**Name: Anish Gade**

UID: 2021700022

Batch: BE CSE(DS)

AIM:

Create advanced charts using Tableau / Power BI / R / Python / Plotly or Chart or 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, Line, Area, Waterfall, Donut, Treemap, Funnel
- Write observations from each chart

1. BOX PLOT AND VIOILIN PLOT
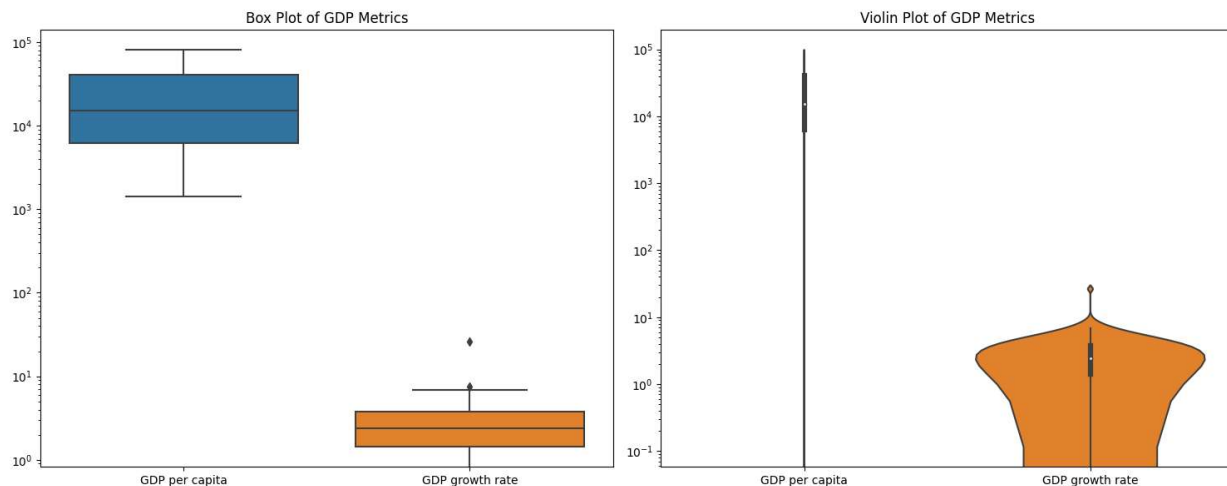
The box and violin plots reveal significant wealth inequality, with GDP per capita showing a broad spread from $1,500 to $80,000 and a right-skew due to a few wealthy outliers. Most countries fall within lower-income brackets, while only a few achieve high GDP per capita. In contrast, GDP growth rates are more uniformly distributed, with a median of 2–3% and most countries within the 0–5% range, indicating moderate growth across economies. The rounded and symmetric violin shape for growth rates, with a slight bimodality, suggests differing growth patterns between developed and developing nations. These patterns indicate that while absolute wealth varies widely, growth rates tend to converge, suggesting potential for economic catch-up by developing countries.
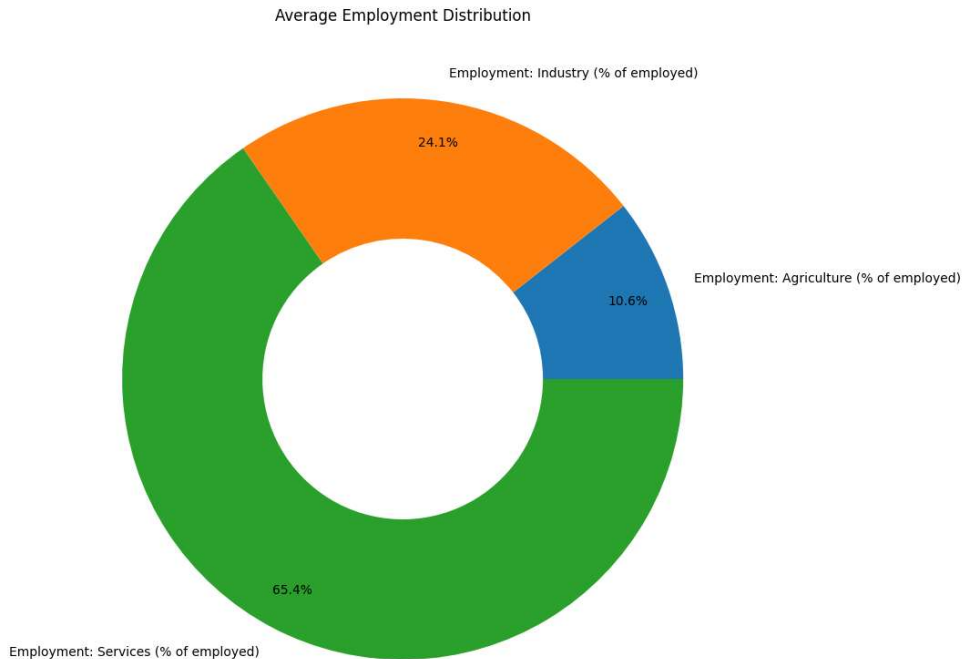
2. Regression Plot - GDP vs Life Expectancy

GDP per Capita vs Life Expectancy
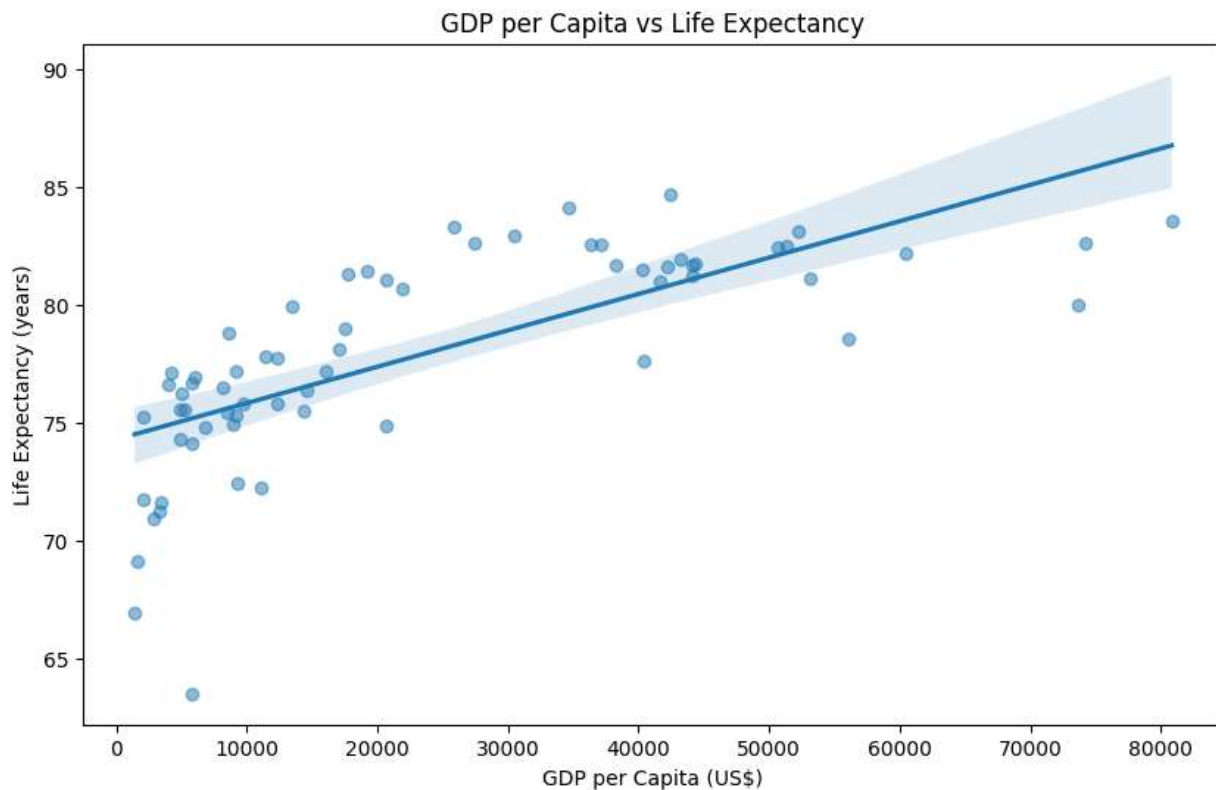
The relationship shows a moderately strong correlation (R-squared ~0.6-0.7) with distinct phases of development: rapid improvement in life expectancy up to $20,000 GDP per capita, moderate gains from $20,000-40,000, and diminishing returns beyond $40,000. The data reveals uneven distribution, with most countries concentrated in the $0-40,000 GDP range. Notable patterns include higher variance in life expectancy at lower GDP levels, a critical transition point around $10,000-15,000 GDP where steep improvements occur, and a plateau around 85 years in the highest GDP ranges. From a policy perspective, this suggests that economic development initiatives would have the highest impact in lower-income countries, while high-income nations need to focus on non-economic factors to improve life expectancy. The presence of overperforming middle-income countries (achieving high life expectancy with moderate GDP) indicates opportunities for knowledge transfer and system efficiency improvements across nations.

3. 3D Scatter Plot - GDP, Life Expectancy, and Internet Usage

3D Relationship: GDP, Life Expectancy, and Internet Usage
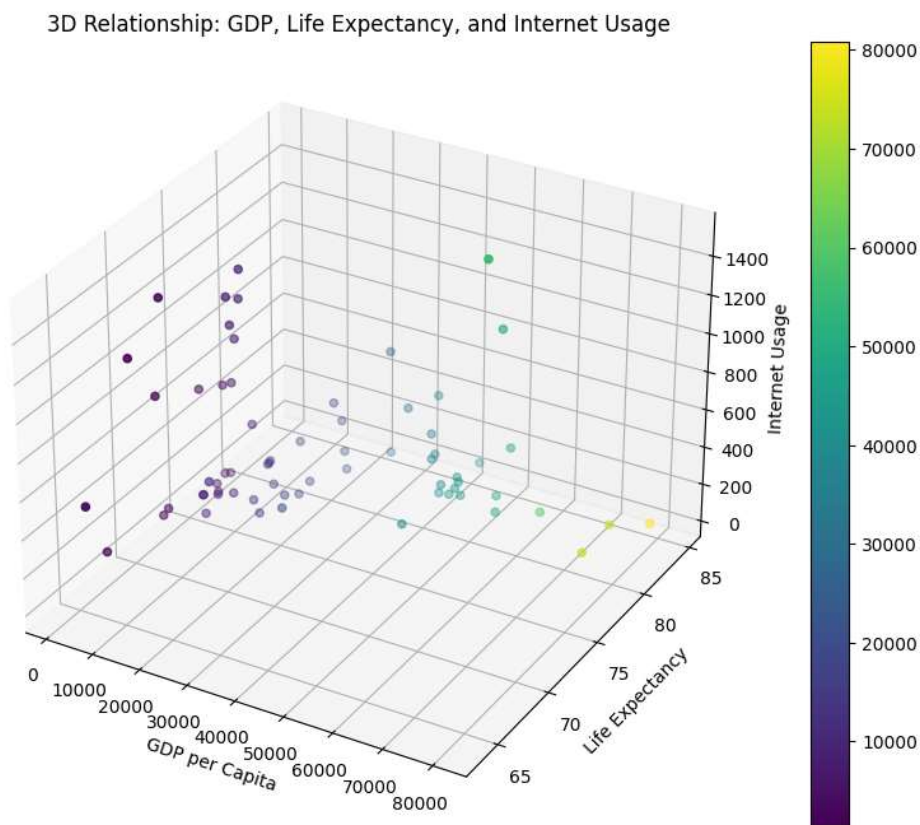
The visualization reveals strong positive correlations among GDP, life expectancy, and internet usage, with developed nations (indicated by yellow/green points) clustering in the upper right quadrant showing high values across all metrics, while developing nations (purple points) cluster in the lower left with lower values. A notable "development gap" exists between these clusters, with few countries in the middle range. Internet usage emerges as the most sensitive indicator of development, showing the greatest variation between low and high GDP countries. While some outliers demonstrate high internet usage despite moderate GDP, few countries achieve high life expectancy without corresponding high GDP, suggesting that while economic development generally drives improvements across all metrics, some countries may find alternative paths to specific aspects of development.

4. Donut Chart - Average Employment Distribution

Average Employment Distribution

Employment: Industry (% of employed)

24.1%

Employment: Agriculture (% of employed)

10.6%

65.4%

Employment: Services (% of employed)

The labor market structure shows a mature, service-oriented economy, with 65.4% of employment in the services sector, typical of advanced economies where healthcare, finance, and education dominate. Industry employs 24.1%, indicating a solid manufacturing base, though the economy is not heavily focused on production. Agriculture, with just 10.6% of employment, reflects highly efficient practices, characteristic of developed economies that rely on mechanized farming. This distribution—high in services, moderate in industry, and low in agriculture—suggests a post-industrial economy requiring a skilled, educated workforce, with ongoing shifts from traditional sectors to meet the demands of a knowledge-driven, service-dominant economy.

5. Jitter Plot - Education vs Internet Usage

Education Expenditure vs Internet Usage (with Jitter)
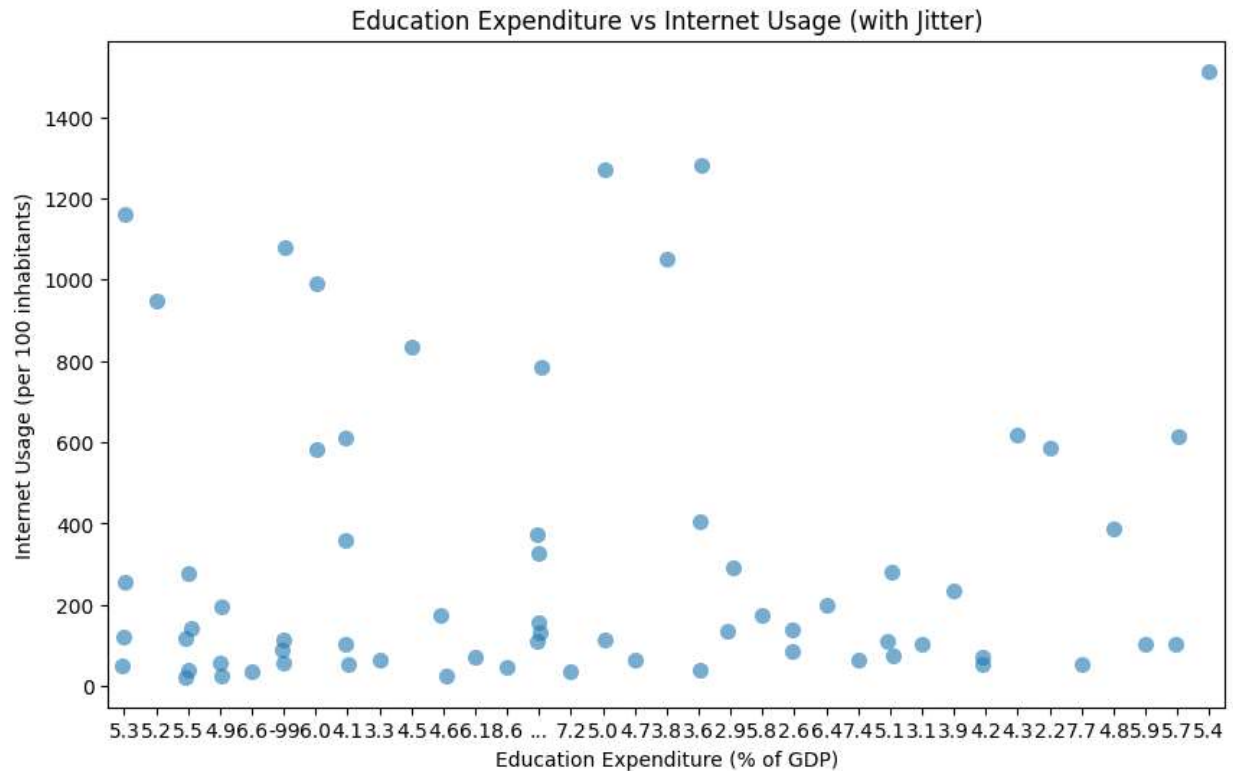
The analysis of education expenditure versus internet usage reveals a notably weak linear relationship, with countries showing widely varying internet penetration rates despite similar education spending levels (typically 2-8% of GDP). The data exhibits significant clustering in the lower range of internet usage (0-400 per 100 inhabitants), with several notable outliers exceeding 1000 per 100 inhabitants. This scattered distribution pattern, combined with high vertical variance, suggests that internet adoption is influenced by multiple factors beyond education spending alone, such as infrastructure development, income levels, and digital policies. The varying levels of internet usage among countries with similar education expenditure indicates that successful digital development likely depends more on efficient resource allocation and targeted digital infrastructure investments rather than solely on education spending.