

# REPORT ON DEATH MORTALITY RATE OF WORLD

## ABSTRACT

Mortality defines 'to be in the state of being prone to death'. The mortality rate is the measure of the frequency of occurrence of death in a defined population during a specified interval. There are many reasons for the occurrences of death and analysis of the reasons is very important for a better survival of human. In this report death mortality rate of the world from 1990 to 2017 analysis is done on various factors which are the causes of death and how those factors are inter-related to each other.

## DATA COLLECTION

The analysis is based on a dataset collected from [www.kaggle.com](http://www.kaggle.com). The dataset provides information about the world death analysis from 1990 to 2017. It consists of the number of deaths caused by different factors.

## APPROACH

- ❖ The various python libraries such as NumPy, pandas, Matplotlib are used for the purpose of mathematical calculations, extraction of data and visualization respectively.
- ❖ 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

Dataset contains the data as follows:

	Entity	Code	...	Diet low in whole grains	Diet low in nuts and seeds
0	Afghanistan	AFG	...	11381.377350	7299.867330
1	Afghanistan	AFG	...	11487.832390	7386.764303
2	Afghanistan	AFG	...	11866.235570	7640.628526
3	Afghanistan	AFG	...	12335.961680	7968.311853
4	Afghanistan	AFG	...	12672.950190	8244.368430
...	...	...	...	...	...
6463	Zimbabwe	ZWE	...	2687.636261	2409.930182
6464	Zimbabwe	ZWE	...	2654.381923	2399.261581
6465	Zimbabwe	ZWE	...	2635.950107	2398.525219
6466	Zimbabwe	ZWE	...	2641.376815	2417.422521
6467	Zimbabwe	ZWE	...	2664.132572	2449.546229

[6468 rows x 32 columns]

The dataset is found to have null values for the columns 'High total cholesterol' and 'Outdoor air pollution'.

```

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

```

The null values are been handled to achieve accurate analysis rather than removing the columns from the dataset. The null values are been replaced by calculating the mean of the values which are present in that specific column.

```

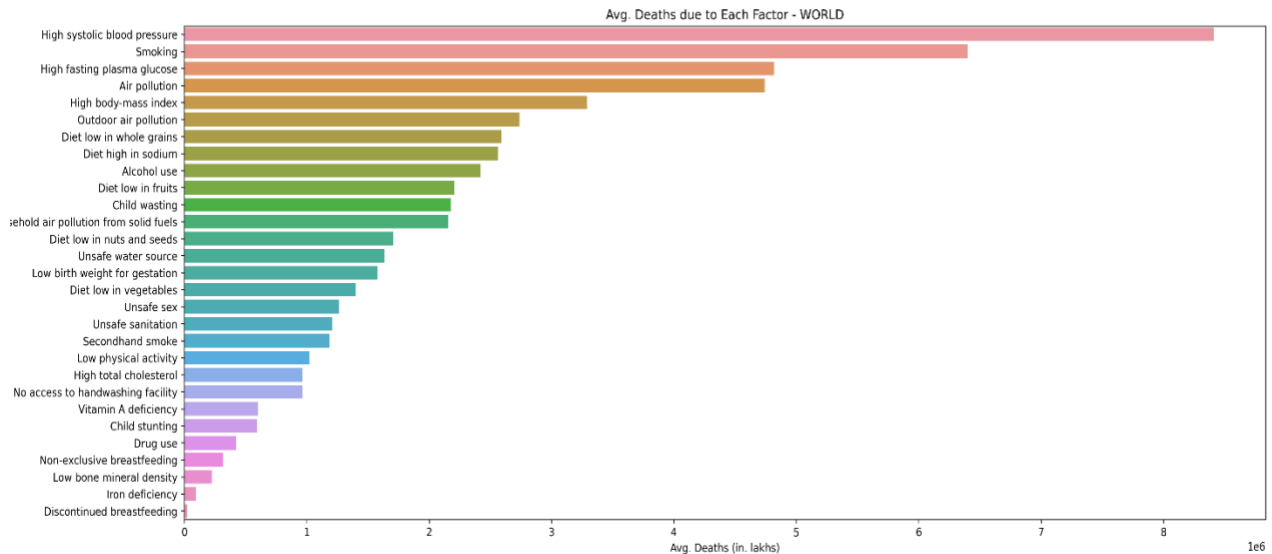
Year 0
Unsafe water source 0
Unsafe sanitation 0
No access to handwashing facility 0
Household air pollution from solid fuels 0
Non-exclusive breastfeeding 0
Discontinued breastfeeding 0
Child wasting 0
Child stunting 0
Low birth weight for gestation 0
Secondhand smoke 0
Alcohol use 0
Drug use 0
Diet low in fruits 0
Diet low in vegetables 0
Unsafe sex 0
Low physical activity 0
High fasting plasma glucose 0
High total cholesterol 0
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 0
Diet high in sodium 0
Diet low in whole grains 0
Diet low in nuts and seeds 0
dtype: int64

```

## 1. Average Deaths caused due to each factor in world.

### PURPOSE

→ To find out the average deaths caused by each factor in the entire world.



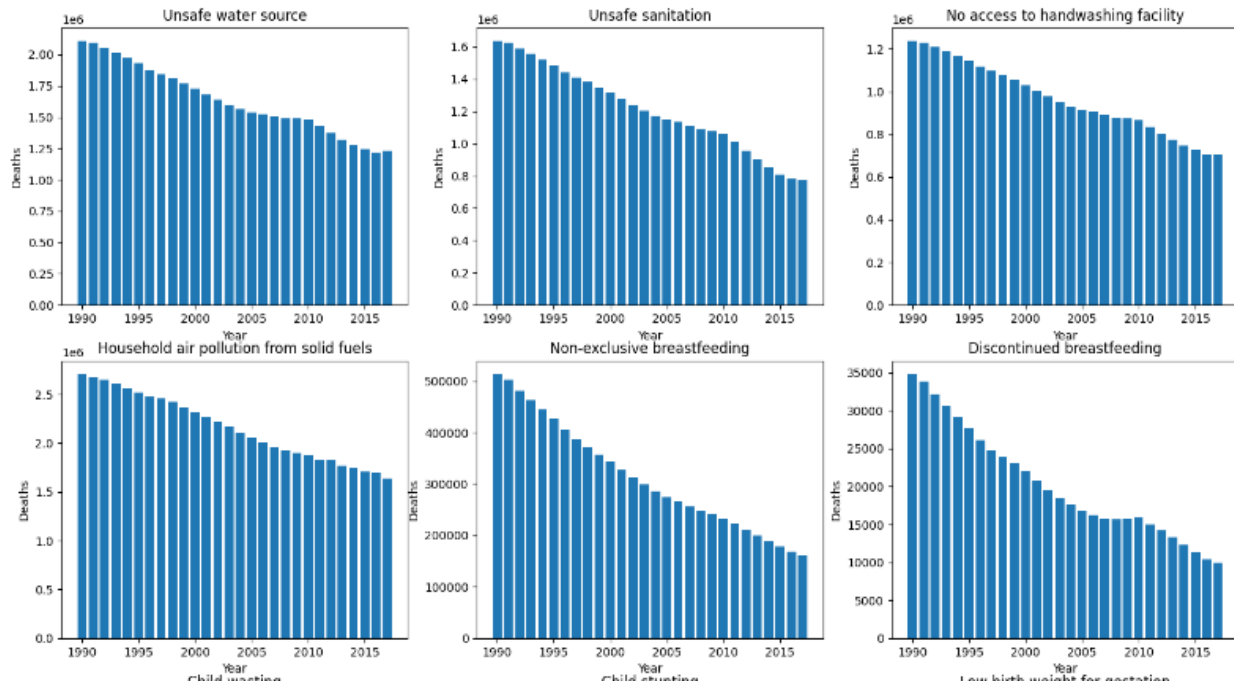
### INFERENCE

→ Most of the deaths in the world are mainly caused due to High Systolic blood pressure and least number of deaths are caused by discontinued breastfeeding.

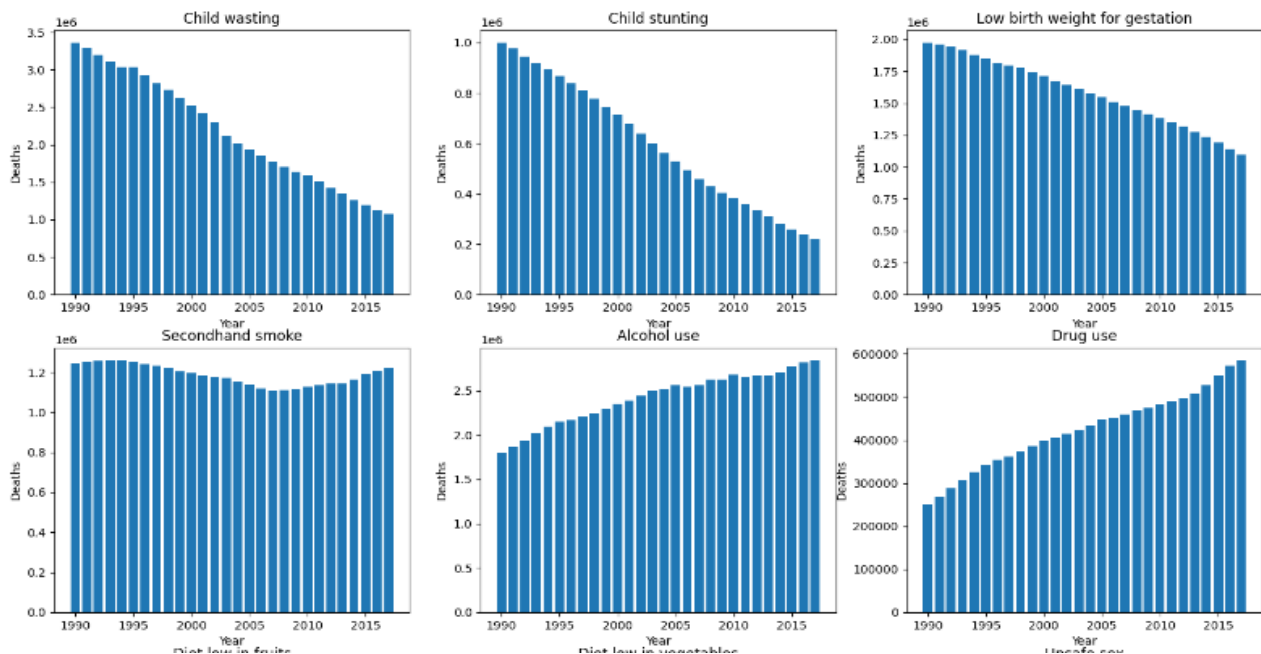
## 2. Number of deaths caused by each factor from 1990 to 2017.

### PURPOSE

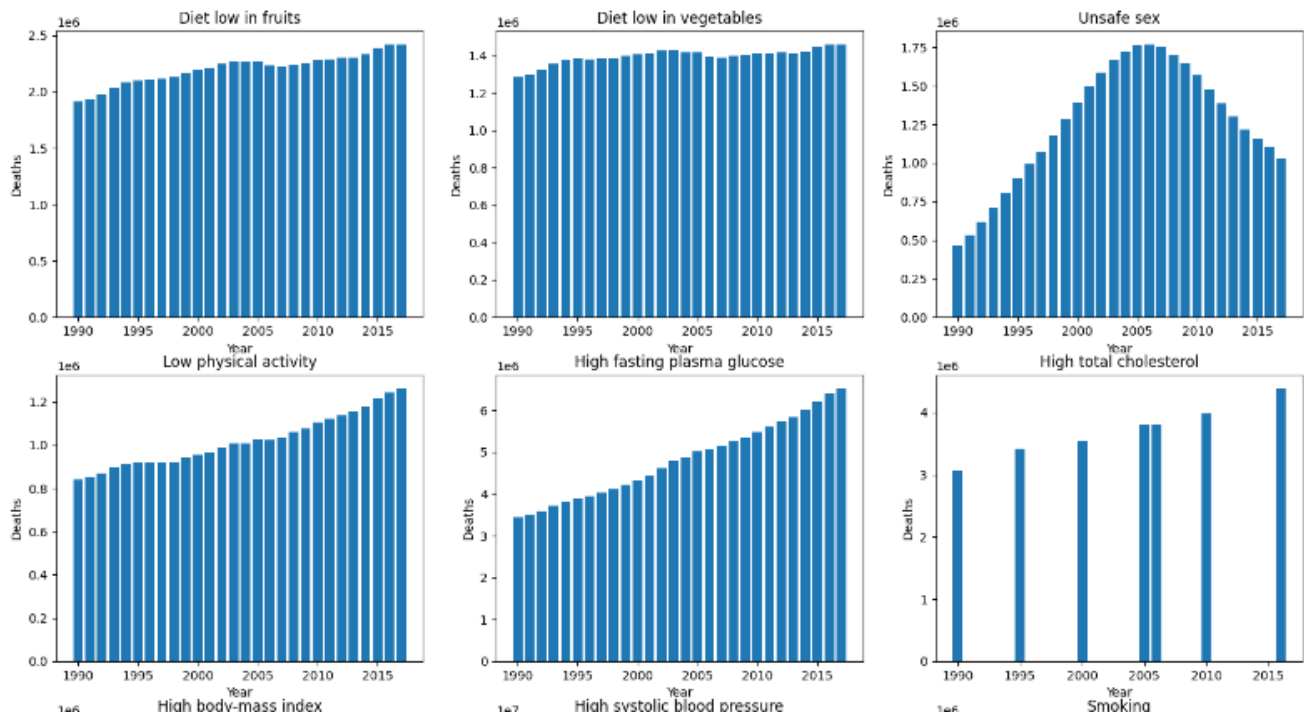
→ To find whether the total number of deaths caused due to various factors in the entire world are increasing or decreasing from 1990 to 2017.



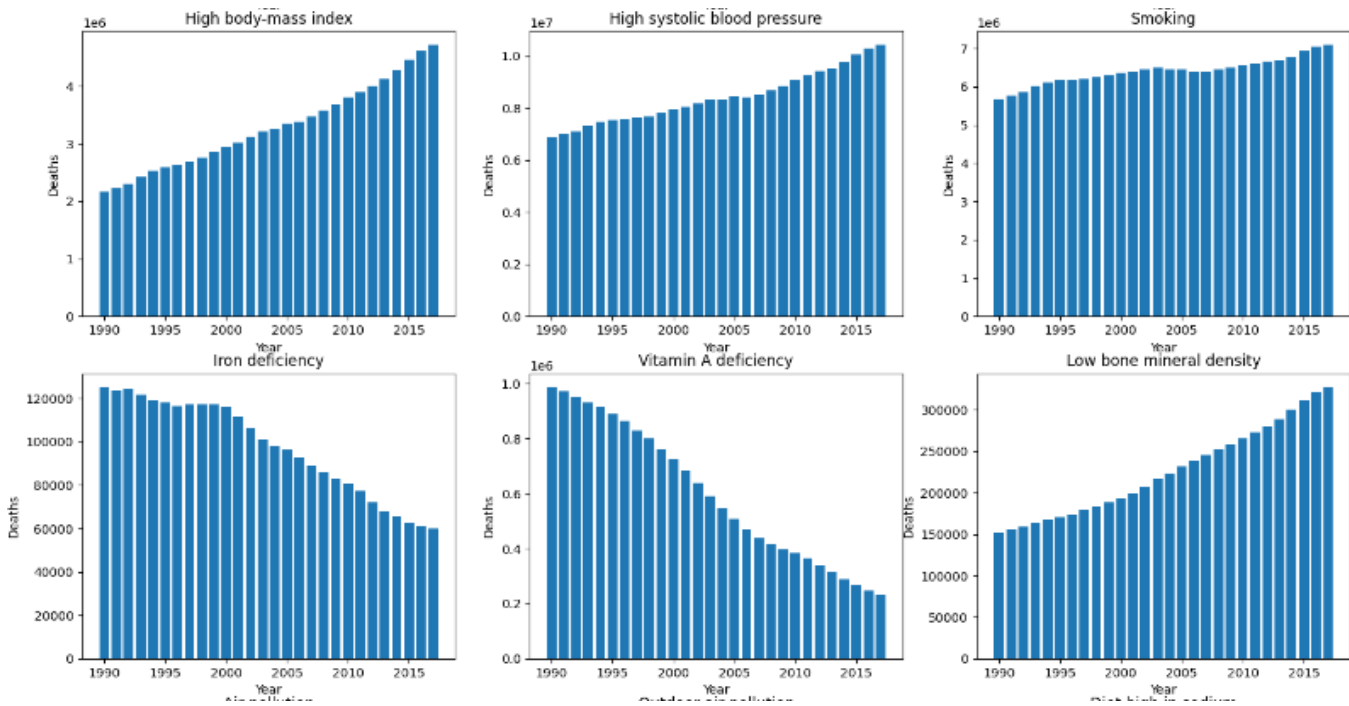
→ The number of deaths caused by Unsafe Water Source, Unsafe Sanitation, no access to handwashing facility, Household air pollution from solid fuels, Non-exclusive breastfeeding and Discontinued breastfeeding is been decreasing almost linearly from 1990 to 2017 due to many advancements in medical field.



- The number of deaths caused by Child wasting, Child Stunting and low birth for gestation is decreasing linearly from 1990 to 2017.
- The number of deaths caused due to Secondhand Smoke is almost similar from 1990 to 2017.
- The number of deaths caused due to Alcohol use and Drug usage is been increasing linearly.

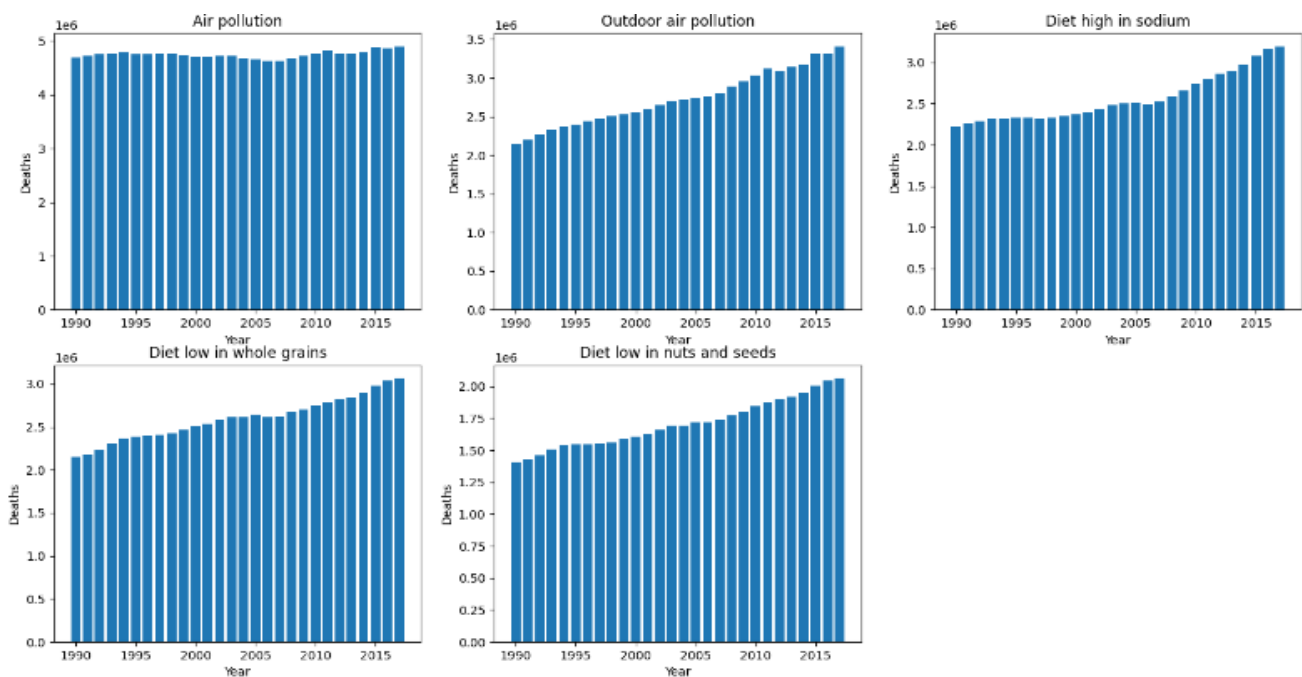


- The number of deaths caused by diet low in fruits, diet low in vegetables, low physical activity, High fasting plasma glucose and high total cholesterol is been increasing linearly from 1990 to 2017.
- The number of deaths caused by Unsafe sex has increased from 1990 to 2005 and from 2005 it has been decreasing.



→ The number of deaths caused by high body mass index, high systolic blood pressure, smoking and low bone mineral density is been increasing from 1990 to 2017.

→ The number of deaths caused by Iron deficiency; Vitamin A deficiency is been decreasing from 1990 to 2017.



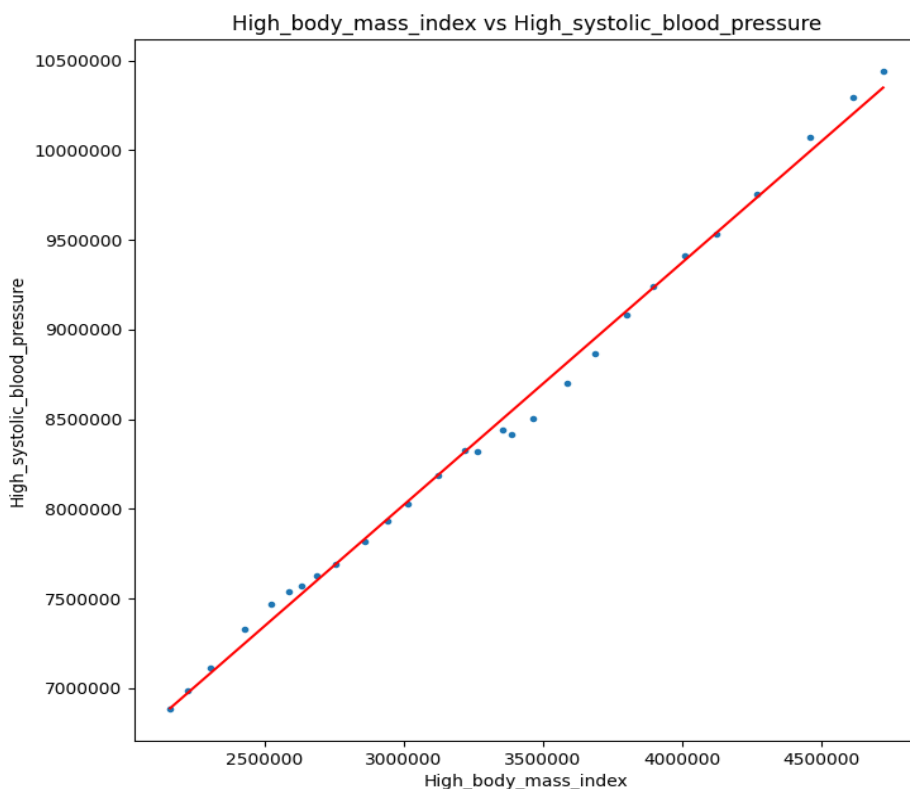
→ The number of deaths caused due to Air Pollution is almost similar from 1990 to 2017 as the number of people and as well as the pollution is increasing in the same frequency.

→ The number of deaths caused due to Outdoor air pollution, Diet high in sodium, Diet low in whole grains and diet low in nuts and seeds is been increasing linearly from 1990 to 2017.

### 3. Analysis on High body mass index and High Systolic Blood Pressure.

#### PURPOSE

→ To find out the relationship between High body mass index and high systolic blood pressure.



#### INFERENCE

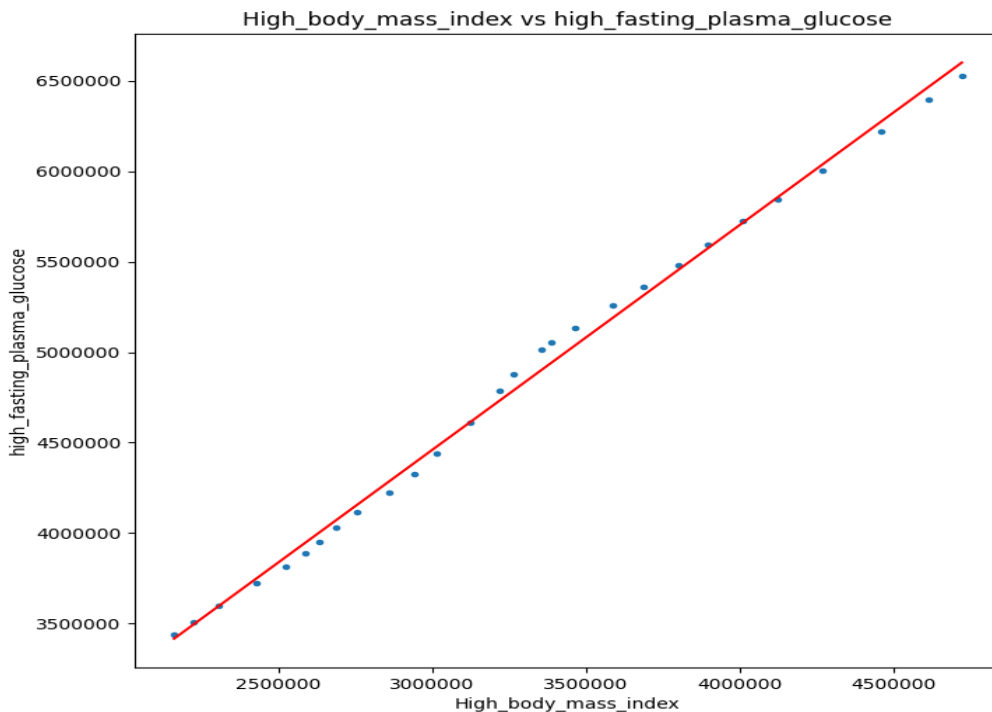
- Dependent variable: High systolic blood pressure
- Independent variable: Low physical Activity
- Linear Regression technique is used to find out how high systolic blood pressure depends on high body mass index. The deaths due to high systolic blood pressure have not increased exponentially with the increase of deaths due to high body mass index as the data is very close to the regression line.

→ The regression coefficient obtained is 1.34 which is close to 1 implies that deaths due to high systolic blood pressure can be predicted using deaths due to high body mass index.

#### 4. Analysis on High body mass index and High Fasting plasma glucose.

##### PURPOSE

→ To find out the relationship between High body mass index and high fasting plasma glucose.



##### INFERENCE

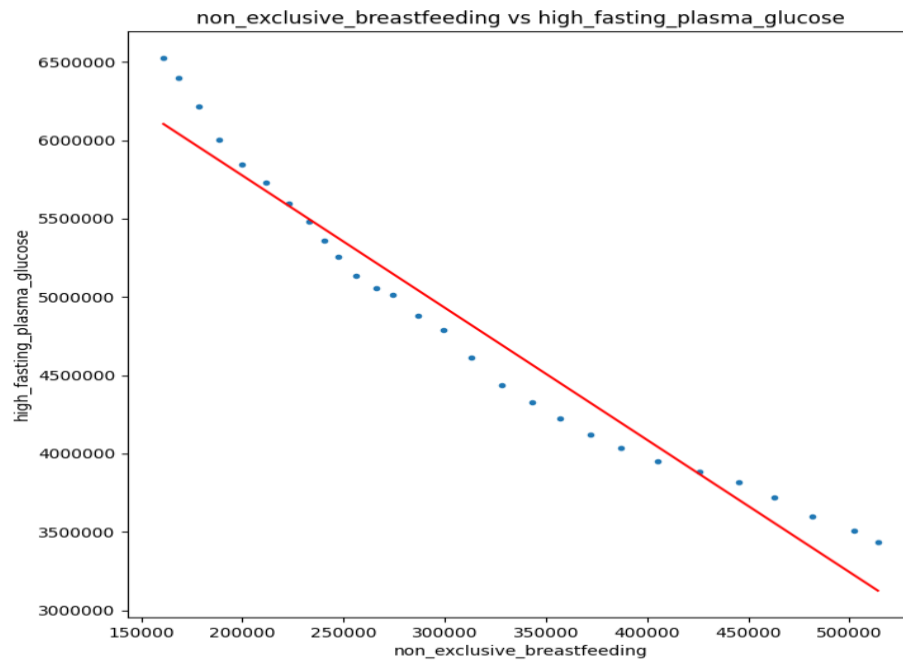
- Dependent variable: High fasting plasma glucose
- Independent variable: High body mass index
- Linear Regression technique is used to find out how high fasting plasma glucose depends on high body mass index. The deaths due to high fasting plasma glucose have not increased exponentially with the increase of deaths due to high body mass index as the data is very close to the regression line.
- The regression coefficient obtained is 1.24 which is close to 1 implies that deaths due to high fasting plasma glucose can be predicted using deaths due to high body mass index.



## 5. Analysis on Non-exclusive breastfeeding and High Fasting plasma glucose.

### PURPOSE

→ To find out the relationship between non-exclusive breastfeeding and high fasting plasma glucose.



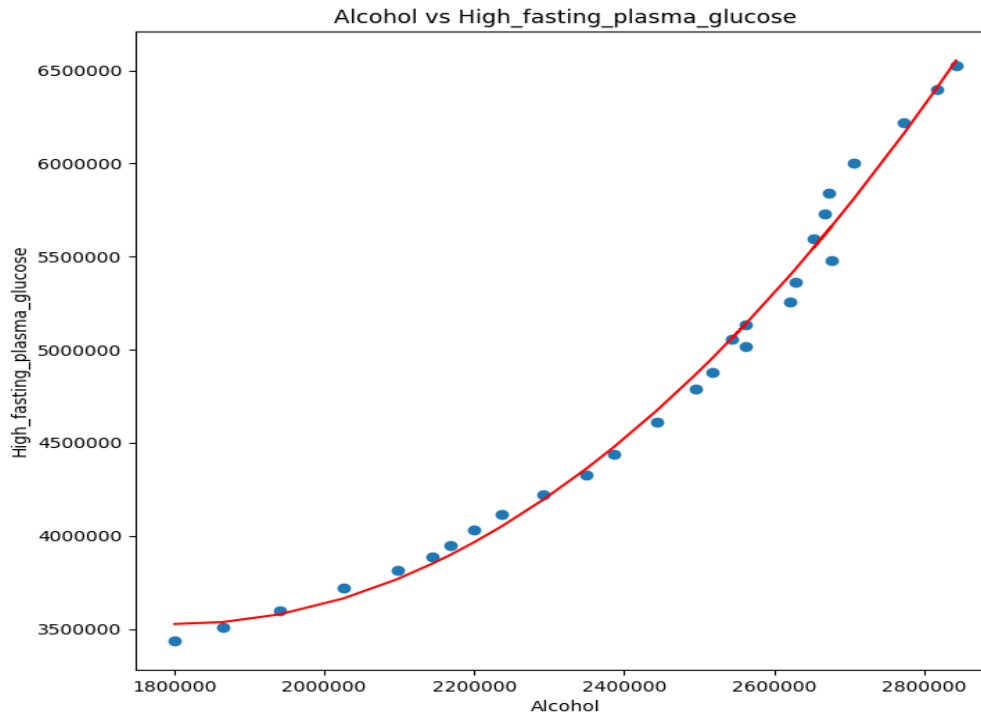
### INFERENCE

- Dependent variable: High fasting plasma glucose
- Independent variable: non-exclusive breastfeeding
- Linear Regression technique is used to find out how high fasting plasma glucose depends on high body mass index. The deaths due to high fasting plasma glucose have not decreased exponentially with the increase of deaths due to non-exclusive breastfeeding as the data is very close to the regression line.
- The regression coefficient obtained is  $-8.4$  implies that deaths due to high fasting plasma glucose can be predicted using deaths due to non-exclusive breastfeeding..

## 6. Analysis on Alcohol and High fasting plasma glucose.

### PURPOSE

→ To find the relationship between Alcohol and High fasting plasma glucose.



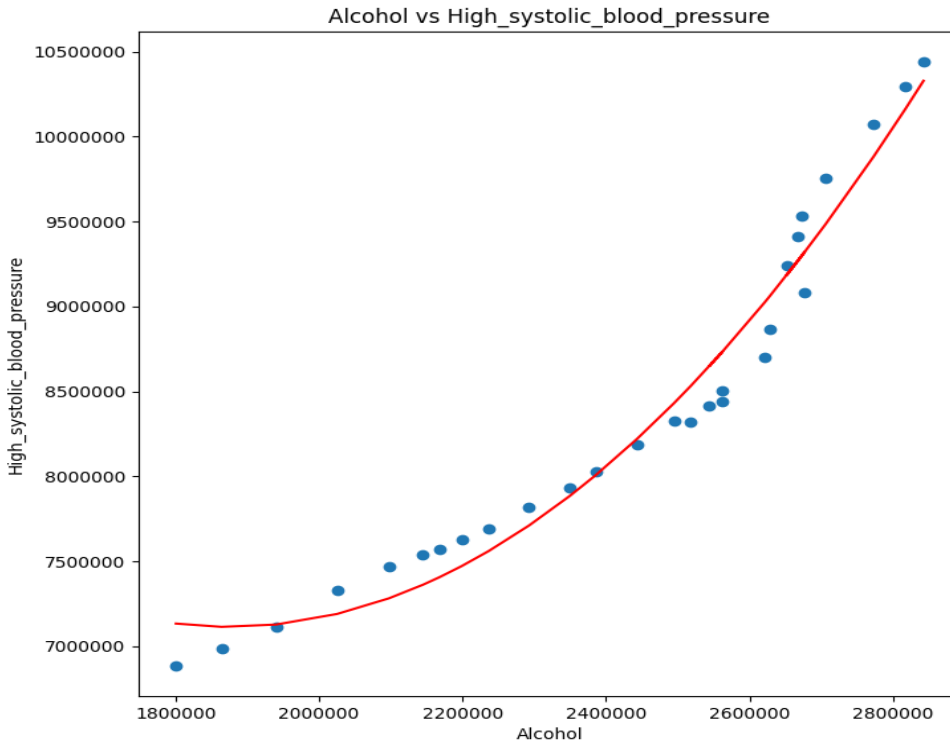
### INFERENCE

- Dependent variable: High fasting plasma glucose
- Independent variable: Alcohol
- Polynomial regression technique is applied on number of deaths caused due to alcohol and number of deaths caused due to high fasting blood glucose.
- The  $r^2$  score obtained is 0.9926 which is almost close to 1 that implies that deaths caused due to high fasting plasma glucose is depending on deaths caused due to alcohol.

## 7. Analysis on Alcohol and High systolic blood pressure.

### PURPOSE

→ To find the relationship between Alcohol and High systolic blood pressure.



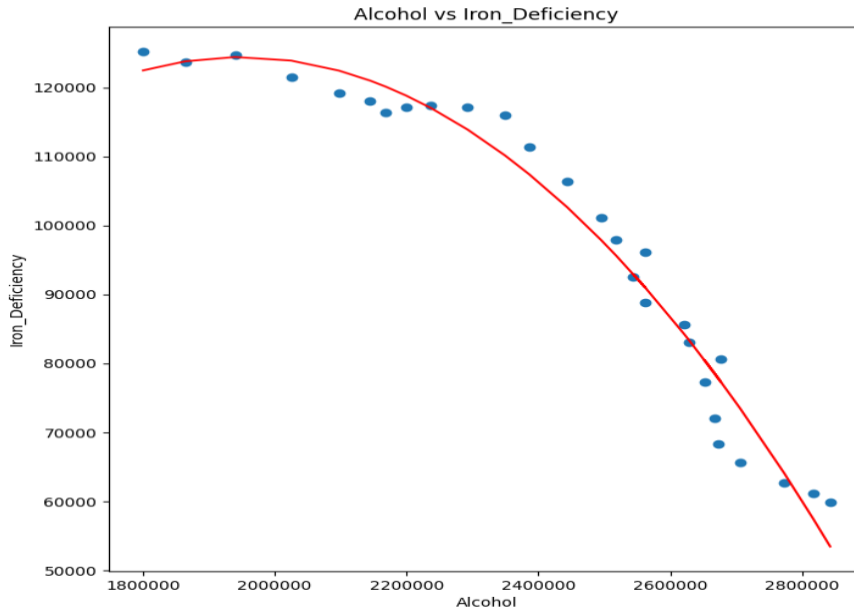
### INFERENCE

- Dependent variable: High systolic blood pressure
- Independent variable: Alcohol
- Polynomial regression technique is applied on number of deaths caused due to alcohol and number of deaths caused due to High systolic blood pressure.
- The  $r^2$  score obtained is 0.9832 which is almost close to 1 that implies that deaths caused due to high systolic blood pressure is depending on deaths caused due to alcohol.

## 8. Analysis on Alcohol and Iron deficiency.

### PURPOSE

→ To find the relationship between Alcohol and Iron deficiency.



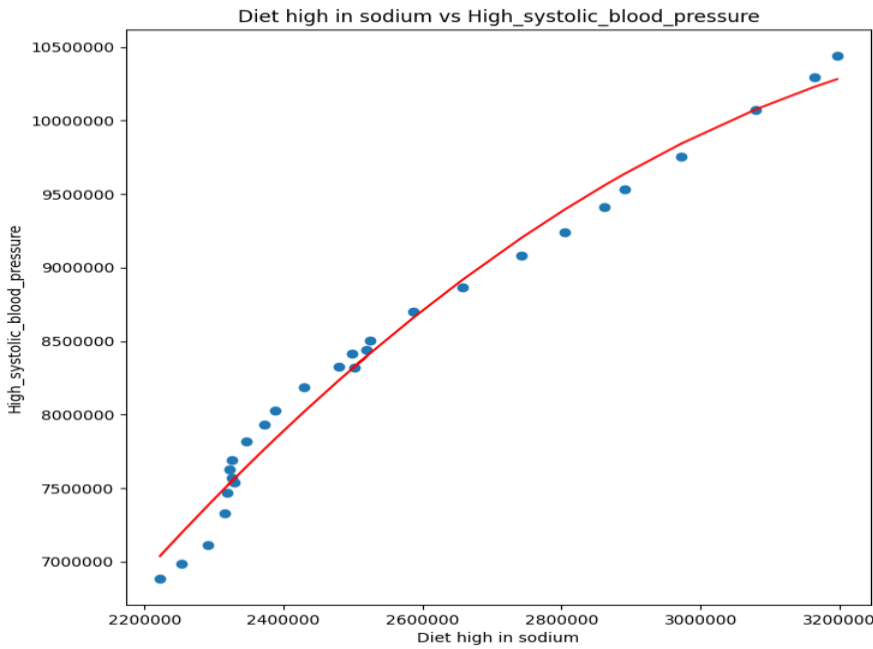
### INFERENCE

- Dependent variable: Iron deficiency
- Independent variable: Alcohol
- Polynomial regression technique is applied on number of deaths caused due to alcohol and number of deaths caused due to Iron deficiency.
- The  $r^2$  score obtained is 0.9806 which is almost close to 1 that implies that deaths caused due to Iron deficiency is depending on deaths caused due to alcohol.

## 9. Analysis on Diet High in Sodium and High systolic blood pressure.

### PURPOSE

→ To find the relationship between Deaths caused by diet high in sodium and high systolic blood pressure.



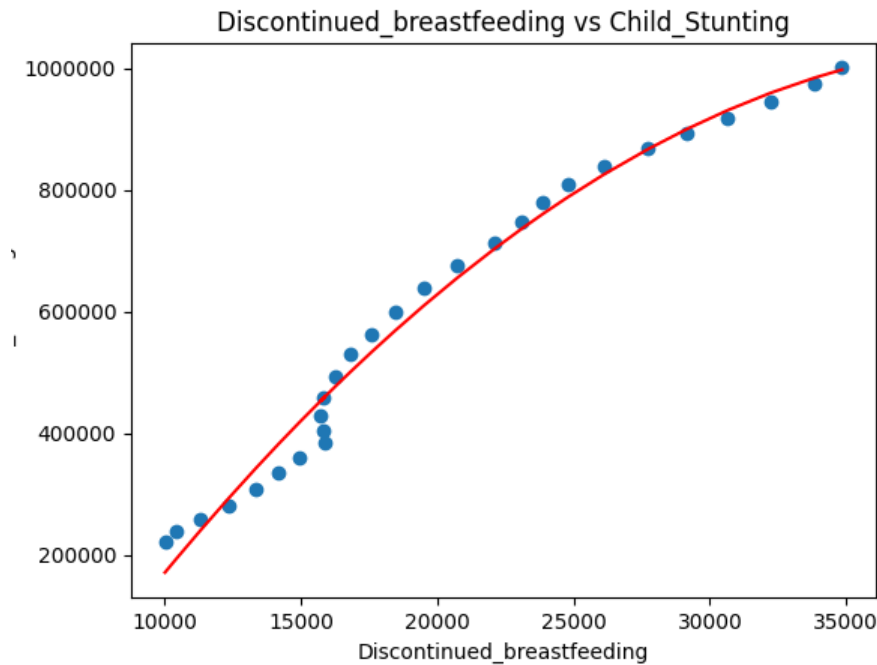
### INFERENCE

- Dependent variable: High systolic blood pressure.
- Independent variable: Diet high in Sodium
- Polynomial regression technique is applied on number of deaths caused due to Diet high in Sodium and number of deaths caused due to High systolic blood pressure.
- The  $r^2$  score obtained is 0.9937 which is almost close to 1 that implies that deaths caused due to High systolic blood pressure is depending on deaths caused due to diet high in sodium.

## 10. Analysis on deaths caused by Discontinued breastfeeding and Child Stunting.

### PURPOSE

→ To find the relationship between Deaths caused by discontinued breastfeeding and child stunting.



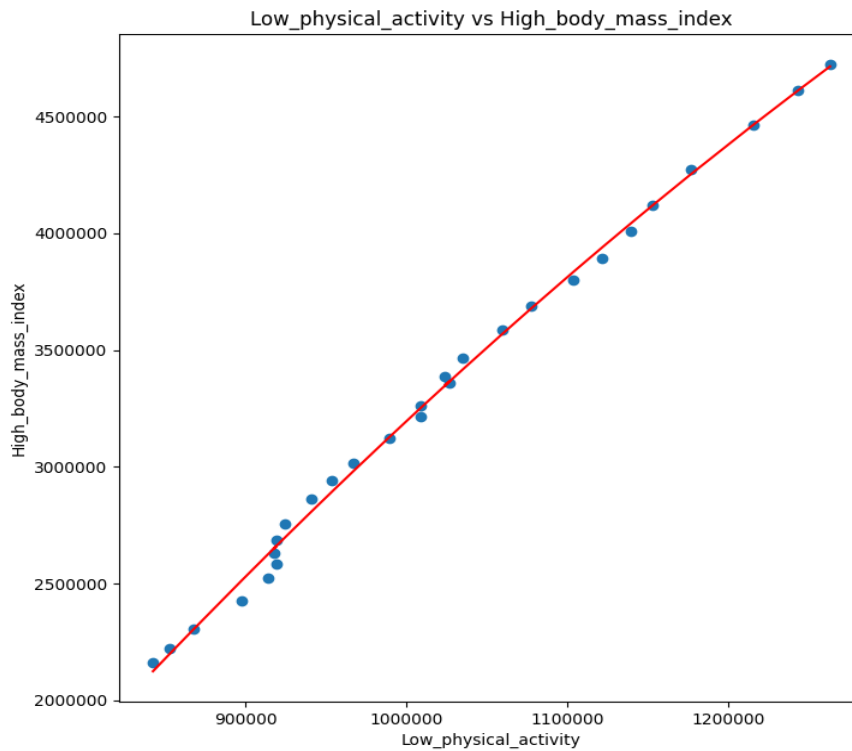
### INFERENCE

- Dependent variable: Child Stunting
- Independent variable: Discontinued breastfeeding
- Polynomial regression technique is applied on number of deaths caused due to Discontinued breastfeeding and number of deaths caused due to Child Stunting.
- The  $r^2$  score obtained is 0.9926 which is almost close to 1 that implies that deaths caused due to Child Stunting is depending on deaths caused due to Discontinued breastfeeding.

## 11. Analysis on deaths caused by Low Physical activity and high body mass index.

### PURPOSE

→ To find the relationship between Deaths caused by low physical activity and high body mass index.



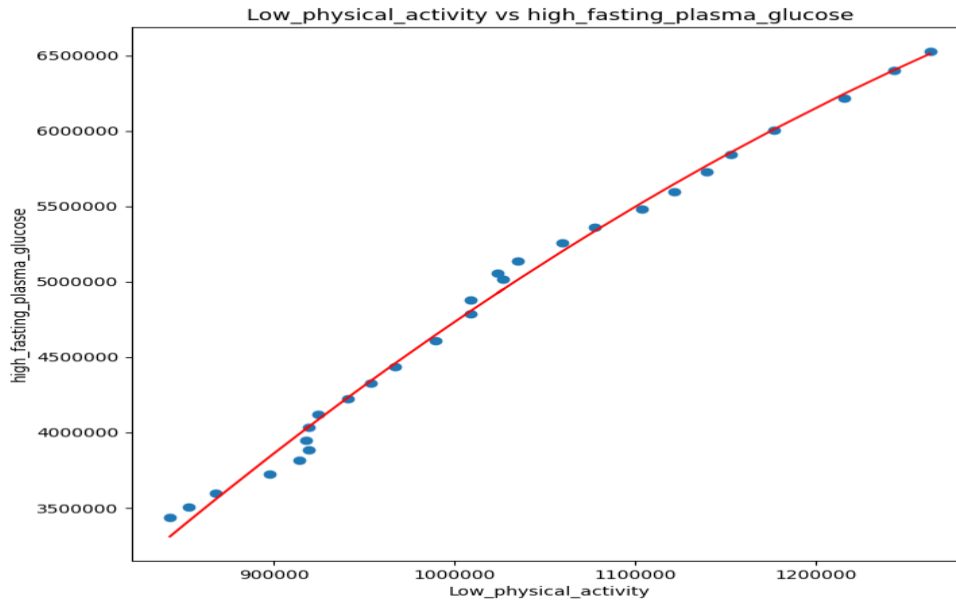
### INFERENCE

- Dependent variable: High body mass index
- Independent variable: Low physical Activity
- Polynomial regression technique is applied on number of deaths caused due to High body mass index and number of deaths caused due to low physical activity.
- The  $r^2$  score obtained is 0.9971 which is almost close to 1 that implies that deaths caused due to high body mass index is depending on deaths caused due to low physical activity.

## 12. Analysis on deaths caused by Low Physical activity and high body mass index.

### PURPOSE

To find the relationship between Deaths caused by low physical activity and high body mass index.



### INFERENCE

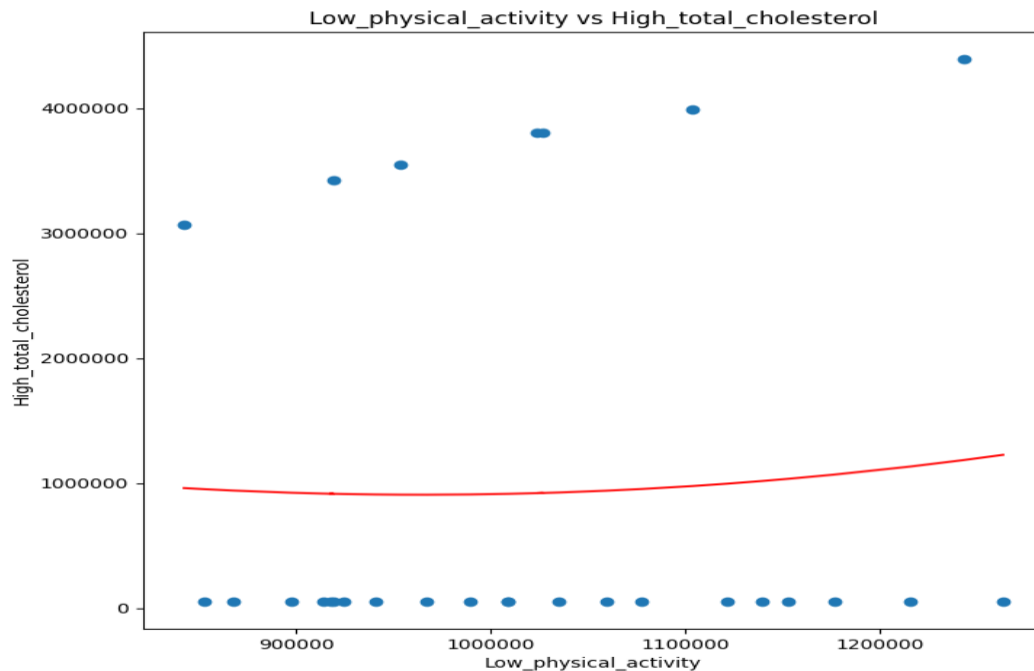
- Dependent variable: High fasting plasma glucose
- Independent variable: Low physical Activity
- Polynomial regression technique is applied on number of deaths caused due to High fasting plasma glucose and number of deaths caused due to low physical activity.
- The  $r^2$  score obtained is 0.9966 which is almost close to 1 that implies that deaths caused due to High fasting plasma glucose is depending on deaths caused due to low physical activity.



### 13. Analysis on deaths caused by Low Physical activity and high total cholesterol.

#### PURPOSE

→ To find the relationship between Deaths caused by low physical activity and high total cholesterol.



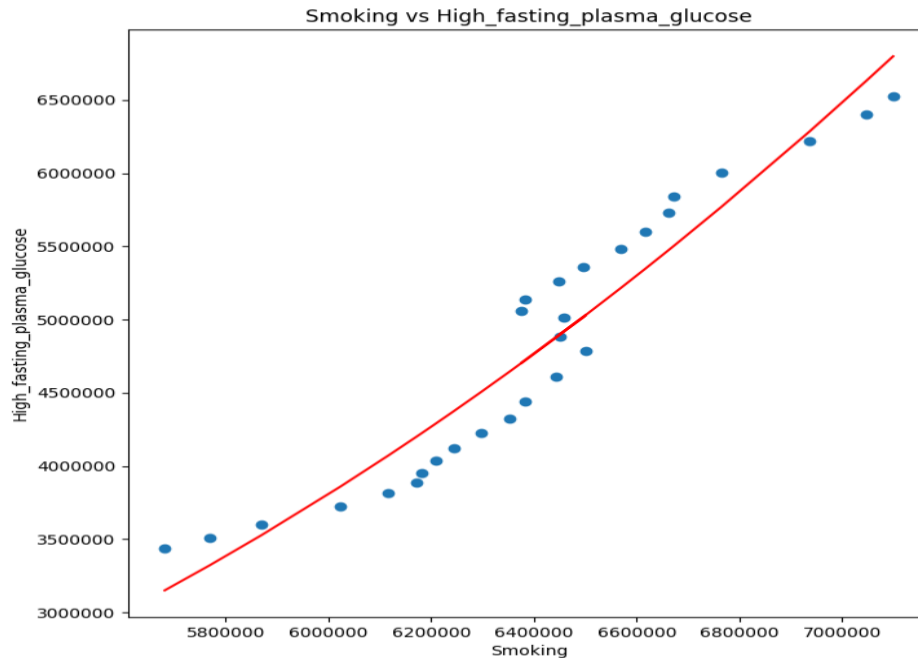
#### INFERENCE

- Dependent variable: High total cholesterol
- Independent variable: Low physical Activity
- Polynomial regression technique is applied on number of deaths caused due to low physical activity and number of deaths caused due to High total cholesterol.
- The  $r^2$  score obtained is 0.04 close to 0 which implies that there is no appreciable relation. deaths caused due to High total cholesterol is not depending on deaths caused due to low physical activity.

## 14. Analysis on deaths caused by Smoking and High fasting plasma glucose

### PURPOSE

→ To find the relationship between Deaths caused by Smoking and high fasting plasma glucose.



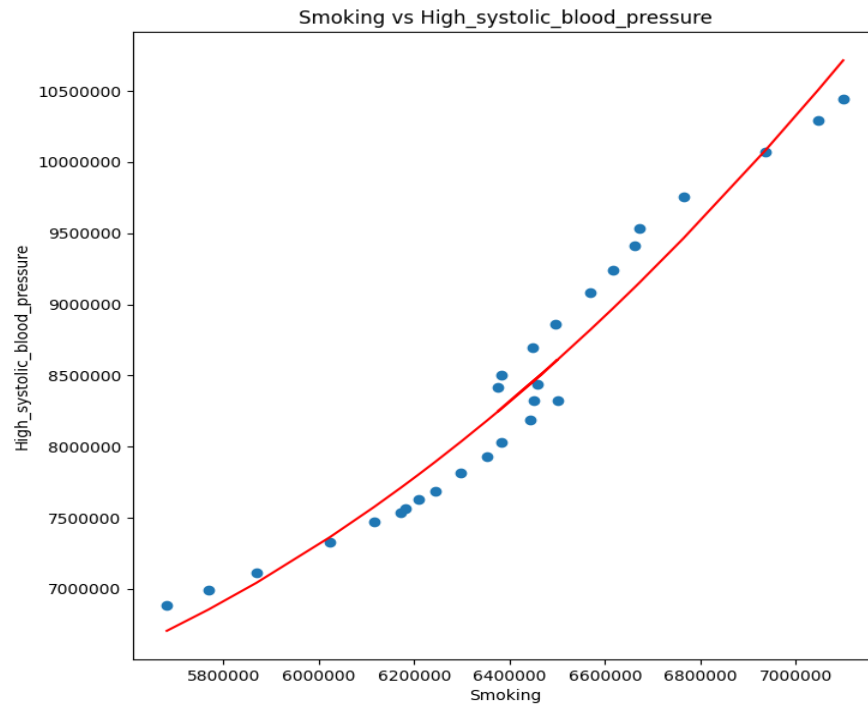
### INFERENCE

- Dependent variable: High fasting plasma glucose
- Independent variable: Smoking
- Polynomial regression technique is applied on number of deaths caused due to smoking and number of deaths caused due to High fasting plasma glucose.
- The  $r^2$  score obtained 0.955 which is close to 1 which implies that there is appreciable relation. Deaths caused due to High fasting plasma glucose is depending on deaths caused due to smoking.

## 15. Analysis on deaths caused by Smoking and High systolic blood pressure.

### PURPOSE

→ To find the relationship between Deaths caused by Smoking and High systolic blood pressure



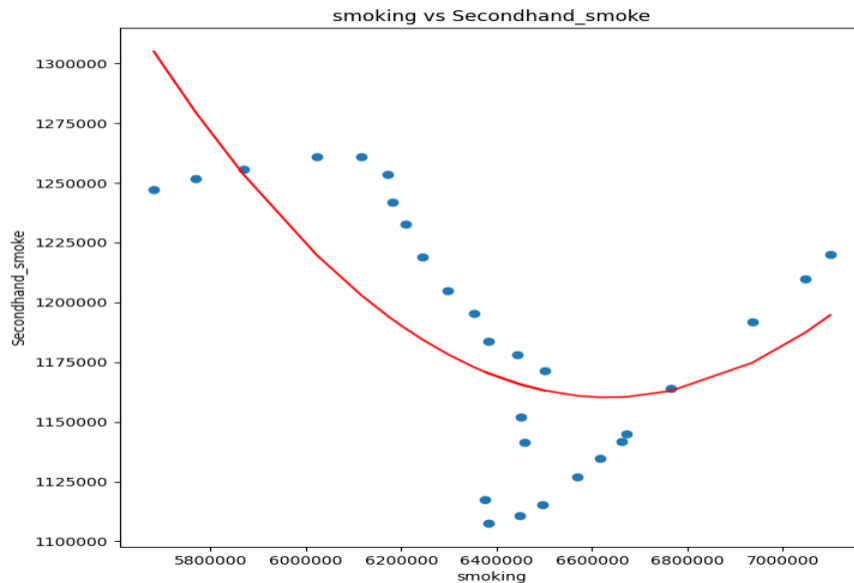
### INFERENCE

- Dependent variable: High systolic blood pressure.
- Independent variable: Smoking
- Polynomial regression technique is applied on number of deaths caused due to smoking and number of deaths caused due to High systolic blood pressure.
- The  $r^2$  score obtained is 0.9768 which is close to 1 which implies that there is appreciable relation. Deaths caused due to High systolic blood pressure is depending on deaths caused due to smoking.

## 16. Analysis on deaths caused by Smoking and Secondhand smoking.

### PURPOSE

→ To find the relationship between Deaths caused by Smoking and secondhand smoking.



### INFERENCE

- Dependent variable: Secondhand smoking
- Independent variable: Smoking
- Polynomial regression technique is applied on number of deaths caused due to smoking and number of deaths caused due to Secondhand smoking.
- The  $r^2$  score obtained is 0.746 which is close to 1 which implies that there is appreciable relation. Deaths caused due to Secondhand smoking is depending on deaths caused due to smoking.

## 17. Analysis on Unsafe Water Source, No access to handwashing and Unsafe Sanitation.

### PURPOSE

→ To find the relationship between unsafe water source, no access to handwashing and unsafe sanitation.

### INFERENCE

- Dependent Variable: Unsafe Sanitation
- Independent Variable: Unsafe Water Source, No access to handwashing

- The coefficient of multiple regression is [0.62798671 0.20555965].
- Deaths caused by Unsafe Sanitation is more dependent on Unsafe Water Source. For one unit increase of deaths due to Unsafe water source there will be 0.67 units of increase in deaths caused by Unsafe sanitation.
- Deaths caused by Unsafe Sanitation is less dependent on no access to handwashing. For one unit increase of deaths due to Unsafe water source there will be 0.20 units of increase in deaths caused by Unsafe sanitation.

## **18. Analysis on Household air pollution from solid fuels, Outdoor air pollution and Air Pollution.**

### **PURPOSE**

- To find the relationship between Household air pollution from solid fuels, Outdoor air pollution and Air Pollution

### **INFERENCE**

- Dependent Variable: Air Pollution
- Independent Variable: Household air pollution from solid fuels, Outdoor air pollution.
- The coefficient of multiple regression is [0.9699 0.9609]
- Deaths caused by Air Pollution is more 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 0.9699 units of increase in deaths caused by Air Pollution.
- Deaths caused by Air pollution is also more dependent on Outdoor air pollution. For one unit increase of deaths due to Outdoor air pollution there will be 0.20 units of increase in deaths caused by Air Pollution.

## **19. Analysis on Alcohol use, Smoking and high systolic blood pressure.**

### **PURPOSE**

To find the relationship between Alcohol use, Smoking and high systolic blood pressure using multiple regression.

### **INFERENCE**

- Dependent Variable: High systolic blood pressure
- Independent Variable: Alcohol use and Smoking.
- The coefficient of multiple regression is [1.7269 0.6494]
- Deaths caused by High systolic blood pressure is more dependent on Alcohol use. For one unit increase of deaths due to Alcohol use there will be 1.7269 units of increase in deaths caused by High systolic blood pressure.

- Deaths caused by High systolic blood pressure is less dependent on Smoking. For one unit increase of deaths due to Smoking there will be 0.6494 units of increase in deaths caused by High systolic blood pressure.

## **20. Analysis on Diet low in fruits, Diet low in vegetables and Vitamin A deficiency.**

### **PURPOSE**

To find the relationship between Diet low in fruits, Diet low in vegetables and Vitamin A deficiency using multiple regression.

### **INFERENCE**

- Dependent Variable: Vitamin A deficiency
- Independent Variable: Diet low in fruits, Diet low in vegetables.
- The coefficient of multiple regression is [-0.6441 1.4091]
- Deaths caused by Vitamin A deficiency is more dependent on Diet Low in vegetables. For one unit increase of deaths due to Diet low in vegetables there will be 1.4091 units of increase in deaths caused by Vitamin A deficiency.
- Deaths caused by Vitamin A deficiency is less dependent on Diet Low in fruits. For one unit increase of deaths due to Diet low in fruits there will be 0.6441 units of increase in deaths caused by Vitamin A deficiency.

## **21. Analysis on Low physical activity, high body mass index and High fasting plasma glucose.**

### **PURPOSE**

To find the relationship between Low physical activity, high body mass index and High fasting plasma glucose using multiple regression.

### **INFERENCE**

- Dependent Variable: High fasting plasma glucose
- Independent Variable: Low physical activity, high body mass index
- The coefficient of multiple regression is [1.5823 0.9308]
- Deaths caused by High fasting plasma glucose is more dependent on Low physical activity. For one unit increase of deaths due to Low physical activity there will be 1.5823 units of increase in deaths caused by High fasting plasma glucose.
- Deaths caused by High fasting plasma glucose is less dependent on high body mass index. For one unit increase of deaths due to high body mass index there will be 0.9308 units of increase in deaths caused by High fasting plasma glucose.

## **22. Analysis on Low physical activity, high body mass index and High total cholesterol.**

### **PURPOSE**

To find the relationship between Low physical activity, high body mass index and High total cholesterol using multiple regression.

### **INFERENCE**

- Dependent Variable: High total cholesterol
- Independent Variable: Low physical activity, high body mass index
- The coefficient of multiple regression is [0.9898 -0.0911]
- Deaths caused by High total cholesterol is more dependent on Low physical activity.  
For one unit increase of deaths due to Low physical activity there will be 0.9898 units of increase in deaths caused by High total cholesterol.
- Deaths caused by High total cholesterol is less dependent on high body mass index.  
For one unit increase of deaths due to high body mass index there will be 0.0911 units of decrease in deaths caused by High total cholesterol.