**AIML – ASSIGNMENT – 2**

**LIFE EXPECTANCY**

**Simple Linear Regression:**

* In Simple Linear Regression only one independent variable should be considered.
* The Dataset contains 22 variables. So, to choose the best variable we must run correlation.
* But before doing that we need to clean the dataset.
* The data set contains ‘0’ values and ‘Nan’ values.
* Before removing the ‘Nan’ values 
* After removing the ‘Nan’ values 

**Correlation:**

**A screenshot of a computer

Description automatically generated**

* From the correlation values we can infer that Schooling has the highest correlation.
* So, we have taken Schooling as Independent variable and Profit as Dependent variable.
* After running the regression, the R-squared value came as 0.5052.
* The 50% of the variance in Life expectancy can be explained by Schooling.

**Multiple Regression**:

* Here the we can taken 5 independent variables for regression.
* Independent variables are Schooling, Income composition of resources, BMI, Adult Mortality, HIV/AIDS.
* Adult Mortality and HIV/AIDS have high negative correlation.
* After running Multiple regression we got R-squared value as 0.7959
* The 79% of the variance in Life expectancy can be explained by Schooling, Income composition of resources, BMI, Adult Mortality, HIV/AIDS.
* So these are the factors that influence the Life expectancy for an individual.
* So, if a person has a better Schooling and education he will be eventually living in higher standards and will have better healthcare facilities.
* A Healthy Body Mass Index will lead to better life expectancy.
* A higher income will have a better resource in terms of healthcare.