

<b>Name</b>	Rucha Kulkarni
<b>Class</b>	BE Computer Engineering (Batch F)
<b>UID</b>	2021300067
<b>Exp No.</b>	2

**Aim:** Create advanced charts using Tableau / Power BI / R / Python / D3.js to be performed on the dataset - Socio economic data

- Advanced - Word chart, Box and whisker plot, Violin plot, Regression plot (linear and nonlinear), 3D chart, Jitter
- Write observations from each chart

## Description:

Dataset used is Life Expectancy Dataset available at <https://www.kaggle.com/datasets/lashagoch/life-expectancy-who-updated>

Following are the columns in the dataset:

Country: Name of the country.

Region: Geographical region to which the country belongs (e.g., Middle East, European Union, Asia).

Year: The year for which the data is recorded.

Infant\_deaths: Number of deaths of infants per 1,000 live births.

Under\_five\_deaths: Number of deaths of children under the age of five per 1,000 live births.

Adult\_mortality: Probability of dying between the ages of 15 and 60 per 1,000 population.

Alcohol\_consumption: Alcohol consumption per capita (in liters).

Hepatitis\_B: Percentage of individuals vaccinated against Hepatitis B.

Measles: Number of reported measles cases per 1,000,000 population.

Polio: Percentage of individuals vaccinated against Polio.

BMI: Average Body Mass Index (BMI) of the population.

Diphtheria: Percentage of individuals vaccinated against Diphtheria.

Incidents\_HIV: Number of new HIV infections per 1,000 uninfected population.

GDP\_per\_capita: Gross Domestic Product per capita in current USD.

Population\_mln: Population of the country in millions.

Thinness\_ten\_nineteen\_years: Percentage of thinness among individuals aged 10-19 years.

Thinness\_five\_nine\_years: Percentage of thinness among individuals aged 5-9 years.

Schooling: Average number of years of schooling.

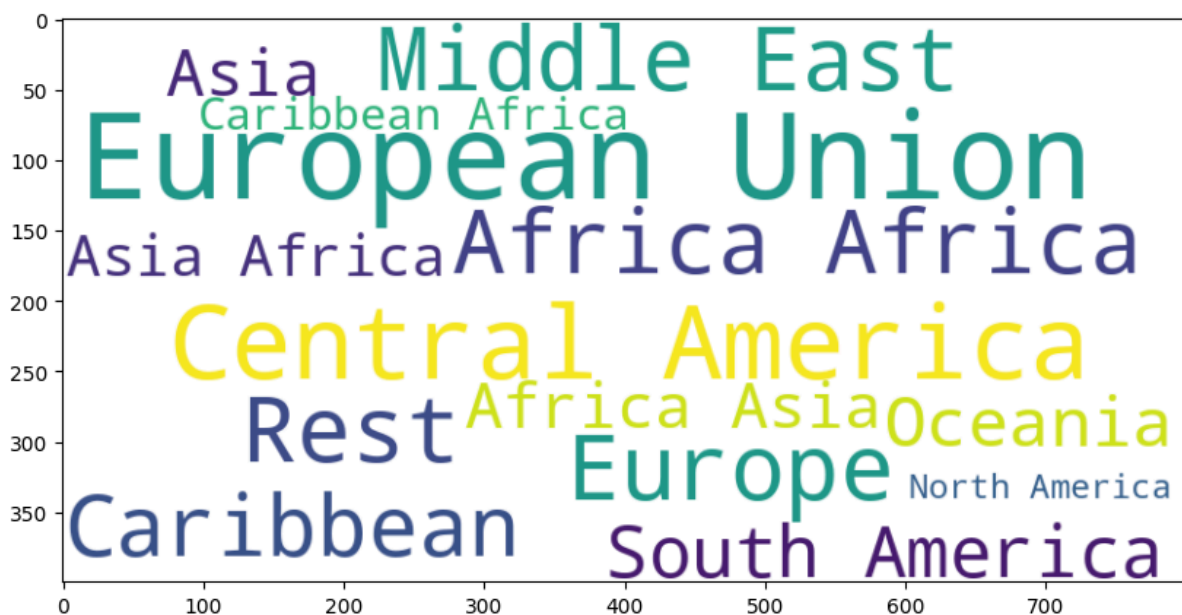
Economy\_status\_Developed: Indicator variable (1 for developed economies, 0 otherwise).

Economy\_status\_Developing: Indicator variable (1 for developing economies, 0 otherwise).

Life\_expectancy: Average life expectancy at birth in years.

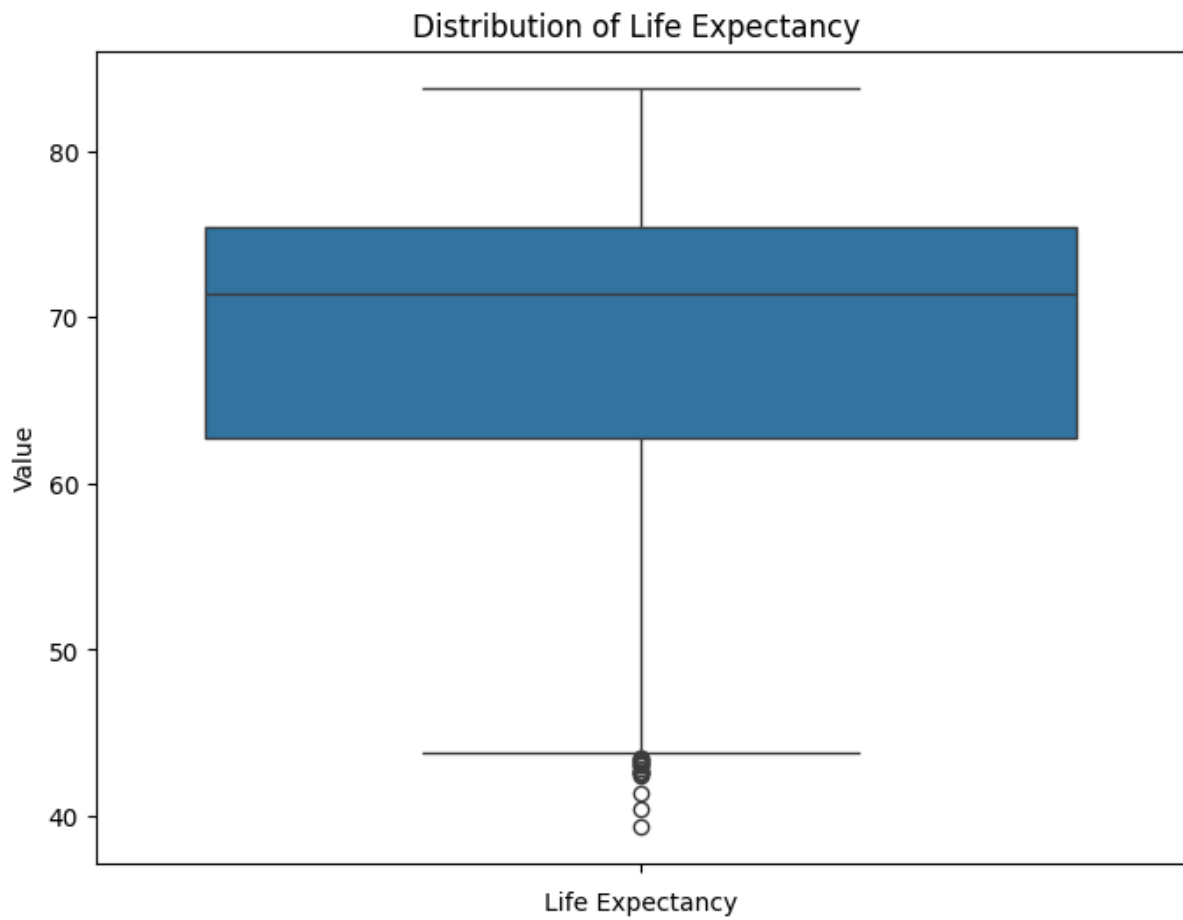
## Graphs and Observation:

### Word Cloud:



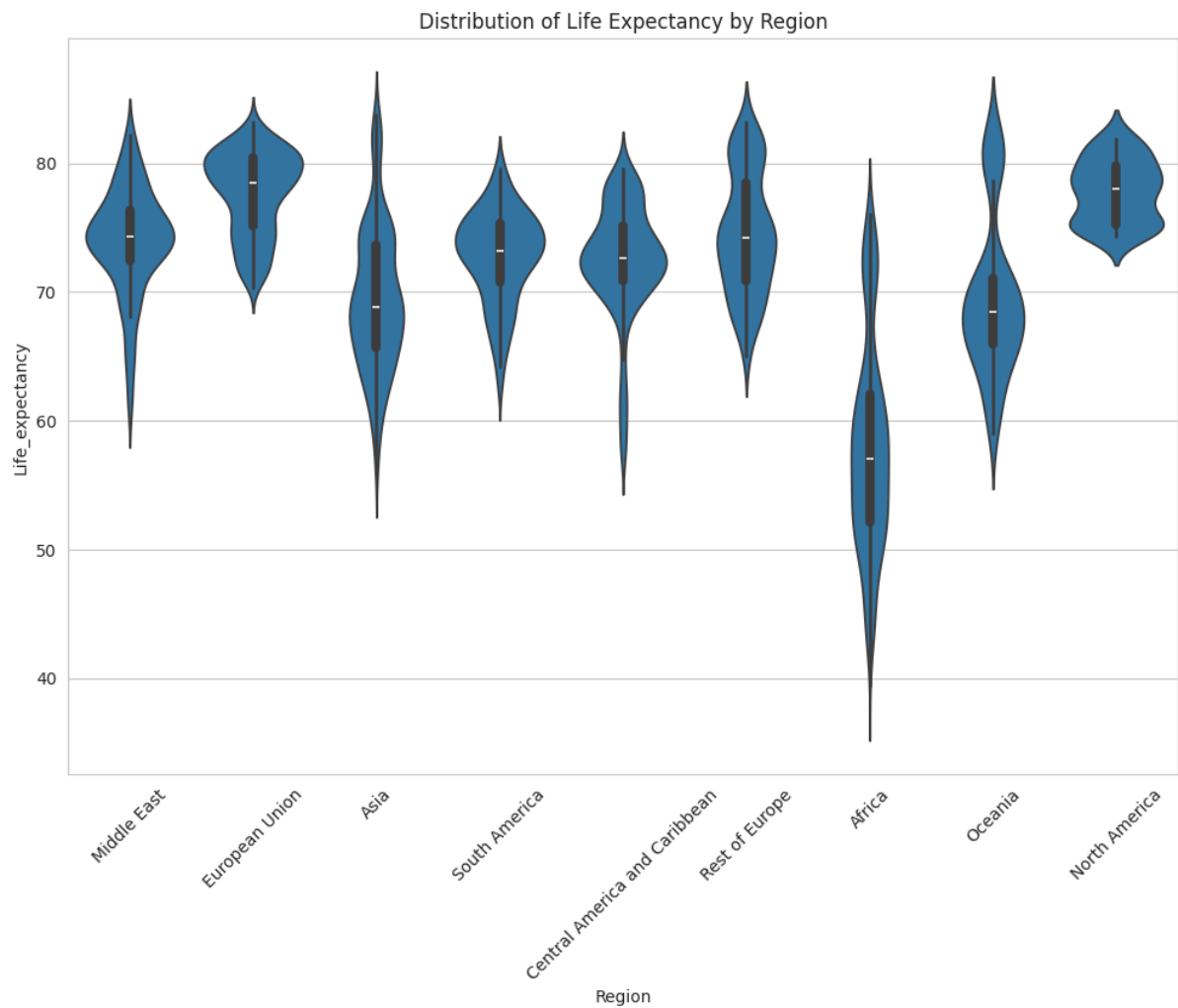
In this word cloud, the size of the 'European Union' word is largest indicating that most of the details about life expectancy belong to the region of the European Union.

### Box Plot:



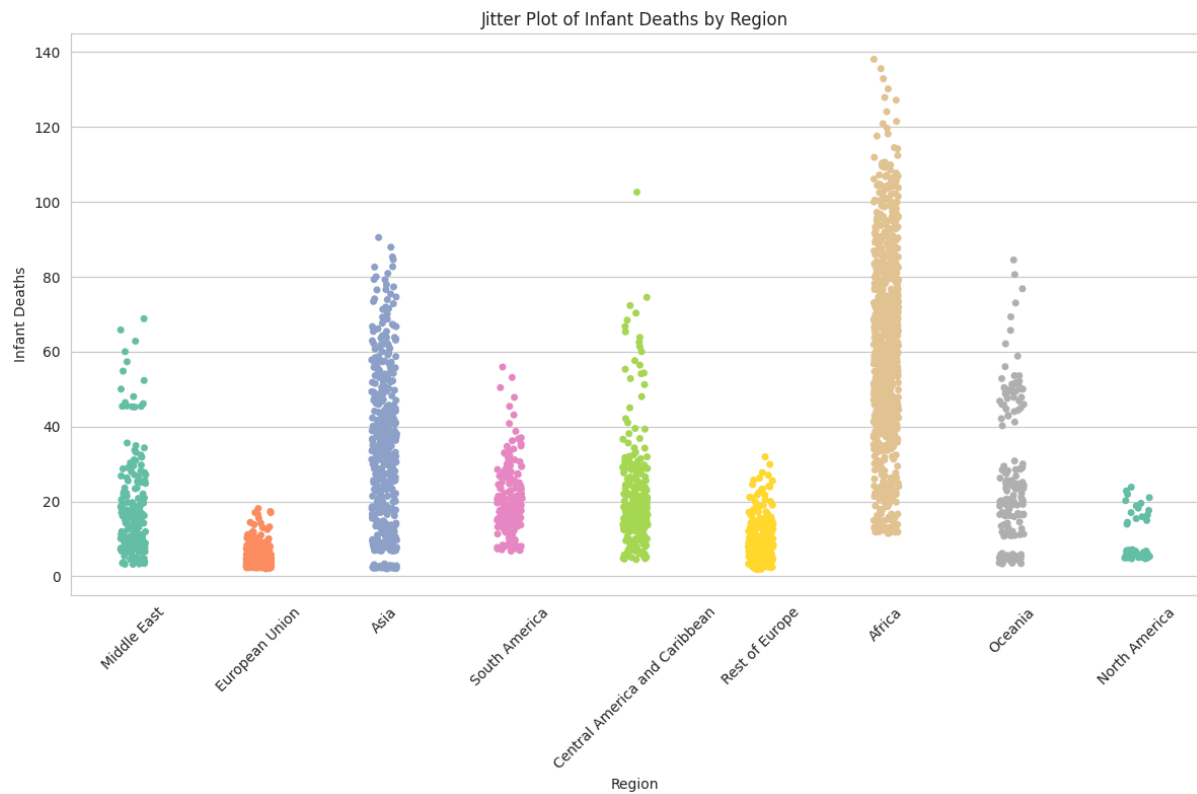
From the above box plot of life expectancy values we observe that most of the values are closer to 70 while values lesser than 40 are outliers.

## Violin Plot:



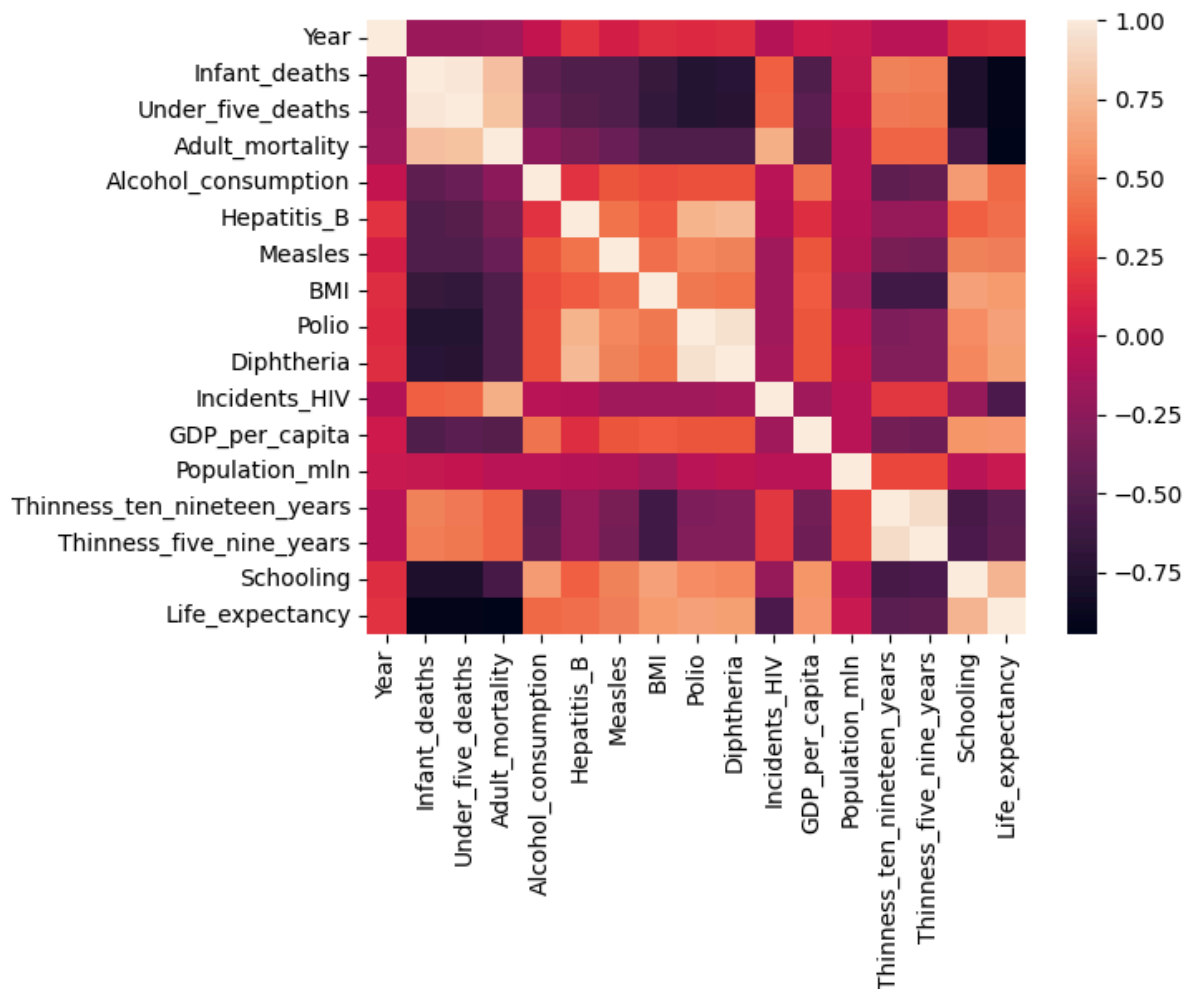
The above graph shows distribution of Life Expectancy around the nine regions. We observe that in the Africa region , the life expectancy values are spread evenly between 40 and 80. In the North America Region, the values are not spread much and lie mostly between 70 and 90. So there is less variance in the life expectancy values in North America compared to other regions.

## Jitter Plot:



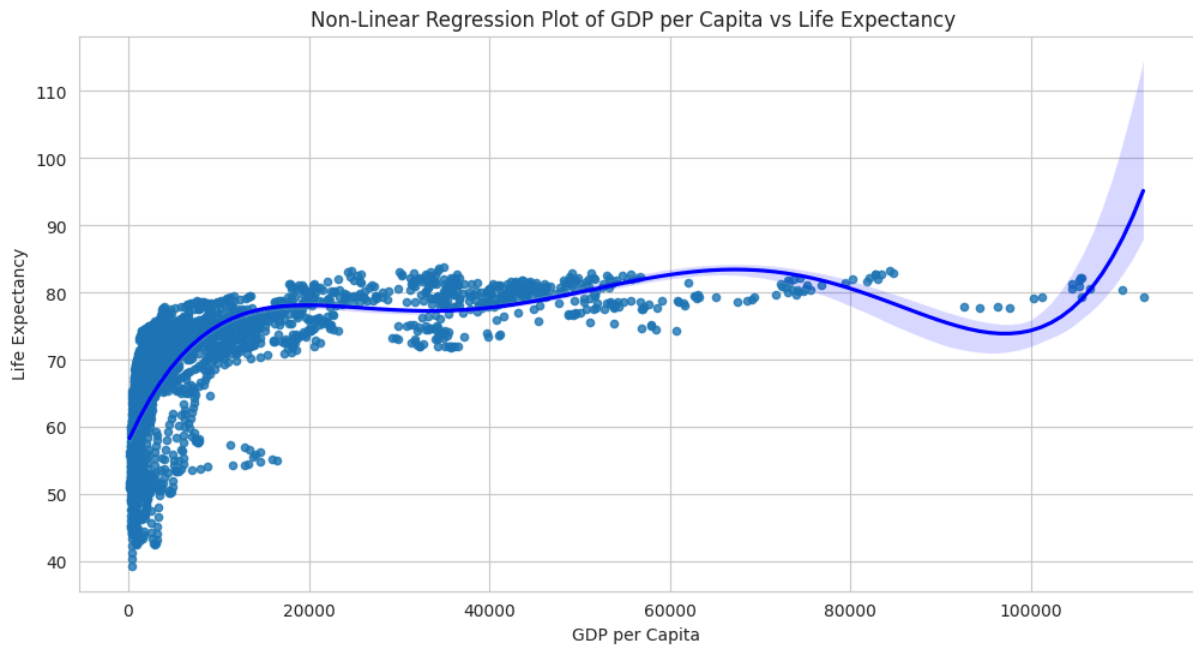
The above jitter plot shows that Africa has the highest number of infant deaths and North America has the least number of infant deaths.

## Heat Map:



This shows the correlation between all the variables in the dataset. We observe that Life expectancy is highly correlated with infant deaths, Under five deaths and adult mortality.

## Regression Plot:



From the above graph we observe that a non linear relationship exists between GDP per capita and Life Expectancy values.

## Conclusion:

From this experiment, we learnt about the charts which can be used to perform advanced data visualization. I was able to make and visualize the various advanced charts on the socio economic dataset.