Team Extreme

Badar Nasser Hilal Alkharusi , Dilpa Rao , Divya Arora , Inshal Khan , Shahzeb Jahangir , Tayyab Ijaz

06/12/2019

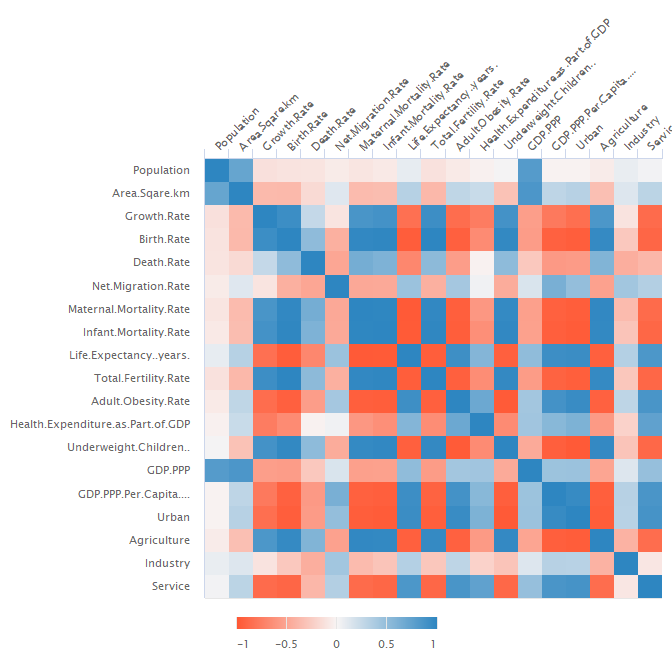
# Introduction

GDP is one of the most important indicators in determining the performance of country economy. The significant factors affecting GDP are population, agriculture, service, industry, health, migration, urban and obesity which are recorded in the dataset.

Null hypothesis is that population,agriculture,service,industry,health,migration,urban and obesity are the factors not affecting the gdp per capita of the countries. By analyzing the data we have to find out whether or not these factors affecting gdp per capita of the country.

For undergoing the analysis, we have used CIA\_Factsheet dataset that records different factors affecting gdp of the countries for the year 2013.In this report, we will briefly descibe the factors and validate the stated hypothesis.

We conducted a preliminary analysis based on the Correlation matrix which is showing the relationship of the various variables with the gdp per capita of the country.We can clearly see from the matrix given below that some variables play a very important role in determing the GDP and the GDP Per Capita of a country.



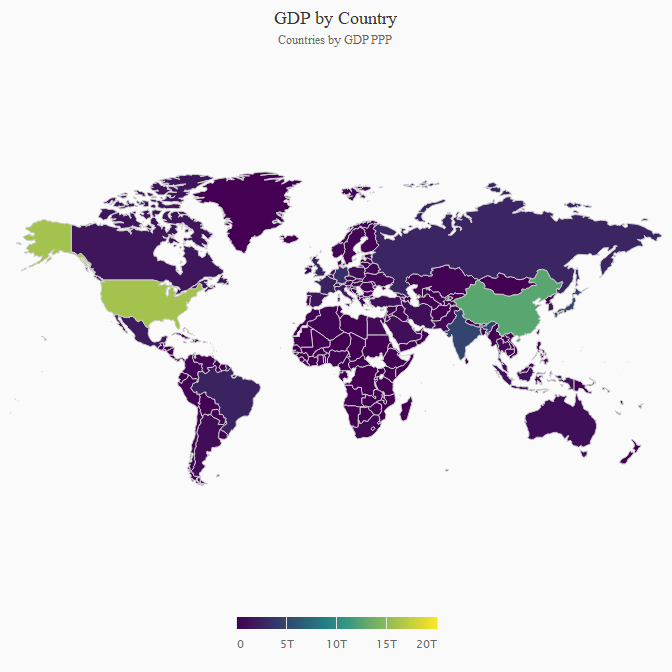
## Data

The dataset consists of twenty one factors with respect to the countries in 2013.The format of the dataset is in CSV and contains information regarding the following variables : Countries,Continent,population,Area Sqare km,Growth Rate,Birth Rate,Death Rate,Net Migration Rate,Maternal Mortality Rate,Infant Mortality Rate,Life Expectancy (years),Total Fertility Rate,Adult Obesity Rate,Health Expenditure as Part of GDP,Underweight Children %,GDP PPP,GDP PPP Per Capita,Urban,Agriculture,Industry,Service.

## Methods

We have used descriptive statistics to describe our data. Based on our previous observation from the correlatopn matrix we have used various statistical and mapping tools like correlation graph,world map, bar charts,treemap and heatmaps to understand the key factors affecting the GDP and the GDp Per Capita of a country and will try to answer as many as questions related to our analysis.

### World Map by GDP.PPP



**Observations from Gdp by country,continent,population,urban**

* United States of America have a very high GDP followed by China,India,Japan,Germany,Russia while South Africa,Egypt have very low GDP .It is observed that Brazil and United Kingdom,Saudi Arabia and Taiwan have equal GDP.
* North America Continent typically have very strong GDP followed by Asia and Europe while other countries are lacking behind.
* It is witnessed that Asian countries like India and china are leading in terms of population and as a result their GDP is high.
* Urbanisation is directly proportional to GDP Per Capita, the countries which have high gdp per capita results in higher urbanisation.

**Observations from health expenditure,migration rate,adult obesity rate,life expectancy:**

* United States of America and Sierre leone have high expenditure on health while Italy,Sweden and United kingdom have almost equal and lowest expenditure on health.
* Qatar,United Arab Emirates gdp per capita are highly affected by migration while Bermuda,Malta,Netherlands are least affected.
* Adult obesity rate is directly dependent on gdp per capita.The higher per capita the population will be more obese.
* Estimation of average age that members of corresponding population survived is directly dependent on the gdp per capita of the country.

**Observations from agriculture,industry and service**

* Agriculture is dominating the GDP in countries such as Sierra Leone, Somalia and Chad whereas Qatar and Luxembourg have least.
* Bermuda,Turkmenistan, The Bahames are having high rate of Services which in turn increases the GDP per capita of the countries, while North Korea and Angola are have very few Services.
* Industry is highly correlated with GDP, Angola Kuwait and Brunei are the leading players whereas Bermuda and Sudan are lacking behind in industrial sector.

## Results and Discussions

We are working on the factors affecting gdp of country world wide for the year 2013 based on the CIA dataset. On analyzing the data list of factors came into forefront which bears a positive relation with the GDP.

Our preliminary analysis was done using correlation matrix which sheds light on interesting relationship between various factors.It was analysed further by using various stastical and mapping tools between response and the predictor variables that population urban,health,migration,obesity,service sectors plays a significant role in determining the GDO Per Capita of the world by showing a strong correlation.

As clearly interpreted by the correlation matrix many factors like area,growth rate,birth rate,death rate,mortality rate,life expectancy,fertility rate and underweight children are insignificant.So these variables were not analysed further. We are investigating what are the factors affecting gdp per capita of the country.After analysing, we have got eight factors i.e population,agriculture,service,industry,health,migration,urban and obesity affecting the gdp per Capita.

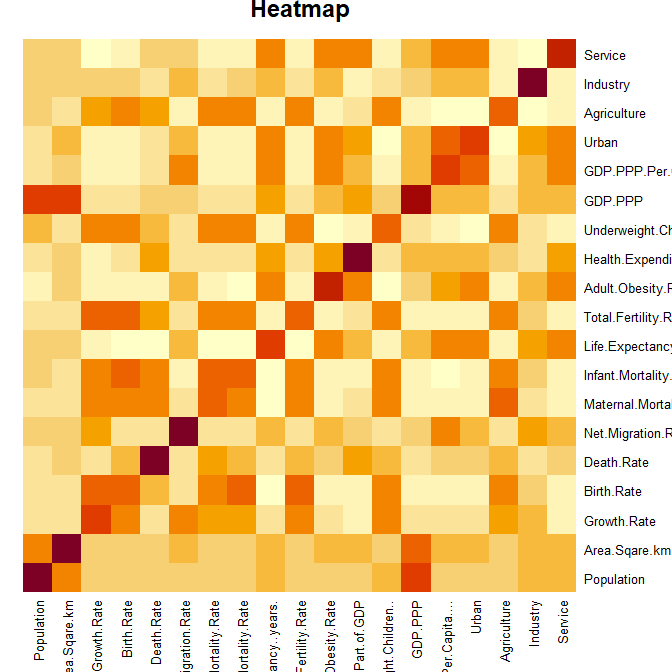
After doing exploratory data analysis and visualization,we are rejecting the null hypothesis by saying that population,agriculture,service,industry,health,migration,urban and obesity are the factors that affects the gdp per capita of the countries.

## Interesting Questions

* What are the most significant factors affecting the gdp per capita of country ?
* What are the factors having negative correlation with the gdp per capita of the country ?
* Which country and continent has the highest gdp ?
* Which sector is affecting the gdp most ?
* Which countries population are having a positive correlation with the gdp per capita of the country ?
* Which countries health expenditure is affecting the gdp of the country ?
* How is it that obesity is affecting the gdp per capita of a country?

## Conclusions

We have used correlation graph,principal component analysis and heat maps for concluding our analysis.As,correlation matrix is used to show the association between two variables,heat maps are the basically the color coded system used to determine the variables which matters the most and principal component graphs is a statistical procedure used to reduce number of factors in our data by extracting important factors affecting gdp per capita among all factors.



```

Correlation matrix and heatmap is showing the relationship of the other factors with the gdp per capita of the country clearly.it is illustrated that Industry,Net Migration Rate,Service,Life Expectancy,Adult obesity are majorly affecting the gdp per capita of the country while some of the factors are least affecting the gdp such as Infant Mortality Rate and Birth Rate.

## References

<http://jkunst.com/highcharter/>

<https://www.datanovia.com/en/lessons/heatmap-in-r-static-and-interactive-visualization/>

<https://davetang.org/muse/2018/05/15/making-a-heatmap-in-r-with-the-pheatmap-package/>

<https://www.highcharts.com/blog/data-science/highcharts-for-r-users/>