Exploratory Data Analysis
On the World Population growth – 2020

Riya Dey
DS January 2021
STS-1

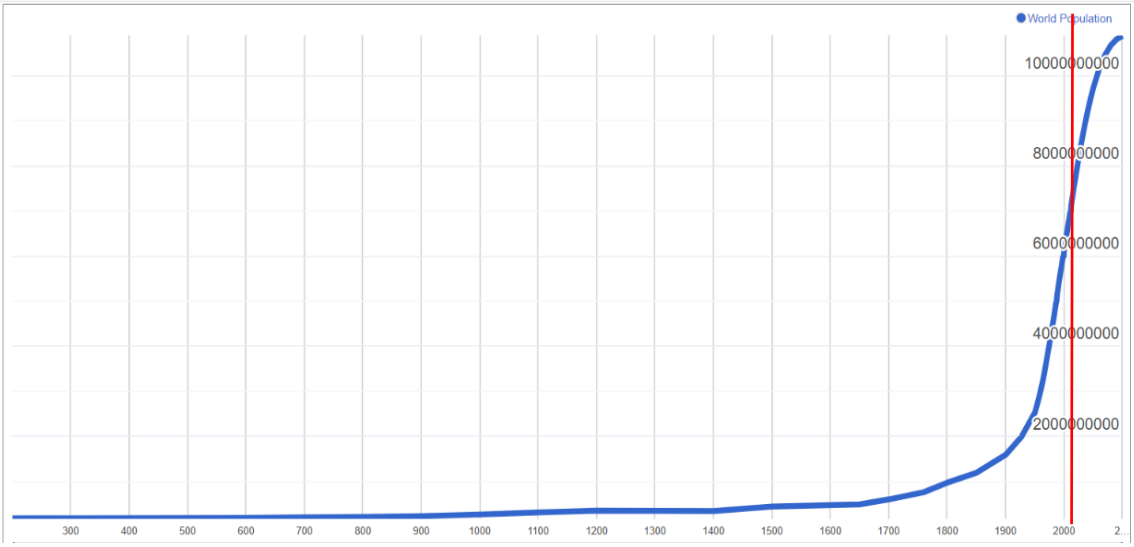
Contents Page No.

INTRODUCTION

The world population have reached 7,800,000,000 people as of March 2020. It took over 2 million years of human prehistory and history for the world's population to reach 1 billion, and only 200 years more to reach 7 billion.

The global population is still increasing, but there is significant uncertainty about its long-term trajectory due to changing rates of fertility and mortality. The UN Department of Economics and Social Affairs projects

between 9-10 billion people by 2050, and gives an 80% confidence interval of 10-12 billion by the end of the 21st century.



Population increases since last 2000 years

OBJECTIVE

Population growth is always a much-debated topic we come across. Even during COVID19 pandemic also lot of emphasis given on world population.

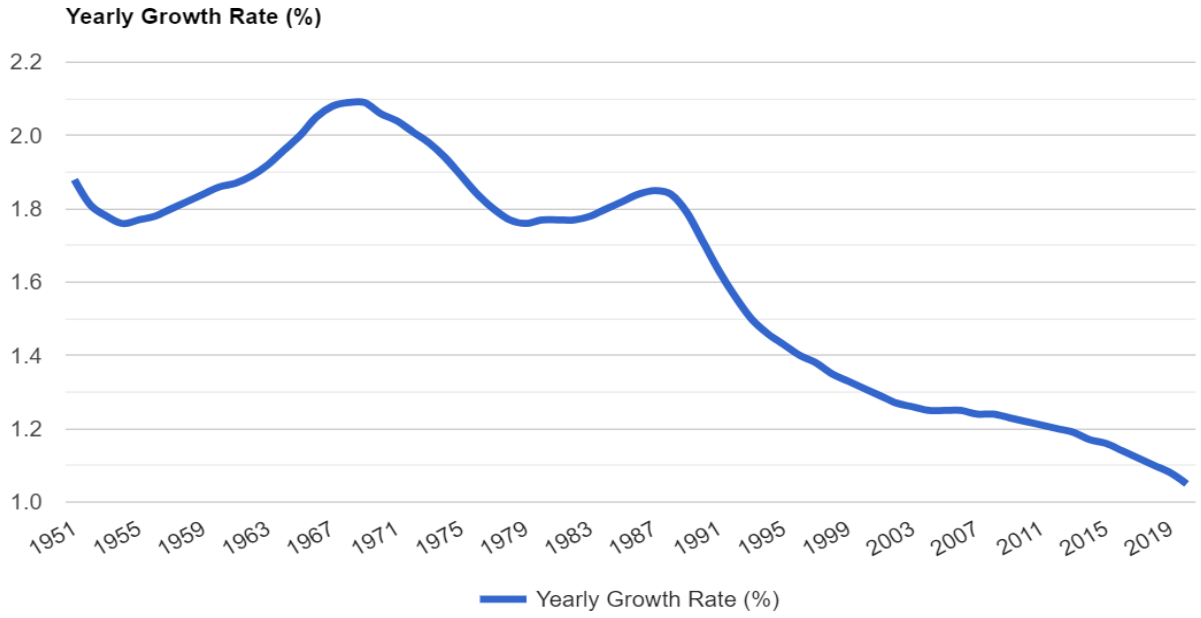
My objective here to explore country-wise world population, population growth trends and identify the countries sitting on population time bomb.

As we have limited resources in the earth. Population growth is a major concern worldwide. Population explosion in future can cause various catastrophe in certain region – ranging from famine, pandemic to civil war.

Parameters to determine Population Growth

Yearly Population Growth:

Total annual births were highest in the late 1980s at about 139 million, and as of 2011 were expected to remain essentially constant at a level of 135 million, while deaths numbered 56 million per year and were expected to increase to 80 million per year by 2040. The median age of the world's population was estimated to be 30.4 years in 2018.



Total Fertility Rate:

Fertility rate, average number of children born to women during their reproductive years. For the population in a given area to remain stable, an overall total fertility rate of 2.1 is needed, assuming no immigration or emigration occurs.

If, on average, women give birth to 2.1 children and these children survive to the age of 15, any given woman will have replaced herself and her partner upon death. A TFR of 2.1 is known as the replacement rate. Generally speaking, when the TFR is greater than 2.1, the population in a given area will increase, and when it is less than 2.1, the population in a given area will eventually decrease, though it may take some time because factors such as age structure, emigration, or immigration must be considered.

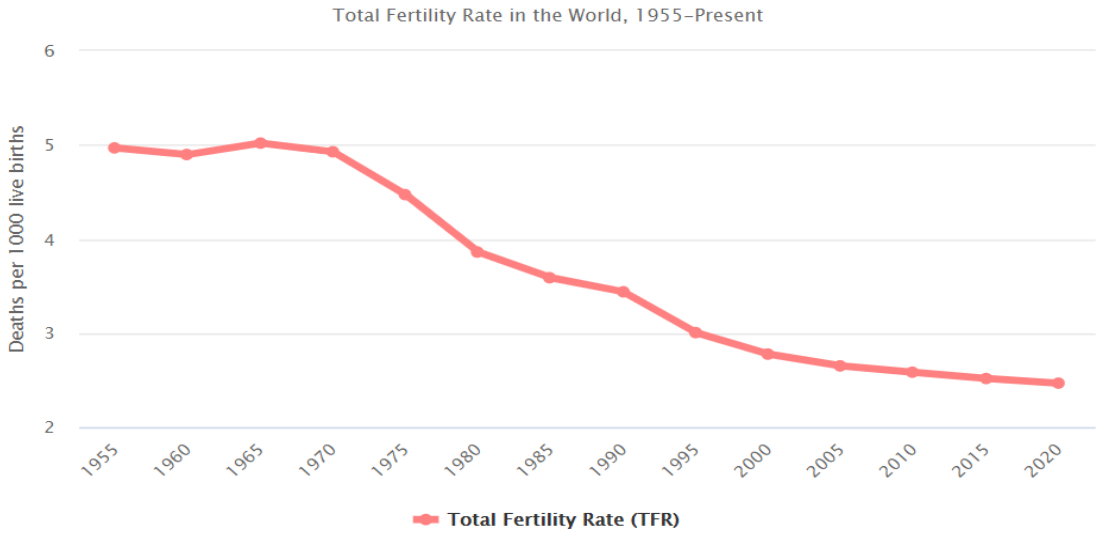
Tracking fertility rates allows for more efficient and beneficial planning and resource allocation within a particular region. If a country experiences unusually high sustained fertility rates, it may need to build additional schools or expand access to affordable child care.

The TFR is calculated as:

$$TFR = n \sum ASFR_a \text{(for n-year age groups)}$$

Where, n = Age group interval

ASFR = age-specific fertility rate



Data Description

The dataset contains list of countries by their population- 2020. There are 235 countries along with their population. And there are 11 columns each representing different component of Population. This is dataset of 2020.

Exploring the Dataset: Exploratory Data Analysis (EDA)

In statistics, exploratory data analysis is an approach of analysing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

Descriptive Statistics (Non-Graphical)

	Yearly Change (%)	Density (P/KmÂ²)	Fert. Rate	Migrants (net)
No. of Sample	50	50	50	50
Average or Mean	1.35%	150.64	2.74	26581.24
Median	1.24%	93.50	2.25	5606.50
Mode	0.0004	25	1.8	-10000
Standard Deviation	0.010126263	198.1567824	1.295222	211530.6131
Range	3.91%	1261.00	4.90	1487493.00
25% (1st Percentile)	0.0061	46.5	1.8	-51064.25
50% (2nd }Percentile or Median)	0.0124	93.5	2.25	5606.5
75% (3rd Percentile)	0.022475	190.5	3.675	66170
Inter Quartile Range (IQR)	0.016375	144	1.875	117234.25
Lower Range (Q1-1.5IQR)	-0.0184625	-169.5	-1.0125	-226915.625
Upper Range (Q3+1.5IQR)	0.0470375	406.5	6.4875	242021.375

From this analysis we can understand that in more than half of the most populous countries fertility rate is more that replacement rate (2.1). Those need population control measures.

The population growth and TFR data range is quite high. Even with the SD we can understand that population growth rate and fertility rate vary significantly country to country.

Though the Descriptive Statistics gave few insights, they are not enough to get the knowledge of countries need population control. This is where graphs and plots come into the picture.

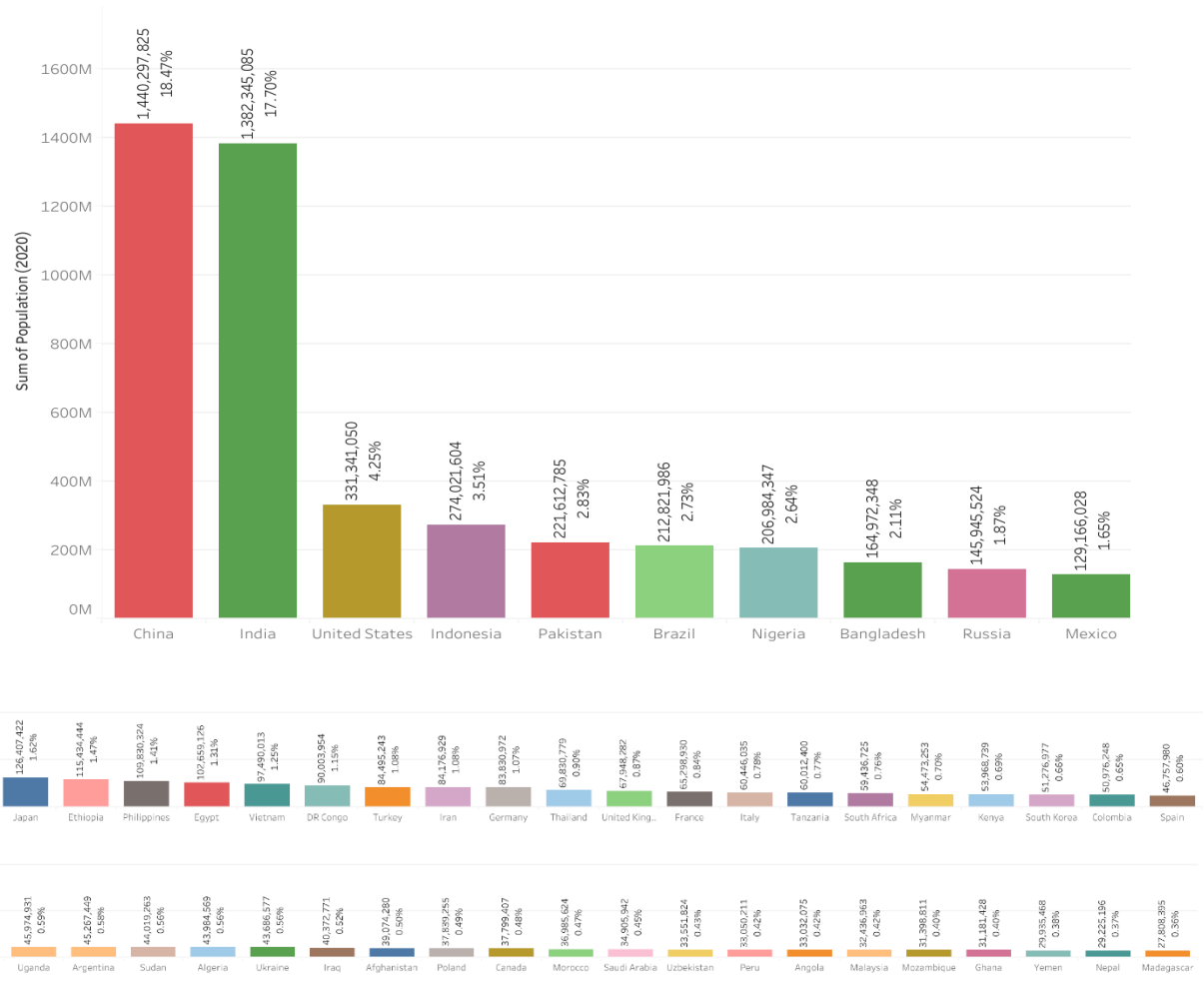
Visualising the Data (Graphical)

In this section we will be exploring the data through some visualizations such as Stacked Bar Chart, Horizontal Bar chart, Maps and scatter plots. Every chart has their own set of advantages and disadvantages; therefore, we use multiple chats/graphs to get the full idea about the dataset.

First, we take a sample data set of 50 countries, and based on those countries we generate some graphs, those graphs will describe my entire work.

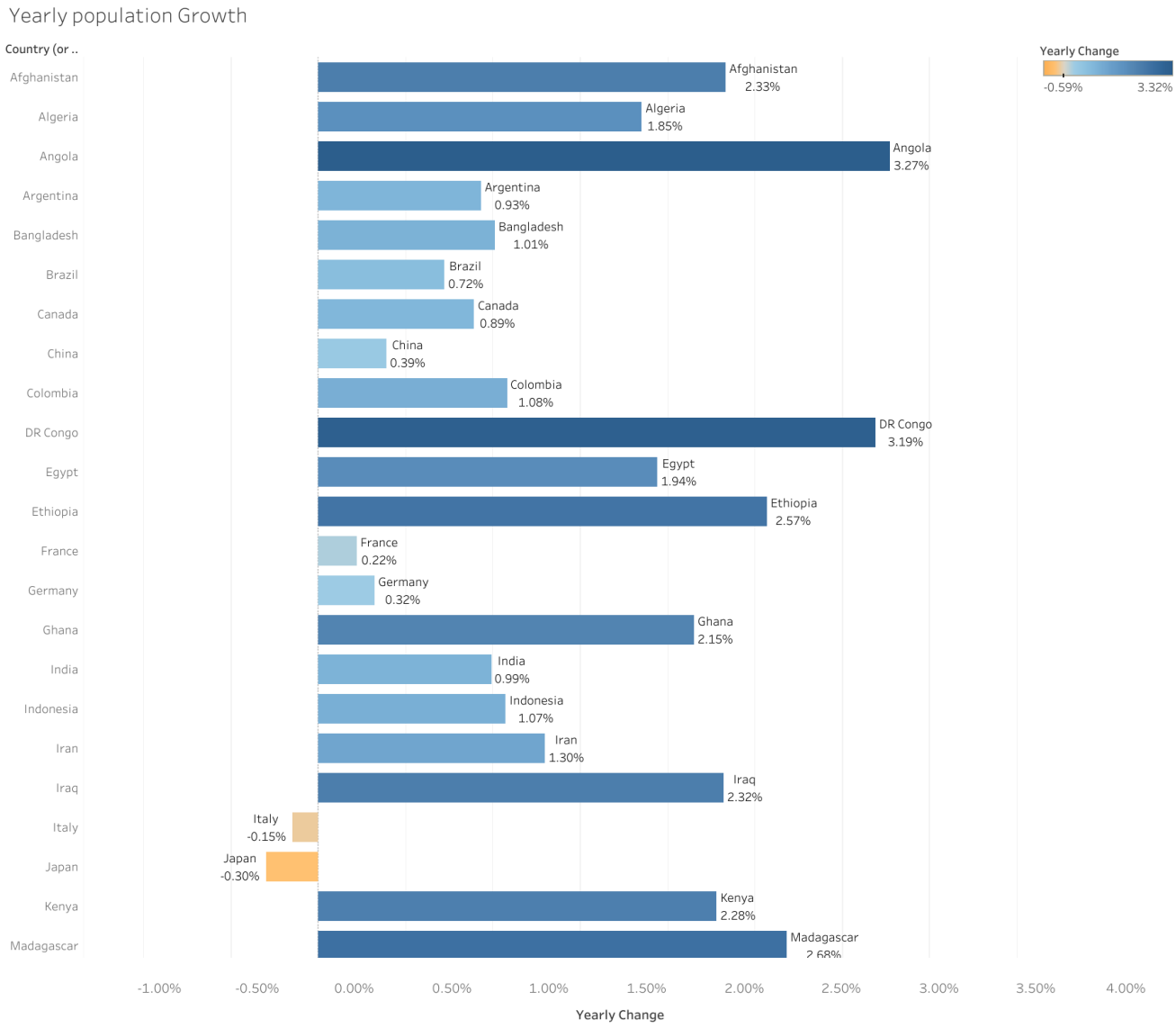
Now we will be visualizing the graphs for each component one by one.

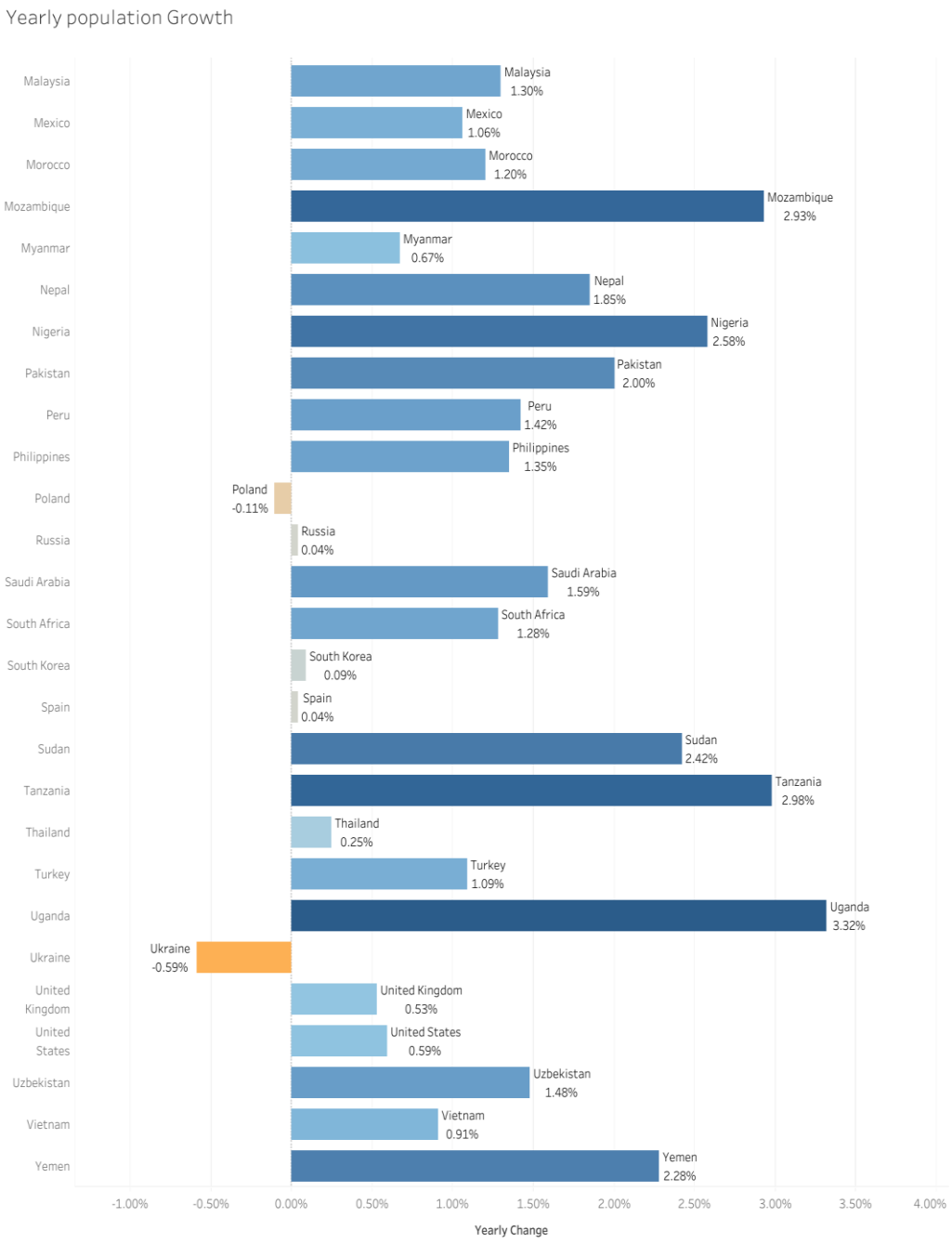
Population Data – top 50 populous countries



- The graph gives us proper idea about the world population share by top 50 countries.

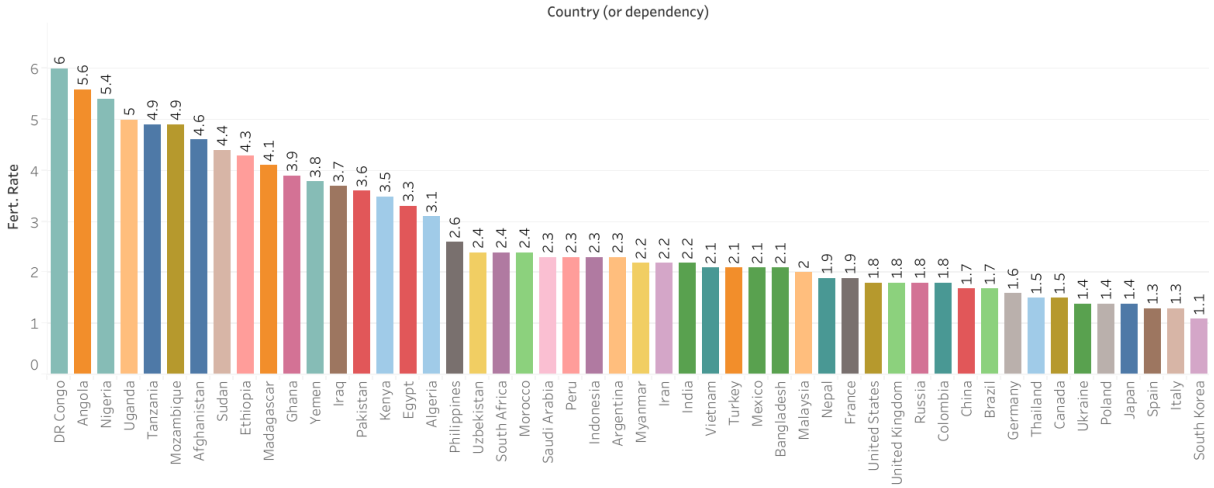
Population Growth – Top 50 Populous Countries



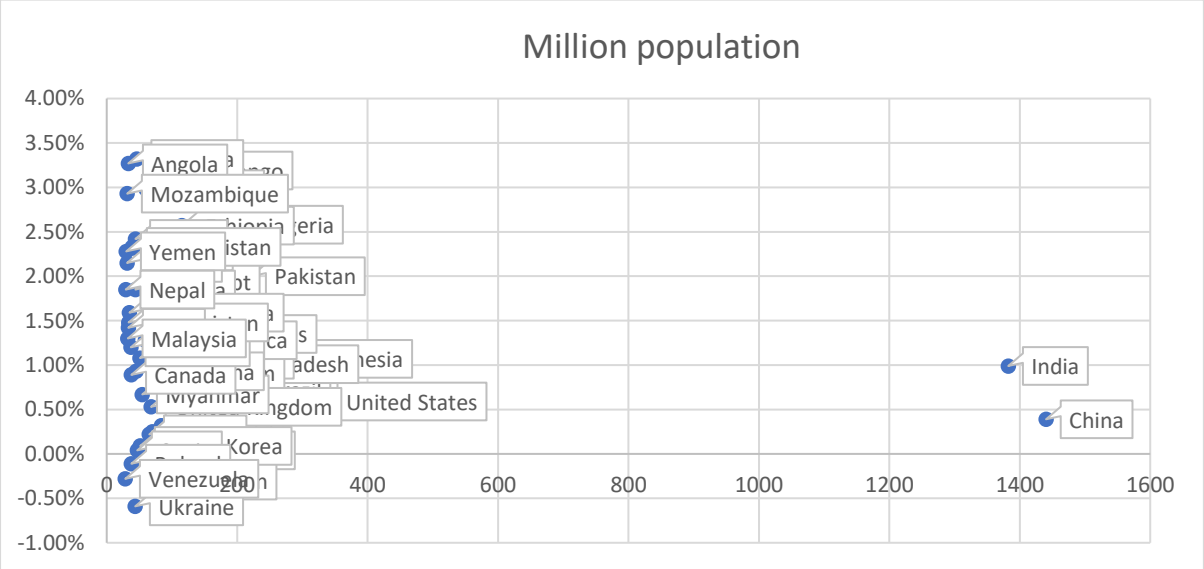


- From this graph we can understand percentage growth of population in top 50 countries.

Total Fert. Rate data – Top 50 Populous Countries

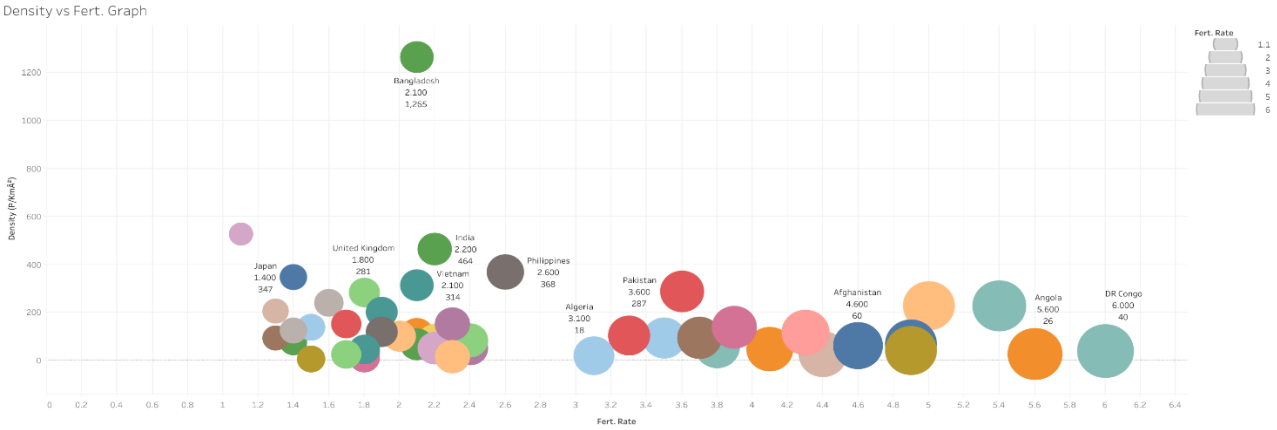


Actual Population increase and Growth Percentage



- The above graph indentify countries as per the to actual population (X) and growth percentage (Y). The coutries with actual high population and high percentage increase of population have risk of rapid population growth. Similarly countries with low population low percentage population growth – can see rapid population decline.

Population Density and TFR Graph

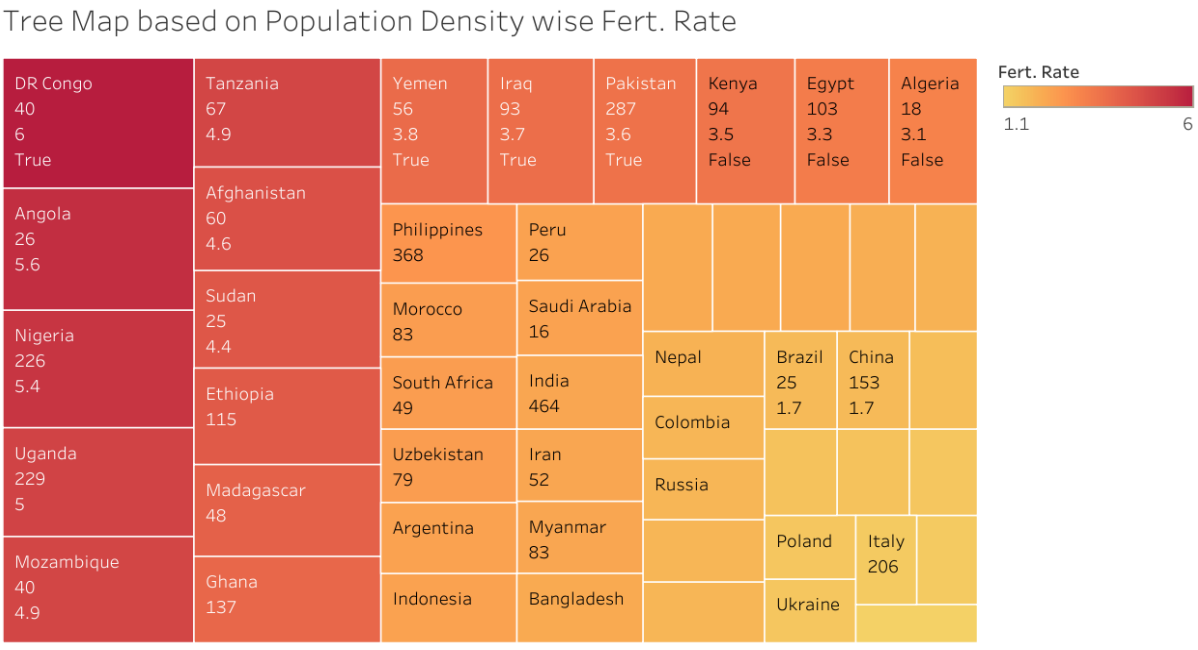


The above graph depicts population density (Y) and TFR (X)

High population density and high fertility rate can face resource scarcity in the future.

Conclusions

From the above data and graphical analysis – we can find out that many west African and middle eastern countries have high risk of population growth and resource scarcity. We have identified the countries need urgent measures of population control.

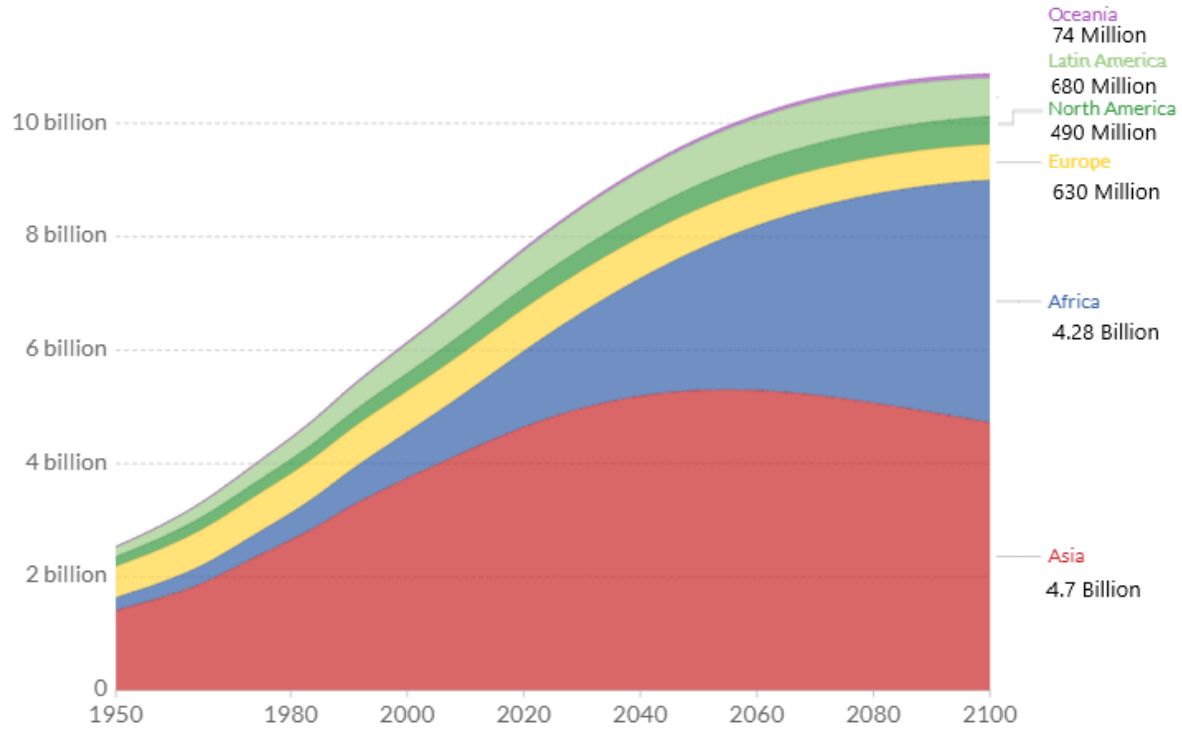


Criteria 1 = Density >200 AND Fert Rate >2.1

Criteria 2 = Fert Rate> 3.5

List of countries
India
Pakistan
Nigeria
Ethiopia
Philippines
DR Congo
Tanzania
Uganda
Sudan
Iraq
Afghanistan
Angola
Mozambique
Ghana
Yemen
Madagascar

As we can see the most of the country in this list are from Africa. As per the current population growth and fertility rate by end of this century the population of Africa and Asia will be almost equal.



Bibliography:

The data has been downloaded from kaggle.com

Link- https://www.kaggle.com/tanuprabhu/population-by-country-2020

https://www.worldometers.info/demographics/world-demographics/

https://www-doh.state.nj.us/doh-shad/view/sharedstatic/TotalFertilityRate.pdf

https://www.measureevaluation.org/prh/rh_indicators/family-planning/fertility/total-fertility-rate