

# CRACKING THE CODE TO DIABETES

Data-Driven Solutions



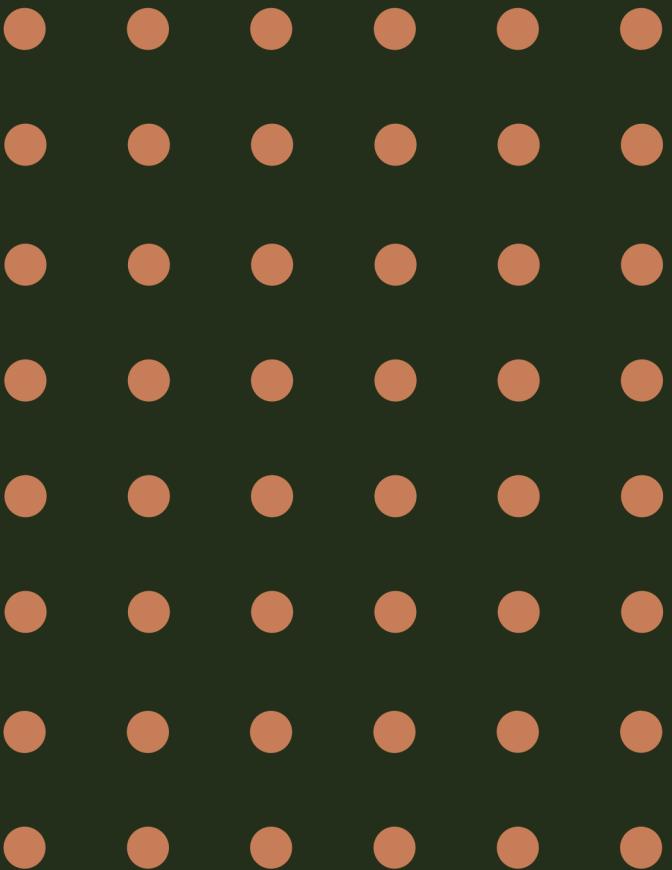
Presented by  
*The Sugar Control Squad*

# MEET THE TEAM

**Jimcollins Wamae (Data Scientist)**

**Rosemary Mburu (Data Scientist)**

**Stella Kitur (Data Scientist)**

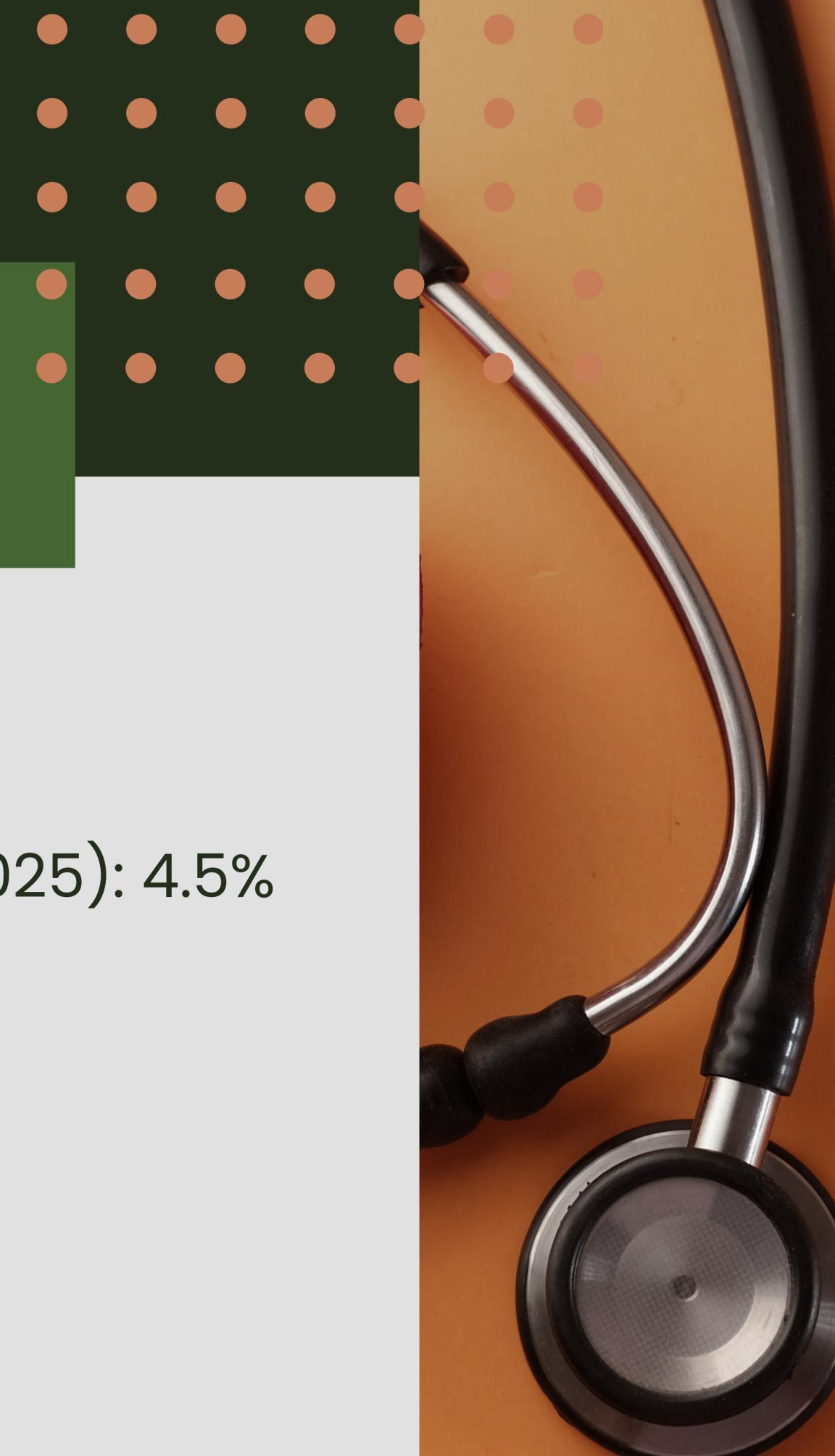
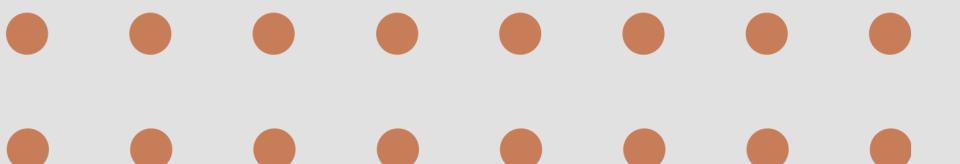


# OVERVIEW

## STATISTICS

1. Diabetes affects millions worldwide.
2. Kenya's projected diabetes prevalence (2025): 4.5%
3. Estimated diabetes expenditure in Kenya
  - a. 2030: KES 27.3 million
  - b. 2045: KES 42.5 million

source



# CHALLENGES

1. Education
2. Delayed Screening
3. Insulin Supply
4. Access to Consumables

source

# PROBLEM STATEMENT

- Develop a predictive model using machine learning and data to accurately identify diabetes risk in individuals.

Focus on finding meaningful features for diabetic status, with the main success metric being the F1 score, crucial for medical cases.

# METHODOLOGY

## CRISP-DM

Business  
Understanding

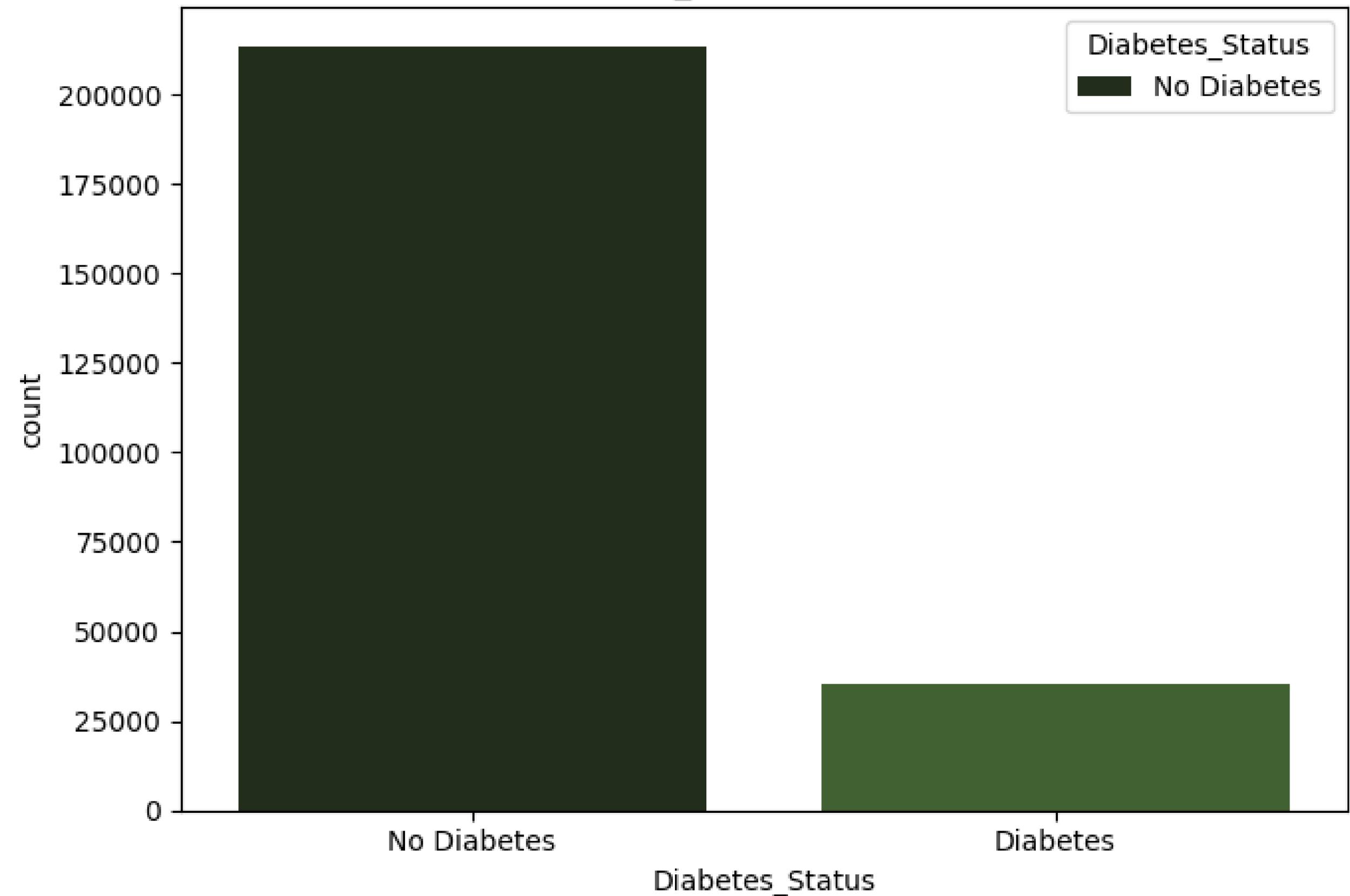
Data  
Understanding

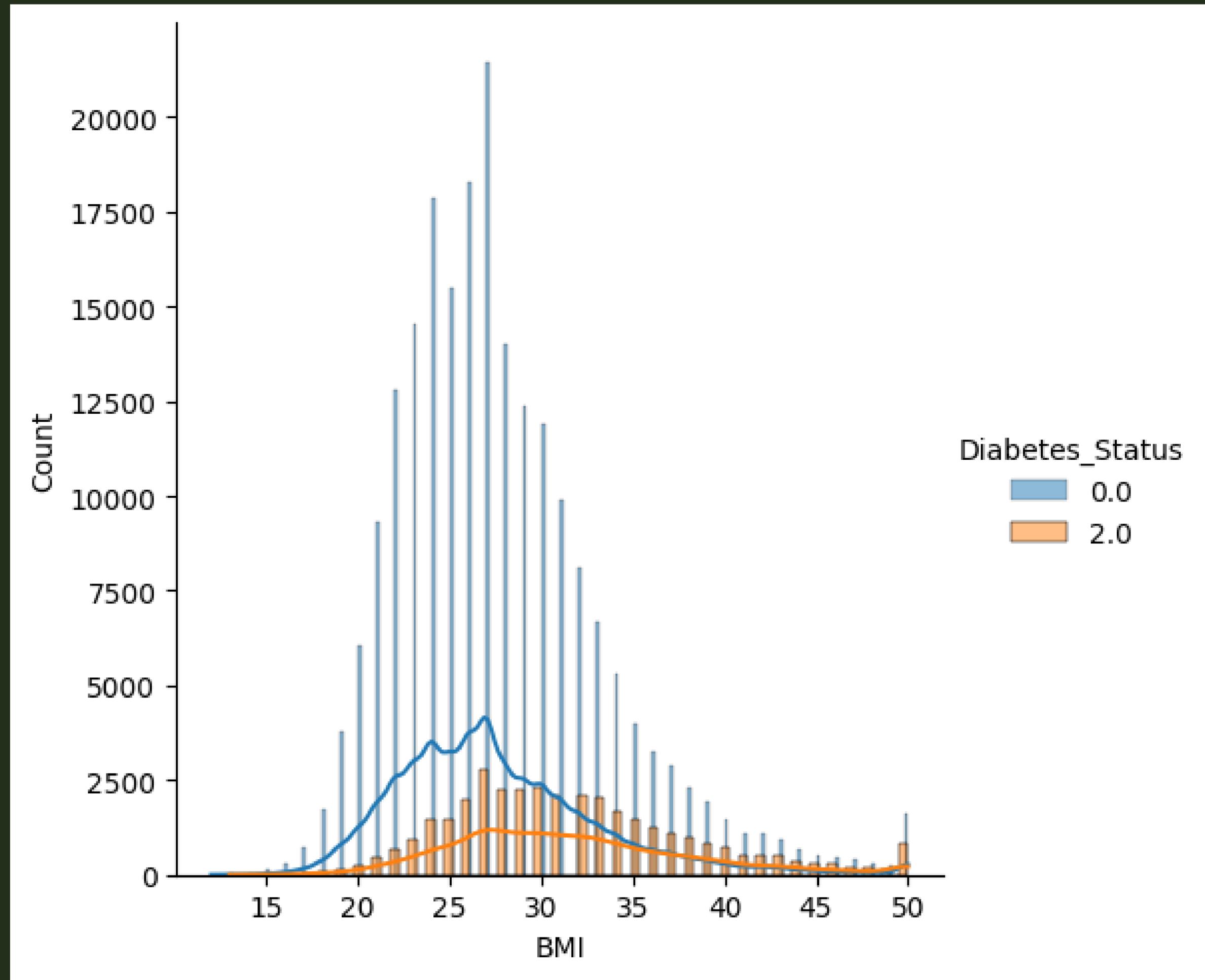
Data  
Preparation

Modeling

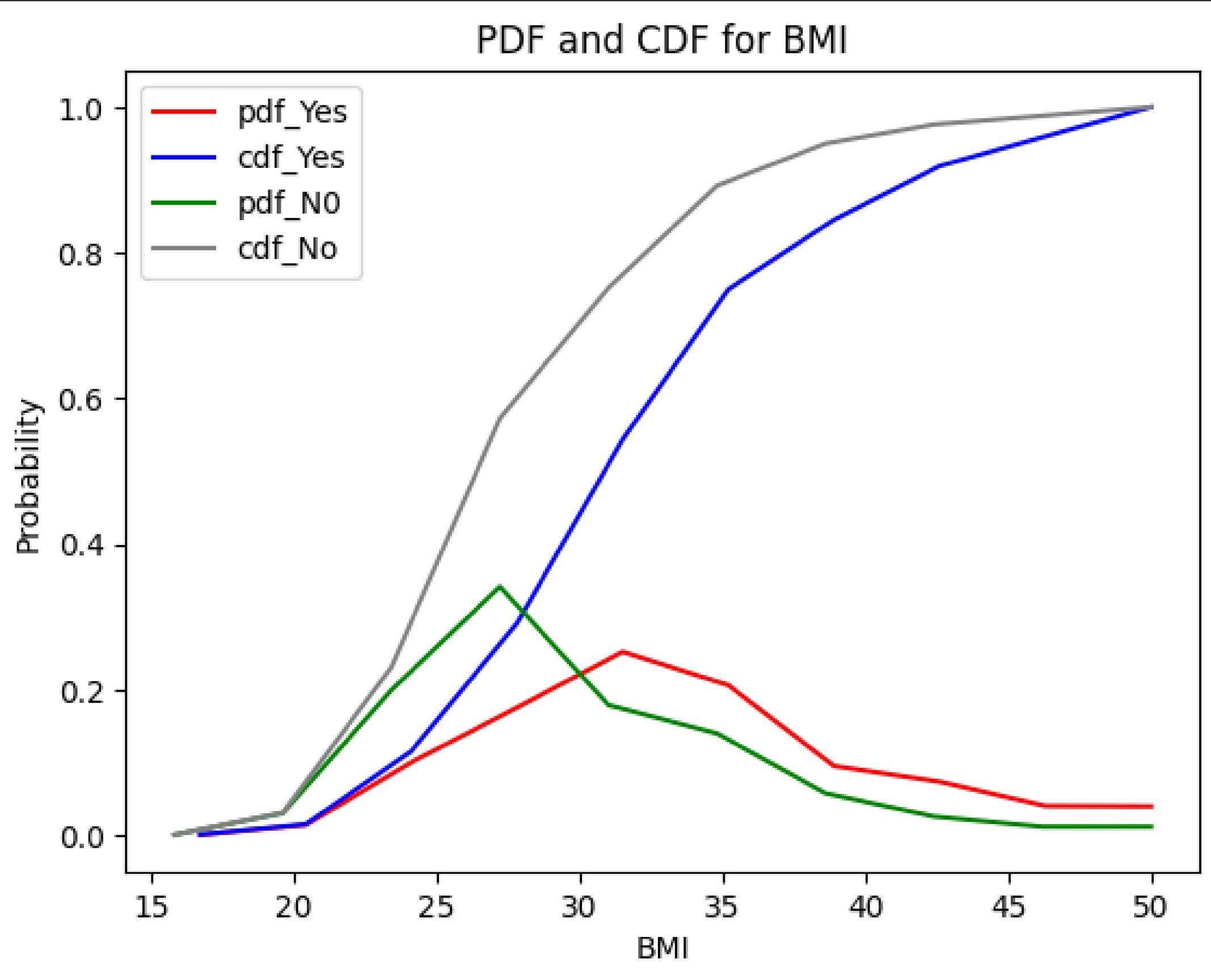
Evaluation

### Diabetes\_Status - Count Plot





PDF and CDF for BMI



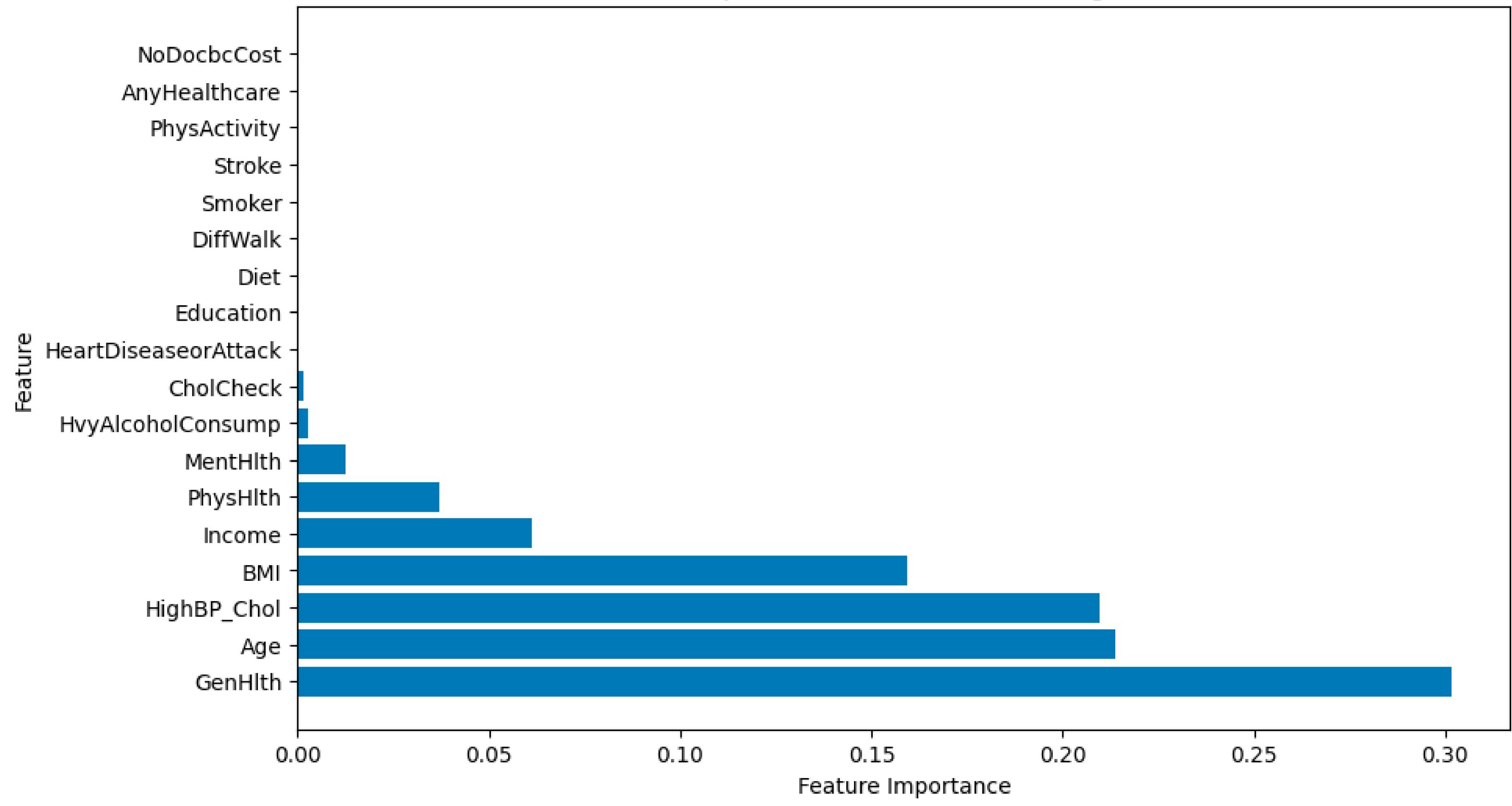
- 1. Grey line > Blue line: Non-diabetic individuals have higher BMI.**
- 2. Green line > Red line: Non-diabetic individuals have higher BMI.**

# MODEL EVALUATION

3 Machine Learning models were used to determine the best performing model that would accurately predict one's diabetic status.

- The best performing model was the Gradient Boost.
- It had an F1 Score of 84%

## Feature Importance of GradientBoostingClassifier



# CONCLUSION

Key Factors to look out for:

- General Health
- Age
- High Blood Pressure & Cholesterol
- Income
- Physical Health

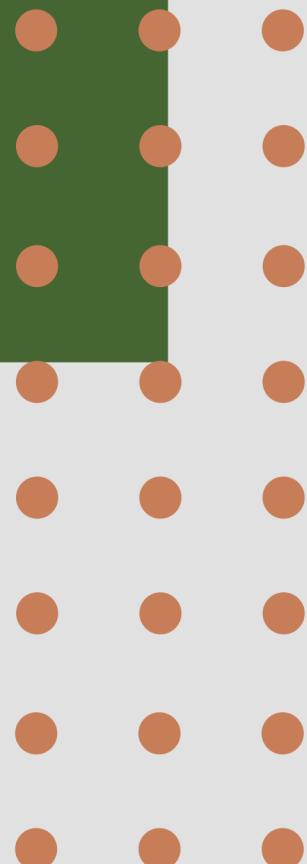


# NEXT STEPS

Leverage  
on the  
Gradient  
Boost  
Model

Increase  
Awareness  
/Education

Enhance  
on data  
collection



# THANK YOU

