Project Report – John Enright

GitHub URL

https://github.com/DeerparkMul/Project-UCDPA_John-Enright

Abstract

"Today, we live in the most prosperous time in human history. Poverty, sicknesses, and ignorance are receding throughout the world, due in large part to the advance of economic freedom"(Hertitage Foundation 2022) https://www.heritage.org/index/about. This is the opening line in the Heritage Foundations explanation of what is the Index of Economic Freedom. Therefore, I wanted to put this to the test and see if "Economic Freedom" really does lead to a better economy. Also I wanted to ascertain what is the most important economic factor related to fighting sickness. I accordingly used the three datasets(below) to try and answer these questions.

Introduction

The state of the economy affects everyone. Whether its people looking for employment or individuals dependent on state welfare or social programs like the health service which are all funded out of tax revenue. The question as to how best to manage an economy is therefore one of the most important issues for a society to agree on. However there, are substantial differences between individuals and political parties in terms of how they believe the economy should best be structured and managed. In this project I will look at one theory. The theory being that the more open and free an economy the better key indicators such as GDP. The project will look at how freedom in areas such as government spending, property rights, taxation and judicial effectiveness effect these indicators. Finally I will look at whether there is any relationship between covid deaths and economic indicators. After all the purpose of the economy is to increase the wellbeing of the people and we will see if there is a relationship between economic indicators and covid deaths per 100,000.

Dataset

I used three datasets from the website kaggel. The main data set I used was " The Economic Freedom Index". I used the main dataset as it is from The Hertitage Foundation which is a well funded policy think tank in the United States. Therefore, the figures in the data set should be reliable. The second dataset I used relates to death figures from different country's called "WHO-COVID-19-global-table-data". This is from the WHO and shows up to date information on covid deaths per 100,000. The third dataset I used contained GDP per Capita and was from the world bank. I used the 2018 figures from the world bank as there were a number of missing values from the 2019 figures. There would not have been any major changes between these figures in just a year. I used these sources because they are all well established organizations and should be reliable.

Implementation Process

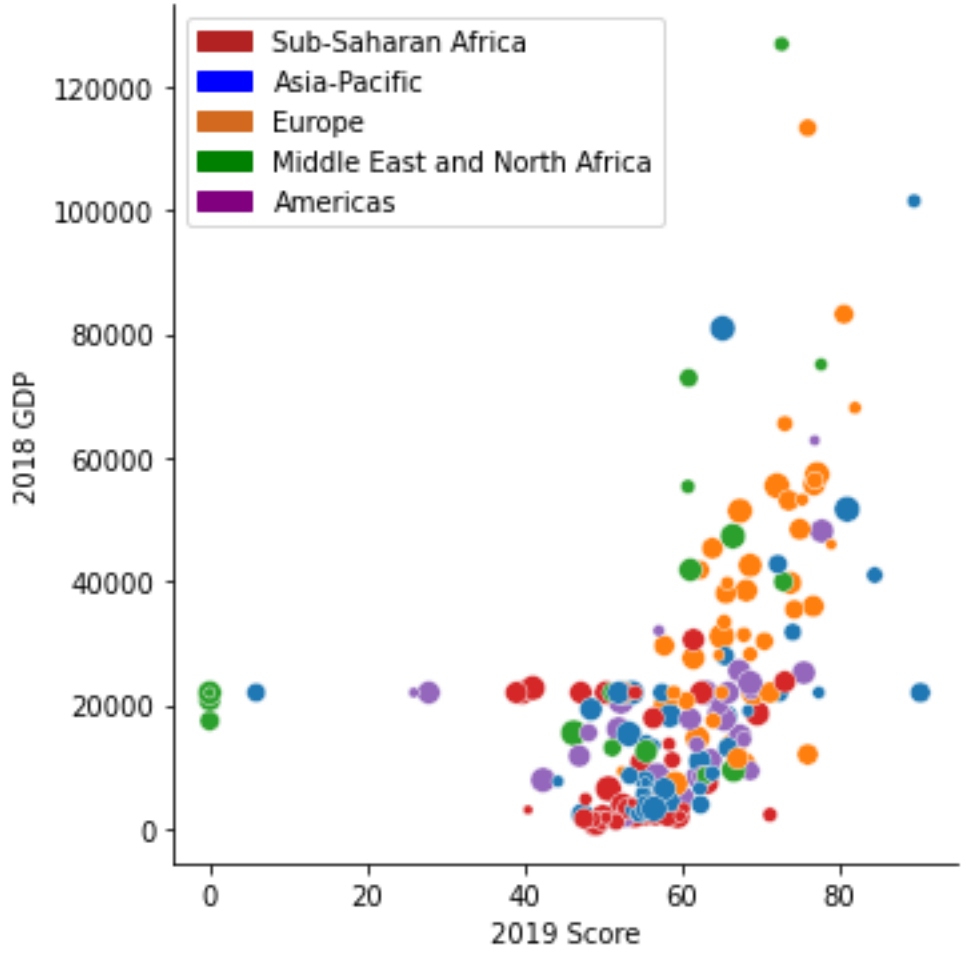
- I installed packages like numpy, pandas, matplotlib, sql connector and seaborn.
- I retrieved data from SQL and called the resulting dataframe df_SQL
- I then imported the main dataframe to do with economic freedom
- I used .head() and .tail() to check out first and last five
- I used .describe() to get a number of statistics of each column
- I then used loc and a list of columns to get specific columns and rows of the dataframe
- I then checked whether the world rank was equal to the regional rank for first two countries in dataframe and if the two country's were in the top 100
- I checked how many NA's each column in the main dataframe has
- I used groupby to get the mean for each region and created a new dataframe
- I used groupby to get the mean and median of inflation of each region
- I removed the dollar sign from the GDP per Capital (PPP) column and created a new column without the dollar sign
- I used dtypes() to find out the types of each column
- I used astype() to change the type of the GDP per Capital no Dollars
- I used a groupby to get the mean/median of GDP growth per region

- I did the same for business freedom
- I used a scatter plot to compare region to business freedom
- I imported the dataset showing gdp per country over time. I then created a new dataframe with just the country and 2018 figure.
- I then merged the two datasets together after I checked there was no duplicates in them.
- I then filled the missing column 2018 with the mean figure
- I repeated the process with the dataframe imported from SQL
- I then used a groupby to the mean and median of 2018 GDP per region
- I used a for loop to get a list of column names
- I used iterrows() to get the index and row for each line of the first two lines of the dataframe and also to create lines like “Afghanistan has a property rights scorce of 19.6 which partly explains its GDP of $69.6 billion” for each country
- I then created a function and used .apply() to passed it through the entire dataframe
- I created a new column called 'Ability to service National Debt' by dividing 'Tax Burden % of GDP' by 'Public Debt (% of GDP)'. I then checked which countries had a value higher than 1.
- I filled the NA’s in the cumulative death column with its mean and removed the column “ï»¿Name” as it was not necessary
- I then used iloc to get certain country’s and columns
- I then created a scatter plot
- I then created a line plot
- I then created a scatter plot
- I then did the scatterplot of inflation vs monetary freedom but outlier in the way
- I then got the nlargest and then droped the row with the outlier that was way off
- I used sort_values to get the highest ranked country’s in the world.
- I then got the scatter plot without the outlier
- I got 3 bar charts

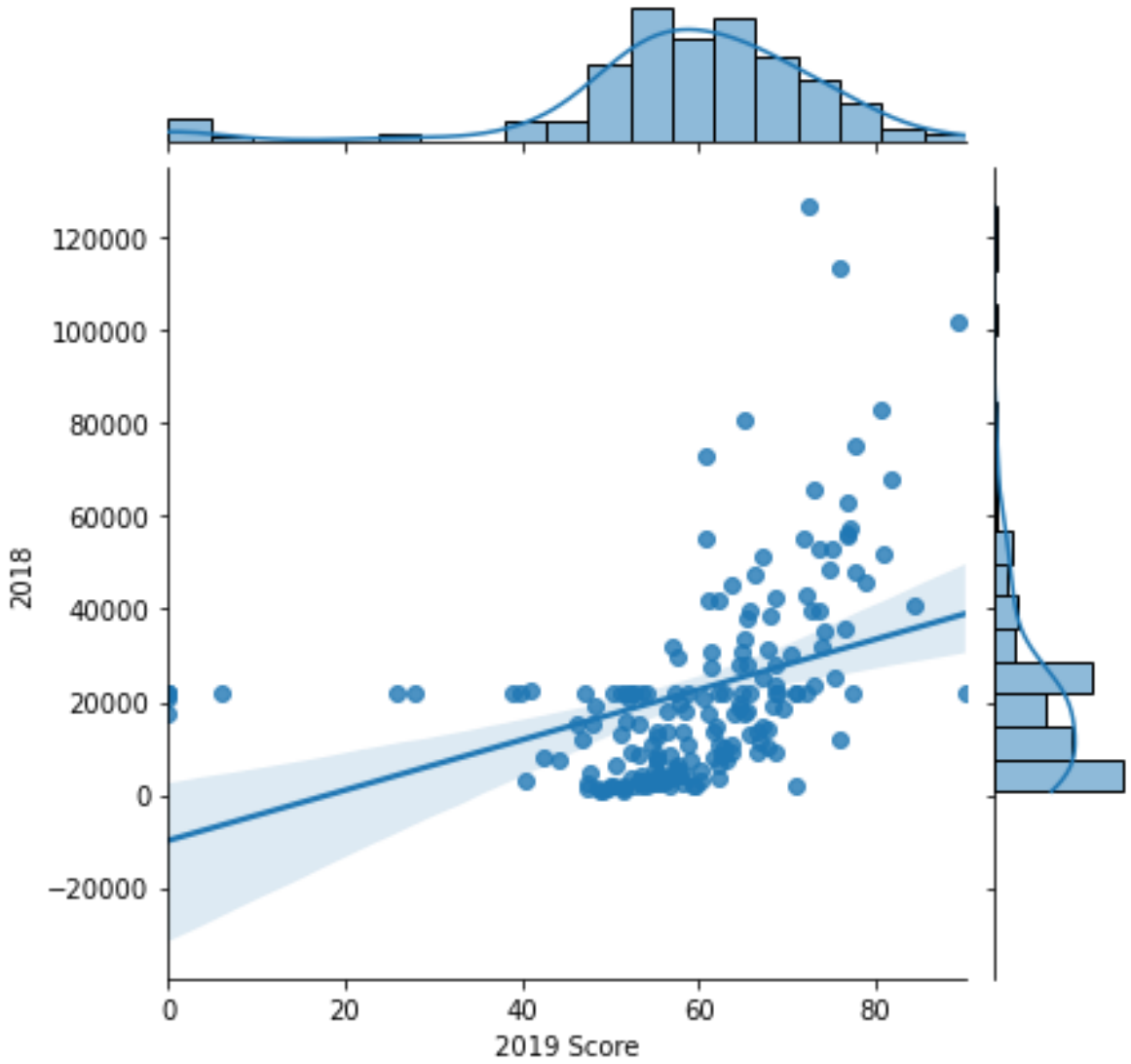
- I then got a histogram of 2019 score
- I then got a scatter plot of 2019 vs cumulative deaths
- I then got a (find name) chart of corporation tax vs 5 year growth rate
- I then got a pairplot of Gov't Spending, Inflation (%), 2018, Deaths - cumulative total per 100000 population
- I then got 3 boxplots
- I then created two new dataframe- one with high inflation and one with high GDP. I then used the describe function.
- Created heat map for govt spending, inflation, GDP and covid deaths.
- I then got the correlation for the entire dataframe
- I then created a function called Score that predicts a country's GDP based on its score in the Freedom index.

Results

Chart1- GDP vs Score

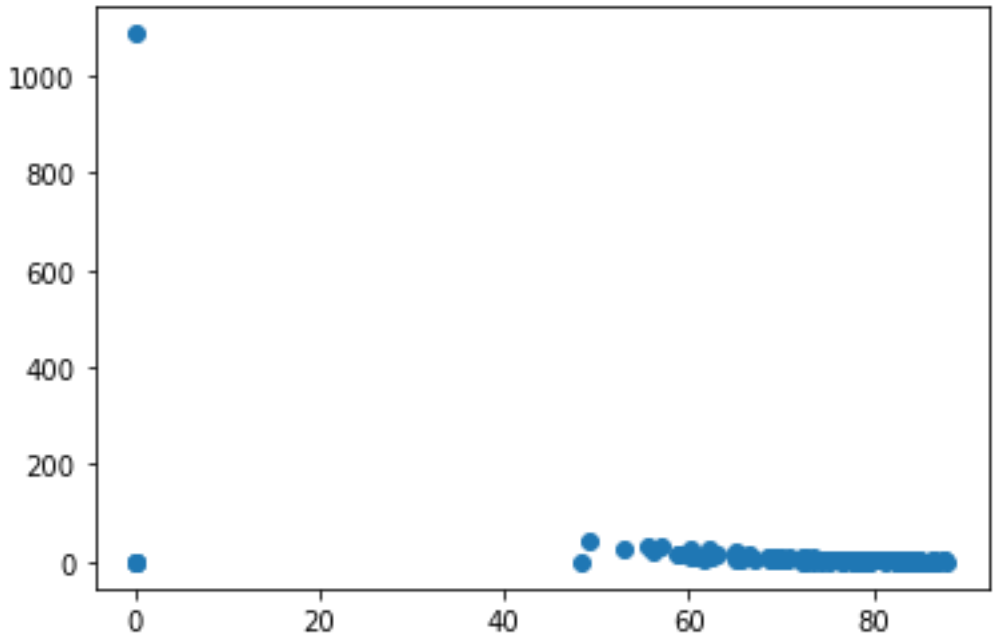


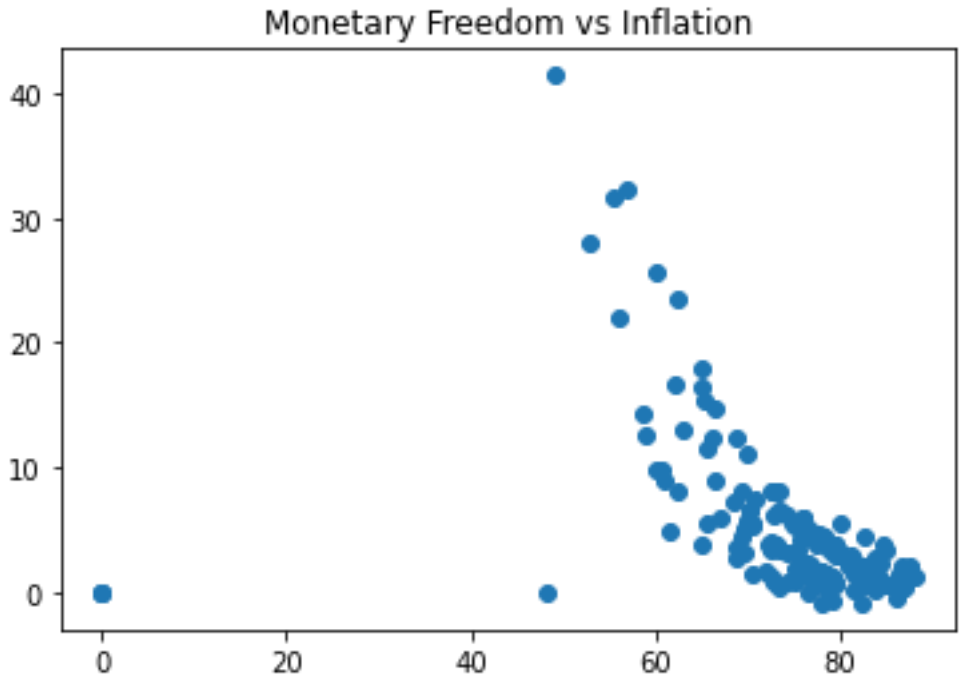
This chart shows 2019 score against 2018 GDP. Each bubble represents a country and the different colors are for each region and the size of each bubble represents the population of each country. By using color and size I added extra information to the graph. For example you can see Sub-Saharan Africa has the lowest GDP while Europe seems to have both the highest GDP and score.



This is a simpler version of the previous graph except it has a model going through it showing a positive relationship between score and gdp. It also has histograms showing the distribution of each variable.

Chart 2-Inflation Vs Monetary Freedom

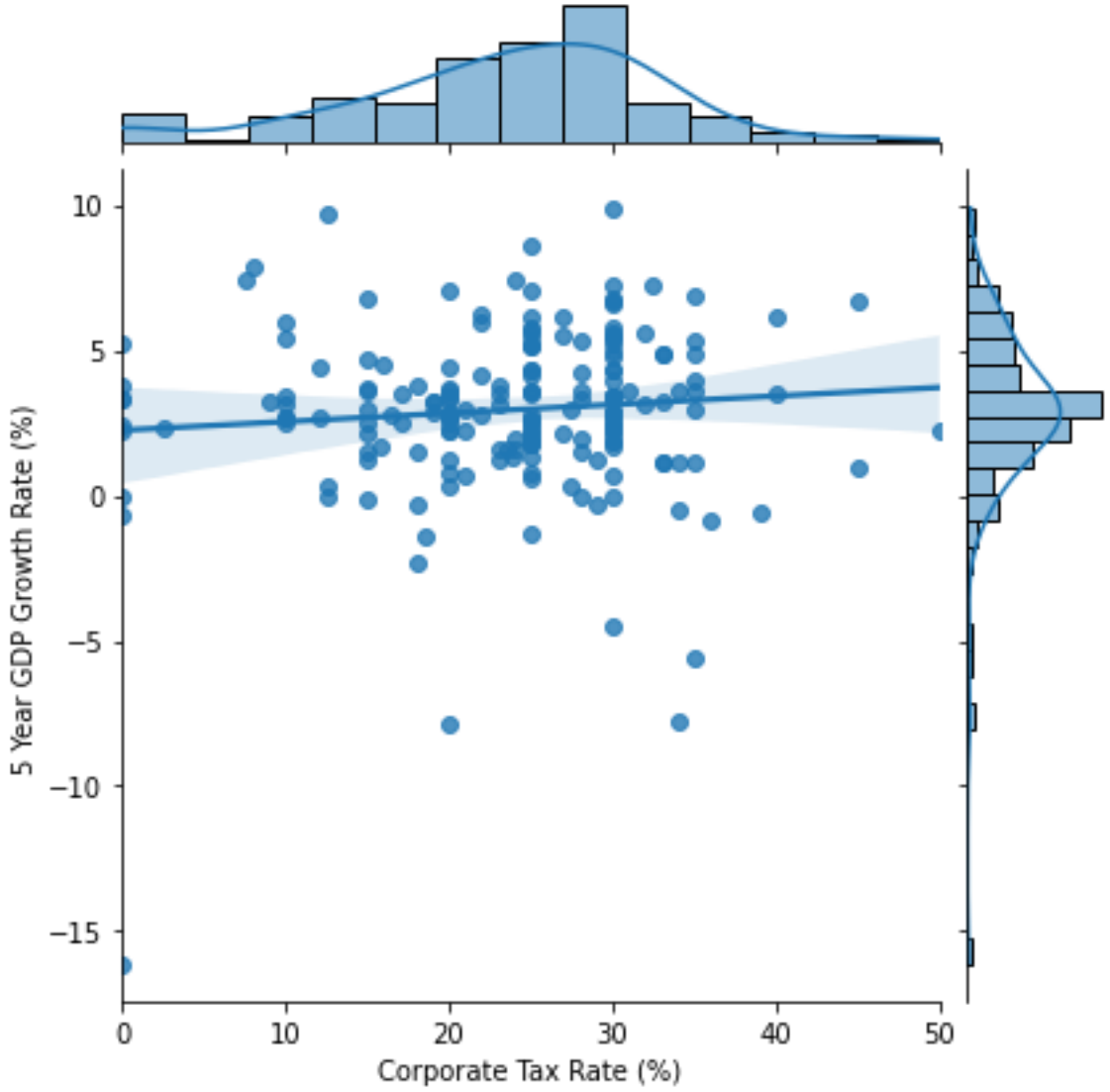




These two graphs show monetary freedom vs Inflation. The first graph includes an outlier and to make the graph clearer I removed this outlier to create the second graph.

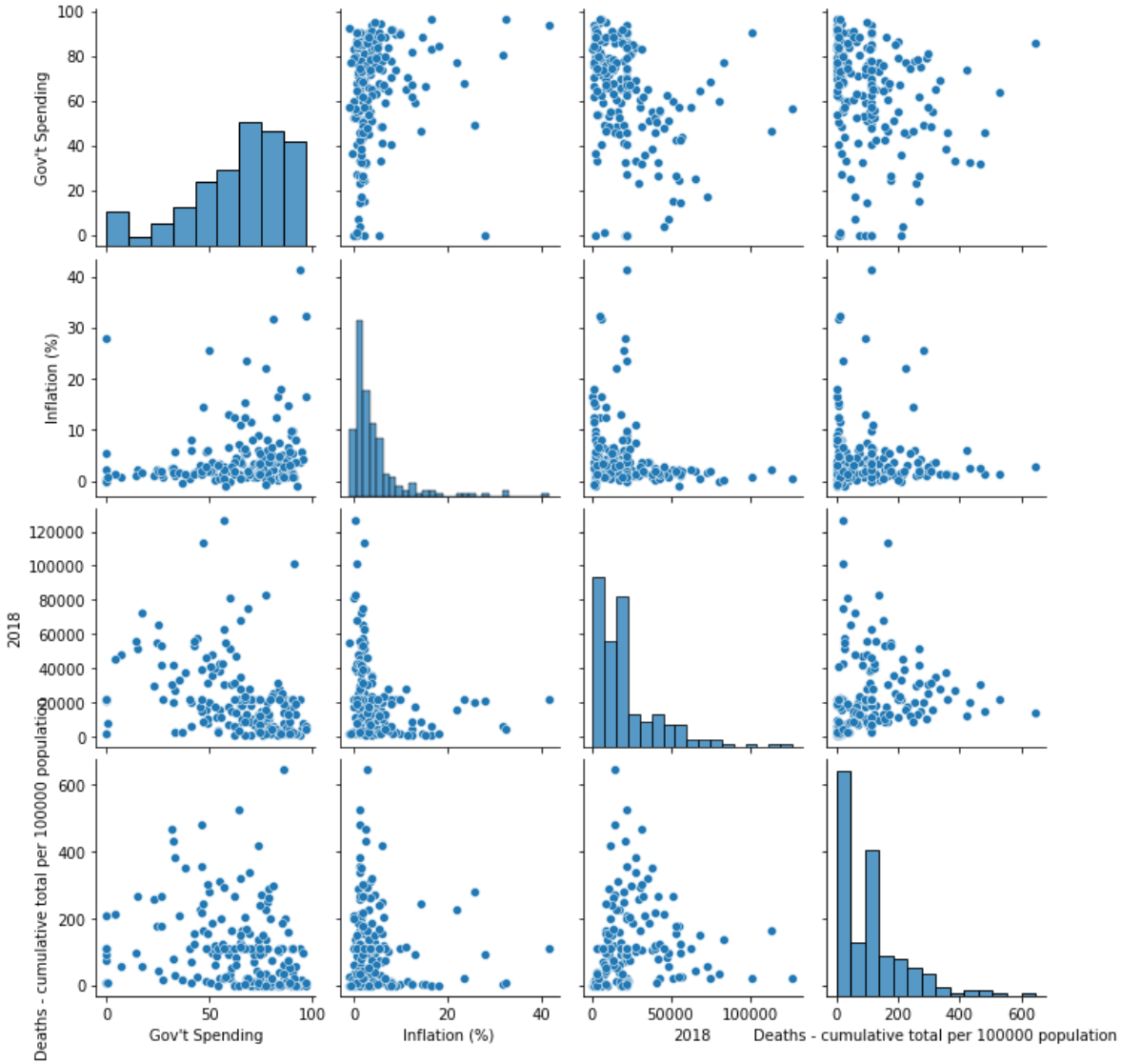
I used a scatter plot to see any correlation and I removed the outlier as it was making the graph unclear.

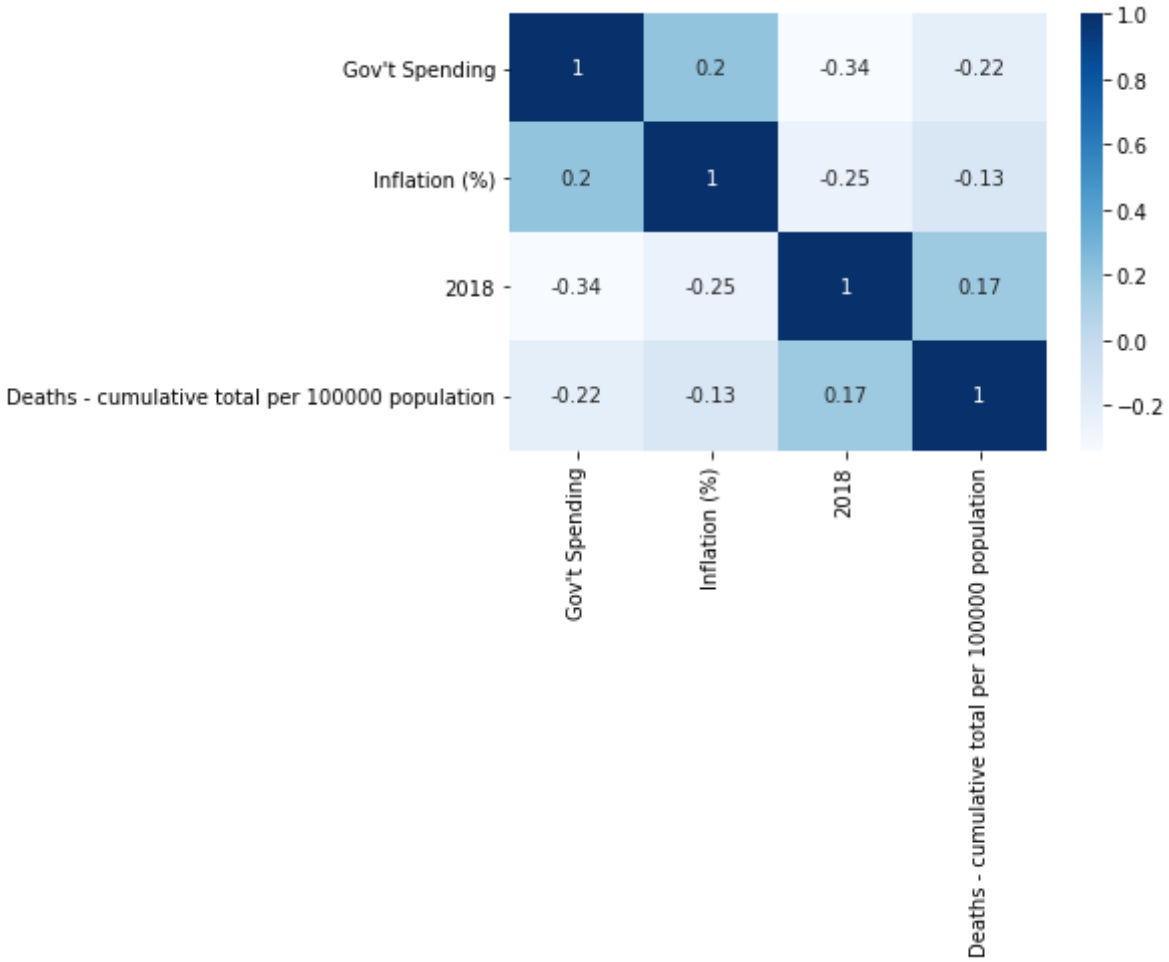
Chart 3 – 5 year growth vs Corporate Tax rate



This shows a scatterplot between corporation tax rate and 5 year gdp growth rate. It also shows histograms showing the distributions of these factors. I used a scatter plot to see if there was any correlation between these two variable and I used the histogram to see the distribution of the variables.

Chart 4-Heat Map and Pairplot





These show scatterplots and also the correlation between 4 variables- Covid deaths per 100,000, Gdp, Inflation, Gov't Spending. I used the pairplot and heatmap to get a more in depth look at the these four variables.

Insights

- Chart 1 shows relationship between GDP and 2019 score. There appears to be a positive relationship between the score in the Index of Economic Freedom and GDP. This means that if countries wish to increase there GDP it may be worth trying to maximize their score on the Index of Economic Freedom
- Chart 3 shows Corporation Tax vs 5 year growth and the chart shows that there doesn't seem to be a relationship between these two and if there was it would be very slightly positive. Therefore while Irelands reduction in corporation tax is generally considered to have increased economic growth this may not be a good strategy in general for countries to adopt and they should perhaps focus on other economic policies.
- Chart 4 shows a pair plot and heatmap of 4 variables – Gov't Spending, Inflation, GDP and Covid Deaths per 100,000. The strongest correlation is a negative relationship between Gov't spending and GDP. This would indicate that if the government of a country wants to have a high GDP then it should control its government spending.
- I got the correlation for entire main data frame- Look into what has highest correlation with GDP. Government integrity has the highest correlation with GDP. This means that if a country wishes to improve its GDP the area it should focus on first is government intergrity. This is defined as the absence of "systemic corruption of government institutions by such practices as bribery, nepotism, cronyism, patronage, embezzlement, and graft." (https://www.heritage.org/index/pdf/2022/book/02_2022_IndexofEconomicFreedom_12-ECONOMIC-FREEDOMS.pdf) (Heritage Foundation, 2022))
- Looking at the correlations in the main data frame for Covid Deaths. Tax Burden as a % of GDP has the highest correlation with covid deaths. This may be because the lower the tax burden the lower a country's ability to fight a pandemic. Governments should consider having an adequate amount of taxation or even a 'rainy day fund' to fund responses to pandemics. Otherwise it may cost lives.
- Chart two shows Inflation vs monetary freedom and the negative relationship between the two. This is in line with how the heritage foundation complies the monetary freedom statistic as it is a measure of price stability (Hertitage Foundation,2022). This therefore is an indication that the index is what it says it is.

- The histogram in the joint plot in chart 1 shows that most countries score between 40 and 80 with the distribution skewed slightly in flavor of the lower end of that spectrum.
- I used machine learning to predict GDP based on the country's score in the Freedom index. I used regression as I was predicting numerical values
- I calculated that there is 16 countries who have a higher annual tax take that national debt. These countries should be in a good position to borrow money if the need arose. However some of them may not be able to borrow money due to poor economy's or other issues. Therefore we should probably exclude the countries with low GDP per Capita.

References

Heritage Foundation 2022, Available at https://www.heritage.org/index/about (Accessed 12th April 2022)

Heritage Foundation 2022, Available at https://www.heritage.org/index/pdf/2022/book/02_2022_IndexofEconomicFreedom_12-ECONOMIC-FREEDOMS.pdf. (Accessed 12th April 2022)

Heritage Foundation 2022, Available at https://www.heritage.org/index/pdf/2022/book/02_2022_IndexofEconomicFreedom_12-ECONOMIC-FREEDOMS.pdf. (Accessed 12th April 2022)