BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]
Department of Computer Engineering

Name	Ms Neerja Doshi
UID	2021300029
DIV	BE COMPS [ADV -> BATCH F]`
ADV EXP 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
Dataset Particulars	Name: Life Expectancy [WHO] Dataset Link: https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who Dataset Features : 1) Country: The name of the country. 2) Year: The year for which the data is recorded. 3) Status: Indicates whether the country is classified as "Developed" or "Developing." 4) Life expectancy: The average lifespan of individuals in the country. 5) Adult Mortality: The probability of dying between the ages of 15 and 60 per 1000 population. 6) Infant Deaths: The number of infant deaths per 1000 population. 7) Alcohol: Recorded per capita consumption of alcohol (in liters of pure alcohol) for individuals aged 15 and older. 8) Percentage expenditure: Expenditure on health as a percentage of Gross Domestic Product (GDP) per capita. 9) Hepatitis B: Immunization coverage against Hepatitis B among 1-year-olds. 10) Measles: Immunization coverage against measles among 1-year-olds.

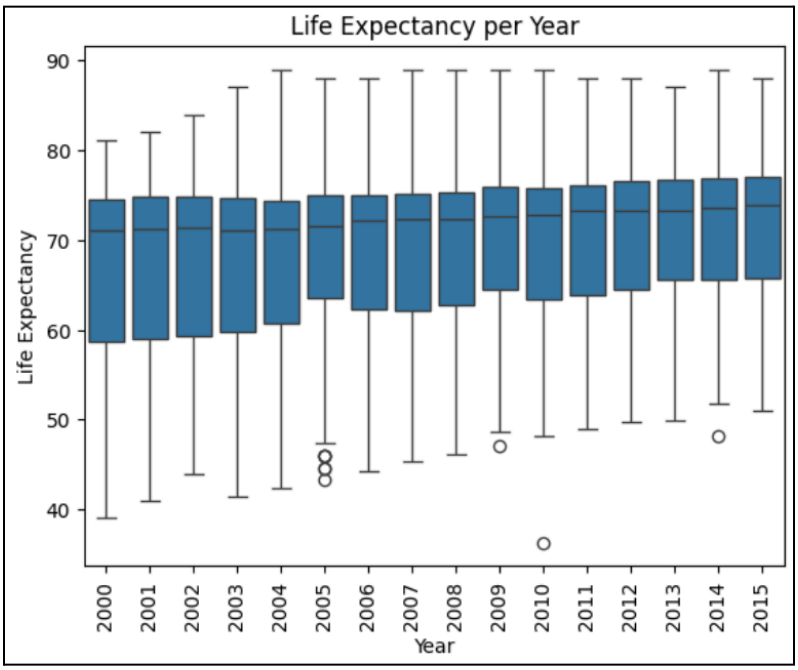
	11) BMI: The Body Mass Index (BMI) of the population. 12) Under-Five Deaths: The number of deaths of children under five years old per 1000 population. 13) Polio: Immunization coverage against polio among 1-year-olds. 14) Total expenditure: Total expenditure on health as a percentage of GDP. 15) Diphtheria: Immunization coverage against diphtheria among 1-year-olds. 16) HIV/AIDS: Prevalence of HIV/AIDS among adults aged 15-49. 17) Population: Total population of the country. 18) Thinness 1-19 years: Percentage of the population aged 1-19 years who are underweight. 19) Thinness 5-9 years: Percentage of the population aged 5-9 years who are underweight. 20) Infant Mortality Rates: Number of infant deaths per 1000 live births. 21) Schooling: Average years of schooling completed by the population aged 25 and older. The shape of the dataset is : (2938, 22).
Data Visualisations `	**Word Cloud**  1) As seen aboev the highest word count is seen of "Developing" followed by "Developed" Countries. 2) This suggests that uptil 2015 , the count if developing countries was highest. —----------

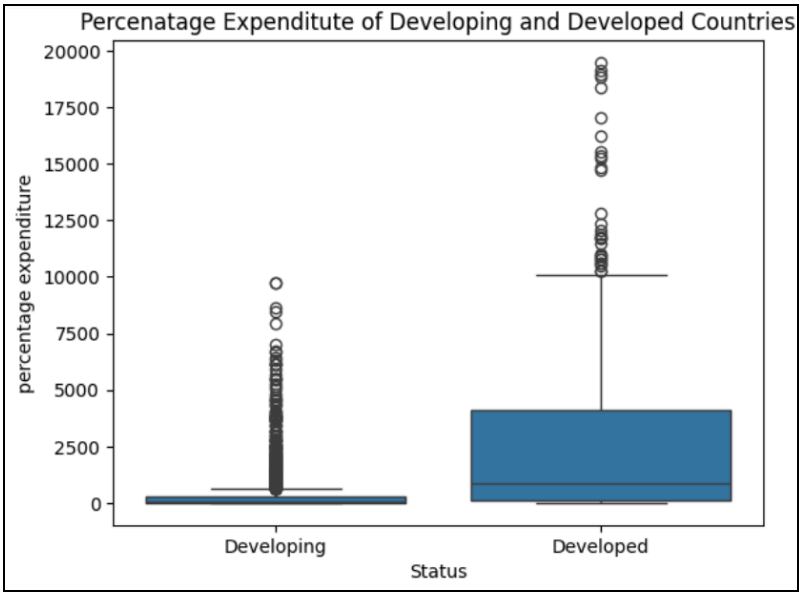
Box Plot

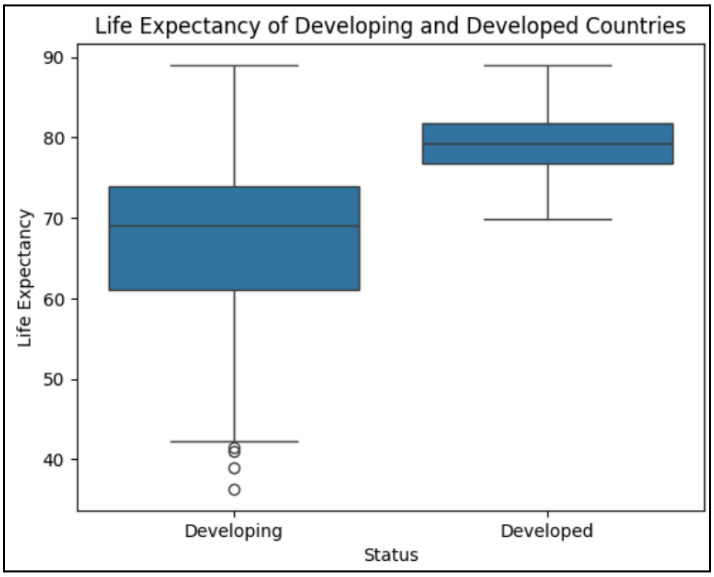
1) Life Expectancy per Year

2) Status vs Percentage expenditure

3) Status vs Life Expectancy







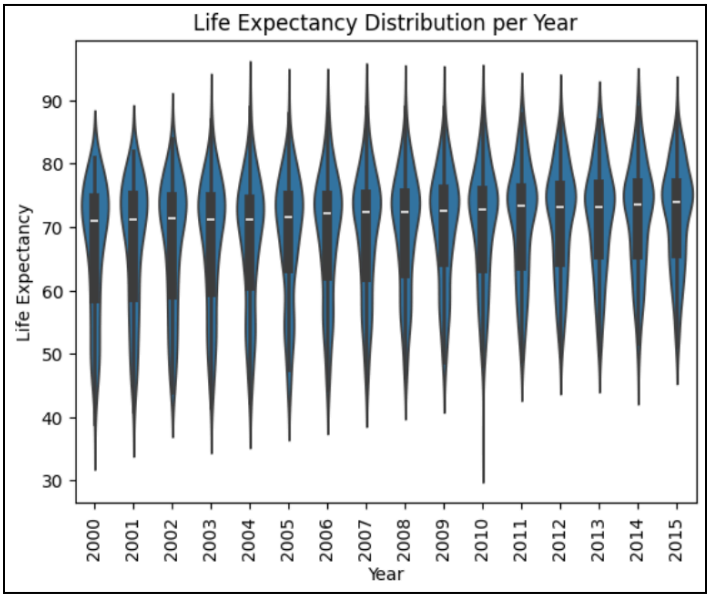
1) The median life expectancy has been almost constant throughout thr years .
2) Same goes with 75% percentile line.
3) Post 2003 , there has been a increase in the min and max value of Life Expectancy . The second box plots gives an answer to this .. as there has been and increas in Percentage expenditure seen as well owing to high quality of life.
4) Devloped nations spend more and hence have higher life expectancies contrary to deveolping nations.

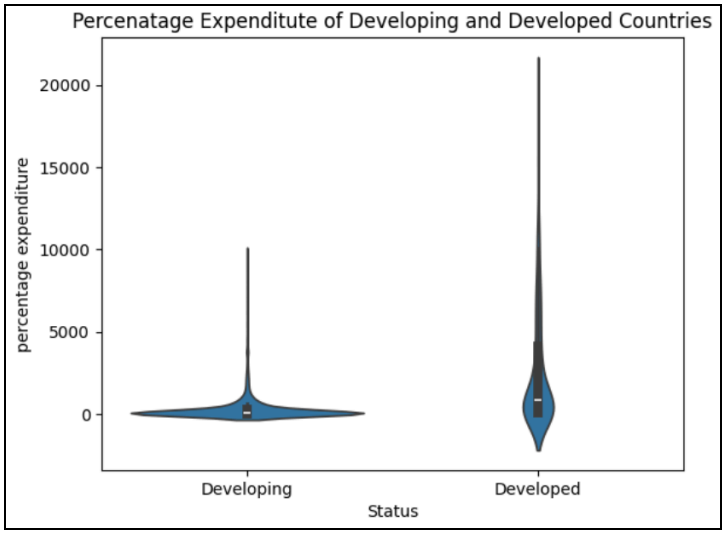
Violin Plot

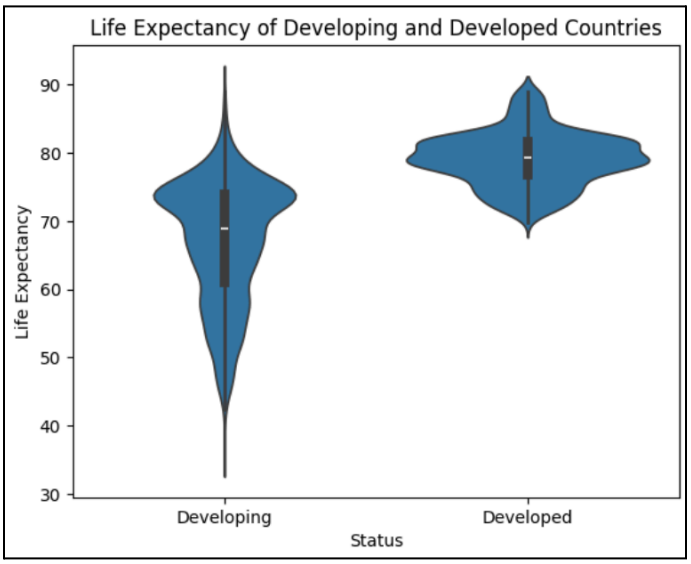
1) Life Expectancy per Year

2) Status vs Percentage expenditure

3) Status vs Life Expectancy





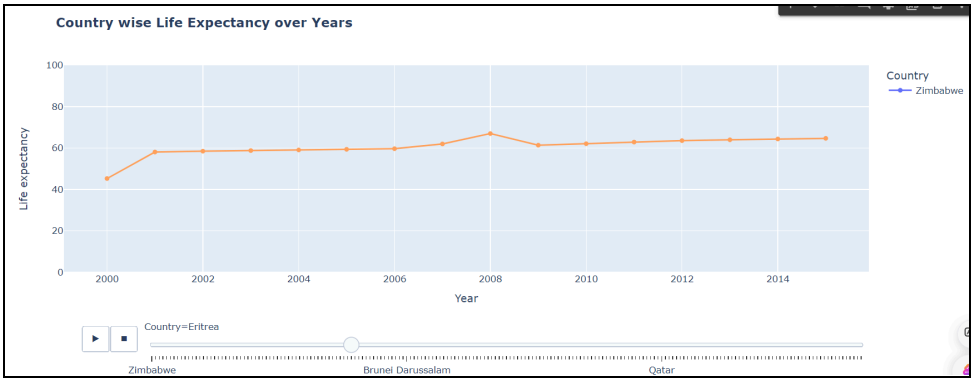


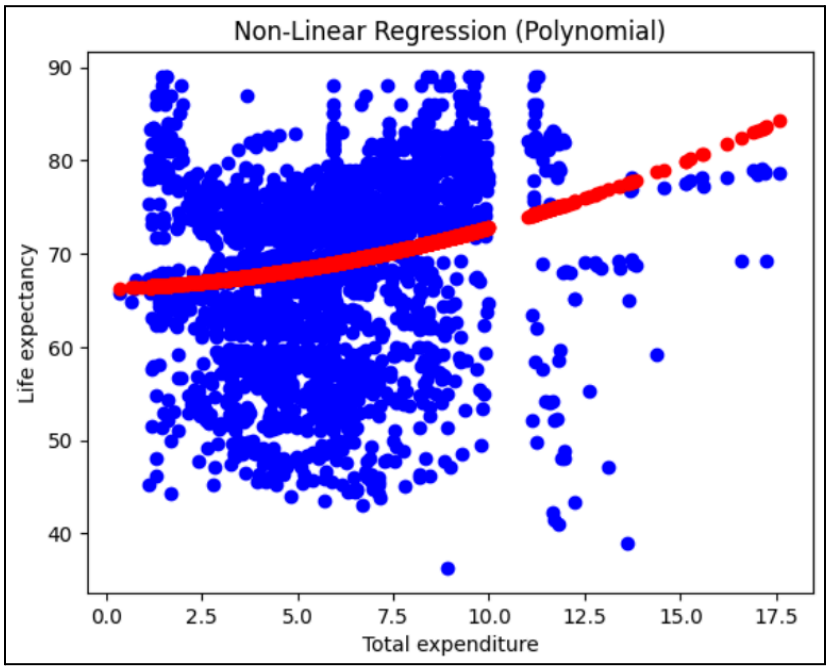
1) **The median life expectancies is concentrated around 72 yrs.**
2) **Majority of the developing nations spend nil amount**
3) **Even though developed nations spend more , there is only a slight increase in median life expectancies.**

Regression Lines

1) Country wise Life Expectancy -> Linear

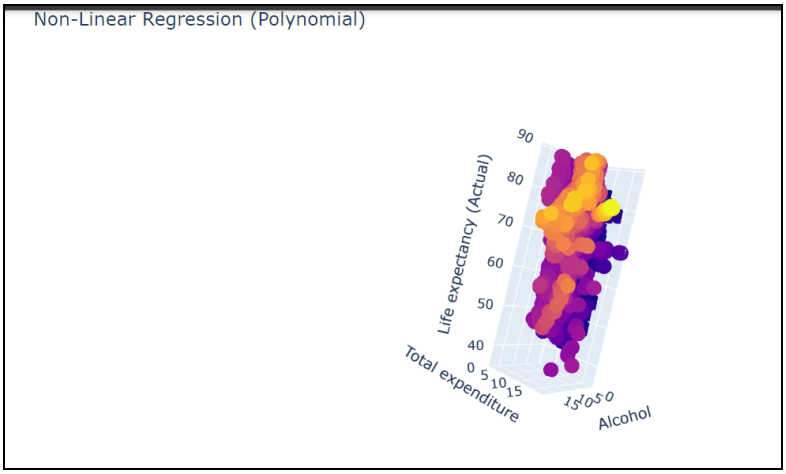
2) Life Expectancy on Total Expenditure and Alcohol -> Nonlinear





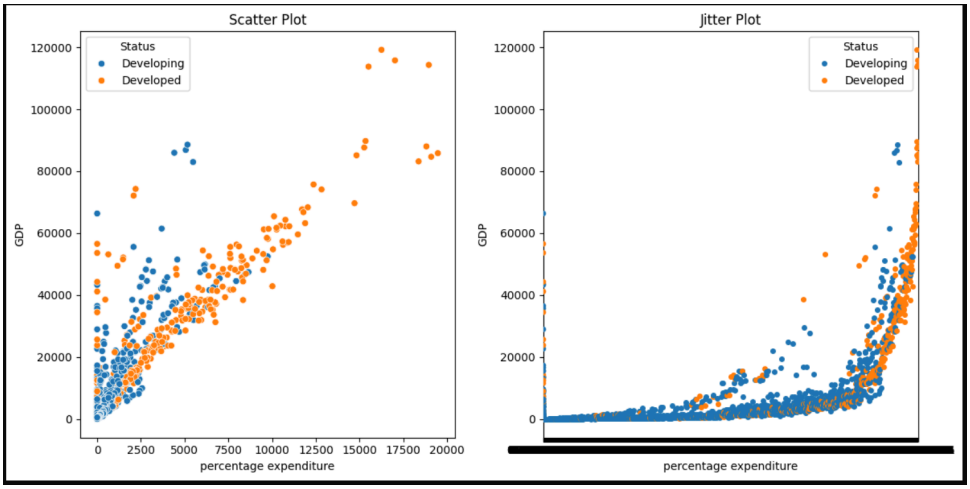
3D Plot

Dependent Variable : Life Expectancy Independent Variables : Alcohol and Total expenditure



Jitter Plot

Sr. No.	Scatter Plot	Jitter PLot
1	overalapping data is seen	To reduce Overlapping .. a little jitter (offset) is added



1) As seen in the correlation matrix , percentage expenditure and GDP have a direct correlation
2) On Plotting both of them we understand that because of this direct relation , there is a Linear plot which is seen.

	3) For many developing nations, spending more is not increasing the GDP alot 4) Developed nations have high GDP coz of their high expenditure.
Conclusion	By performing this , i understood better chart visualizations on Socio economic dataset