PROJECT 4 - FINAL

GROUP 6

Team: Mark Flascas, Anjana Sanjeev, Ang Li

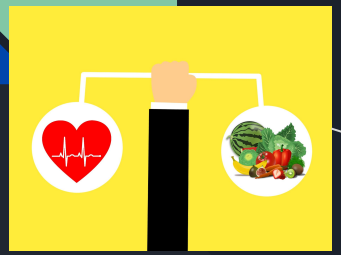
DIABETES SELF ASSESSMENT APP

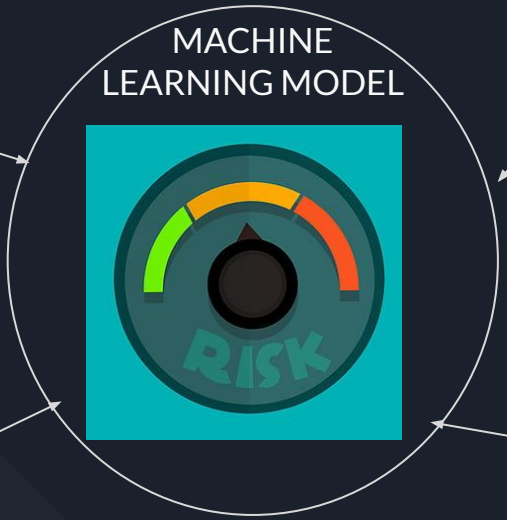
OVERVIEW

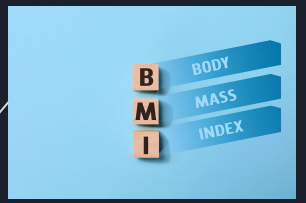
- INTRODUCTION - Mark
- DATA SOURCING - KAGGLE DATA SOURCE - Anjana
- DATA REVIEW - CHARTS - Anjana
- MACHINE LEARNING MODEL - Mark
- WEBPAGE AND LIVE DEMO - Ang

DIABETES SELF ASSESSMENT APP

Enable web based self assessment based on key criteria to determine if you are at an increased risk







DATA SOURCING - KAGGLE DATA SOURCE

KAGGLE DATA - 70692 scientific

Important Risk Factors - 21 factors identified

https://www.kaggle.com/code/alexteboul/diabetes-health-indicators-dataset-notebook

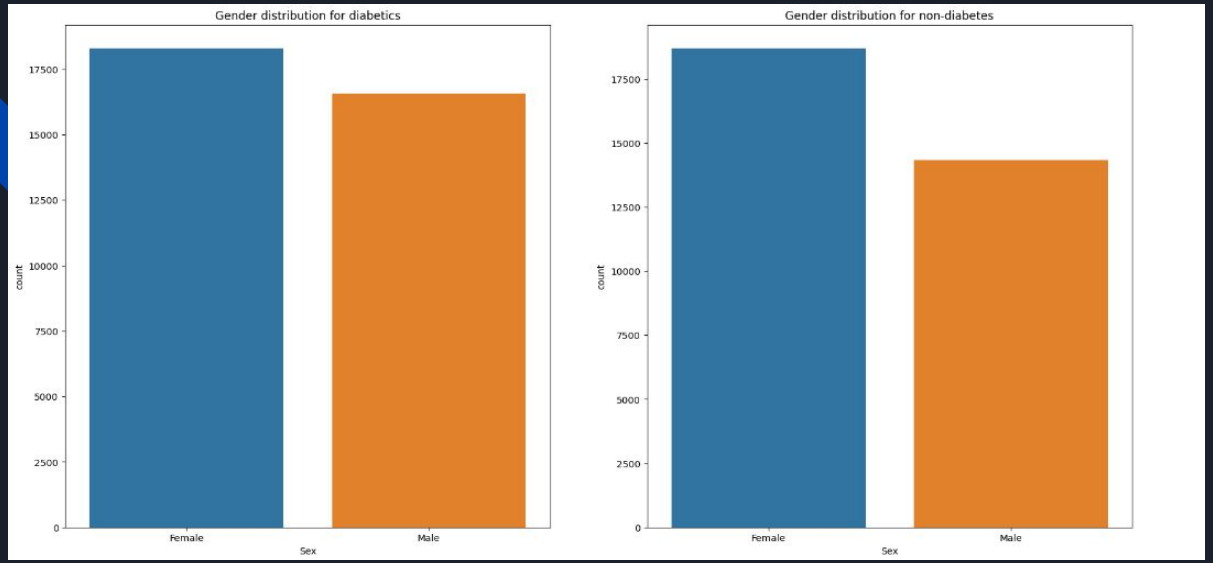
DATA REVIEW

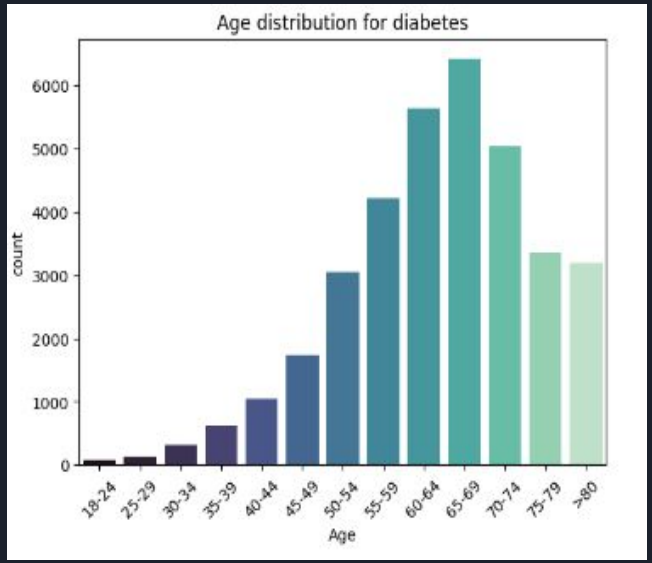
Data Cleaning

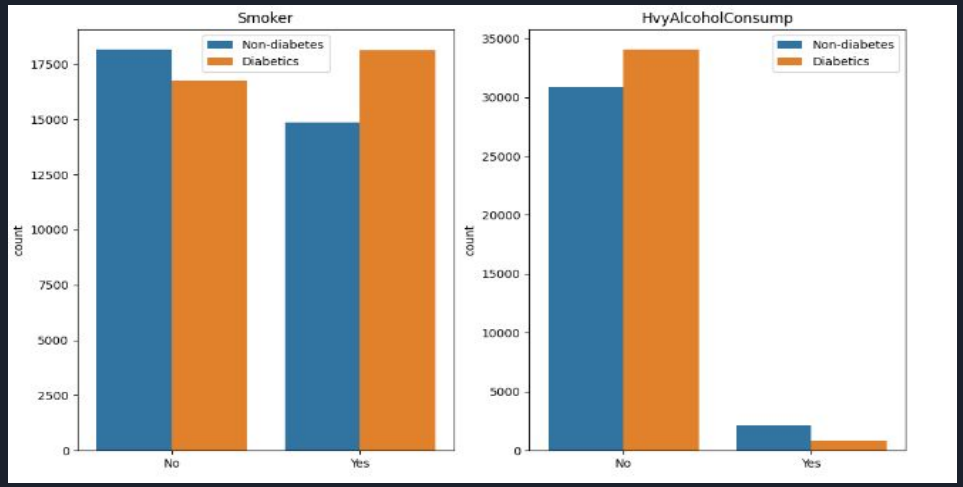
- Drop null and duplicates values
- Convert to appropriate data types

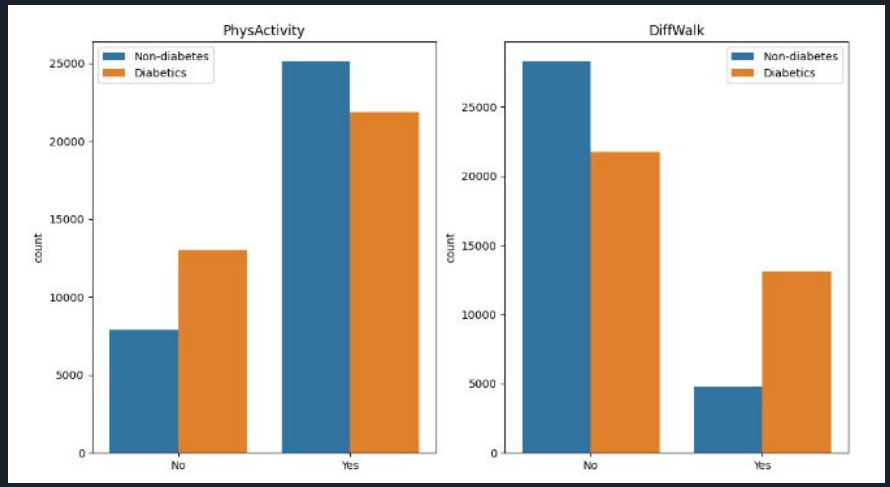
Data Analysis

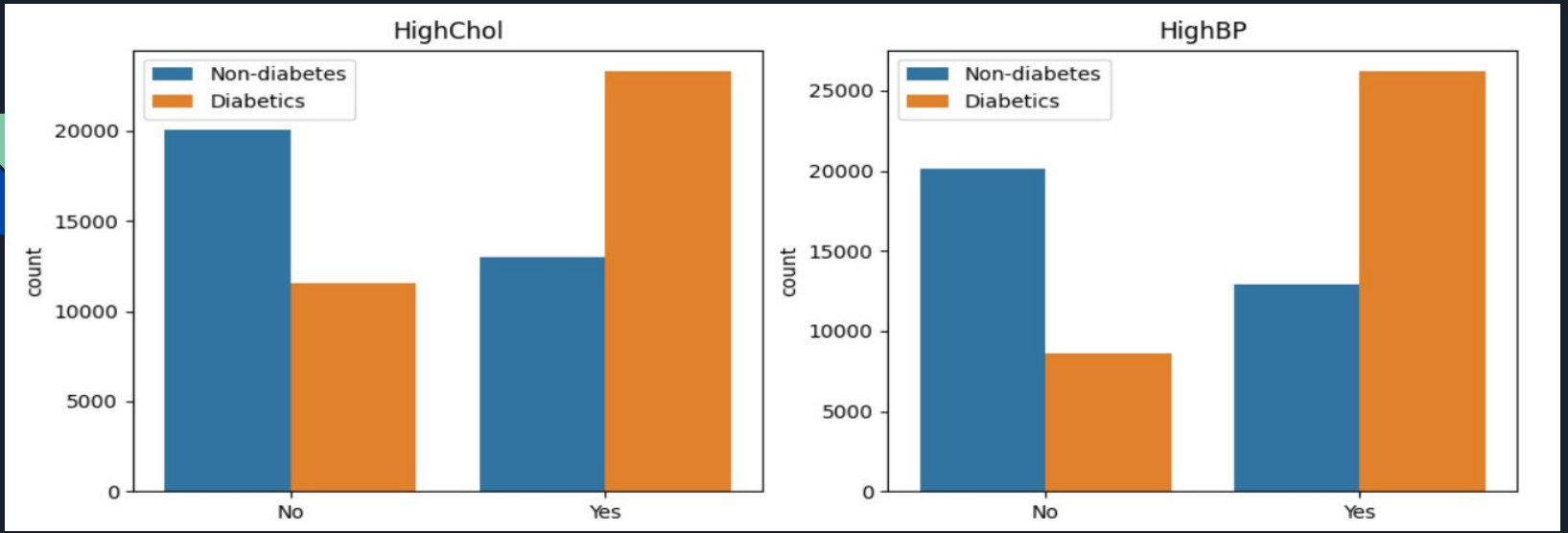
- Understanding the dataset using the few variables which are dependent with diabetes

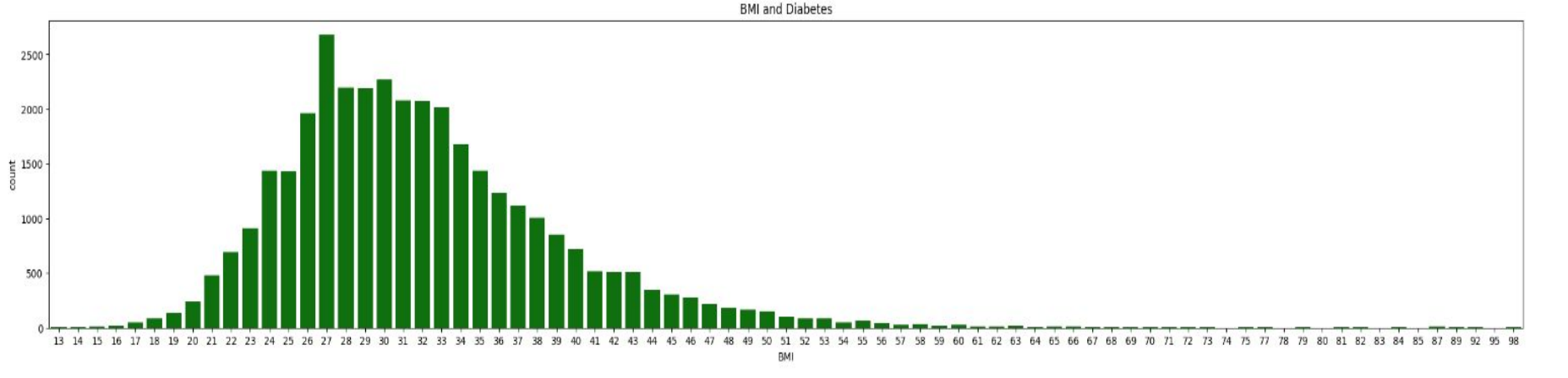

Gender distribution for diabetics
count
Female
Male
Sex
Gender distribution for non-diabetes
count
Female
Male
Sex


Age distribution for diabetes
count
18-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
>80
Age


Smoker
Non-diabetes
Diabetics
count
No
Yes
HvyAlcoholConsump
Non-diabetes
Diabetics
count
No
Yes


PhysActivity
Non-diabetes
Diabetics
count
No
Yes
DiffWalk
Non-diabetes
Diabetics
count
No
Yes


HighChol
Non-diabetes
Diabetics
count
20000
15000
10000
5000
0
No
Yes
HighBP
Non-diabetes
Diabetics
count
25000
20000
15000
10000
5000
0
No
Yes


BMI and Diabetes
count
2500
2000
1500
1000
500
0
13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 77 78 79 80 81 82 83 84 85 87 89 92 95 98
BMI

MACHINE LEARNING TO CREATE A MODEL

Attempt 1

Deep neural net

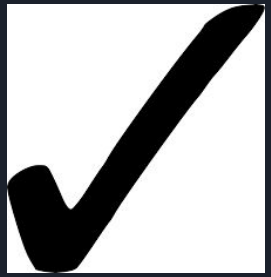
- Using, Tensorflow keras

Why it failed?

- Too many zero results ~ 85%

Attempt 2

Decision tree



- Addressed classification problem in the data through layers and branching.
- Model was much faster than the Keras attempted earlier and provide a reasonable accuracy score providing more reliable results when testing

DIABETES SELF ASSESSMENT APP

LIVE DEMO