Ashoka Horizons

Final Project

Pima Indians Diabetes Database

By Suhaani Singh

Introduction

The Pima Indians Diabetes Database, originally from the National Institute of Diabetes and Digestive and Kidney Diseases, aims to diagnostically predict whether or not a patient has diabetes based on certain diagnostic measurements.

I have chosen the question, **“How does Body Mass Index (BMI) affect the likelihood of diabetes in Pima Indian females?”**
I chose this topic since this question is the first to get asked when someone is inquiring about diabetes. Generally, obese people are seen to have diabetes and I wanted to statistically understand if that is true.

Conda Environment Setup

To ensure a consistent and isolated environment for the analysis, a Conda environment named "diabetes_env" was created. This environment included essential libraries such as pandas, numpy, matplotlib, seaborn, and scipy.

Data Loading and Cleaning

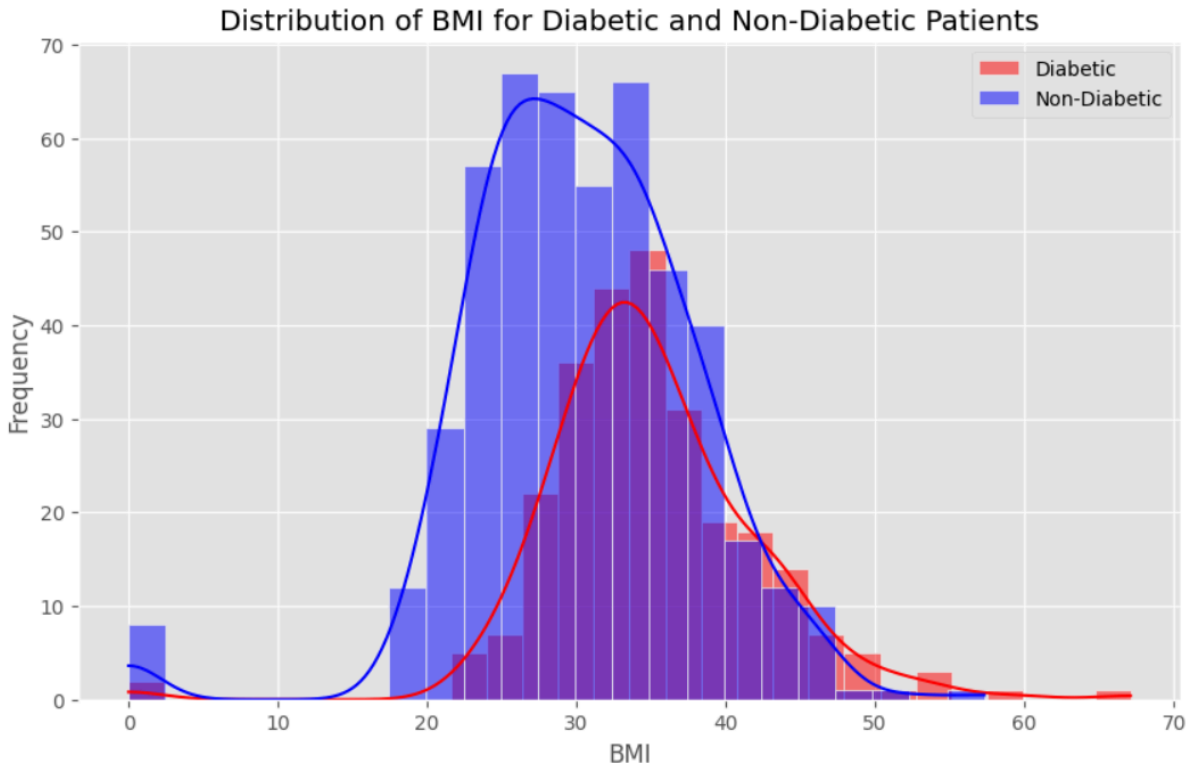
The dataset was loaded using pandas and checked for null values. Rows with any null values were removed to maintain data quality.

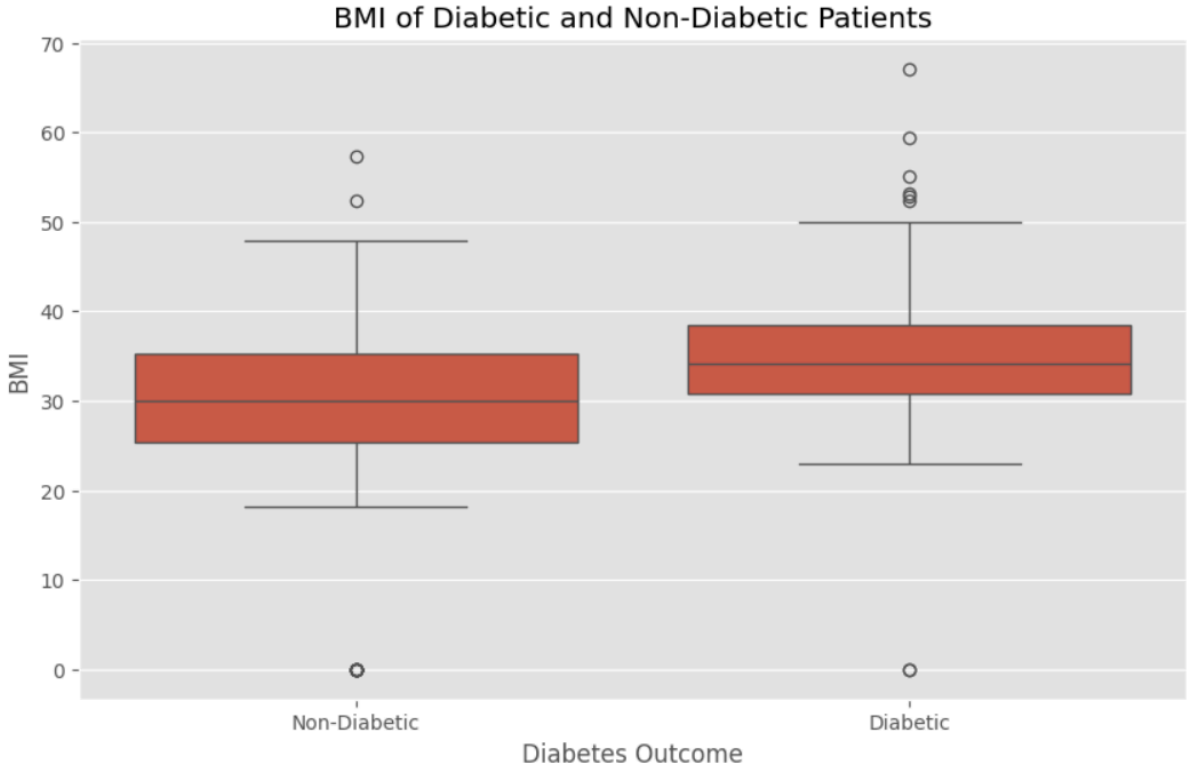
Exploratory Data Analysis (EDA)

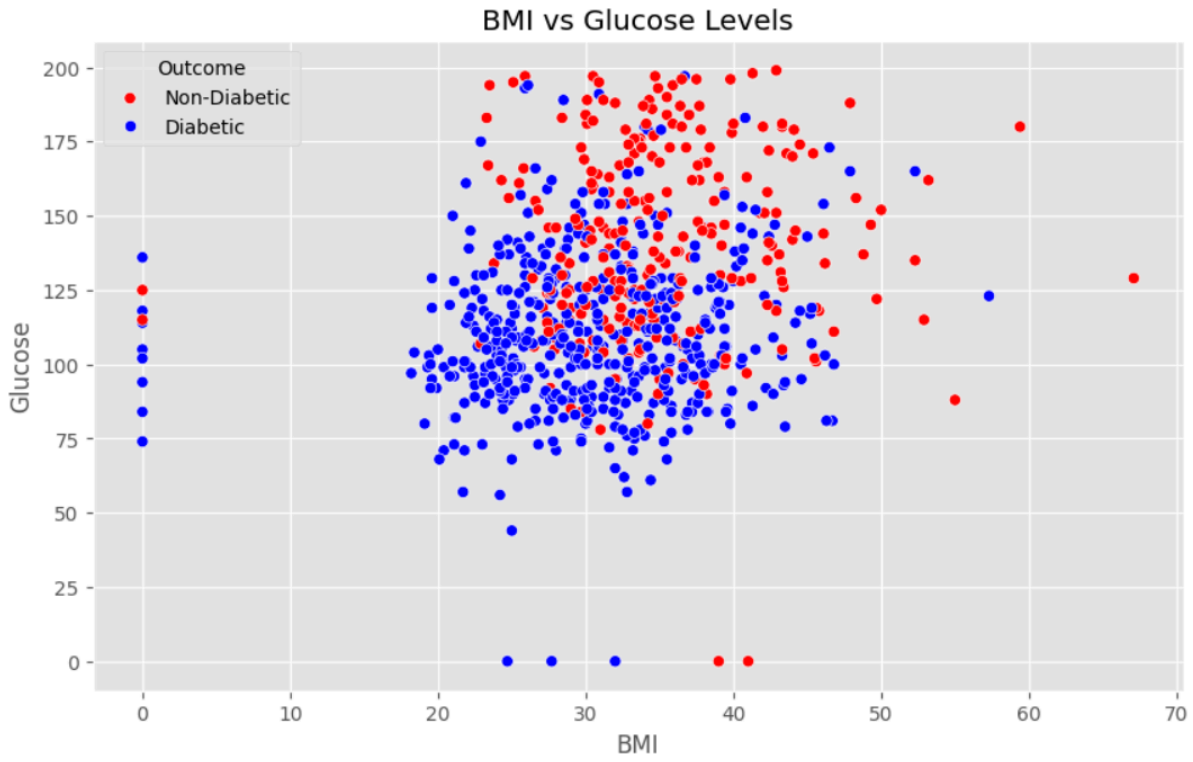
Descriptive statistics were computed for BMI grouped by diabetes outcome. The analysis revealed that the average BMI for diabetic patients is higher compared to non-diabetic patients, suggesting a potential association between higher BMI and increased likelihood of diabetes.

Visualizations

Frequently used methods of data visualizations such as histograms, box plots, and scatter plots were used to illustrate the relationships between BMI, glucose levels, and diabetes outcome. These visualizations provided further evidence supporting the association between higher BMI and the likelihood of diabetes.


Distribution of BMI for Diabetic and Non-Diabetic Patients
Diabetic
Non-Diabetic
Frequency
BMI
0
10
20
30
40
50
60
70


BMI of Diabetic and Non-Diabetic Patients
BMI
Non-Diabetic
Diabetic
Diabetes Outcome
0
10
20
30
40
50
60
70



Interpretation

The analysis of the Pima Indians Diabetes Dataset indicates a strong association between higher BMI and the likelihood of diabetes. The descriptive statistics, and visualizations all support this conclusion. Interventions aimed at reducing BMI, such as promoting healthy eating habits and regular physical activity, could be effective in lowering the risk of diabetes in this population. Additionally, BMI can be considered an important factor in the early screening and diagnosis of diabetes.