it would also be beneficial for India to look for various investment opportunities that fall outside of robust infrastructure such as investments in education and technology.

Many unprecedented decisions were made during the global crisis in 2020, which lead to additional funds being withdrawn from countries acting as borrowers. Our recommendation in this instance is to always ensure that enough liquid reserves are set aside as an emergency response fund. Additionally, investments during a global pandemic should be monitored. The type of infrastructure investment can determine the level of risk on an investment. These recommendations will also ensure that if another global crisis were to happen again, lower-income and middle-income countries would be able to guarantee a more effective response than what was seen in 2020.

Our third analysis question targets various regions and their respective income levels. In the case of the Sub-Saharan African region, in order to grow its economy and increase the relative GDP for the region, our recommendation is to increase its borrowing with IMF loans. These loans are designed for lower income countries and target countries that need additional financial stability to assist them in a crisis.

The International Monetary Fund (IMF) is a great investment tool for countries looking to strengthen their economy like China has. The IMF is already investing in India, however, if India can use that money towards improvements in technology, education, and its infrastructure, it can continue to strengthen its economy and increase their GDP. Additionally, we believe that it would be beneficial for China to consider investing in external debt with official creditors of other countries to continually expand its portfolio and keep it balanced.

Having an appropriate amount of reserves set aside by a country is imperative for growth. A country never wants to risk taking on too much debt for the fear of default on its loans. Additionally, having low cash reserves on hand prevents a country from making quick investments and can be detrimental if there’s not enough cash on hand in case of an emergency. Having an adequate amount of liquid assets on hand, like the Russian Federation, allows a country more freedom in its decision-making and investments and ensures that there is always a fund set aside that can be used in any circumstance.

Lastly, we recommend taking into consideration all of the factors that affect the data, not just the variables that are on paper. Anomalies such as Saudi Arabia’s high debt levels despite low GDP growth serve as a reminder that economic data always needs further analysis.

**References**

The World Bank. (2022, December 7). *International debt statistics.* The world Bank. <https://datacatalog.worldbank.org/search/dataset/0038015/International-Debt-Statistics>

The world Bank (2022, December 7). *IDS excel file*. The World Bank. <https://databank.worldbank.org/data/download/IDS_Excel.zip>

Real Python. *Pythonic Data Cleaning with Pandas and NumPy (by Malay Agarwal)*. Real Python. <https://realpython.com/python-data-cleaning-numpy-pandas/>

Analytics Vidhya (2021, June 14). *Data Cleaning Using Pandas in Python – Complete Guide for Beginners.* Analytics Vidhya <https://www.analyticsvidhya.com/blog/2021/06/data-cleaning-using-pandas/>

Data Camp. *Cleaning Data in Python (Adel Nehme)*. Data Camp. <https://app.datacamp.com/learn/courses/cleaning-data-in-python>

Data Camp. *Data Manipulation with pandas (by Maggie Matsui)*. Data Camp.

<https://app.datacamp.com/learn/courses/data-manipulation-with-pandas>

Net Information. *Combine multiple column values into a single column.* Net Information. <https://net-informations.com/ds/pd/comb.htm>

Data Camp. *Introduction to Data Visualization with Matplotlib (by Ariel Rokem).* Data Camp. <https://app.datacamp.com/learn/courses/introduction-to-data-visualization-with-matplotlib>

Data Camp. *Introduction to Data Visualization with Seaborn (by Mona Khalil & Yashas Roy).* Data Camp. <https://app.datacamp.com/learn/courses/introduction-to-data-visualization-with-seaborn>

Machine Learning Plus (2021, September 05). *Pandas Line Plot | Python.* Machine Learning Plus. <https://www.machinelearningplus.com/pandas/pandas-line-plot/>

GitHub Pages. *Plotting and Programming in Python.* GitHub Pages.

<https://swcarpentry.github.io/python-novice-gapminder/09-plotting/index.html>

Seaborn. *Line plots on multiple facets*. Seaborn.

<https://seaborn.pydata.org/examples/faceted_lineplot.html>

Plotly. *Facet and Trellis Plots in Python.* Plotly. <https://plotly.com/python/facet-plots/>

Geeks for geeks. *Choropleth Maps using Plotly in Python (by devanshigupta1304).* Geeks for geeks. <https://www.geeksforgeeks.org/choropleth-maps-using-plotly-in-python/>

Plotly. *Pie Charts in Python.* Plotly. <https://plotly.com/python/pie-charts/>

Plotly. *Bar Charts in Python.* Plotly. <https://plotly.com/python/bar-charts/>

Dataquest. *How to Plot a DataFrame Using Pandas (2022, June 08)*. Dataquest. <https://www.dataquest.io/blog/plot-dataframe-pandas/>

Seaborn. *Seaborn.FacetGrid*. Seaborn. <https://seaborn.pydata.org/generated/seaborn.FacetGrid.html>