**DATA VISUALIZATION 2 ASSIGNMENT**

Project Name:

**Analysis of G20 + 1(Singapore) Countries key Economic and Social Indicators**

Contents

[1.0 Project Introduction: 2](#_Toc57990966)

[1.1 Scope: 2](#_Toc57990967)

[2.0 Project Objective: 2](#_Toc57990968)

[3.0 Related Work 2](#_Toc57990969)

[4.0 Data 2](#_Toc57990970)

[4.1 Data Cleanup 3](#_Toc57990971)

[4.2 Data Type 3](#_Toc57990972)

[5.0 Exploratory Data Analysis 4](#_Toc57990973)

[5.1 Data Observations 4](#_Toc57990974)

[5.1.1 GDP Data 4](#_Toc57990975)

[5.1.2 Number of Schooling Years 4](#_Toc57990976)

[5.1.3 Number of kids under age 5 years 5](#_Toc57990977)

[5.1.4 Mortality rate 5](#_Toc57990978)

[5.1.5 Per Capita Income 6](#_Toc57990979)

[5.1.6 Percentage of Employment Male vs Female in different sectors 6](#_Toc57990980)

[5.2 Statistical Findings 7](#_Toc57990981)

[5.2.1 'Immunization, DPT (% of children ages 12-23 months)' – Quartile Analysis 7](#_Toc57990982)

[5.2.2 Immunization, measles (% of children ages 12-23 months) – quartile analysis 8](#_Toc57990983)

[5.2.3 G20 Countries Mean Inflation (2019), Standard Deviation 8](#_Toc57990984)

[6.0 Interactive Dashboard 8](#_Toc57990985)

[7.0 Appendix: 9](#_Toc57990986)

[7.1 Deliverables: 9](#_Toc57990987)

[7.1.1 Python files: 9](#_Toc57990988)

[7.1.1 CSV files 9](#_Toc57990989)

# 1.0 Project Introduction:

This project is a part of data visualization assignment. Analyst is expected to apply all previous learnings of data analysis, statistical analysis and visualization with the help of learned tools i.e. python, tableau, excel to dig out insights and present them in the story.

## Scope:

Scope of this work are as follows:

* Dataset ingestion, processing and cleaning
* Data Analysis.
* Statistical Analysis
* Visualization & Dashboard.

# 2.0 Project Objective:

Objective of this project include following:

* To analyze G20 + 1 (Singapore) countries key economic and social indicators.
* To derive insights from the dataset.

# 3.0 Related Work

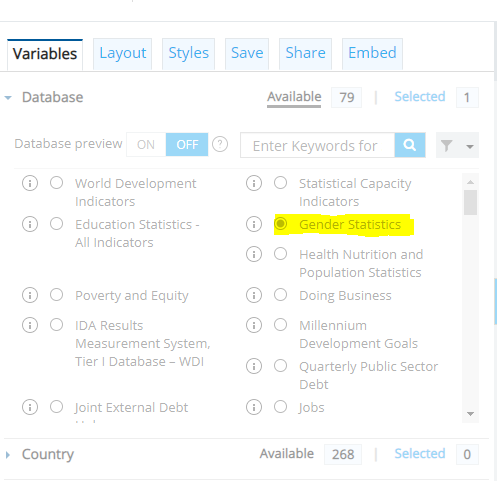
Numerous exercises/activities performed during the classes, along with data story telling blogs have enabled me to execute this project.

One such example is the following blog from world bank site- [The 2020 Atlas of Sustainable Development Goals: Stories and insights through innovative visuals (worldbank.org)](https://blogs.worldbank.org/opendata/2020-atlas-sustainable-development-goals-stories-and-insights-through-innovative-visuals)

# 4.0 Data

Data is pulled from world bank databank. Following reference of data is from “Gender Statistics” database for G20+Singapore countries. Data source snippet below:

Link - [Gender Statistics | DataBank (worldbank.org)](https://databank.worldbank.org/source/gender-statistics)



Dataset downloaded from world bank database is in csv format.

File name - “gender\_statistics.csv” (main file)

Other important files for this project are as follows sliced from the main file.

This is created, processed and cleaned for visualization purpose.

1. G20\_Countries\_economic\_data.csv – hold economic data indicators.
2. G20\_population\_age\_under5.csv – hold demographic information
3. G20Countries\_employment\_stats.csv- hold employment data

## 4.1 Data Cleanup

Following data processing and cleaning-up activities have been performed:

1. Ingestion of data, pre-processing and cleaning activities.
2. Removing duplicates, redundant rows and columns.
3. Removing rows with NAN values.
4. Data value columns with “..” or no value have been cleaned up.
5. Dropping of dimensions with no values
6. Adding derived columns based on usage objective

Overall, circa ~9k bad records have been removed out of 12k total.

## 4.2 Data Type

Three (3) economic indicators and three (3) social indicators are taken for analysis purpose.

Economic Indicator:

* GDP data
* Per Capita Income
* Employment data – percentage of workforce employed in different sectors of an economy, qualified with gender %age.

Social Indicator:

* Demography of Infants and kids under age 5 - to understand Male vs Female population gap.
* Mortality rate of different countries are analyzed
* Male and Female employer %age of the total employee.

# Exploratory Data Analysis

Basic field attributes have been checked to understand level of details.

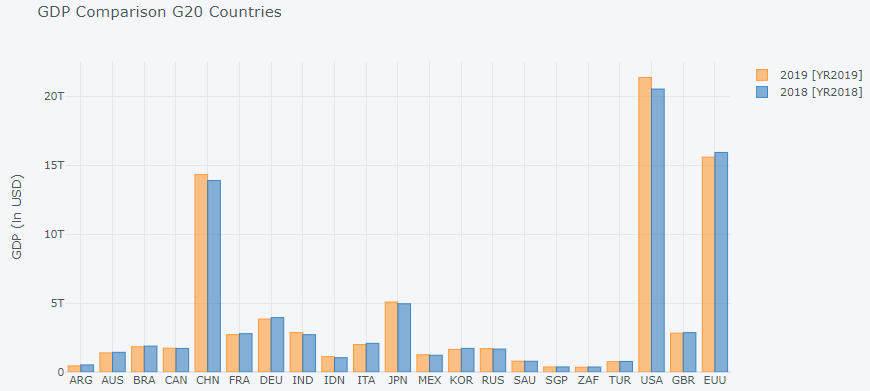
Bar chart, stacked bar chart, pie chart, spatial chart and other charts have been used to provide contextual reference.

G20 countries contribute 80 percent to the world economy. Collectively, the G20 economies account for around 90% of the [gross world product](https://en.wikipedia.org/wiki/Gross_world_product) (GWP), 80% of [world trade](https://en.wikipedia.org/wiki/International_trade), two-thirds of the [world population](https://en.wikipedia.org/wiki/World_population), and approximately half of the [world land area](https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_area).

## Data Observations

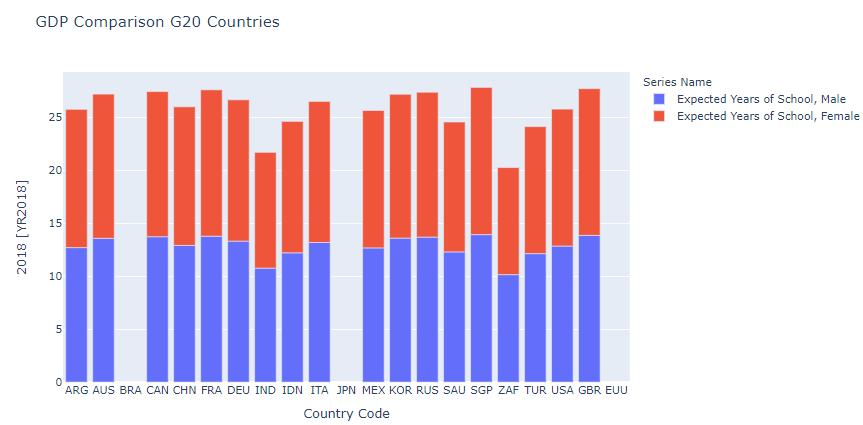
5.1.1 GDP Data – It has been observed that there is no uniformity in GDP of G20 countries. US is the largest economy with ~20 T GDP followed by China ~15 T and Japan ~5T.

South Africa being smallest economy among G20.



### 5.1.2 Number of Schooling Years

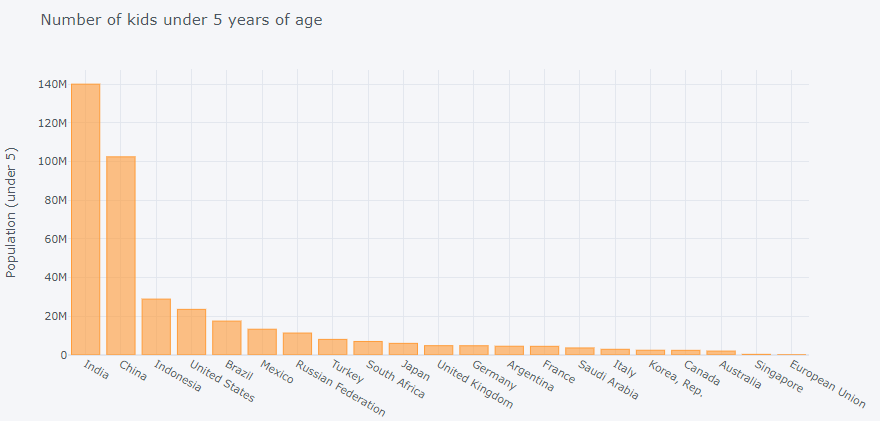
India and South Africa have been noticeably lowest schooling years.



### 5.1.3 Number of kids under age 5 years

India tops the chart with 140 million kids under age 5 followed by China 102 million.

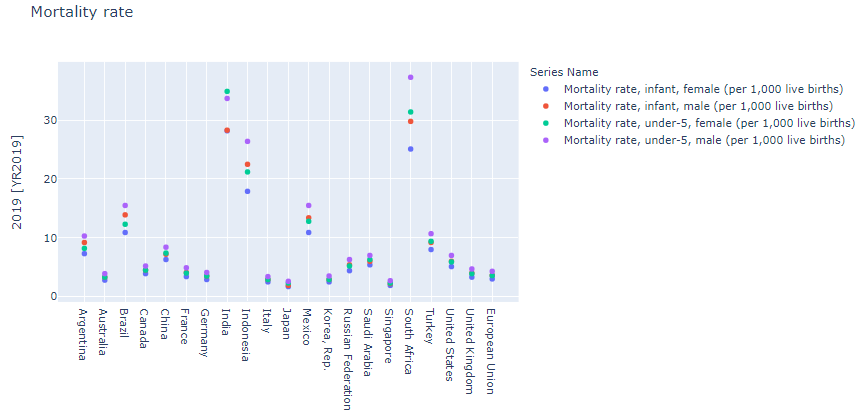
Singapore has lowest number of under age 5 kids.



### 5.1.4 Mortality rate

South Africa and India have highest mortality rate under age 5. {(male, female): (34.9, 33.7) (37.3, 31.4)} respectively.

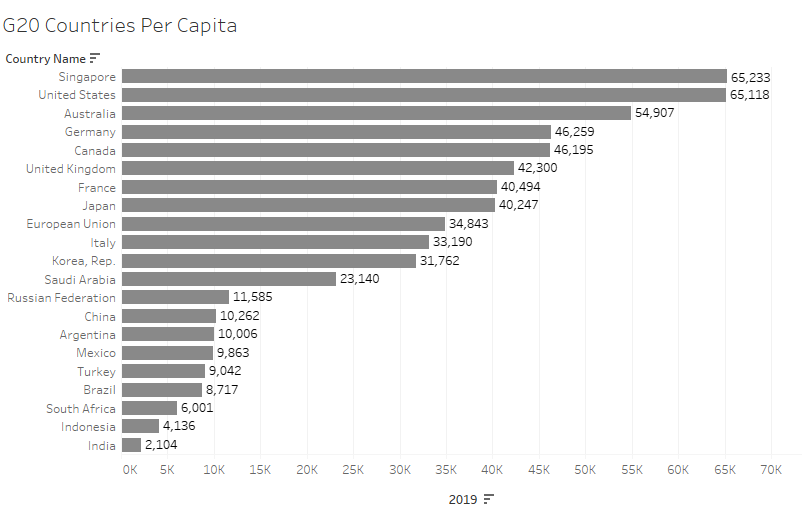
Singapore, Japan among the lowest mortality rate countries.



### 5.1.5 Per Capita Income

Singapore has highest per capita income followed by United states and Australia among G20 countries i.e. ~65k/yr, ~65k/yr, ~55k/yr respectively.

India has lowest per capita income followed by Indonesia and South Africa i.e. 2k/yr,4k/yr,6k/yr respectively.

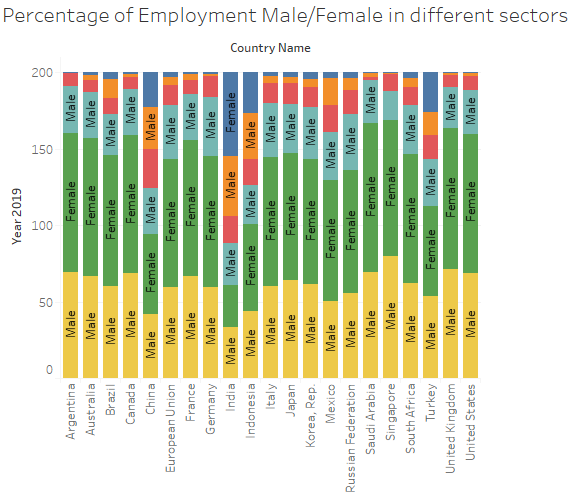


### 5.1.6 Percentage of Employment Male vs Female in different sectors

India has highest percentage of female workers employed in agriculture sector i.e.54.5%.

Top 10 countries based on per capita income have very tiny workforce employed in agriculture sector.

Singapore has 89% of female worker employed in service sector. United state has highest 92% female worker employed in service industry.



## Statistical Findings

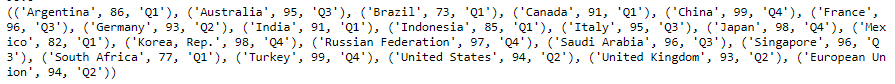
## 'Immunization, DPT (% of children ages 12-23 months)' – Quartile Analysis

Quartile analysis of DTP immunization program suggest that few countries are doing better than others and vice-versa

Result suggest that:

* China, Russian Federation, Turkey are in quartile 4 with maximum percentage of children immunized for DTP.
* However, Argentina, Brazil, Canada, India, Indonesia, Maxico, South Africa, are doing bad with Q1 score.
* Singapore is in Q3, DTP immunization program.

**Output**:



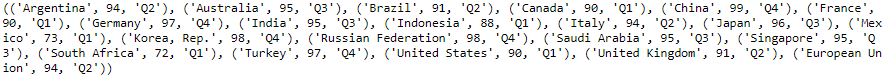
## Immunization, measles (% of children ages 12-23 months) – quartile analysis

Quartile analysis of measles immunization program suggest that few countries are doing better than others and vice-versa

Result suggest that:

* China, Germany, Korea, Russian Federation, Turkey are in quartile 4 with maximum percentage of children immunized for measles.
* However, Canada, France Indonesia, Maxico, South Africa, United Stats are doing bad with Q1 score.
* Singapore is in Q3, Measles immunization program.

**Output**:

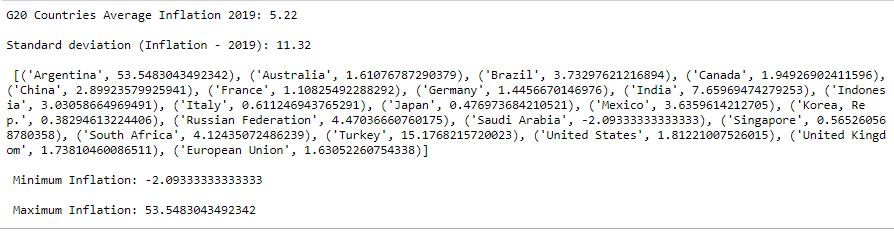


## G20 Countries Mean Inflation (2019), Standard Deviation

G20 countries mean inflation for 2019 was 5.22 with standard deviation of 11.32.

Saudi Arabia stood with negative inflation of (-)2%, Argentina with highest inflation among G20 countries with (+) 52.54% tops the rank.

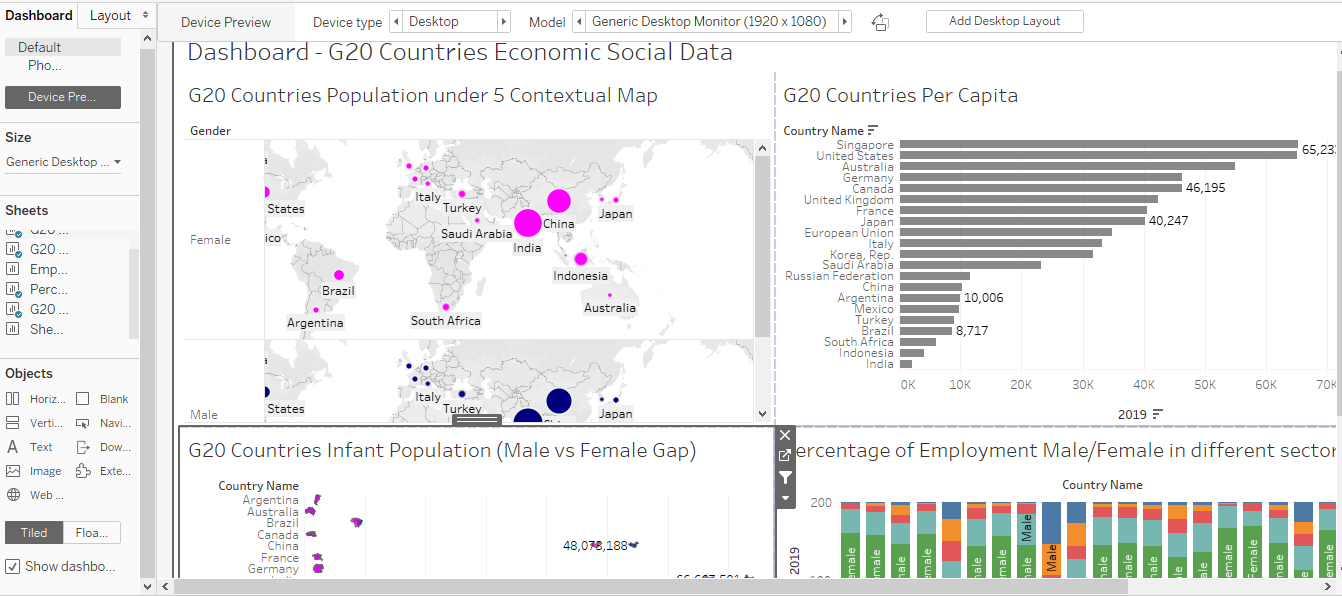
Results below:

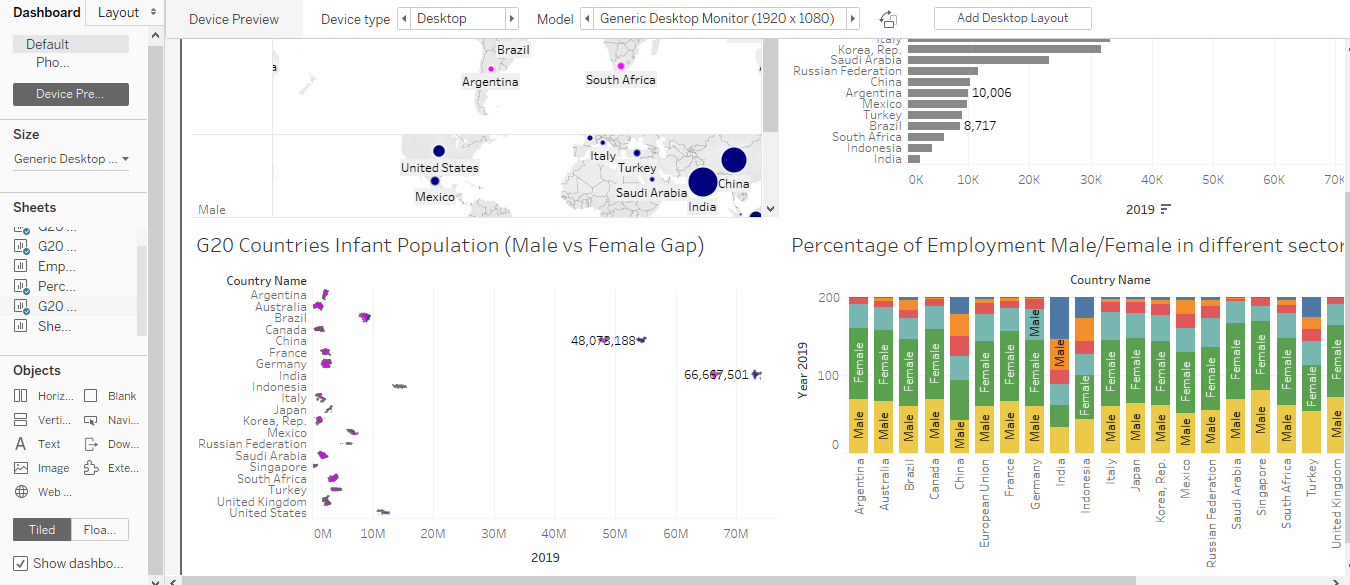


# Interactive Dashboard

Interactive dashboard has been built in Tableau. Data after cleaning have been imported into Tableau. Worksheets are created to have different view. Dashboard is created based on worksheets and combining them to have one consolidated view.

Snapshots of Interactive dashboard.





# Appendix:

## Deliverables:

### 7.1.1 Python files:

* + - DV2\_Vivek\_Sharma.ipynb,
    - Data\_analysis.ipynb

### 7.1.1 CSV files

* + - gender\_statistics.csv
    - G20\_Countries\_economic\_data.csv
    - G20\_population\_age\_under5.csv
    - G20Countries\_employment\_stats.csv