



**L** OVELY  
**P** ROFESSIONAL  
**U** NIVERSITY

# **DATA SCIENCE PROJECT REPORT**

(Project Term January-April, 2019)

## **Project-Countries of the world**

Submitted by

Prithve Kumar P

**Course Code: INT232**

Under the Guidance of

**Mrs. Amanpreet Kaur**

**School of Computer Science and Engineering  
Phagwara**

# Introduction

## Data science

Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Data science is the same concept as data mining and big data

## Data analytics

Data analytics (DA) is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software. Data analytics technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions and by scientists and researchers to verify or disprove scientific models, theories and hypotheses.

Data analytics initiatives can help businesses increase revenues, improve operational efficiency, optimize marketing campaigns and customer service efforts, respond more quickly to emerging market trends and gain a competitive edge over rivals -- all with the ultimate goal of boosting business performance.

## Types of data analytics applications

At a high level, data analytics methodologies include exploratory data analysis (EDA), which aims to find patterns and relationships in data, and confirmatory data analysis (CDA), which applies statistical techniques to determine whether hypotheses about a data set are true or false.

Data analytics can also be separated into quantitative data analysis and qualitative data analysis. The former involves analysis of numerical data with quantifiable variables that can be compared or measured statistically. The qualitative approach is more interpretive -- it focuses on understanding the content of non-numerical data like text, images, audio and video, including common phrases, themes and points of view.

## Open Source Data Tools

**Knime**

**OpenRefine**

**R-Programming**

**Orange**

**Pentaho**

## Data Visualization Tools

**Datawrapper**

**Solver**

**Tableau Public**

**Google Fusion Tables**

## Sentiment Tools

**Opentext**

**Trackur**

**SAS Sentiment Analysis**

## Data Extraction Tools

**Octoparse**

**Content Grabber**

**Import.io**



Interactive, visual analysis lets the untangle thorny business questions and quickly get to the insights that propel the business forward. Powered by the patented VizQL technology, Tableau gives powerful analytics to ask deeper questions and deliver more meaningful answers.

**Collaboration**

**Analytics**

**Content Discovery**

**Data Prep**

## Project - Countries of the world

We classify the countries on the basis of **area and population**

If the population is more, we generally do not prefer the country

The countries with less area and more population



In these countries we generally do not prefer to go

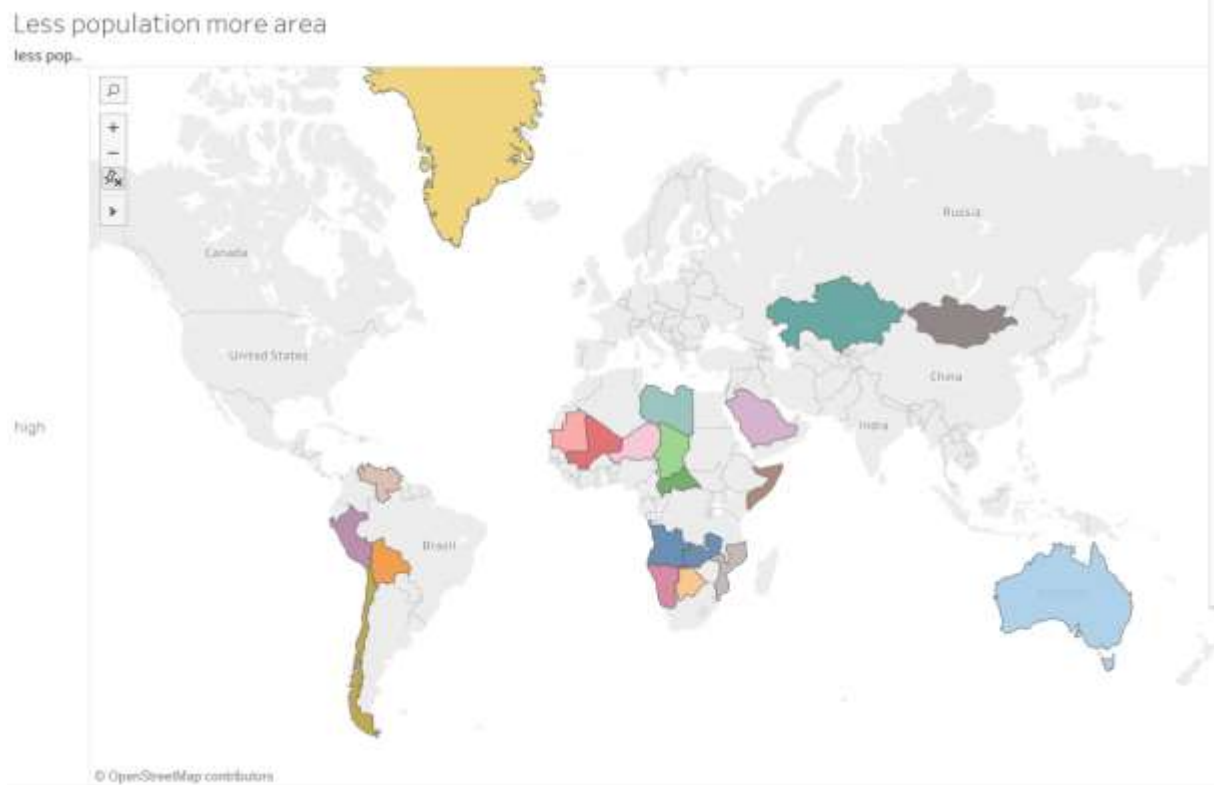
Tourism will be very less

Huge problem of employability

Increased demand for resources such as fresh **water** and food, **starvation** and **malnutrition**

Diseases such as tuberculosis, **malaria**, **HIV**, and dysentery spread faster in overpopulated areas.

## The countries with more area and less population



In these countries we generally prefer to go

Tourism will be more

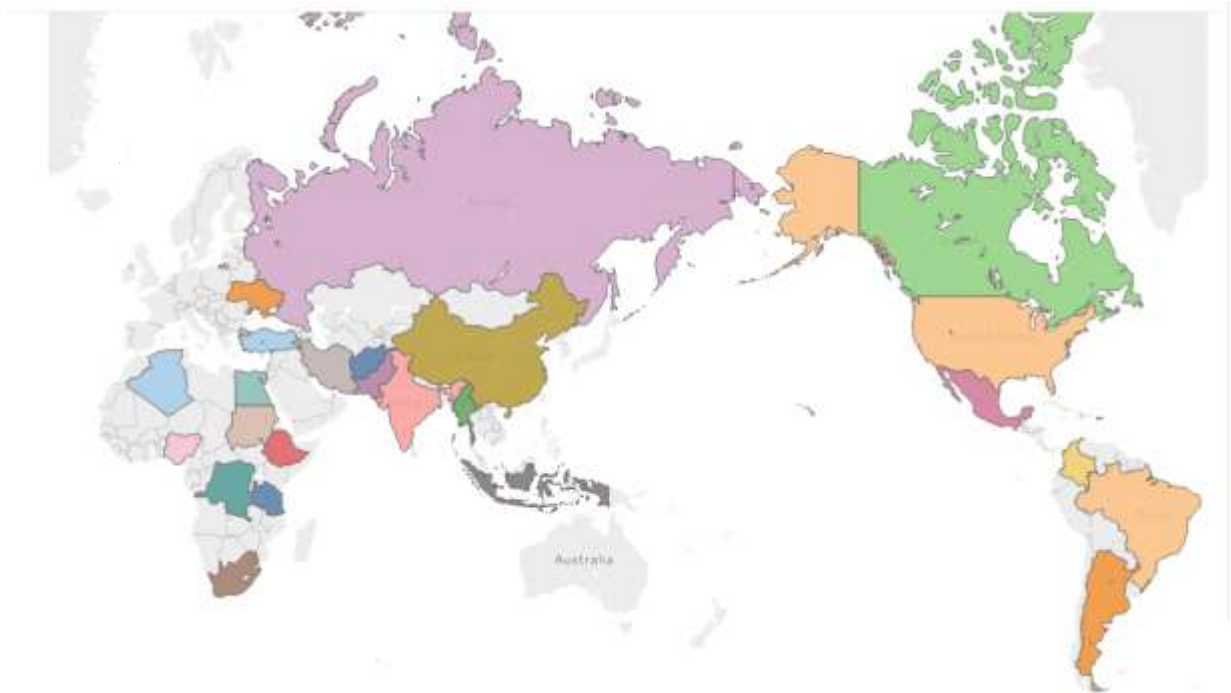
Decreased Air pollution

Less polluted water

Less global warming

Less Deforestation and Hunting of Animals

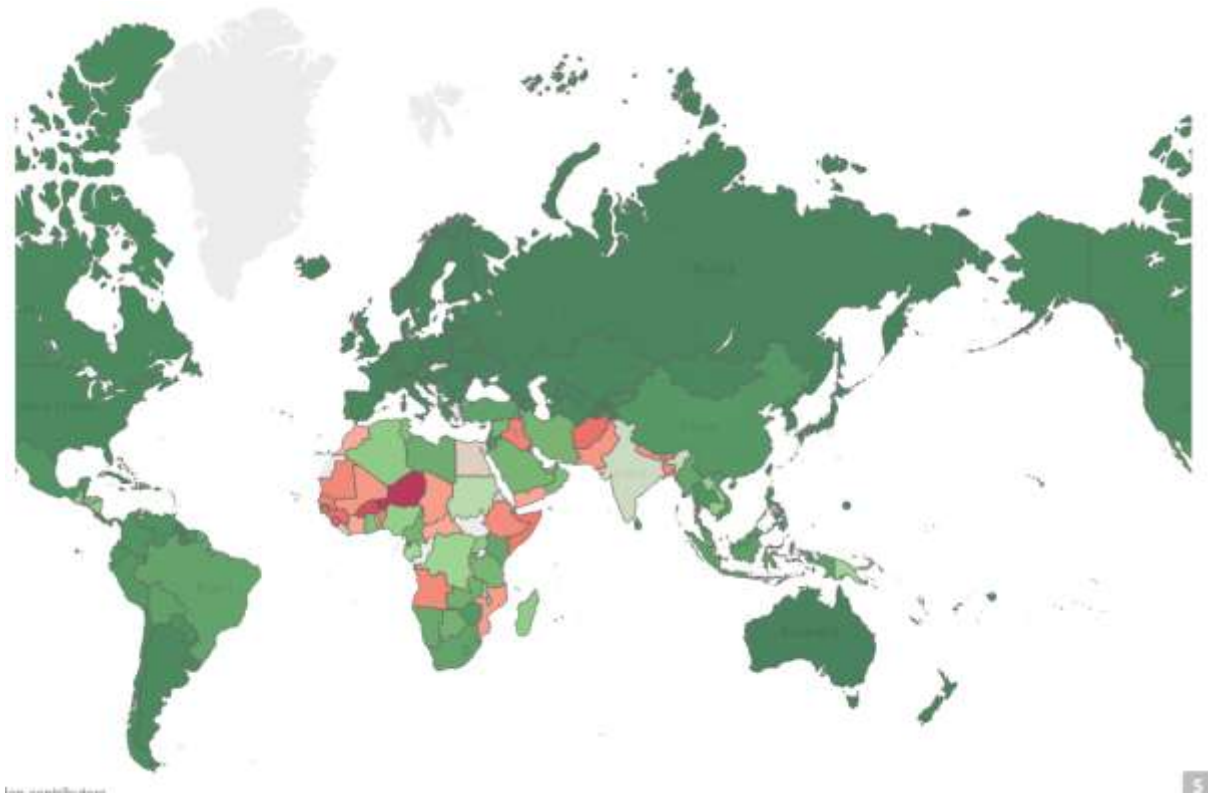
## The countries with more area and more population



## The countries with less area and less population



## Literacy Rate



The countries having Dark Green are having good Literacy Rate ( $> 90$ )

The countries having Light Green are less Literacy rate (50 - 90)

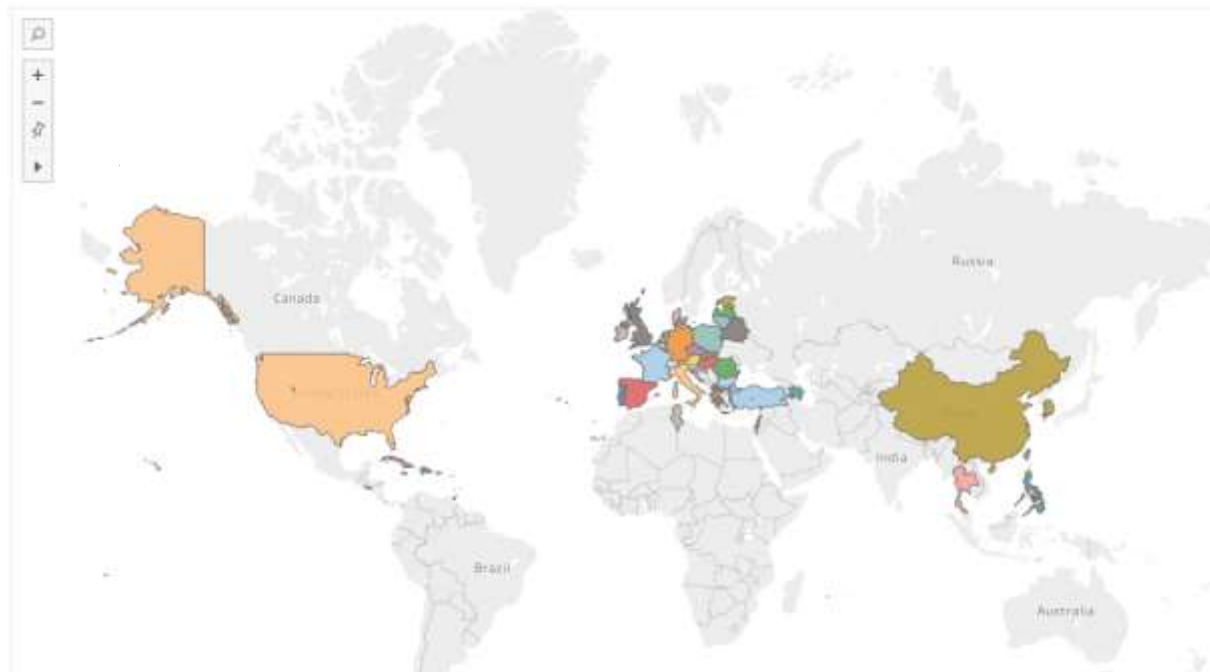
The countries with Orange are below average (20-50)

The countries with Red ( $< 20$ )

### Illiteracy leads to:

- Limited ability to obtain and understand essential information
- Lower income
- Lower-quality jobs
- Unemployment
- many positions remain vacant for lack of personnel adequately trained to hold them
- The difficulty understanding societal issues lowers the level of community involvement and civic participation

## Highly Arable but less agriculture



These are the countries in the world where there are large amount of land for agriculture but do not perform agriculture. They use land for industries or left useless.

Less availability of food

Must Import food products which will in turn increase in price of the products

Decreased agricultural Byproducts ex: cotton and wool



## Large Coastlines

high coastline



Darker region has larger Coastlines

More and more of the world's people live in coastal regions

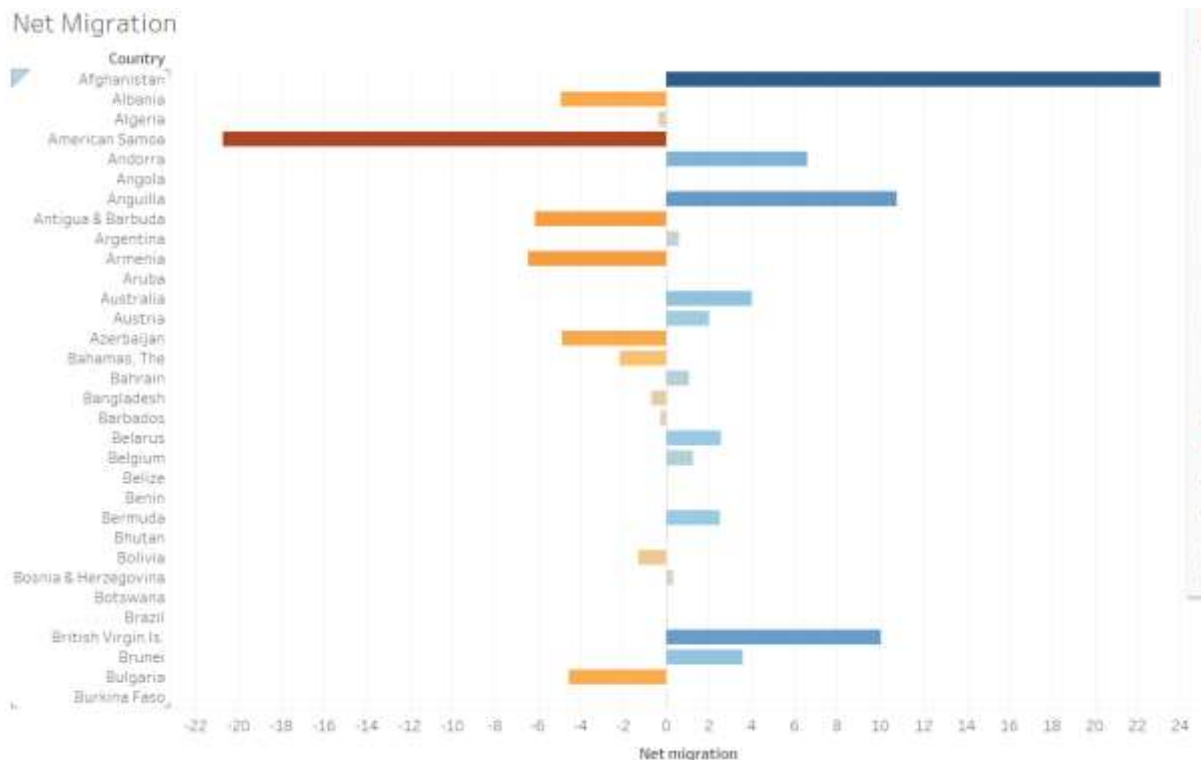
Many major cities are on or near good harbors and have port facilities

The coast is a frontier that nations have typically defended against military invaders  
smugglers and illegal migrants

Coasts, especially those with beaches and warm water, attract tourists

1	Canada	202,080	125,567
2	Norway	58,133	36,122
3	Indonesia	54,720	33,939
4	Russia	37,653	23,396
5	Philippines	36,289	22,549

# Net Migration



Migrations occur over a series of different push and pull factors that revolve around social, political, economic, and environmental factors

The **net migration** rate is the difference between the number of immigrants and the number of emigrants throughout the year

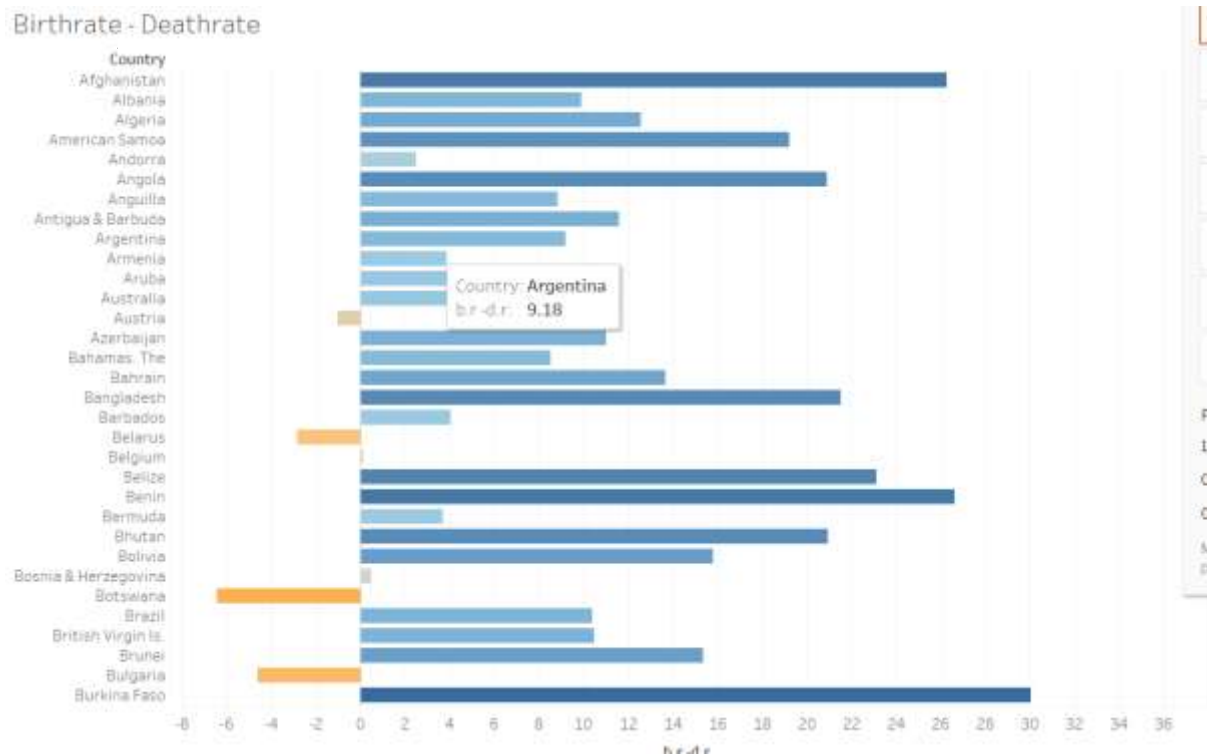
When the number of immigrants is larger than the number of emigrants, a **positive net migration** rate occurs

When more emigrate from a country, the result is a **negative net migration** rate

When there is an equal number of immigrants and emigrants, the net migration rate is balanced.

**In the above fig, Bars on left side is Negative net and on right side is Positive net**

## Birth Rate minus Death Rate



If birth rate minus death rate is positive then there is increase in population

If it is in negative then there is decrease in the population

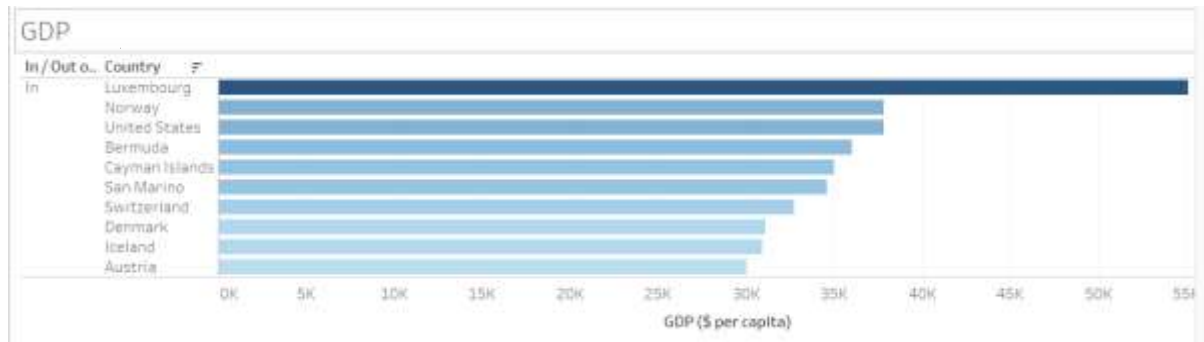
The increase or decrease in population are based on many factors include Health, Society, Environment

If there is any attack of a severe virus then there may be decrease in population since death rate is more..

There are chances of decrease in population due to lack of food (starvation)

**In the above fig, the orange bars are negative rate and blue indicates positive rate**

# Gross Domestic Product



Gross Domestic Product (GDP) is a broad measurement of a nation's overall economic activity

GDP is the monetary value of all the finished goods and services produced within a country's borders in a specific time period.

GDP includes all private and public consumption, government outlays, investments, additions to private inventories, paid-in construction costs etc.

If GDP goes up, the value of these things has gone up

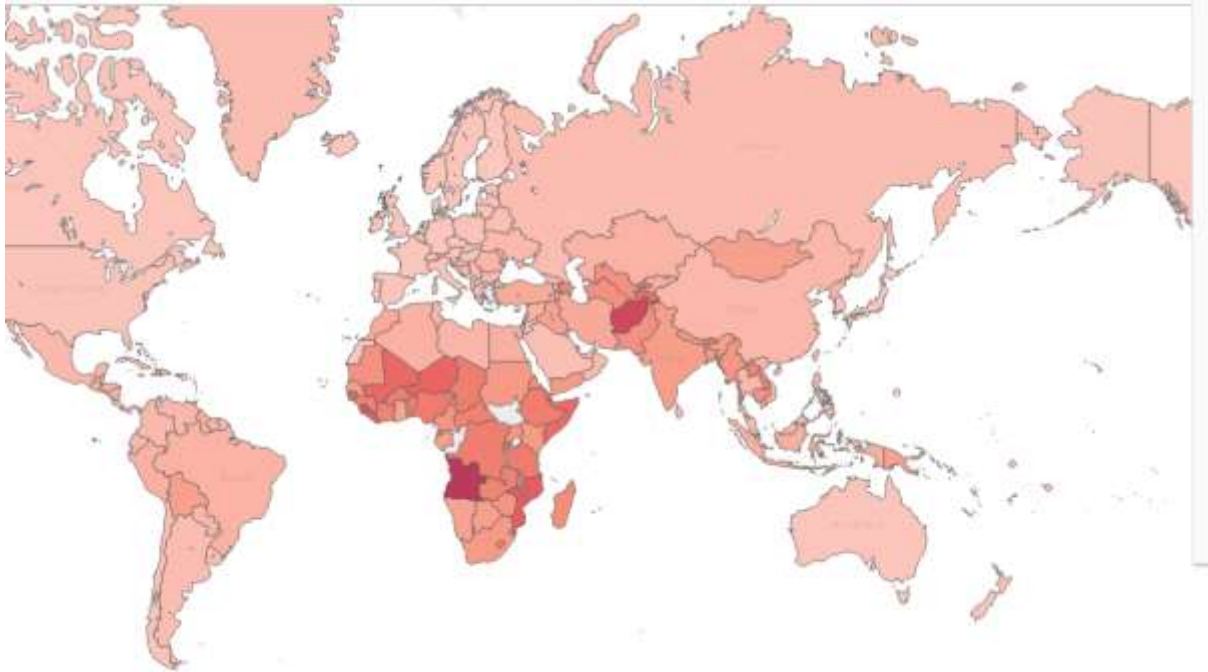
If GDP goes down, layoffs and unemployment and declining business revenues and consumer spending

Shrinking GDP also affects interest rates

## Countries with increase in GDP:

- United States
  - China
  - Japan.
  - Germany.
- United Kingdom.
  - India.
  - France.
  - Italy

## Infant mortality rate



Infant mortality is the death of young children under the age of 1

Premature birth is the biggest contributor to the IMR

Other leading causes of infant mortality are birth asphyxia, pneumonia and congenital malformations

One of the most common preventable causes of infant mortality is smoking during pregnancy

Many factors contribute to infant mortality, such as the mother's level of education, environmental conditions, and political and medical infrastructure

**Countries with darker Red have Higher IMR and countries with light Red has lesser IMR**

## References

**Google**

**Tableau**

**“Countries of the world” Dataset**

**Wikipedia**

---