PIMA Indian Diabetes Analysis

Divyanshi Kumari MDS202322 divyanshi.mds2023@cmi.ac.in

2023-09-30

Introduction

Type-2 Diabetes in women is a chronic metabolic condition characterised by elevated blood sugar levels resulting from insulin resistance and insufficient production of insulin from pancreas. The PIMA Indian Population particularly, the PIMA Indians of Arizona has one of the highest reported prevalence rates of type-2 diabetes in the world. The relationship between PIMA Indian Women and diabetes has been a focus of research due to this alarming rates of diabetes within this community.

Dataset Information

The PIMA Indian Diabetes dataset is named after the PIMA people, a group of Native Americans living in United States, particularly Arizona.The dataset is focused on studying and understanding risk factors and characteristics associated with the development of type-2 diabetes among PIMA women.The dataset comprises of various health related features:Pregnancies, Glucose levels, Blood Pressure, Age etc. The primary objective herein is to develop a binary classification model that can classify individuals in two categories: With Diabetes(Outcome:1) and Without Diabetes(Outcome:0).The dataset comprises values obtained exclusively from female individuals who are above the age of 21.

Variable Descriptions

- **Age:** Age of the individual.
- **Pregnancies:** Number of times Pregnant
- **Glucose:** Plasma glucose concentration a 2 hours in an oral glucose tolerance test
- **Blood Pressure:** Diastolic Blood Pressure (mm Hg).
- **Skin Thickness:** Triceps skin fold thickness (mm).
- **Insulin:** 2-Hour serum insulin (mu U/ml).
- **BMI:** Body Mass Index.
- **Diabetes Pedigree Function:**a measure of the diabetes heredity risk
- **Outcome:** 0: NO Diabetes, 1: Diabetes

Exploratory Data Analysis

```
##                         Minimum
## Pregnancies               0.000
## Glucose                   0.000
## BloodPressure             0.000
## SkinThickness             0.000
```

```
## Insulin                    0.000
## BMI                        0.000
## DiabetesPedigreeFunction   0.078
## Age                       21.000
## Outcome                    0.000
```

Its not physiologically possible for certain health related features like Glucose, Blood Pressure, Skin Thickness, Insulin, BMI to have values minimum as 0.This suggests that these 0 values might be missing or improperly recorded data points.Hence, replacing with mean for the 0 values in specific columns.

```
##   Pregnancies        Glucose       BloodPressure    SkinThickness
##  Min.   : 0.000   Min.   : 44.00   Min.   : 24.00   Min.   : 7.00
##  1st Qu.: 1.000   1st Qu.: 99.75   1st Qu.: 64.00   1st Qu.:20.54
##  Median : 3.000   Median :117.00   Median : 72.00   Median :23.00
##  Mean   : 3.845   Mean   :121.68   Mean   : 72.25   Mean   :26.61
##  3rd Qu.: 6.000   3rd Qu.:140.25   3rd Qu.: 80.00   3rd Qu.:32.00
##  Max.   :17.000   Max.   :199.00   Max.   :122.00   Max.   :99.00
##     Insulin           BMI        DiabetesPedigreeFunction      Age
##  Min.   : 14.0   Min.   :18.20   Min.   :0.0780           Min.   :21.00
##  1st Qu.: 79.8   1st Qu.:27.50   1st Qu.:0.2437           1st Qu.:24.00
##  Median : 79.8   Median :32.00   Median :0.3725           Median :29.00
##  Mean   :118.7   Mean   :32.45   Mean   :0.4719           Mean   :33.24
##  3rd Qu.:127.2   3rd Qu.:36.60   3rd Qu.:0.6262           3rd Qu.:41.00
##  Max.   :846.0   Max.   :67.10   Max.   :2.4200           Max.   :81.00
##     Outcome
##  Min.   :0.000
##  1st Qu.:0.000
##  Median :0.000
##  Mean   :0.349
##  3rd Qu.:1.000
##  Max.   :1.000
```

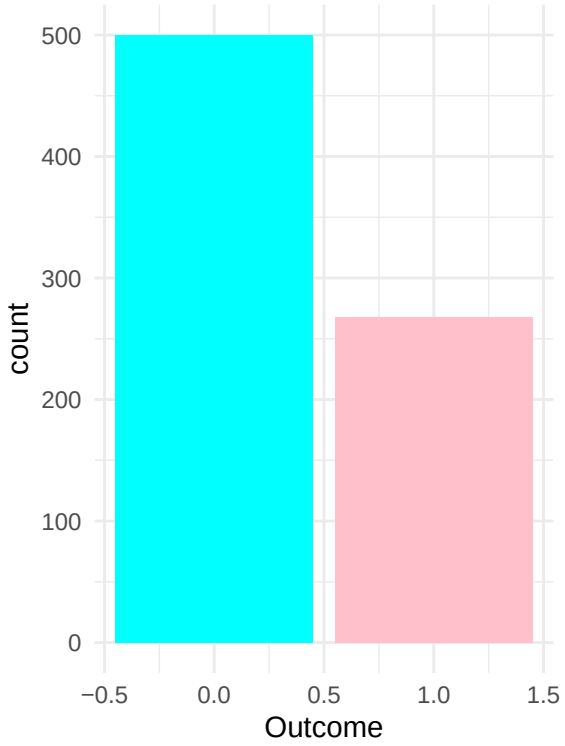

Fig 1.1: The proportion of No Diabetes is greater than Diabetes

Correlation Heatmap

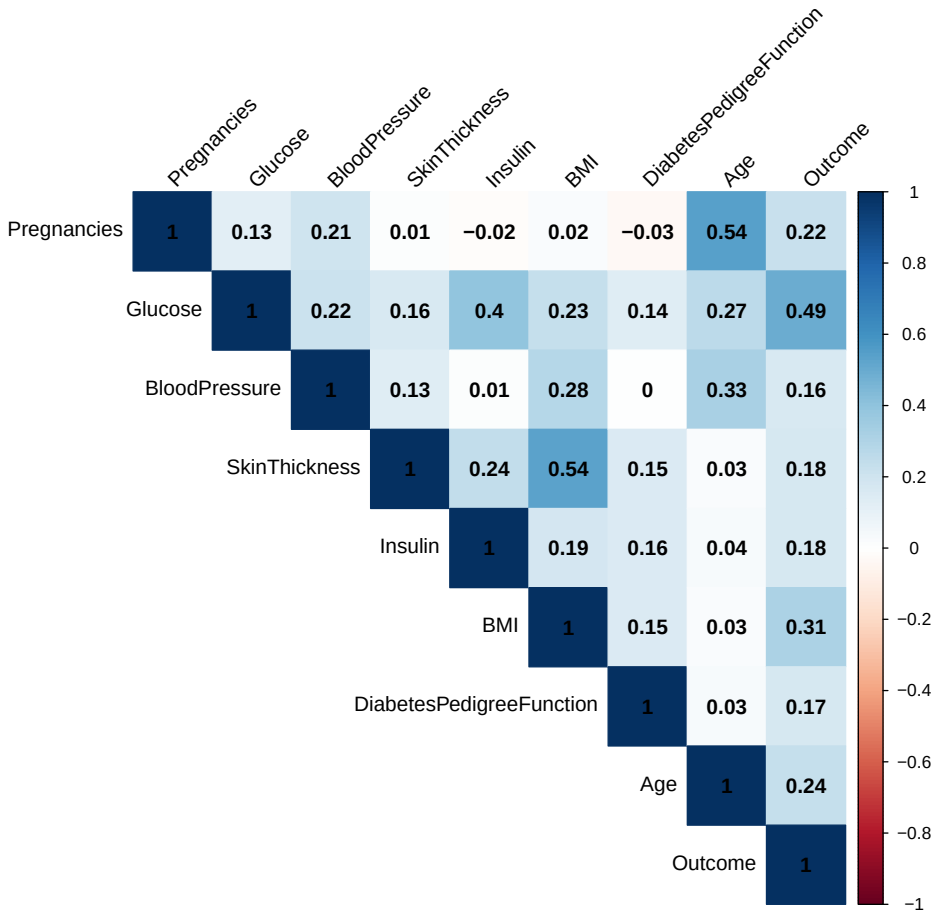

Fig 1.2: The Correlation heatmap of the variables in the dataset

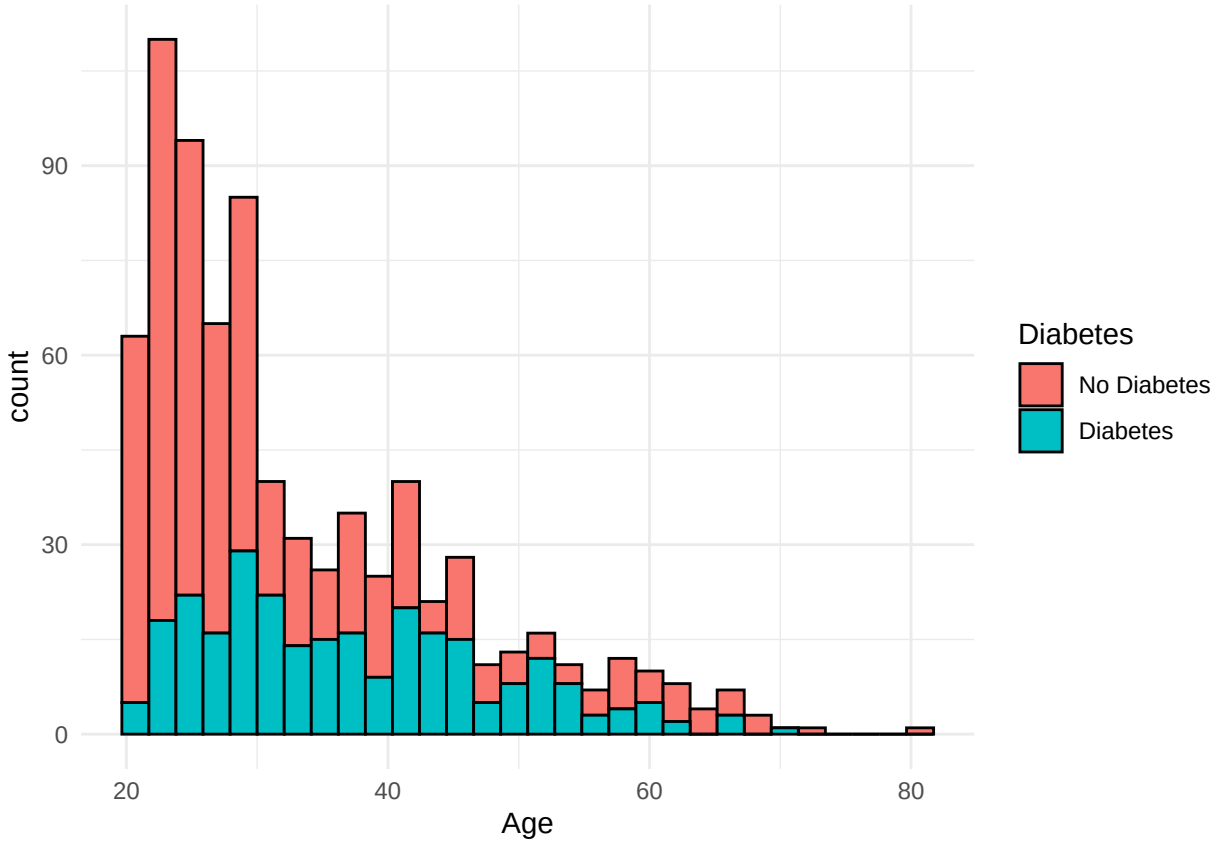

Fig 1.3: With increasing Age, we see proportion of diabetic female increasing

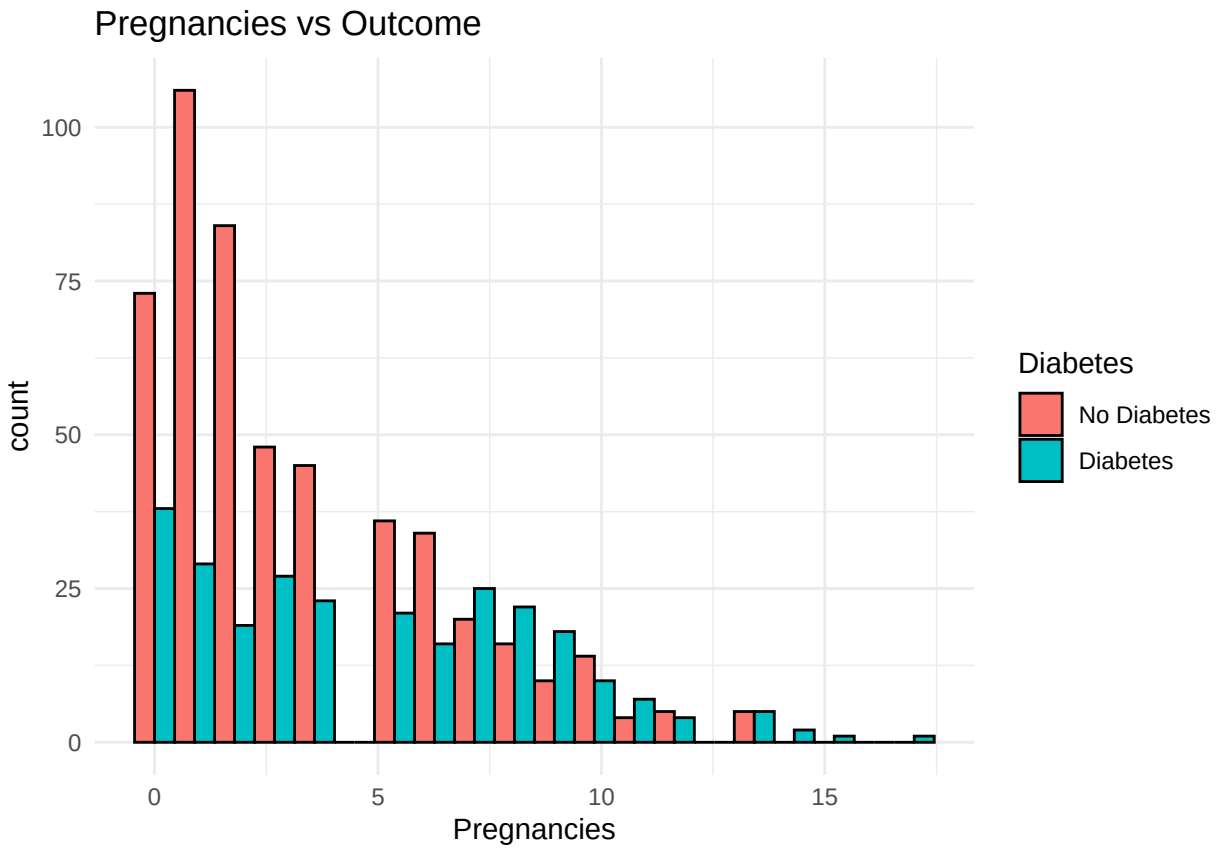

Fig 1.4: Increasing No. of Pregnancies may imply increasing chances of Diabetes

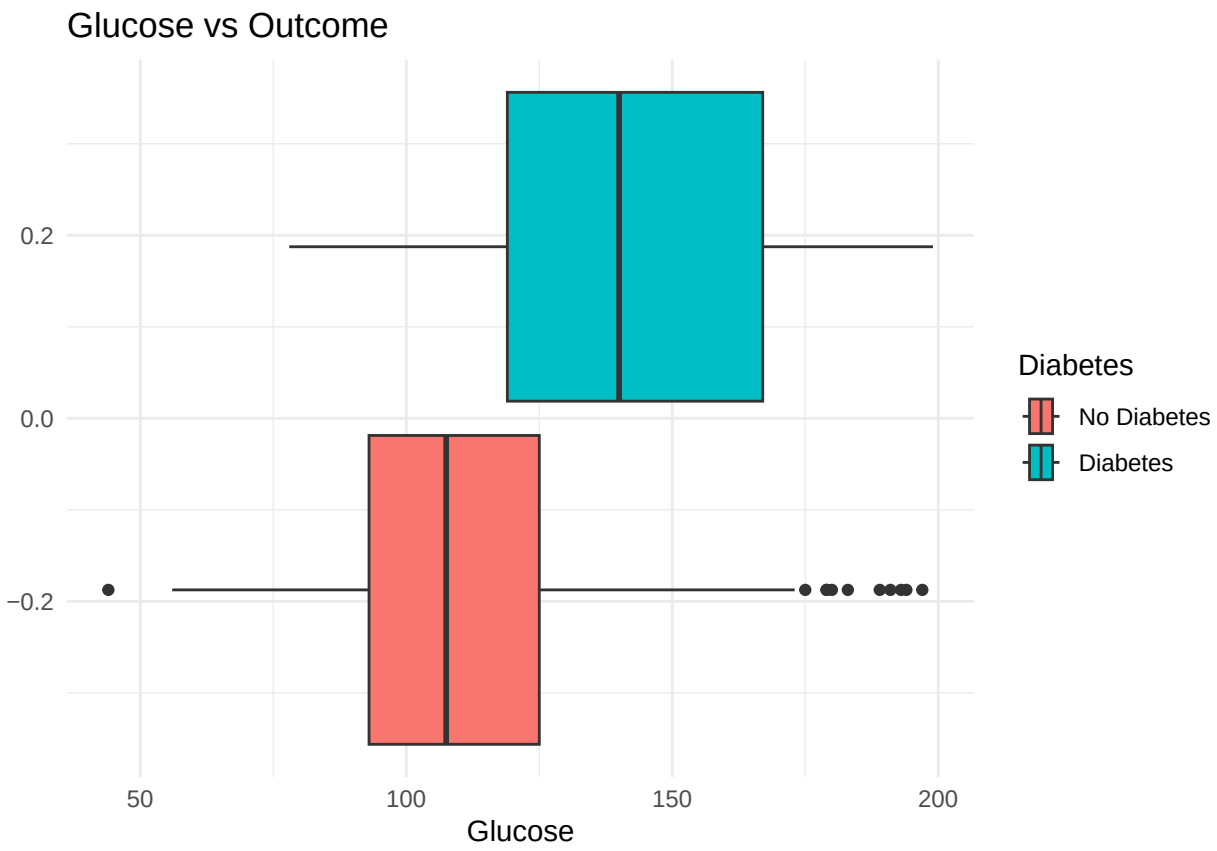

Fig 1.5: Median Glucose levels for Diabetic group is considerably higher than that of No Diabetes Group

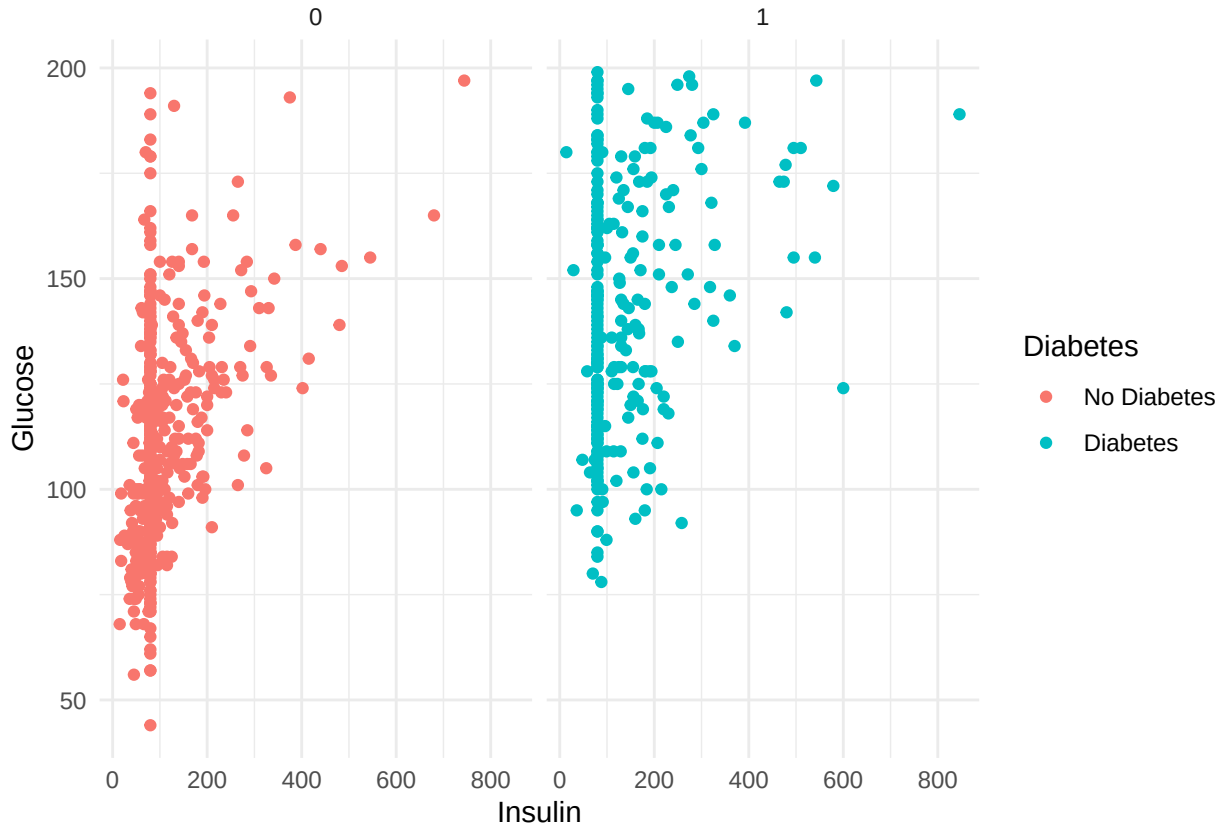

Fig 1.6: With increasing Insulin and Glucose levels, there are high chances of diabetes

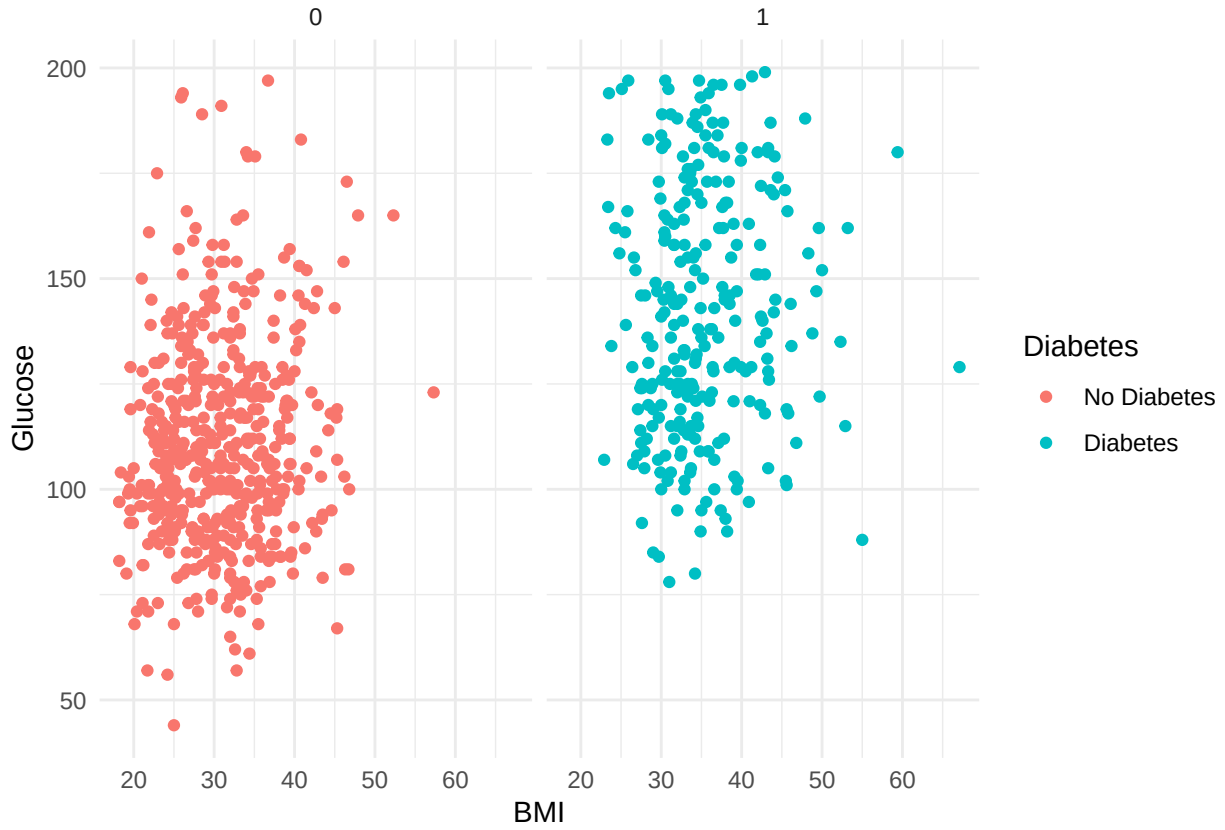

Fig 1.7: There is a positive association between an increase in Body Mass Index (BMI) and higher glucose levels which indicate increasing chances of Diabetes

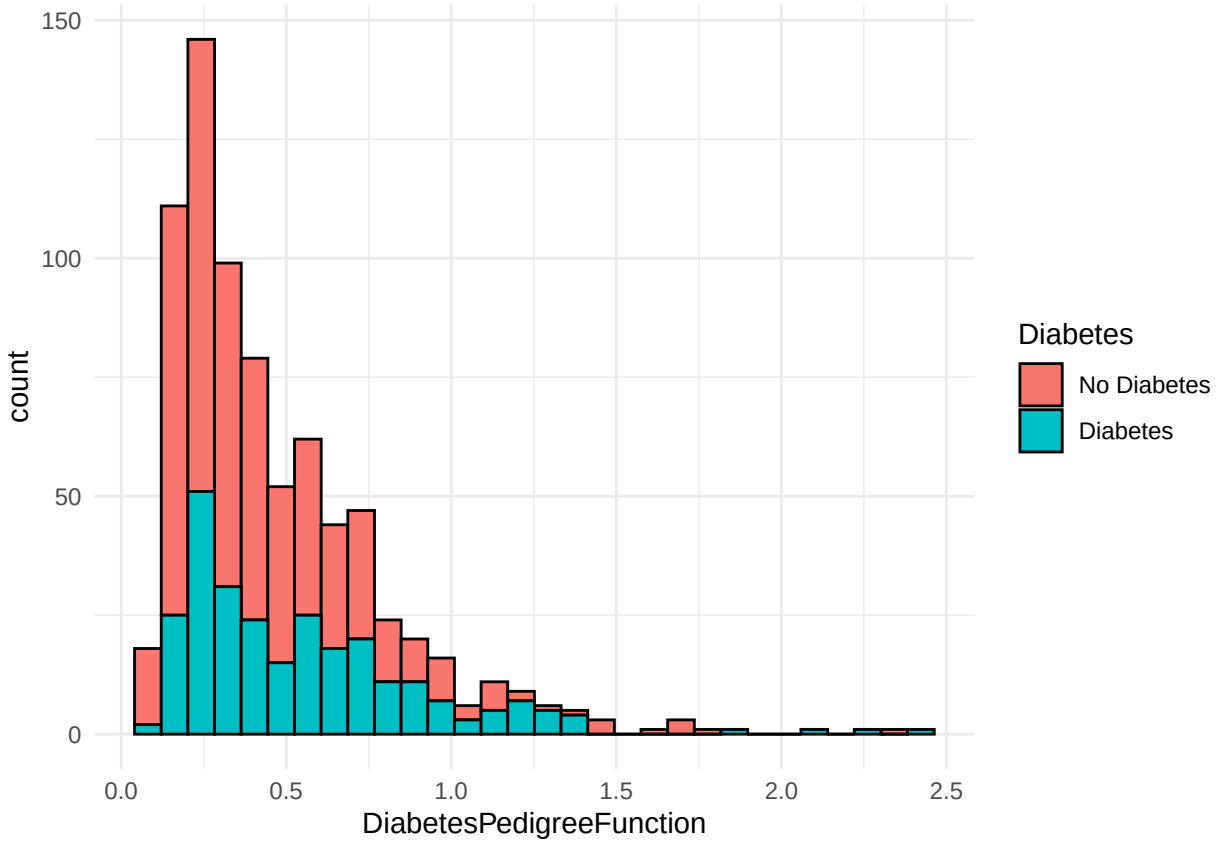

Fig 1.8: Although the data points in higher range is quite relatively small, there is a noteworthy increase in the proportion of patients suffering from diabetes within this subset.

Conclusion

- Fig 1.4 suggest Multiple Pregnancies may be associated with higher risks of type-2 diabetes in female.
- From Fig 1.5 ,a substantial distinction is evident between median Glucose levels for diabetic and non diabetic female. It may suggest that elevated glucose levels are associated with the presence of diabetes.
- Fig 1.7 may suggest increasing BMI and elevated Glucose levels can be associated with higher risks of Diabetes.
- Fig 1.8 may suggest that higher diabetes pedigree function may be associated with higher risks of Diabetes thus unfolding the potential role of genetic traits in development of Diabetes.