

DEATH MORTALITY RATE OF WORLD

Abstract:

Nowadays there are a lot of risk factors that cause death like smoking, Alcohol consumption, Diabetes, Blood pressure, etc. In a year for a risk factor, thousands of people are dying. so, analysing the Relationship between risks factors is essential. In this report mortality rate of world is analysed from the year of 1990-2017. Mortality rate is the ratio of death occurring to the total number of populations. From this analysis, conclusions can be drawn which help to bring awareness to people. Using various Regression like linear, polynomial, multiple regression, and statistics concepts.

Data Collection:

The dataset which is used for analysing the death mortality is taken from Kaggle. The dataset provides information about the death due to a risk factor for an entity in a particular year. The data is from the year 1990 to 2017.

Approach:

- The various python libraries such as Numpy, pandas, Matplotlib are used for the purpose of mathematical calculations, extraction of data and visualization respectively.
- Some other packages are used like sklearn which helps to find relationship between risk factors.
- The extracted dataset was found to have null values which are then effectively handled to achieve accurate analysis rather than removing them from the dataset.

Analysis and visualization:

1. Contains any NULL value or not:

Purpose: Handling NULL is essential because operation can't be performed on NULL data

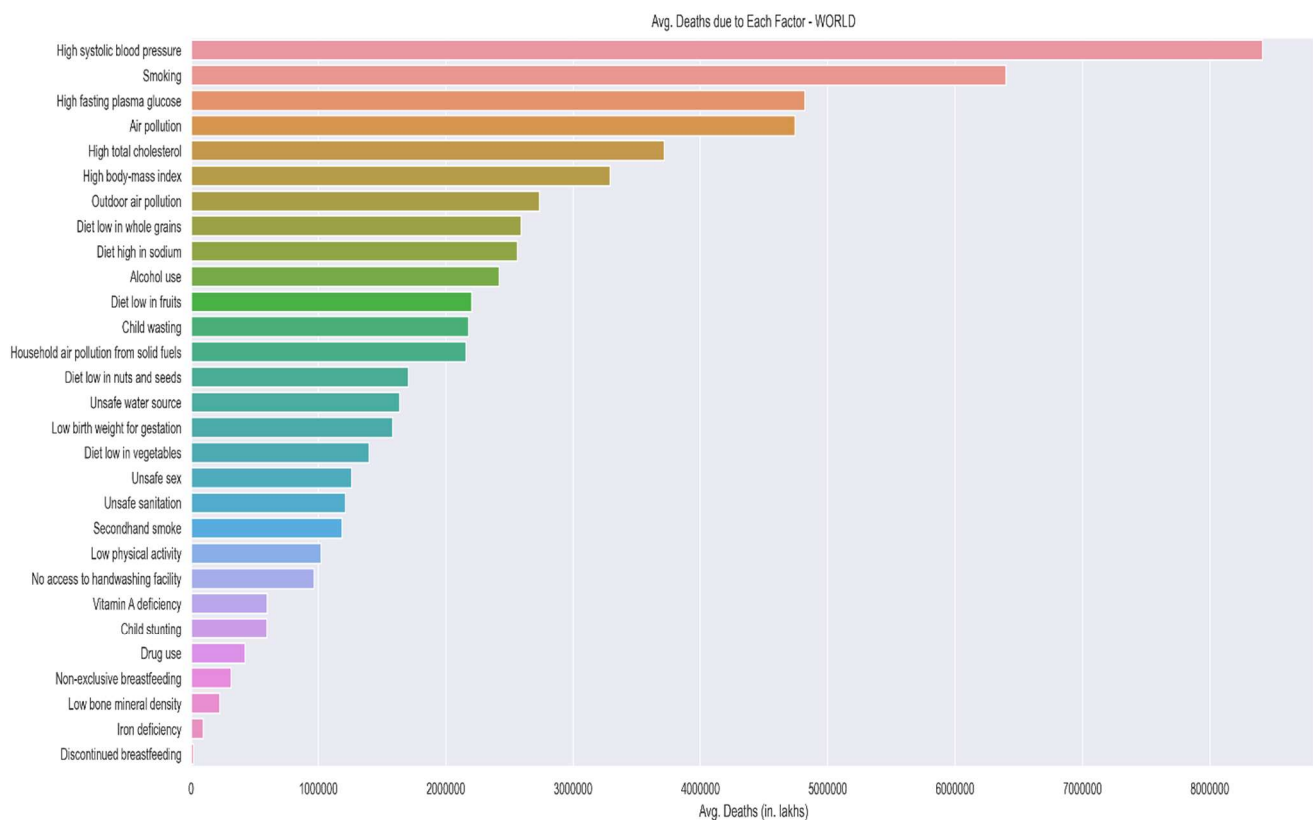
```
Entity      0
Code       980
Year        0
Unsafe water source  0
Unsafe sanitation  0
No access to handwashing facility  0
Household air pollution from solid fuels  0
Non-exclusive breastfeeding  0
Discontinued breastfeeding  0
Child wasting  0
Child stunting  0
Low birth weight for gestation  0
Secondhand smoke  0
Alcohol use  0
Drug use  0
Diet low in fruits  0
Diet low in vegetables  0
Unsafe sex  0
Low physical activity  0
High fasting plasma glucose  0
High total cholesterol  4907
High body-mass index  0
High systolic blood pressure  0
Smoking  0
Iron deficiency  0
Vitamin A deficiency  0
Low bone mineral density  0
Air pollution  0
Outdoor air pollution  1
Diet high in sodium  0
Diet low in whole grains  0
Diet low in nuts and seeds  0
dtype: int64
```

Null values are replaced by mean of that column non values. So that they will be no problem while handle the data and to get good accuracy.

2. Which risk factor causes more death:

Purpose:

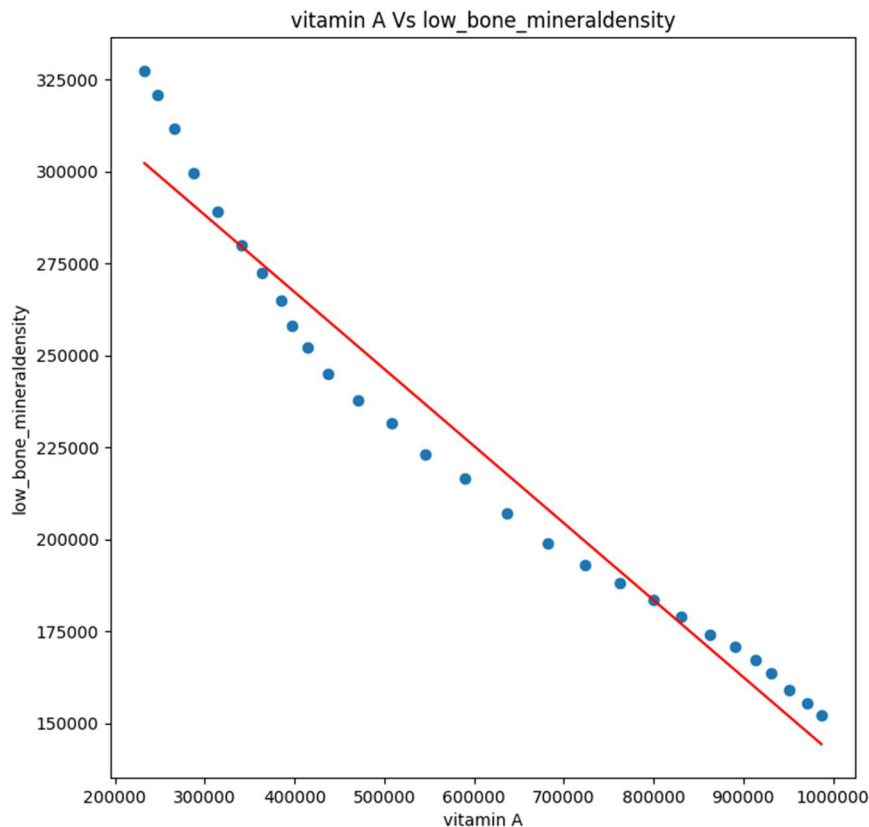
The factor which effect people more is known then create awareness among people is easy



Inference: It is clear that more deaths are causing due to High symbolic blood pressure in world

3. Relationship between Vitamin A and Low bone mineral density:

Hypothesis: Vitamin A deficiency leads to low bone mineral density.

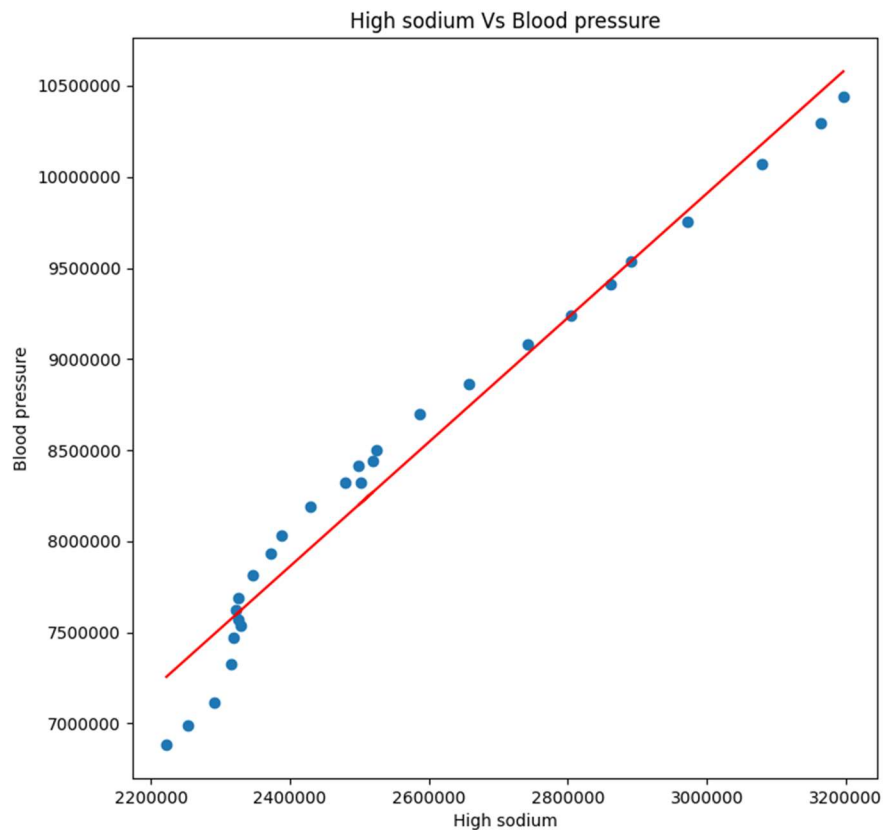


With the Linear Regression technique, the relationship between vitamin A and Low bone mineral density analysed. As a result, the slope is -0.209 and Bias is 350932.75 . The negative value of slope indicates that the right-side angle of the line with the x-axis is more than 90 degrees. The increase in value of low bone mineral density deaths with respect to Vitamin A is 0.2 times with the change of 350932.

Inference: From analysis it is clear that there is relationship between Vitamin A and low bone mineral density.

4. Relationship between High sodium and High systolic blood pressure:

Hypothesis: More sodium in blood leads to increase in blood pressure

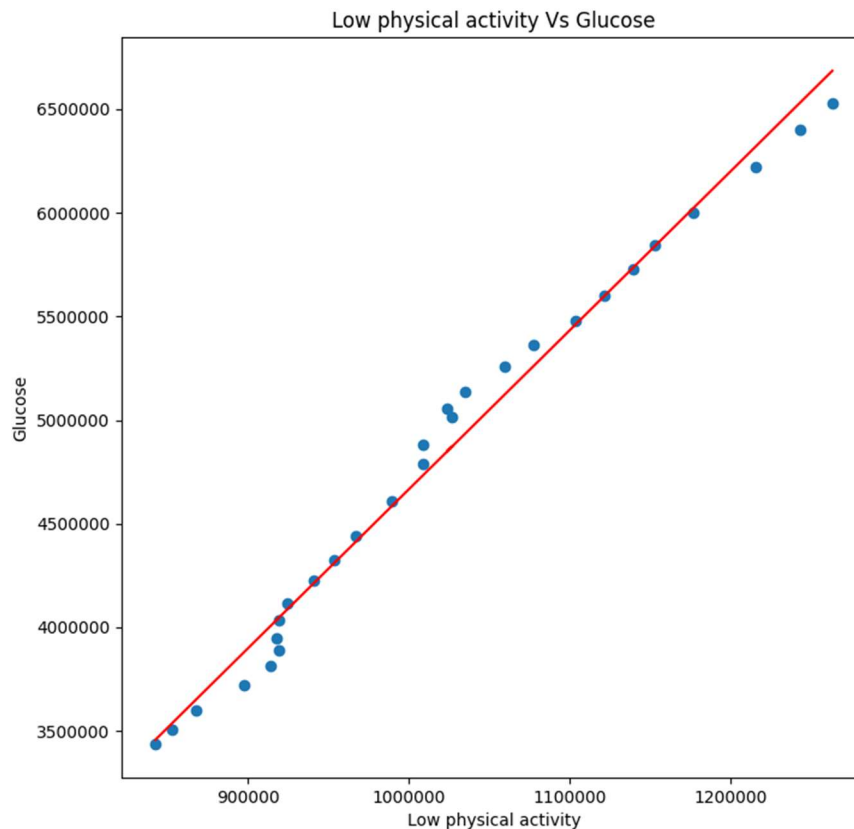


With the Linear Regression technique, the relationship between High Sodium and High systolic blood pressure analysed. As a result, the slope is 3.4111 and Bias is -324570.28. So, the regression line is $y = 3.4X - 324570$. The increase in value of low bone mineral density deaths with respect to High Sodium is 3.4 times with the change of -324570.28.

Inference: The analysis is speaking that there is relationship between Vitamin A and low bone mineral density.

5. Relationship between Low physical activity and High fasting plasma glucose:

Hypothesis: Low physical activity leads to high glucose in blood

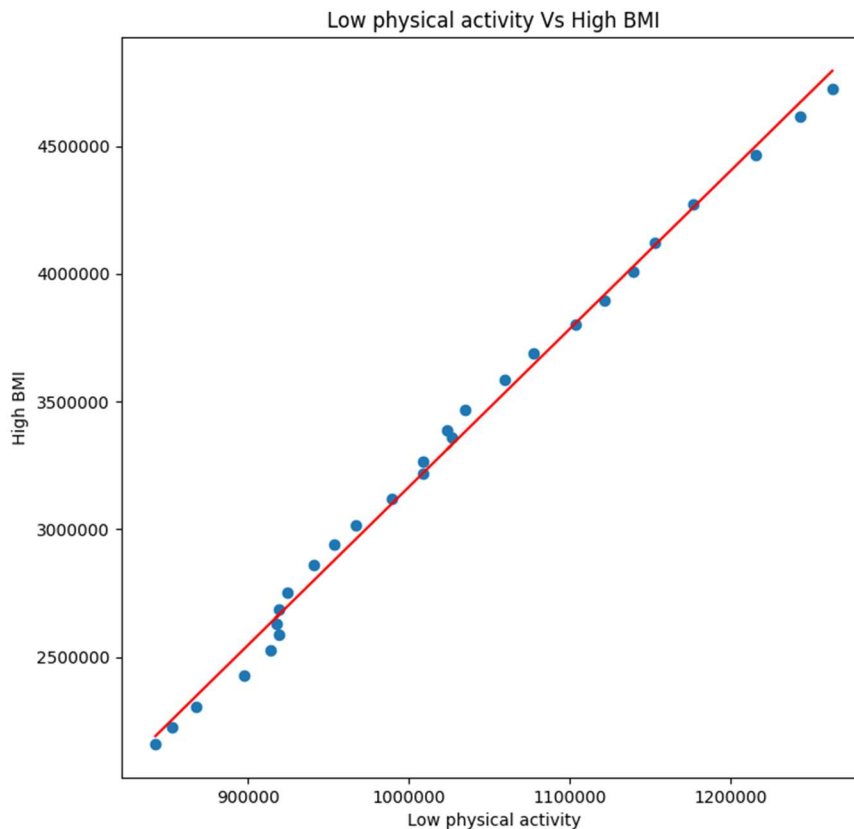


Points are scattered on the graph as the line can be fitted Linear Regression technique can be to analyze the relationship between Low physical activity and High fasting plasma glucose analysed. As a result, the slope is 7.67 and Bias is -3011476.761. So, the regression line is $y = 7.67 X - 3011476.76$. Negative bias indicates negative change in after multiplying slope. The increase in value of Low physical activity deaths with respect to High fasting plasma glucose is 7.67 times with the change of -324570.28.

Inference: The analysis indicates that there is relationship between Low physical activity and High fasting plasma glucose.

6. Relationship between Low physical activity and High body mass-index:

Hypothesis: Low physical activity leads to High Body mass index

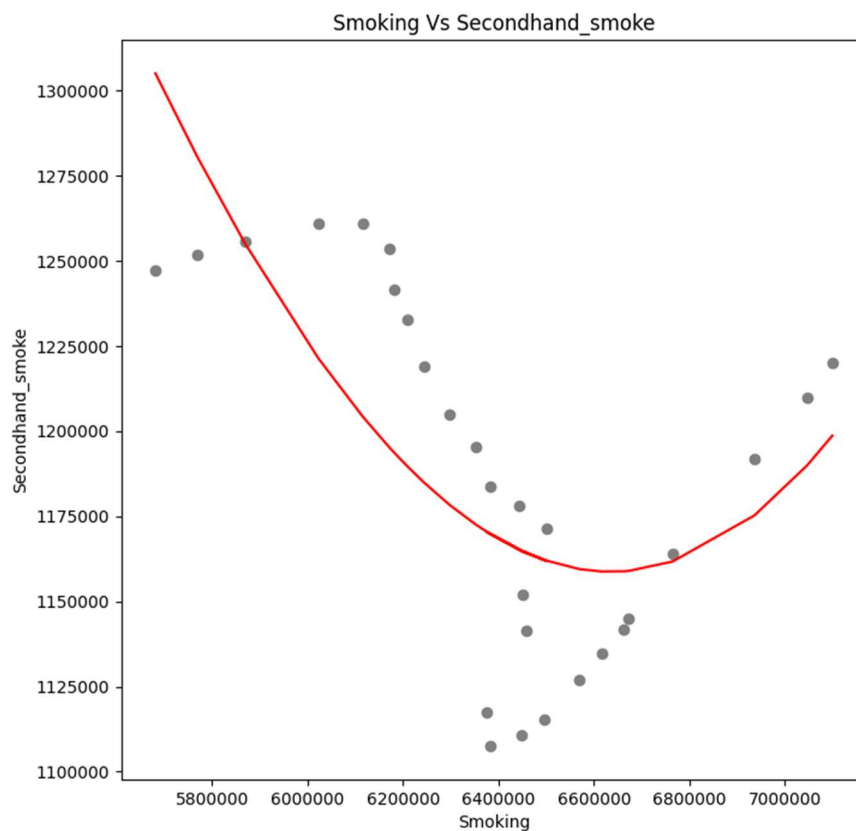


Points are scattered on the graph as the line can be fitted Linear Regression technique can be to analyze the relationship between Low physical activity and High Body mass-index analysed. Taking Independent variable as low physical activity and dependent as High Body mass-index. As a result, the slope is 6.19 and Bias is -3029176.81. So, the regression line is $y = 6.19 X - 3029176.81$. Negative bias indicates negative change in after multiplying slope. The increase in value of Low physical activity deaths with respect to High Body mass-index is 6.19 times with the change of -3029176.81.

Inference: The analysis indicates that there is relationship between Low physical activity and High Body mass-index.

7. Relationship between Smoking and Secondhand smoke:

Hypothesis: More people are dying due to secondhand smoke which are occurred by people smoke.

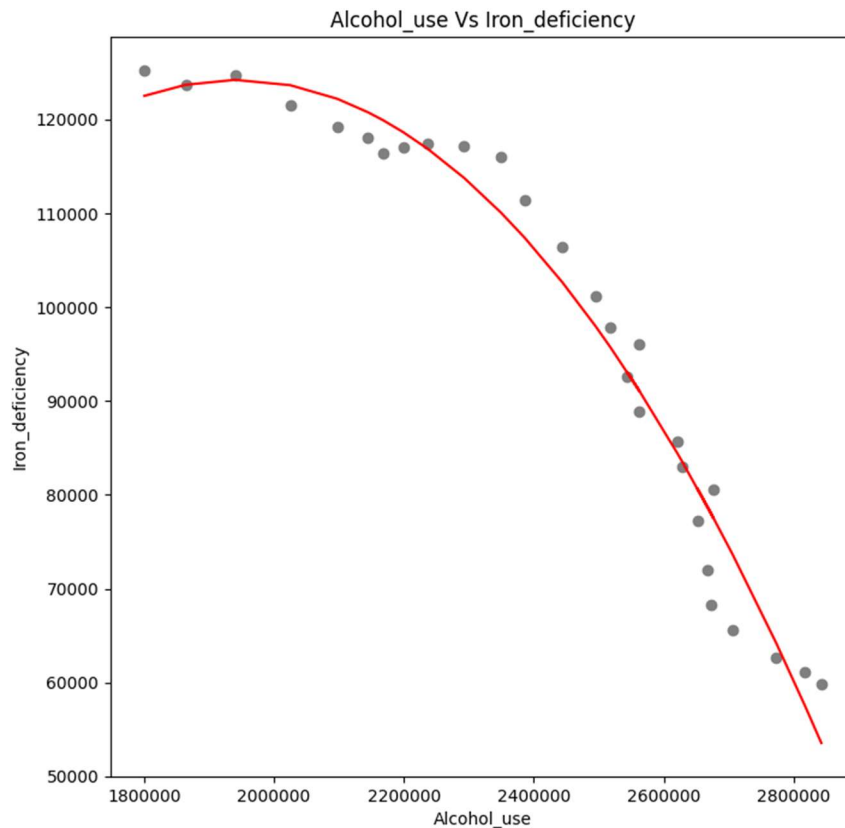


After scattering points on the graph, it will be that the Polynomial regression technique should be applied to find the relationship between the number of deaths caused due to Smoking and Secondhand smoke. The r^2 score obtained is 0.7466 which is almost close to 1 that implies that deaths caused due to Secondhand smoke is depending on deaths caused due to Smoking.

Inference: deaths caused due to second hand smoke is depending on deaths caused due to smoking.

8. Relationship between Alcohol use and Iron deficiency:

Hypothesis: People are dying due to iron deficiency which is depending on who is dying by consuming more alcohol.

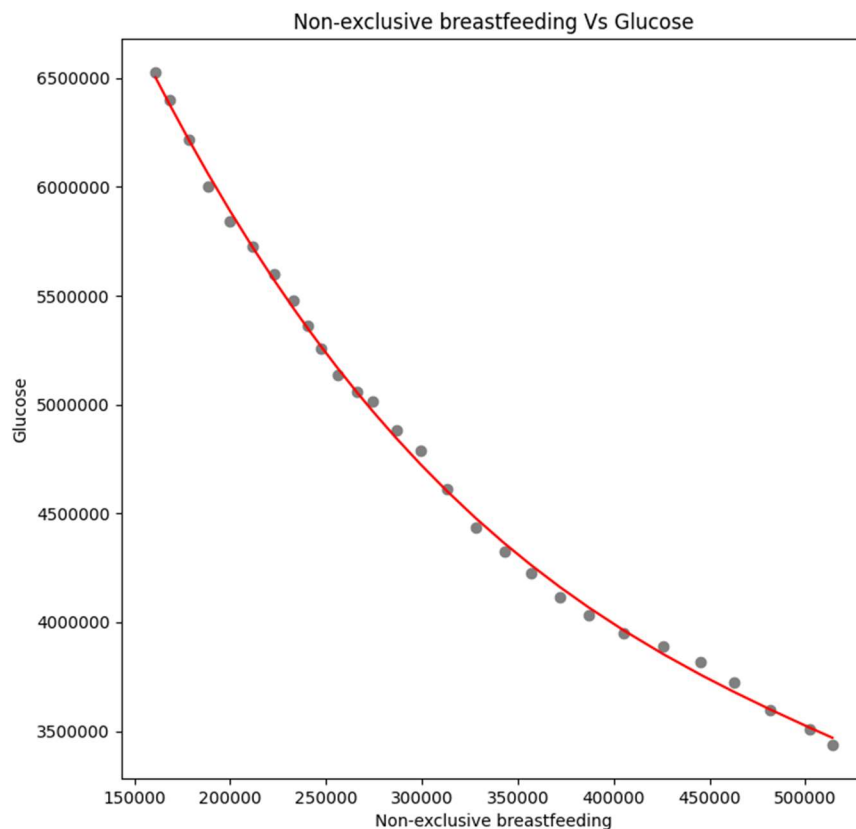


After scattering points on the graph, it will be that the Polynomial regression technique should be applied to find the relationship between the number of deaths caused due to iron deficiency and Alcohol use. The r^2 score obtained is 0.96682 which is almost close to 1 which implies that deaths caused due to iron deficiency is depending on deaths caused due to Alcohol use.

Inference: deaths caused due to iron deficiency is depending on deaths caused due to alcohol consumption.

9. Relationship between Non-exclusive breastfeeding and High fasting plasma glucose:

Hypothesis: People are dying due to High fasting plasma glucose which is depending on who is dying by Non-exclusive breastfeeding.

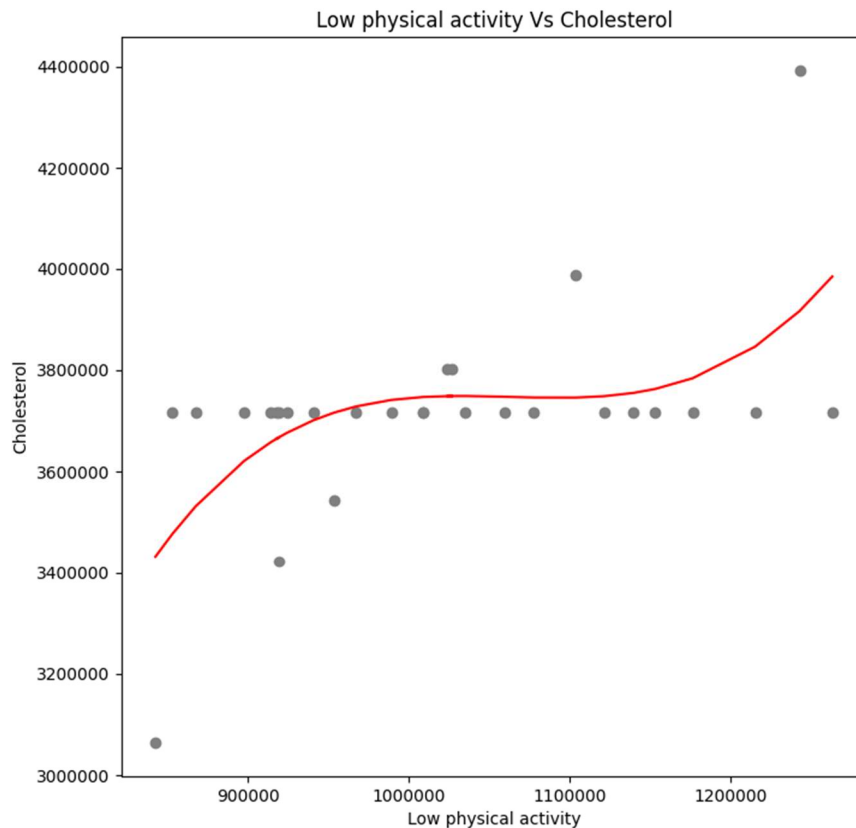


After scattering points on the graph, it will be that the Polynomial regression technique should be applied to find the relationship between the number of deaths caused due to Non-exclusive breastfeeding and High fasting plasma glucose. It is clear that Non-exclusive breastfeeding is independent variable and high fasting plasma glucose is dependent variable. The r^2 score obtained is 0.9986 which is almost close to 1 which implies that deaths caused due to High fasting plasma glucose is depending on deaths caused due to Non-exclusive breastfeeding.

Inference: deaths caused due to High fasting plasma glucose is depending on deaths caused due to Non-exclusive breastfeeding.

10. Relationship between Low physical activity and High total cholesterol:

Hypothesis: People are dying due to High total cholesterol which is depending on who is dying by Low physical activity.



After scattering points on the graph, it will be that the Polynomial regression technique should be applied to find the relationship between the number of deaths caused due to Low physical activity and High total cholesterol. It is clear that Low physical activity is independent variable and High total cholesterol glucose is dependent variable. The r^2 score obtained is 0.3282 which is almost close to 0 which implies that deaths caused due to High total cholesterol is not depending on deaths caused due to Low physical activity.

Inference: Deaths caused due to High total cholesterol is not depending on deaths caused due to Low physical activity.

11. Relation between Household air pollution from solid fuels, Outdoor air pollution Vs Air pollution:

In the compare Household air pollution from solid fuels, Outdoor air pollution and Air pollution. Generally, Air pollution is dependent on Household air pollution and Outdoor air pollution.so, independent variables are Household air pollution, Outdoor air pollution and dependent variable is Air pollution.

The coefficients of multiple regression are [0.92075656,0.94338962]

Deaths caused by Air pollution is less dependent on Household air pollution from solid fuels. For one unit increase of deaths due to Household air pollution from solid fuels there will be 0.9207 units of increase in deaths caused by Air pollution. Deaths caused by Air pollution is more dependent on Outdoor air pollution. For one unit increase of deaths due to Outdoor air pollution there will be 0.9433 units of increase in deaths caused by Air pollution.

12.Relation between Unsafe water source, No access to handwashing facility and Unsafe sanitation:

In the compare Unsafe water source, No access to handwashing facility and Unsafe sanitation. Generally, Unsafe sanitation is dependent on Unsafe water source and No access to handwashing facility.so, independent variables are Unsafe water source and No access to handwashing facility and dependent variable is Unsafe sanitation.

The coefficients of multiple regression are [-0.5706703 2.52404146]

Deaths caused by Unsafe sanitation is less dependent on Unsafe water source. For one unit increase of deaths due to Household air pollution from solid fuels there will be -0.5706 units of increase in deaths caused by Unsafe sanitation. Deaths caused by Unsafe sanitation is more dependent on No access to handwashing facility. For one unit increase of deaths due to No access to handwashing facility there will be 2.524 units of increase in deaths caused by Unsafe sanitation.

13.Relation between High systolic blood pressure, High fasting plasma glucose and Diet low in nuts and seeds:

In the compare High systolic blood pressure, High fasting plasma glucose and Diet low in nuts and seeds. Generally, Diet low in nuts and seeds is dependent on High systolic blood pressure and High fasting plasma glucose.so, independent variables are High systolic blood pressure and High fasting plasma glucose and dependent variable is Diet low in nuts and seeds.

The coefficients of multiple regression are [0.13543987,0.05584921]

Deaths caused by Diet low in nuts and seeds is more dependent on High systolic blood pressure. For one unit increase of deaths due to High systolic blood pressure there will be 0.1354 units of increase in deaths caused by Diet low in nuts and seeds. Deaths caused by Diet low in nuts and seeds is less dependent on High fasting plasma glucose. For one unit increase

of deaths due to High fasting plasma glucose there will be 0.0558 units of increase in deaths caused by Diet low in nuts and seeds.

14.Relation between High systolic blood pressure, High fasting plasma glucose and Smoking:

In the compare High systolic blood pressure, High fasting plasma glucose and Smoking. Generally, Smoking is dependent on High systolic blood pressure and High fasting plasma glucose.so, independent variables are High systolic blood pressure and High fasting plasma glucose and dependent variable is Smoking.

The coefficients of multiple regression are [0.45781432, -0.14149493]

Deaths caused by smoking is more dependent on High systolic blood pressure. For one unit increase of deaths due to High systolic blood pressure there will be 0.4578 units of increase in deaths caused by Smoking. Deaths caused by Smoking is less dependent on High fasting plasma glucose. For one unit increase of deaths due to High fasting plasma glucose there will be-0.1414 units of increase in deaths caused by Smoking.

15.Relation between Low bone mineral density, Diet low in nuts and seeds and Child stunting:

In the compare Low bone mineral density, Diet low in nuts and seeds and Child stunting. Generally, Child stunting is dependent on Low bone mineral density and Diet low in nuts and seeds. So, independent variables are Low bone mineral density and Diet low in nuts and seeds and dependent variable is Child stunting.

The coefficients of multiple regression are [-4.78247808, 0.07025614]

Deaths caused by Child stunting is less dependent on Low bone mineral density. For one unit increase of deaths due to Low bone mineral density there will be -4.7824 units of increase in deaths caused by Child stunting. Deaths caused by Child stunting is more dependent on Diet low in nuts and seeds. For one unit increase of deaths due to Diet low in nuts and seeds there will be 0.0702 units of increase in deaths caused by Child stunting.