**Overview of Data Science Project**

Project name: GDP Analysis

Instructor:Haipeng Wang(王海鹏)

University:Shandong University of Technology

Team Name:Team of Kognetics

Member: Das Polash Chandra, class id-12

Member: Barua Supriyo, class id-5

Member: Chowdhury Faiyaz Karim, class id-26

## **Overview and Motivation**

Economic growth increases state capacity and the supply of public goods. Every country wants to increase their GDP If GDP is rising, the economy is in solid shape, and the nation is moving forward. Last few years, we have seen China’s GDP growth was really amazing compare another country China can be role model for other country.

## **Related work**

Yes of course we have taken idea from your lecture and through internet also we have visited some website like Kaggle and so on.

## **Objective**

The primary goal of this project is to investigate the dataset "Countries of the World" and to focus on the elements that are influencing a Country's GDP per capita.

## **Research questions**

1. What are the factors affect the growth of GDP of a country?
2. What are the other factors other than the major factor which effect a country’s GDP?
3. What is the Co-relation between the dataset elements?
4. Is the low-birth rate Country effect the GDP of a country?
5. Which Countries of the World have the highest GDP and Why?

After the analysis of the project, we have ended up for the above five questions. As the project is about the GDP of the countries so at first, we have to find the major factor which directly affect the growth of GDP of a country. After that we have to find the others factors other than the major factors which effect the growth of GDP. Also, we have to find the Co-relation between the dataset elements of GDP factors. During the Analysis, we have found a fact that a country which have low birth rate will have a higher GDP which is very interesting. As the project progress and the questions evolve then we have to find the top countries which have high GDP.

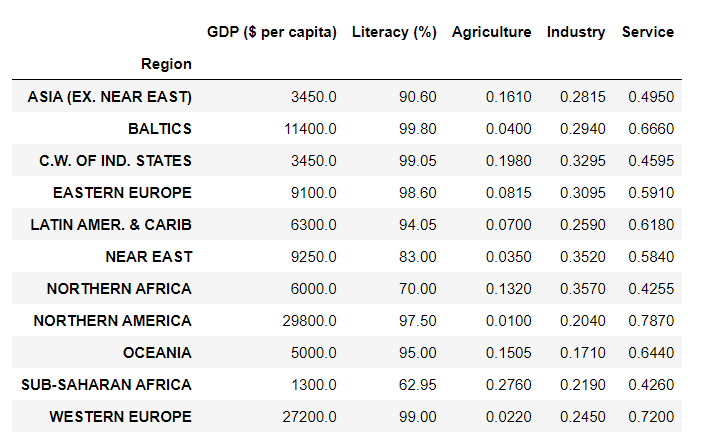
## **Dataset**

For this Project, we first collected the data of the world country GDP by surfing the internet and download the worldGDPperCapital.csv file. In the dataset, there are information about the ten factors which are needed to calculate GDP of a Country.

# Analysis

## **Data Preparation – fill in missing values**

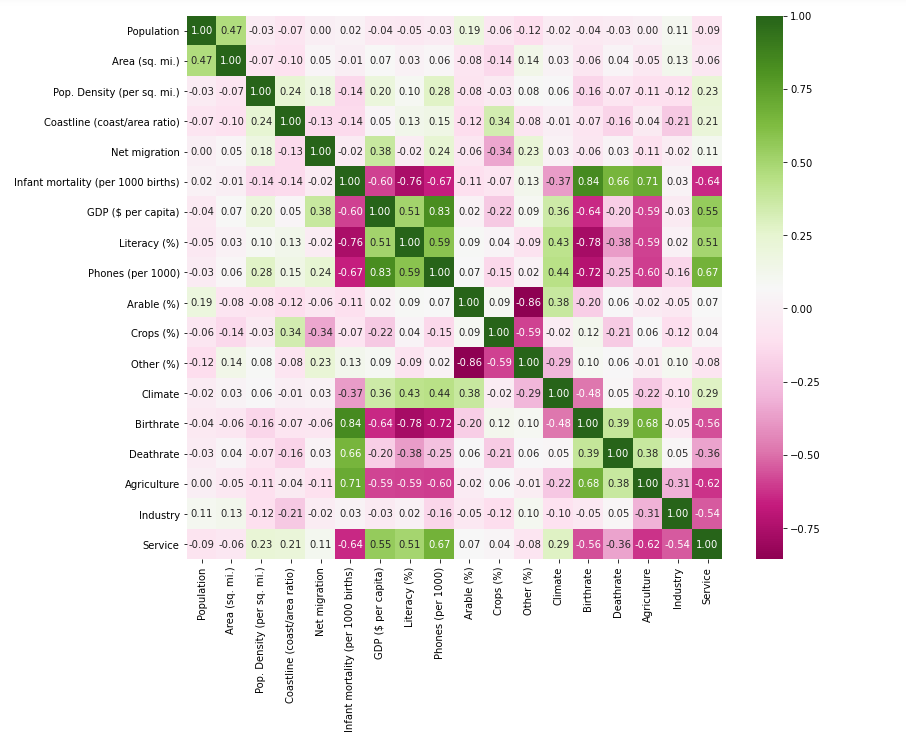
We observed that the table is missing certain information. We'll fill in the missing data with the median of the region that a country belongs to because countries that are geologically adjacent are often similar in many aspects. Let's look at the region medians for 'GDP ($ per capita),' 'Literacy (percent),' Agriculture, Industry, and Service,' for example. Note that we use the mode instead of the median for 'climate' because it appears that 'climate' is a categorical feature here.



## **Data Exploration & Visualization**

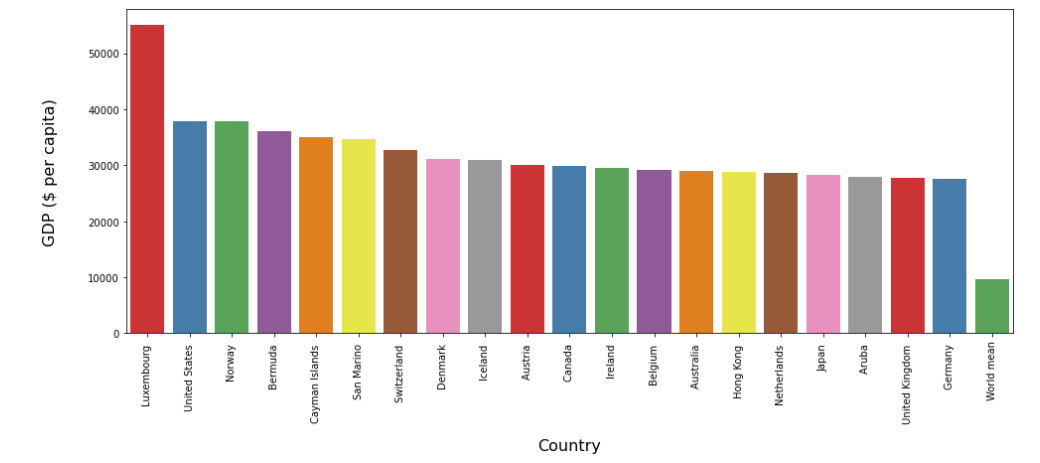
### **Correlation between Variables**

Here we need to show the correlation between all numerical columns. So, here we using Seaborn and Seaborn is a data visualization library that is built on top of matplotlib and contains a direct function to create heatmaps. The heatmap shows the correlation between all numerical columns.



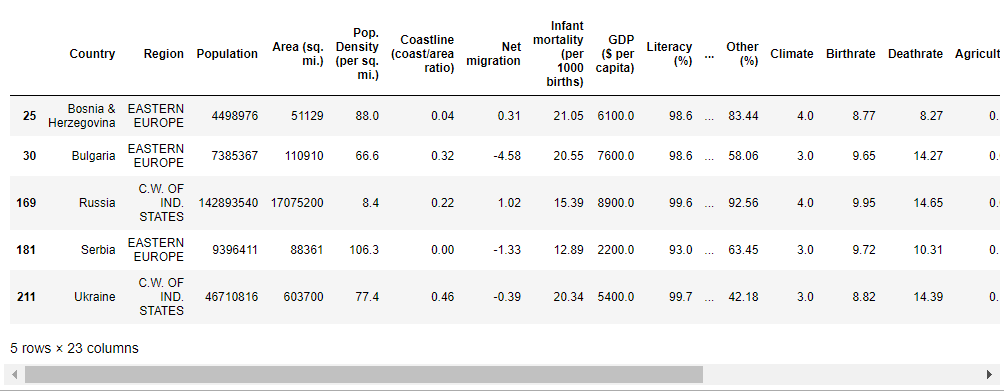
### **Top Countries with the highest GDP per capita**

In this bar chart top 20 countries with the highest GDP per capita. Luxembourg's GDP per capita is around $55000. Luxembourg is higher than the next 19 countries, the next 19 countries are close. Germany, the 20th has about 2.5 times GDP per capita of the world average.

****

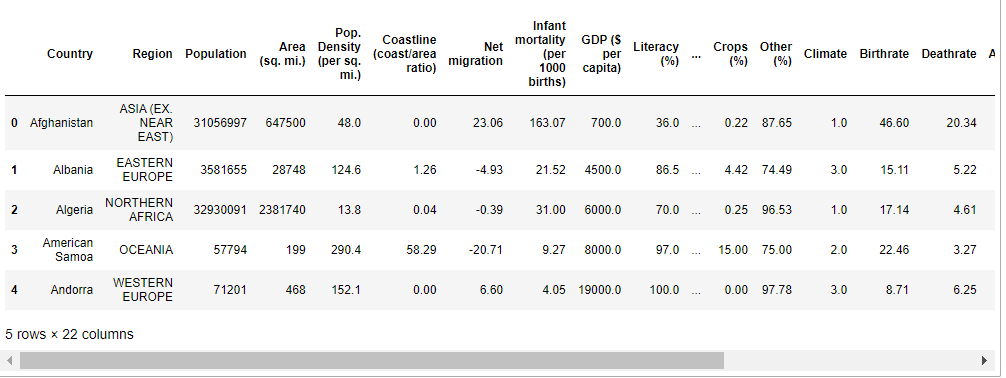
### **Countries with low Birthrate and low GDP per capita**

A few highlights, similar to phones, are connected with the average GDP all the more straightly, while others are not. For instance, High birthrate generally low GDP per capita, however normal GDP in low birthrate countries can change a lot.  
Let’s look at the countries with low birthrate (< 10%) and low GDP per capita (< 10000$). They also have high literacy, like other high average GDP countries. But we hope their other features can help distinguish them from those with low birthrate but high average GDPs, like service, are not quite an important portion in their economy, not a lot phone procession, some have negative net migration. We can see 5 countries (Bosnia & Herzegovina, Bulgaria, Russia, Serbia and Ukraine) here low birthrate and also low GDP per capita from eastern Europe or C.W. of IND. STATES, so the ‘region’ feature may also be useful.

****

## **Training and Testing**

First label encoding the categorical features ‘Region’ and ‘Climate’, and we are just use all features given in the data set without further feature engineering.

****

We are using the linear regression model here. As for metric, we are check both root mean squared error and mean squared log error.

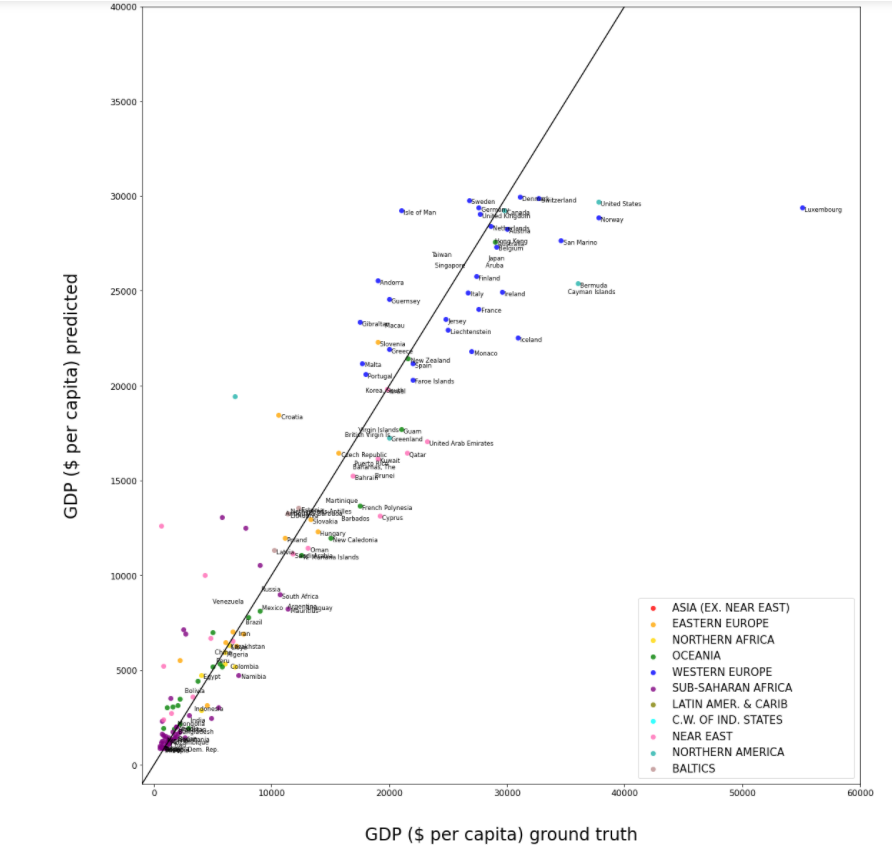
****

As we know, the target isn't linear and has a lot of features, it's worth experimenting with nonlinear models like the random forest model.

****

## **Visualization of Results**

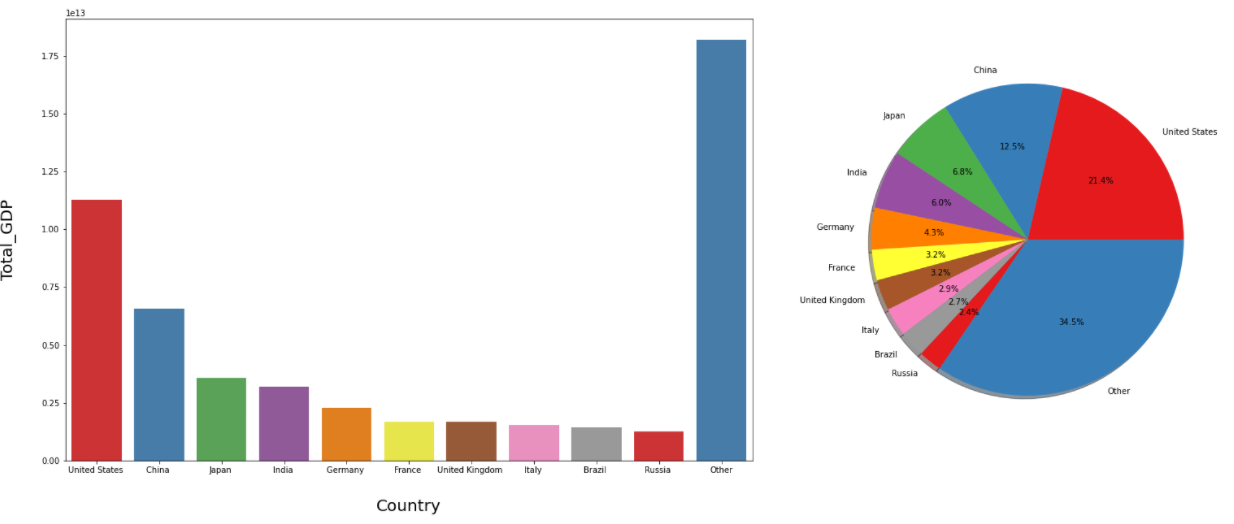
To see how the model is doing. is it far from here, we can make scatter plot of prediction against ground truth? The model gives a reasonable prediction, as the data points are gathering around the line y=x.

****

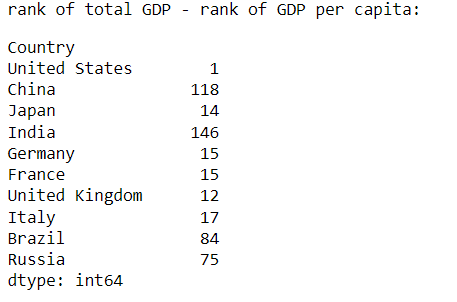
## **Total GDP**

### **Top Countries**

It is additionally fascinating to check out the all-out GDPs, which I take as 'Gross domestic product ($ per capita)' × 'Populace'. Here are the top 10 countries with highest total GDPs, their GDP make up to about 2/3 of the global GDP. So, these top 10 countries are playing a very important role in the world's GDP.

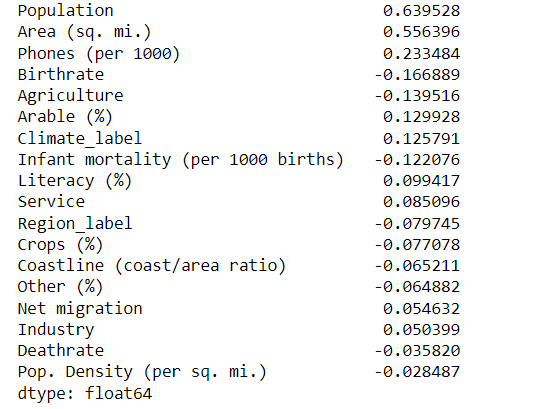
****

Here we compare the above ten countries’ rank in total GDP and GDP per capita.

****

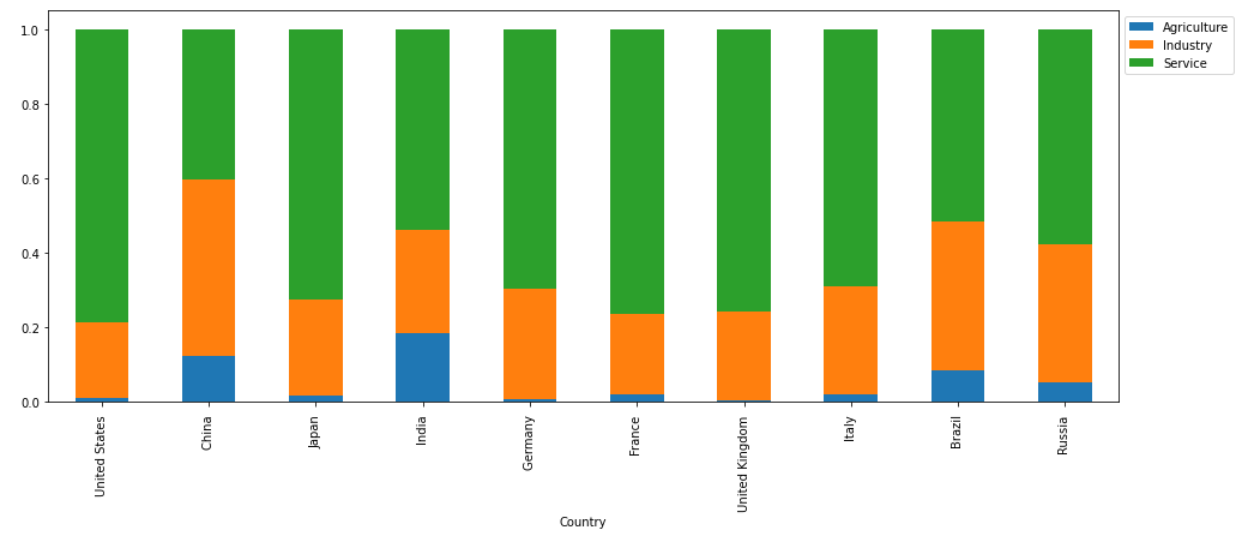
### **Factors affecting Total GDP**

Here we can also check the correlation between total GDP and the other columns. The top two factors are population and area, following many factors that have also been found mostly correlated to GDP per capita.

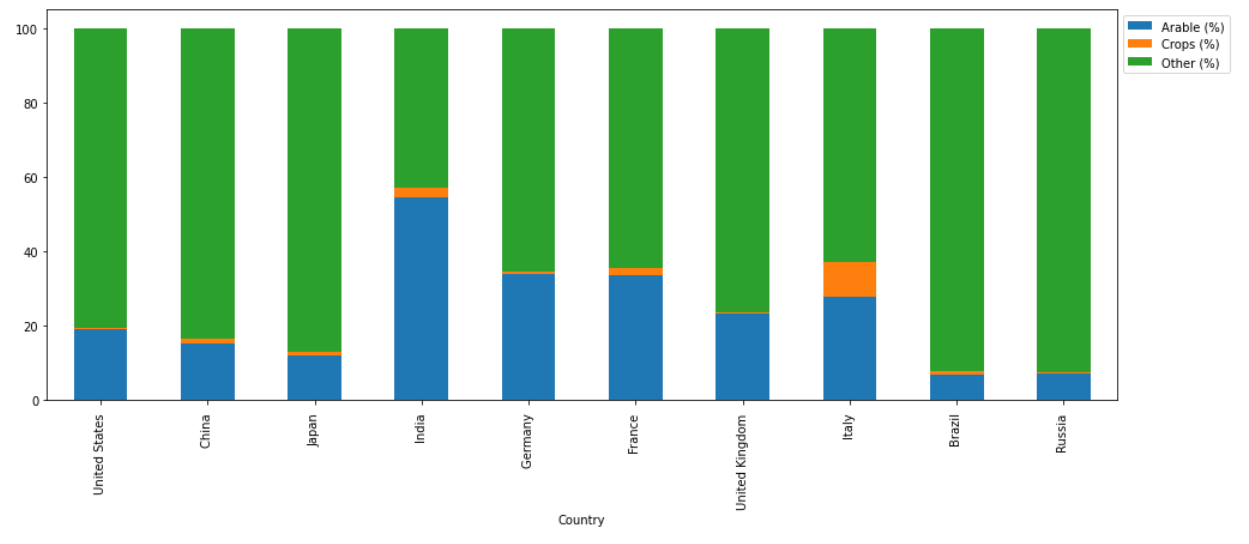
****

### **Comparison of the Top 10 countries**

Finally, let us do a comparison of the economic structure for the ten countries with the highest total GDP. Here we are comparing three major sectors like the primary sector (Agricultural), the secondary sector (Industry), and the tertiary sector (services) all contribute to GDP by producing goods and services (Services).

****

As well as their land usage

****

**Team work overview and contribution**

In this GDP (Gross domestic product) data science project we are working on a month-long. The goal of this project is to analyze the dataset "Countries of the World" and see what factors impact a country's GDP per capita. We'll focus on data collection, data cleaning, data exploration, analyzing, visualizing, extracting, managing, and storing data for this project. Using data science in GDP analysis will enable us to know the factors that will affect the GDP per capita of various countries.

An important part of the project is the technical report (TR). We have made a technical report (TR) and there we have discussed in detail the steps taken in developing solutions, including where did we get the data, describing step-by-step statistical, and including many visualizations.

In our technical report (TR), we are focusing on some important topics like overview and Motivation, Related work, Initial questions, dataset, and analysis.

We have completed this data science project by dividing it into several parts and we have decided through the google meeting who will take which part of the project. We are discussing together how to complete those parts of the project, many of times team members occur some problems while doing their specific tasks and then informs us about it and we meet again to discuss how the work can be done and try to find a solution. Sometimes I also have occurred some problems, so I tell my team about my problem and we all discuss together and solve this problem.

I am really too grateful to our all-team members. In our (Team of Kognetics) has worked hard. And I especially thank Supriya Barua he works hard all the time, and our other team member Faiyaz Karim Chowdhury he has done wonderful works on this project.

**Our every members GitHub link below**

1. Das Polash Chandra

GitHub Link: <https://github.com/polashdas040/GDP-Analysis.git>

2. Barua Supriyo

GitHub Link: <https://github.com/Supriyoba/GDP-Analysis.git>

3. Chowdhury Faiyaz Karim

GitHub Link: <https://github.com/FaiyazKarim/Data-Science-Analysis.git>

Thank You