Which Countries Had the Highest Overall Debt?

Ivan Escalona
Student of Computer Science
Georgia State University
Atlanta, United States
iescalonafaria1@student.gsu.edu

Oseriemen Ivbaze
Student of Computer Science
Georgia State University
Atlanta, United States
oivbaze1@student.gsu.edu

Christian Moronta
Student of Computer Science
Georgia State University
Atlanta, United States
cmoronta1@student.gsu.edu

Bryan Davison
Student of Computer Science
Georgia State University
Atlanta, United States
bdavison7@student.gsu.edu

I. Introduction

Our team had five questions in total. I, Christian Moronta, was responsible for the question:”Which countries have the greatest amount of debt?”. I chose this question because of the conversation in the political sphere about debt, deficits, and trade wars, so I found this question to be appropriate for the times that we are currently in. I was hoping that viewing the numbers would give me more insight about what is being spoken about on the television. Although I am not familiar with many economics concepts, it was still interesting to dive into this question.

II. Dataset Used

World Development Indicators Dataset (1960-2018)

This dataset is what we're basing our initial analysis on. It has information regarding the economics of each country globally as provided by the World Bank. https://www.kaggle.com/theworldbank/world-development-indicators

III. Preprocessing

Preprocessing was perhaps the most difficult stage for my question. The original data was structured in a way where it was very hard to work with. One column had the names of the countries, the next column had the features, and the columns after that were years associated with the value of that feature for that year. This proved difficult to work with, so I had the idea of turning it into a hierarchical index with the outcome looking similar to this:

Country Name	Year	Feature 1	...	Feature 198
United States	1960	...	...	...
	1961	...	...	...
	....	...	...	...
	2018	...	...	...

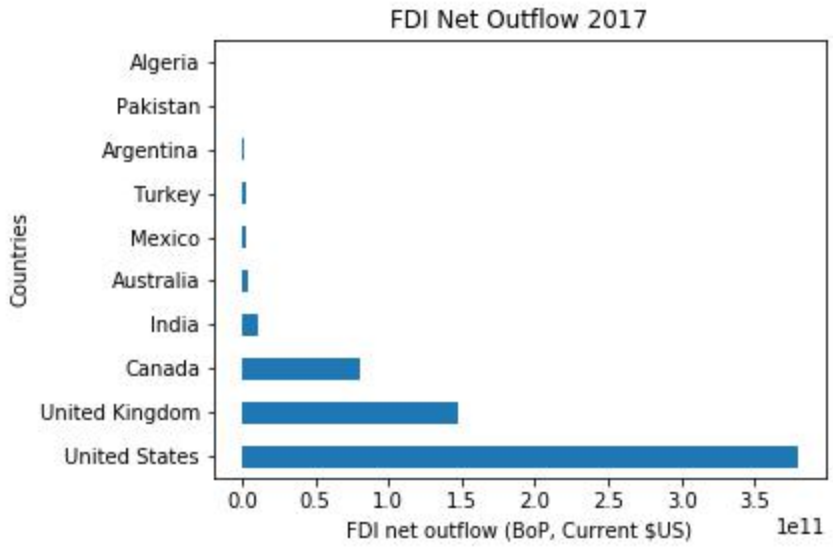
This proved to be incredibly to accomplish. I attribute the difficulty to my newness to using pandas. I had tried for hours to get the data to be like that, but I ended up with errors in the values of the features. The numbers from the original data frame would not much the numbers that were included in my new, transformed data frame. After a lot of time trying to debug the code, I had settled to use Ivan's preprocessed data. It was not in the way that I wanted, but it was structured in a way where it would be quite simple to do the extractions that I had planned to do regarding debt. The data in that preprocessed csv was structured with each row being a country at a specific year (i.e. UnitedStates_2018) and each column (except the first with the country name/year being the index) being a feature.

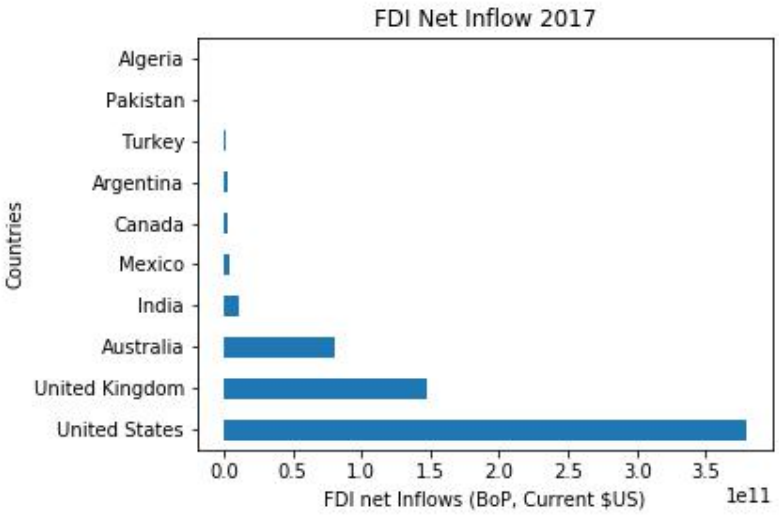
III. Data Extraction to Answer the Question

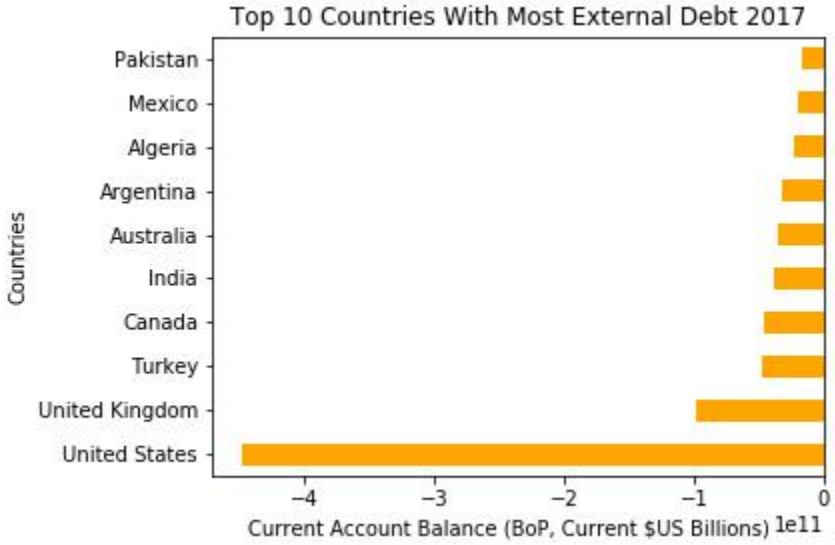
To extract the data that was needed I had looked through the original data set and found that many of the countries were missing data related to the final year in the dataset (2018). So I chose to use data from the most recent year that I could find many had values for (2017). The feature that was related to debt was the feature called “Current account balance (BoP, current US$)”. Balance of payments is a way to describe how much a country owes. BoP is made up of three factors, but the only one that was measured in the data was the current account. This just shows the net imports and exports a country has. I then filtered the data by indexes that had the year 2017 in the index. This data frame was then sorted from greatest to least. After viewing the data, I had come up with a couple more questions that would be interesting to answer such as: of the countries with the highest debt in 2017 how much where they invested in and how much investments did they have in other countries. Gathering the data for this was done in a similar process as described above, but replacing the feature that was selected with “Foreign direct investment, net inflows (BoP, current US$)” which would show much investment by foreign countries was done in that country, and “Foreign direct investment, net outflows (BoP, current US$)” which would show how much money that given country invests in other countries. All of the data from the data frames was plotted using the matplotlib library in python.

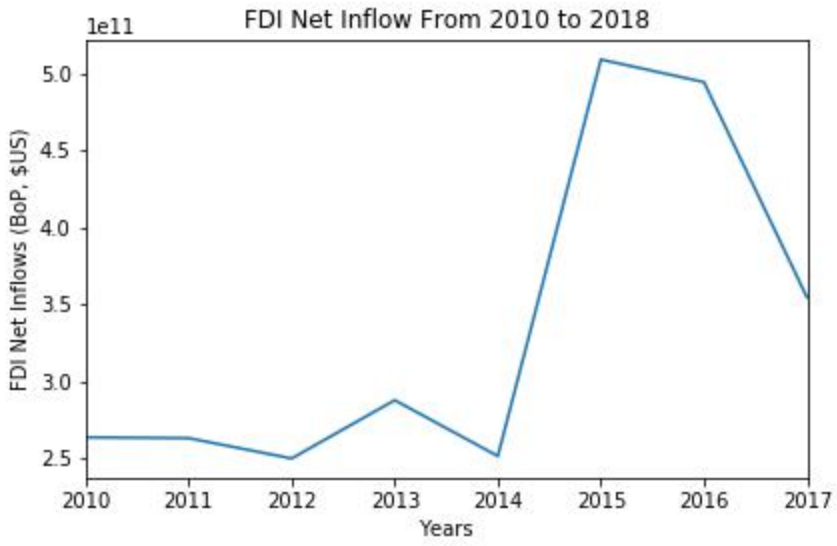
After looking at how much foreign investment was done in the United States. I wanted to see how the 2016 election affected foreign direct investment in the US, especially with all the rhetoric about trade deficits and trade wars. I gathered the US data from 2010-2017 and plotted it in a line plot to see the changes.

IV. Visualisations




FDI Net Inflow 2017
Algeria
Pakistan
Turkey
Argentina
Canada
Mexico
India
Australia
United Kingdom
United States
Countries
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
FDI net Inflows (BoP, Current $US)
1e11


Top 10 Countries With Most External Debt 2017
Pakistan
Mexico
Algeria
Argentina
Australia
India
Canada
Turkey
United Kingdom
United States
Countries
−4
−3
−2
−1
0
Current Account Balance (BoP, Current $US Billions)
1e11



V. Conclusions

From the data and visualisations I had. The top ten countries with the most debt were (descending debt amount): United States, United Kingdom, Turkey, Canada, India, Australia, Argentina, Algeria, Mexico, and Pakistan. Of those countries, Australia, the UK, and the US had the highest amount of foreign investment into their respective countries with a factor greater than all the other countries. The countries with the most investments in foreign countries were the US, the UK, and Canada.

It seems justified to have as much debt as the countries with the highest amount of debt only when that country has high amounts of foreign direct investment inflow/outflow. Otherwise it could prove to be rather troublesome. Interestingly enough, investment in the United States dipped after the 2016 election by quite a large margin as opposed to investments made from 2014-2017. This could be due to the rhetoric on the campaign trail, but there is no way to definitively prove it. As I had come to understand, economics isn’t so black and white and it’s difficult to pinpoint one sole explanation for data.