Name	**Harditya Shah**
UID No.	**2021300114**
Batch	**L**

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

Dataset used:
Global Socio-Economic & Demographic Insights - https://www.kaggle.com/datasets/samybaladram/databank-world-development-indicators

Interpolated Dataset Columns

- `Year`: The time frame of the data recorded.
- `Country`: The country to which the data pertains.
- `Region`, `Sub-region`: Geographical classifications as per the United Nations geoscheme.
- `Agriculture, forestry, and fishing, value added (% of GDP)`: Insights into the primary sector's economic value.
- `Exports of goods and services (% of GDP)`: A measure of a country's integration into the global economy.
- `Population density (people per sq. km of land area)`: Spatial demographics.
- `Urban population growth (annual %)`: The rate at which urban areas are expanding.
- ... and many more covering fertility, mortality, and technology adoption.

Final Imputed Dataset Columns

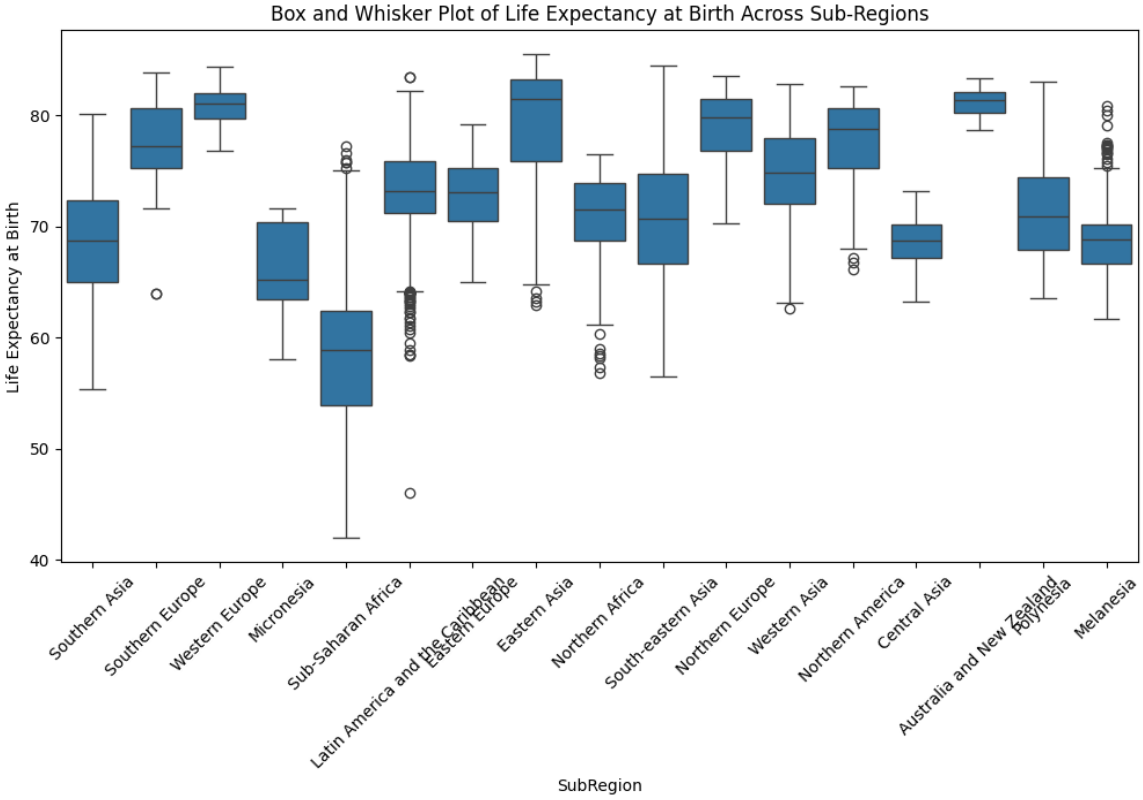
- `Net migration`: The net total of migrants during the year.
- `Surface area (sq. km)`: The total land area of a country.
- `Urban population growth (annual %)`: The annual increase in urban population.
- `Population, total`: The total number of people residing in the country.
- ... and other fundamental indicators.

Charts:

1. Word Cloud - By Country

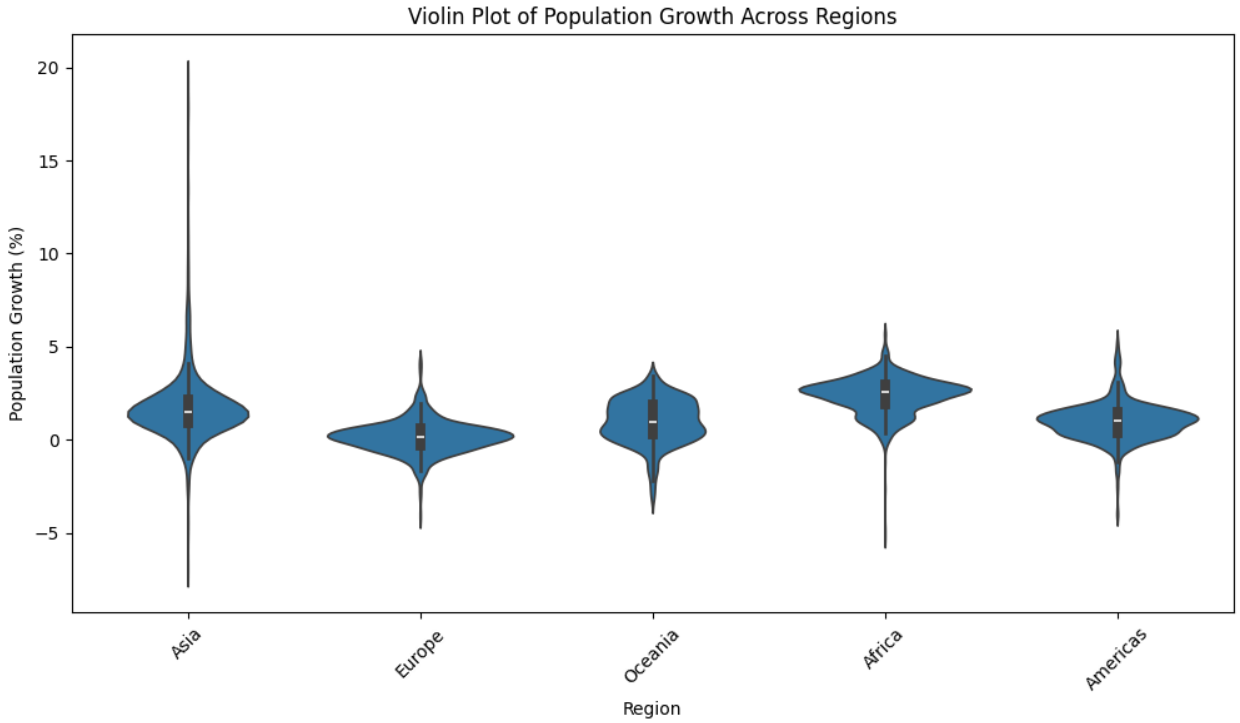
Analysis - The word cloud suggests that countries like China, Belgium, Afghanistan, Mauritania, and Malta have a significant amount of data. This indicates that these countries could have a strong influence on the socio-economic trends observed in the dataset.

2. Box and Whisker Plot



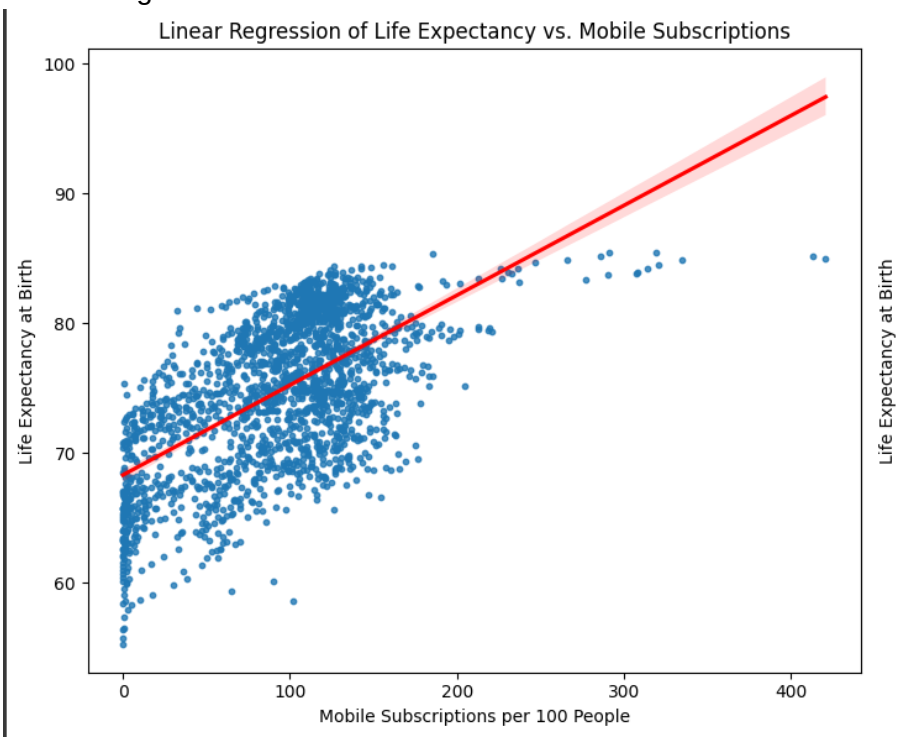
Analysis - This box and whisker plot highlights global disparities in life expectancy at birth. Australia and New Zealand have the highest and most consistent life expectancy, while Sub-Saharan Africa has the lowest with significant variability. Regions like Western Europe and Northern Europe also show high life expectancy with less variability. The wide variability and outliers in regions like Sub-Saharan Africa and Micronesia indicate uneven life expectancy within these regions.

3. Violin Plot



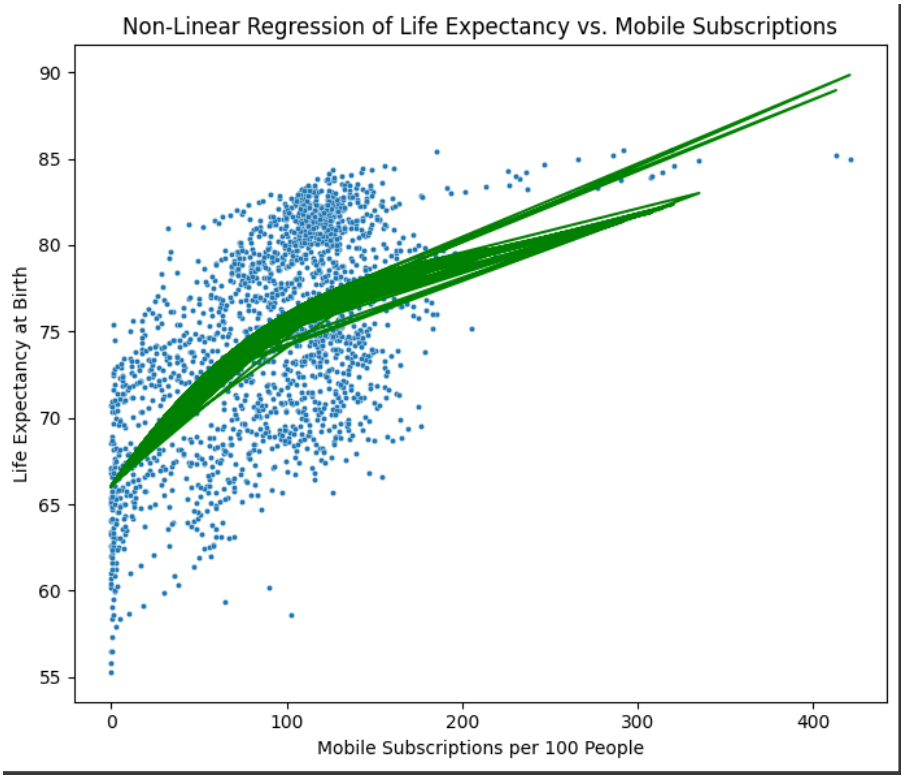
Analysis - Asia has the widest distribution, indicating a high variability in population growth, with some countries experiencing significant growth. Europe and Oceania have more compact distributions, suggesting more consistent and lower population growth rates across countries in these regions. Africa and the Americas have moderate variability in population growth, but their distributions indicate generally positive growth rates.

4. Regression Plot - Linear



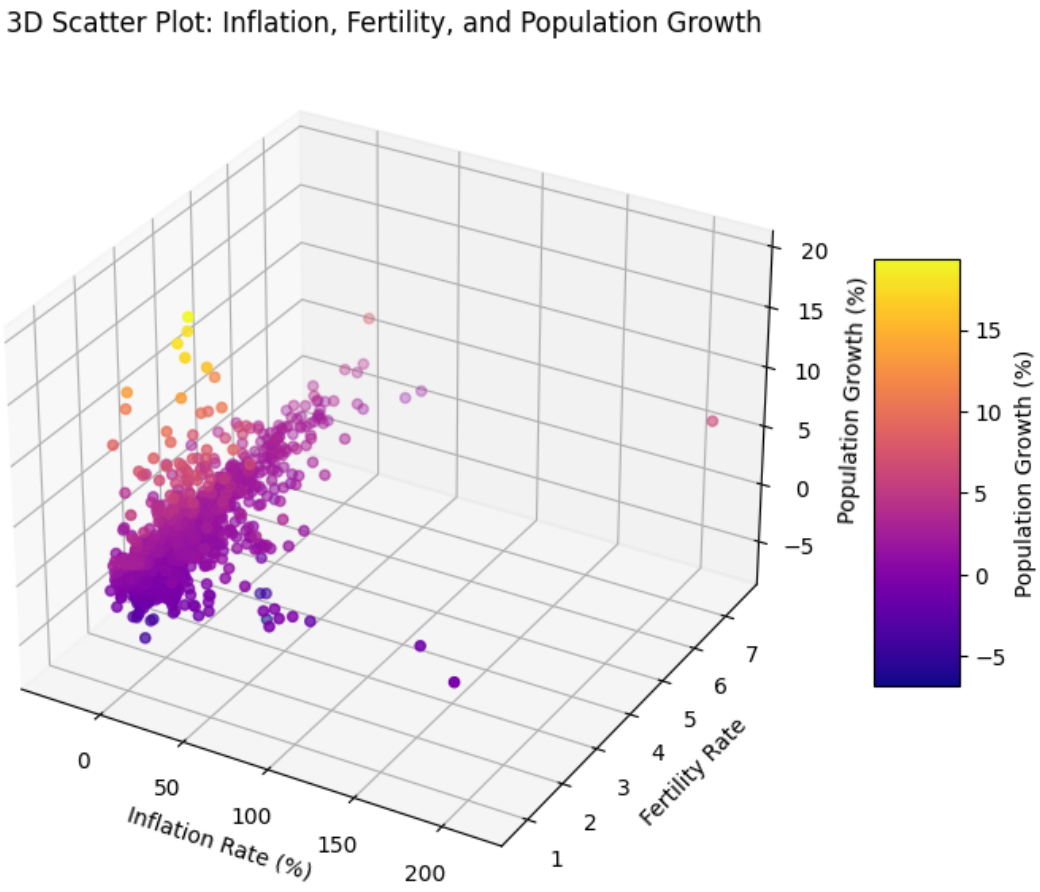
Analysis - The linear regression plot suggests a positive correlation, meaning higher mobile subscriptions tend to be associated with higher life expectancy.

5. Regression Plot - Nonlinear



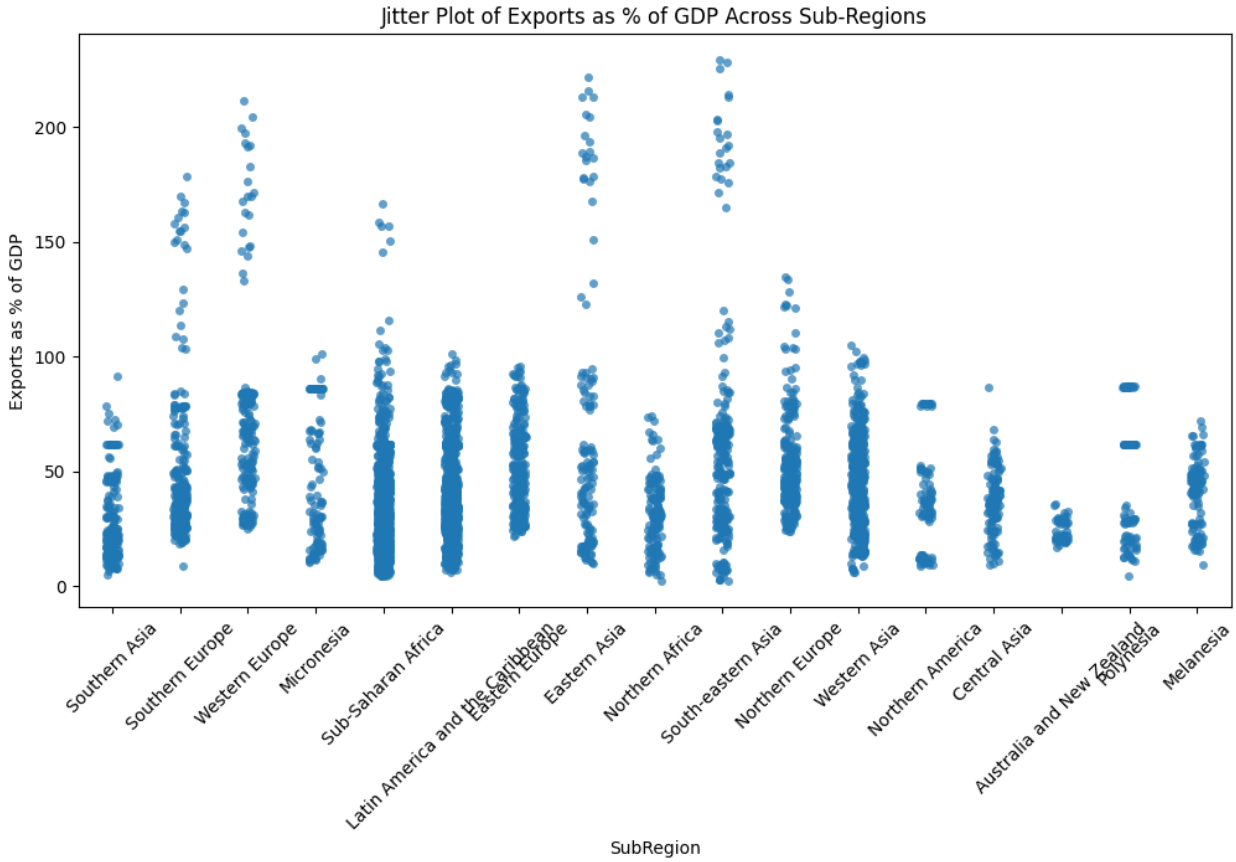
Analysis - The non-linear regression plot provides a more refined view. It captures a potential plateau effect, where the impact of mobile subscriptions on life expectancy might diminish beyond a certain point. This suggests that while mobile phones can improve access to information and services, contributing to better health outcomes, other factors like healthcare infrastructure and socioeconomic conditions also play significant roles in determining life expectancy.

6. 3D Chart



Analysis - The color gradient represents population growth, with yellow indicating high growth and purple indicating low or negative growth. High population growth is typically associated with moderate to high fertility rates and lower inflation rates. Low population growth or negative growth (purple) occurs across a wide range of inflation rates and tends to appear more in countries with lower fertility rates.

7. Jitter Plot



Analysis - Sub-regions like Western Europe and Eastern Asia show a concentration of points towards the higher end of the Exports as % of GDP scale, suggesting a strong reliance on exports for their economies. While regions like Australia and New Zealand have the lowest reliability on exports.