

Risk Factors Decoded: A Predictive Model for Diabetes

By: Joe Ortiz, Max Valladolid,
and Zuleyka Benavides

Data Overview

Contains **253,680 rows** (patients) and **22 columns** (indicators)

Diabetes Type: No Diabetes = 0, Diabetes = 1.

Includes health indicators that may be linked to diabetes.

Data was collected via surveys to get health indicators in the year 2015.

Data consist of **141,974 Females** and **111,706 Males**.

No missing values!

Health Indicators

Health Conditions:

High Blood Pressure, High Cholesterol, Stroke, Heart Disease

General Health Indicators:

BMI, General Health Rating, Mental & Physical Health Days

Lifestyle Factors:

Smoking, Physical Activity, Alcohol Consumption, Diet (Fruits & Vegetables)

Demographics:

Age, Sex, Education, Income

Questions



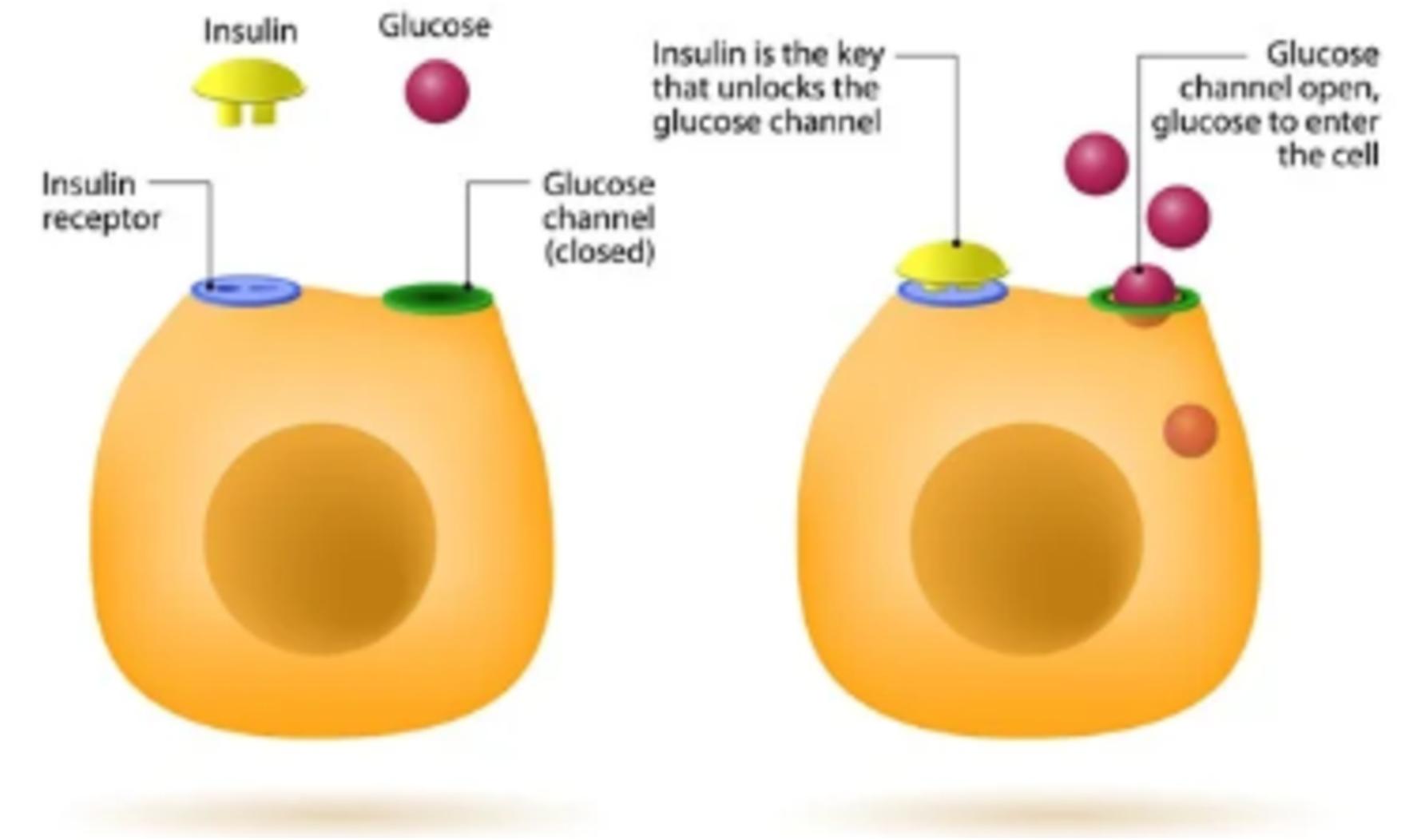
Can the data collected here provide **accurate predictions** of whether an individual has diabetes?

What **risk factors** are most **impactful/predictive** of an individual having diabetes?

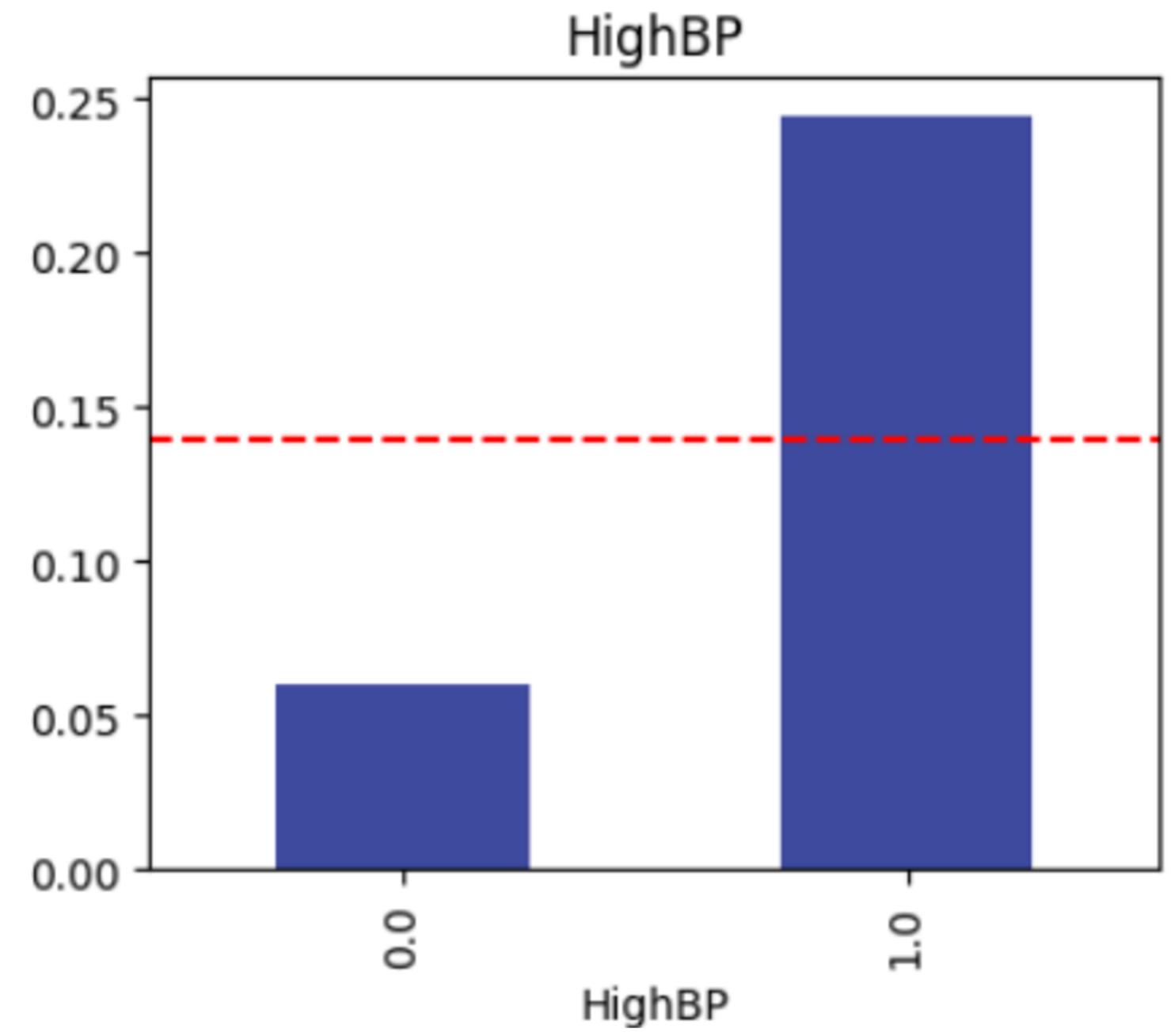
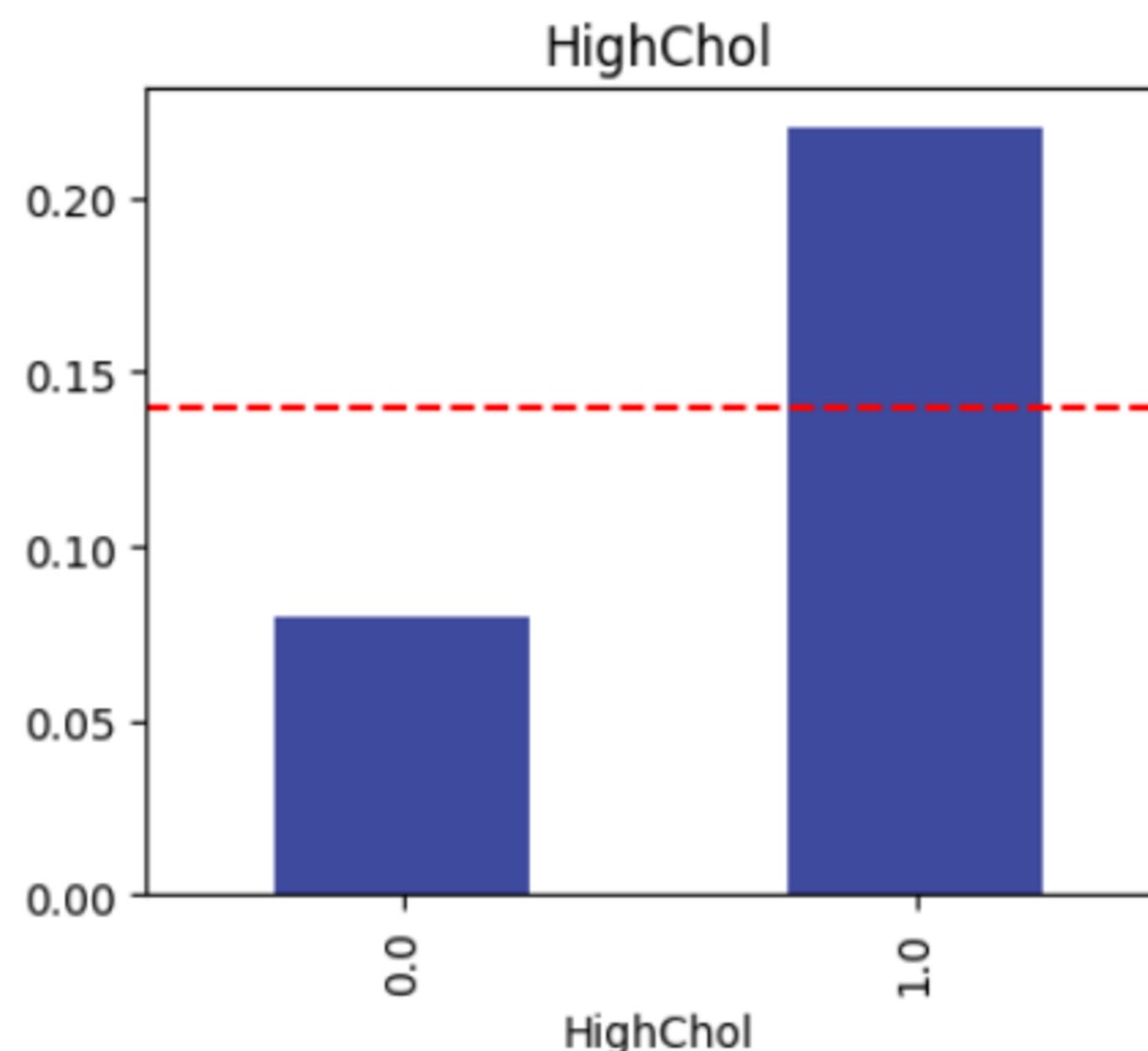
What is Diabetes?

Diabetes is a chronic disease where your body either can't produce insulin or is unable to use the insulin made as effectively as needed. That leads to increased glucose (sugar levels) in the bloodstream.

HOW DOES INSULIN WORK?



Big Factors



Why Blood Pressure and Cholesterol?

When glucose in the bloodstream is not properly regulated:

- it can damage blood vessels, leading to stiffness and narrowing of the arteries. This increases the heart's workload, **raising blood pressure.**
- it disrupts fat metabolism, leading to increased LDL (bad cholesterol). This promotes plaque buildup in the arteries, increasing the risk of **high cholesterol.**

Why Blood Pressure and Cholesterol?

High cholesterol causes plaque buildup in the arteries, making them narrow and stiff, which **increases blood pressure**.

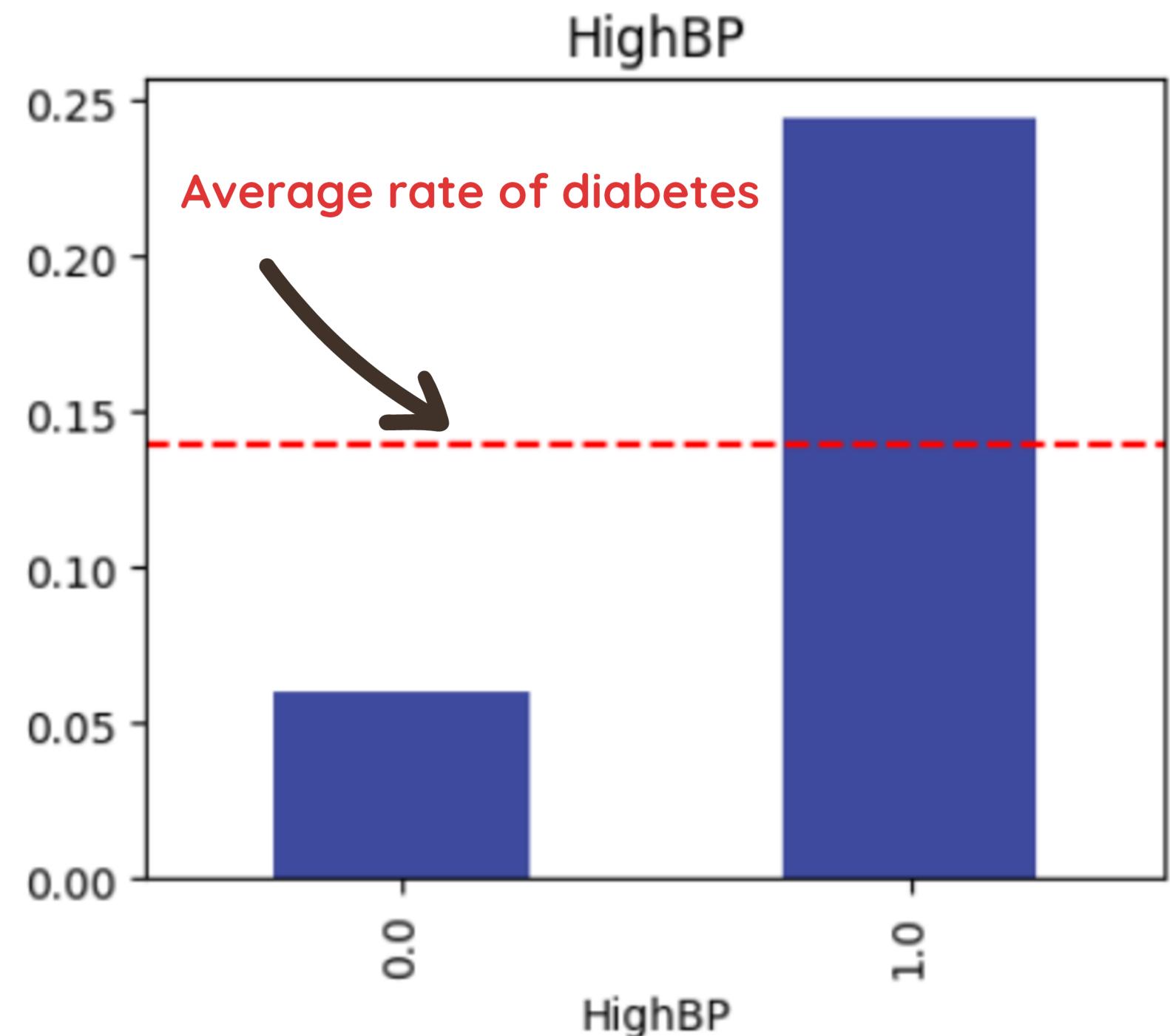
Big Factors

1

Having **High Blood Pressure** and **High Cholesterol** leads to significantly **higher probability** of having Diabetes

2

Inversely Not Having **High Blood Pressure** and **High Cholesterol** leads to significantly **LOWER probability** of having Diabetes



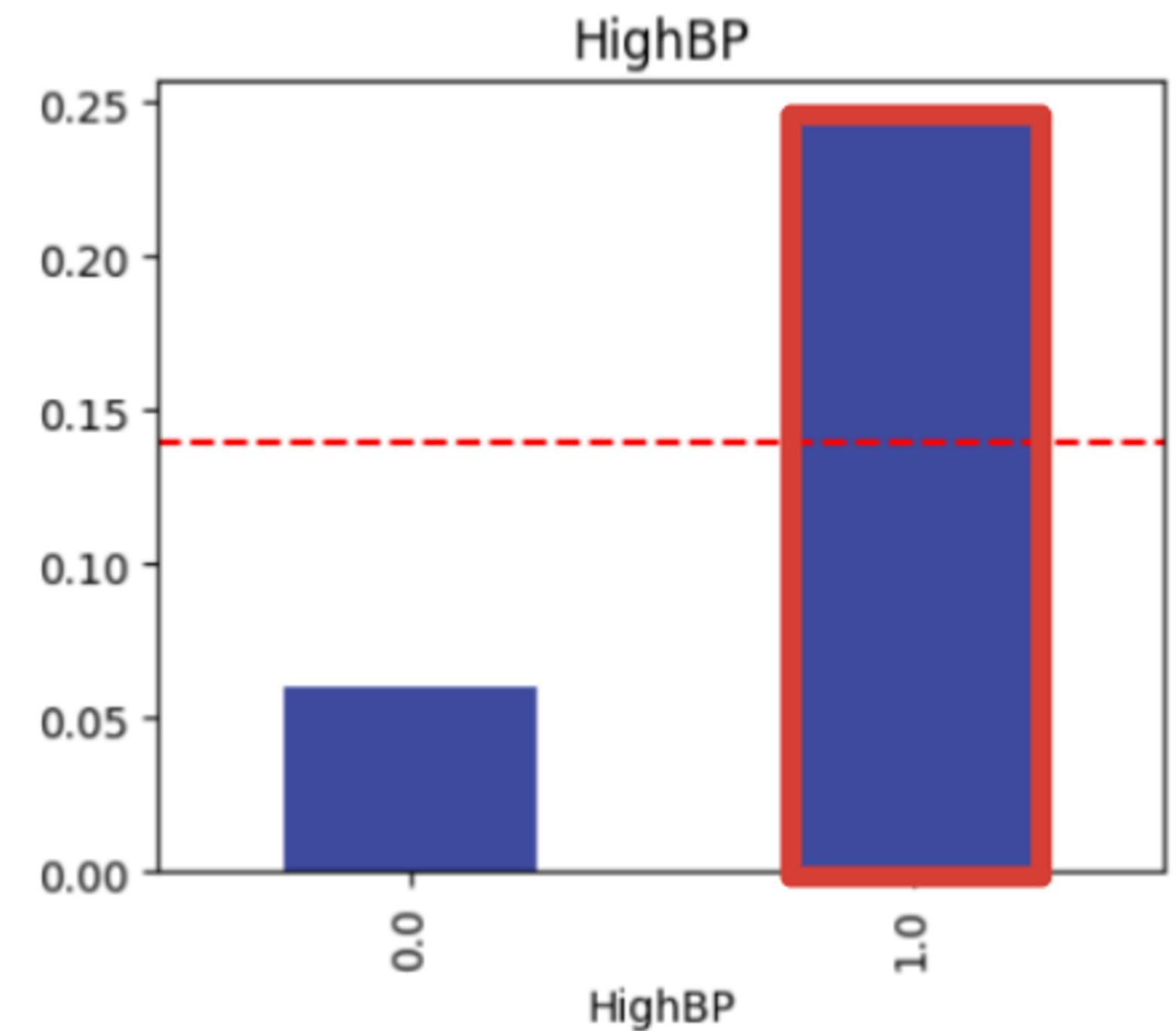
Big Factors

1

Having **High Blood Pressure** and **High Cholesterol** leads to significantly **higher probability** of having Diabetes

2

Inversely **Not Having High Blood Pressure** and **High Cholesterol** leads to significantly **LOWER probability** of having Diabetes



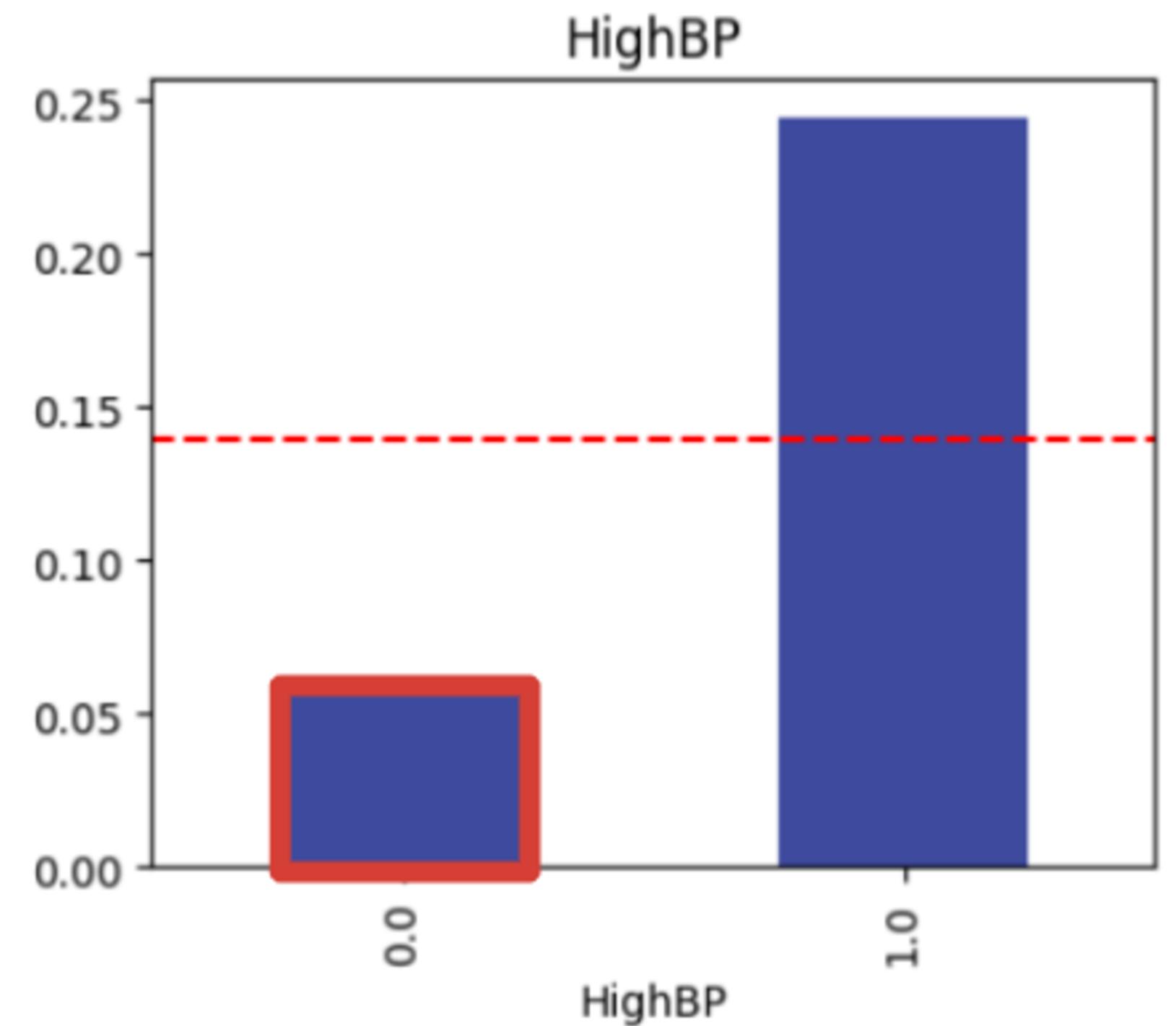
Big Factors

1

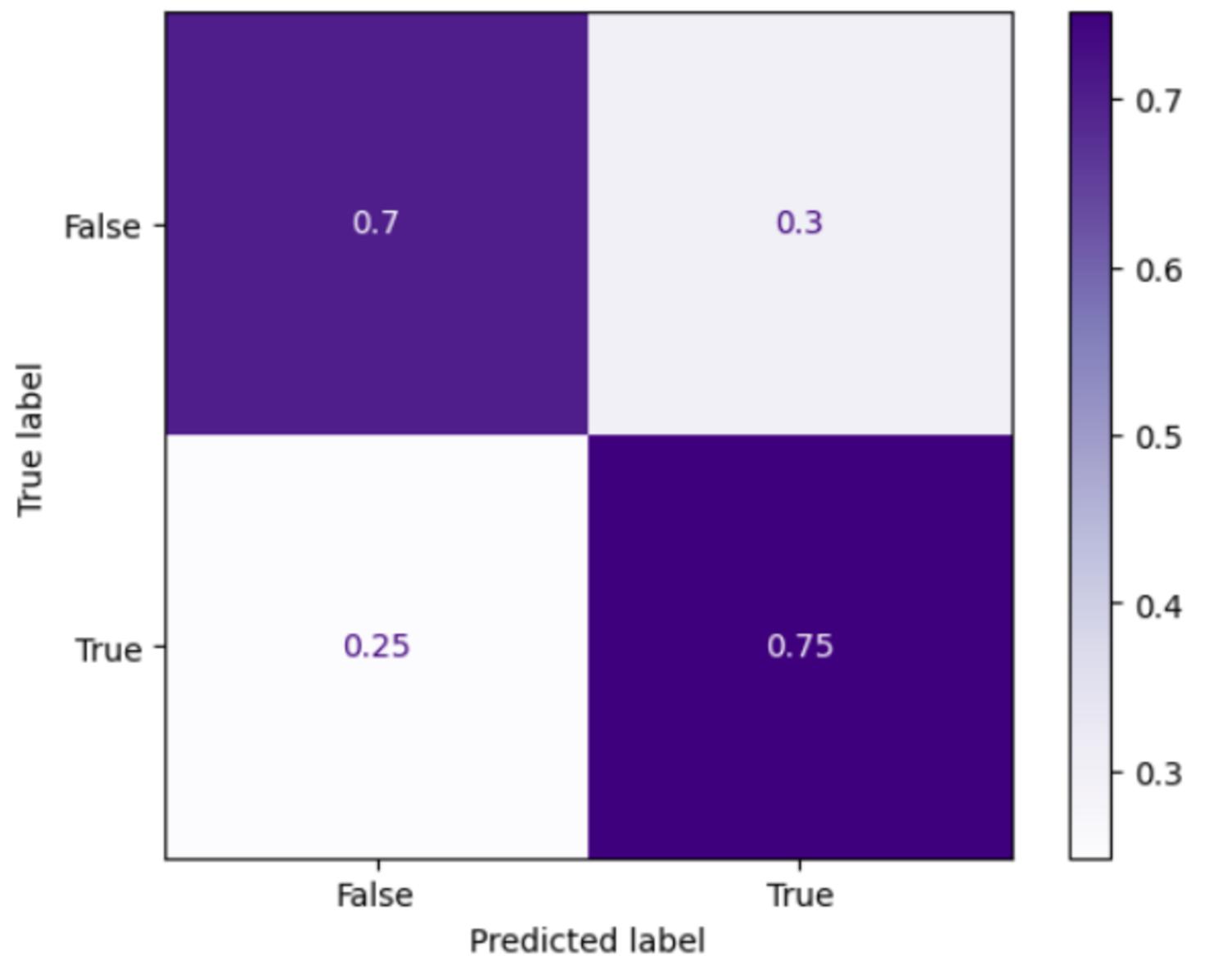
Having **High Blood Pressure** and **High Cholesterol** leads to significantly **higher probability** of having Diabetes

2

Inversely Not Having **High Blood Pressure** and **High Cholesterol** leads to significantly **LOWER probability** of having Diabetes

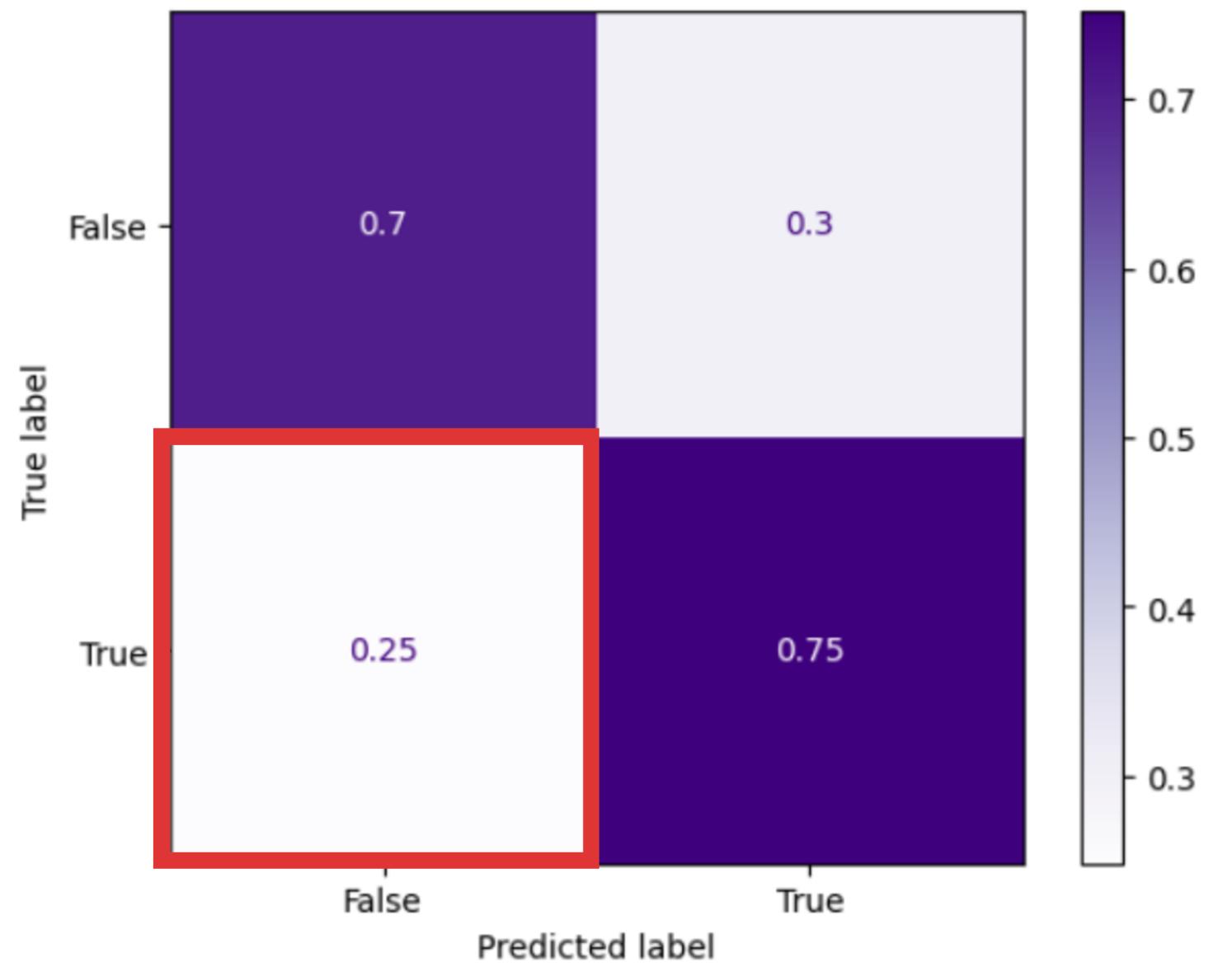


Correlation Matrix



Accuracy Score:
0.72

Correlation Matrix

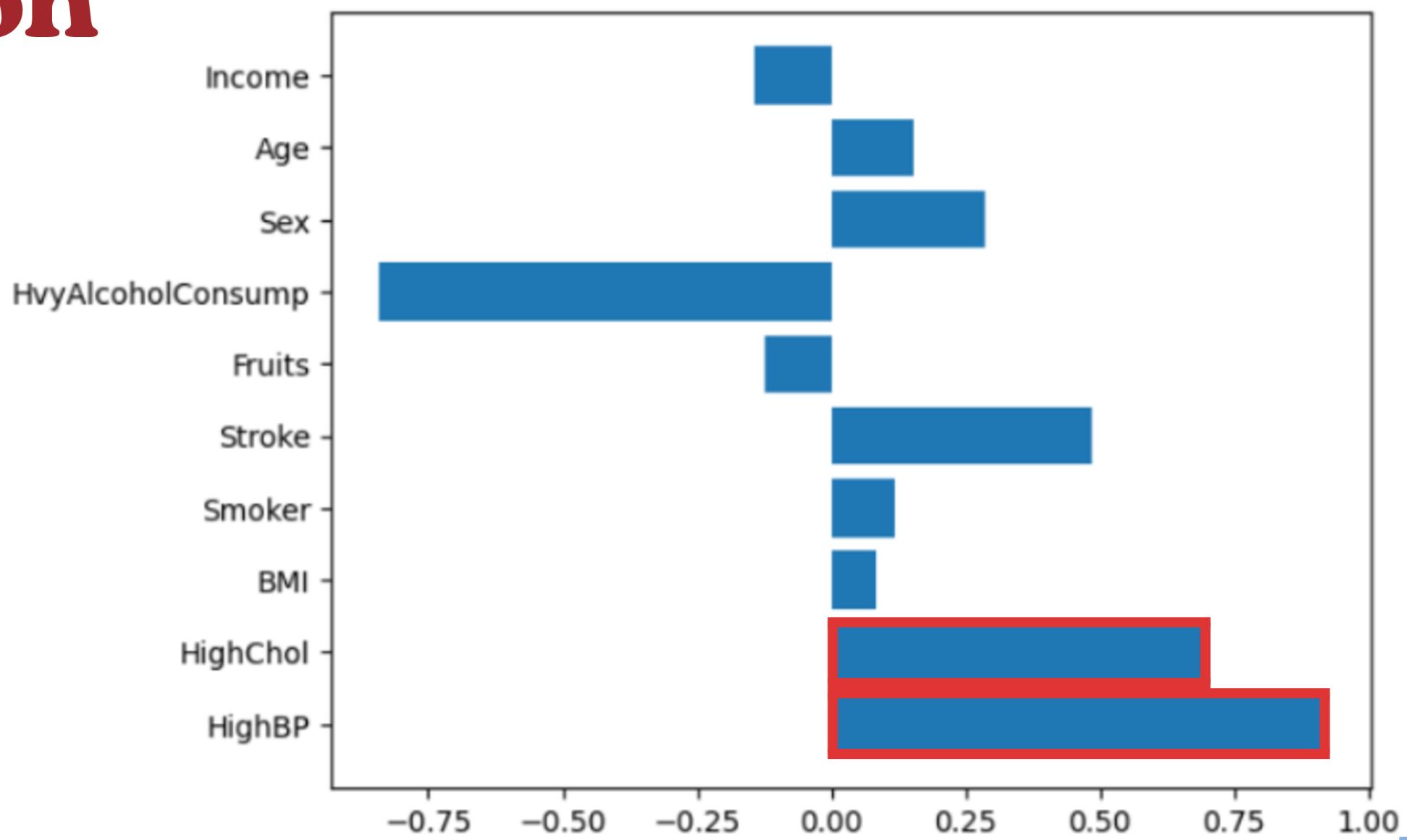


The model does a decent job, but we need to look and see what is causing the **false positives**.

2

Coefficient of Logistic Regression Model

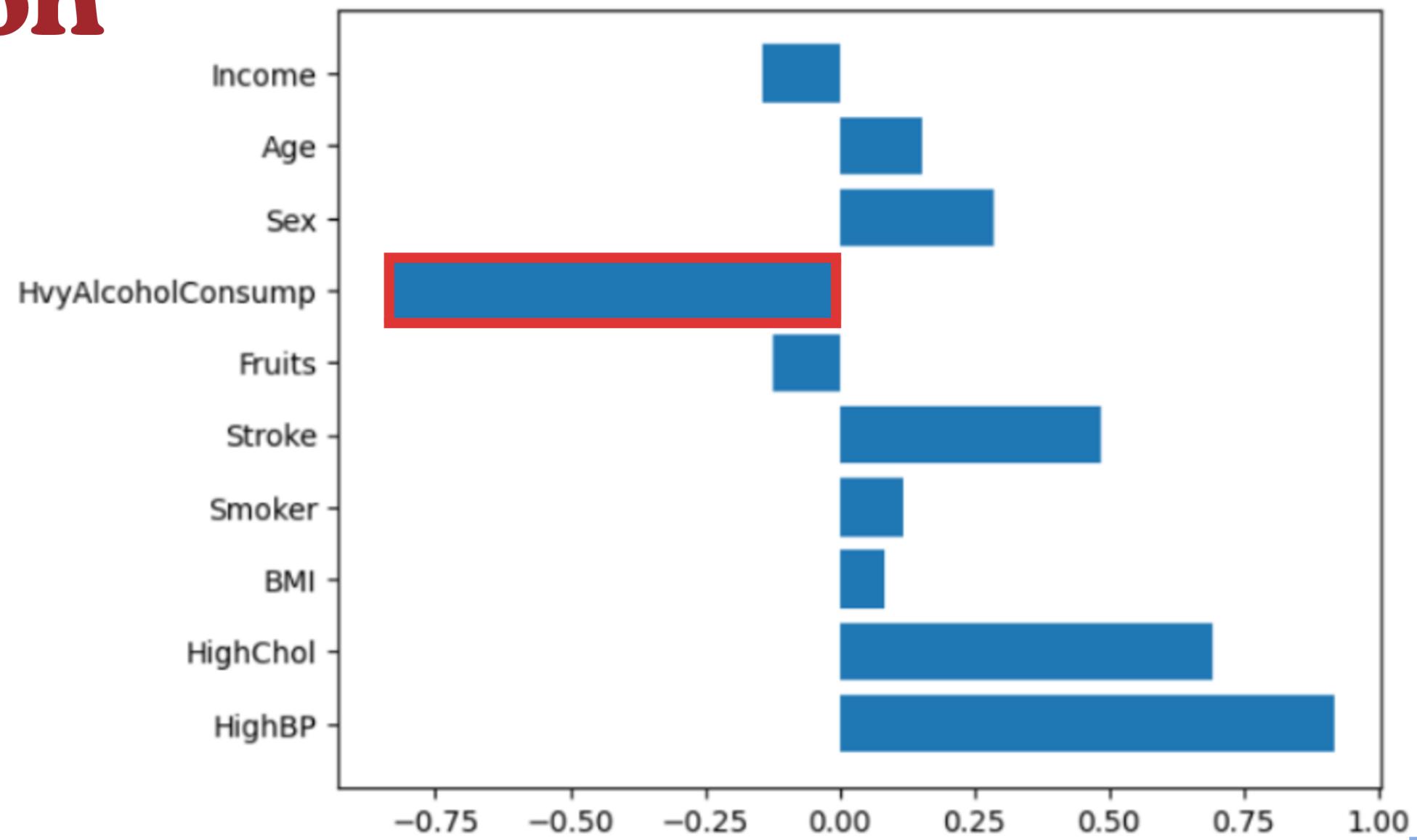
We expect **HighChol** and **HighBP** to have significant weight on prediction.



2

Coefficient of Logistic Regression Model

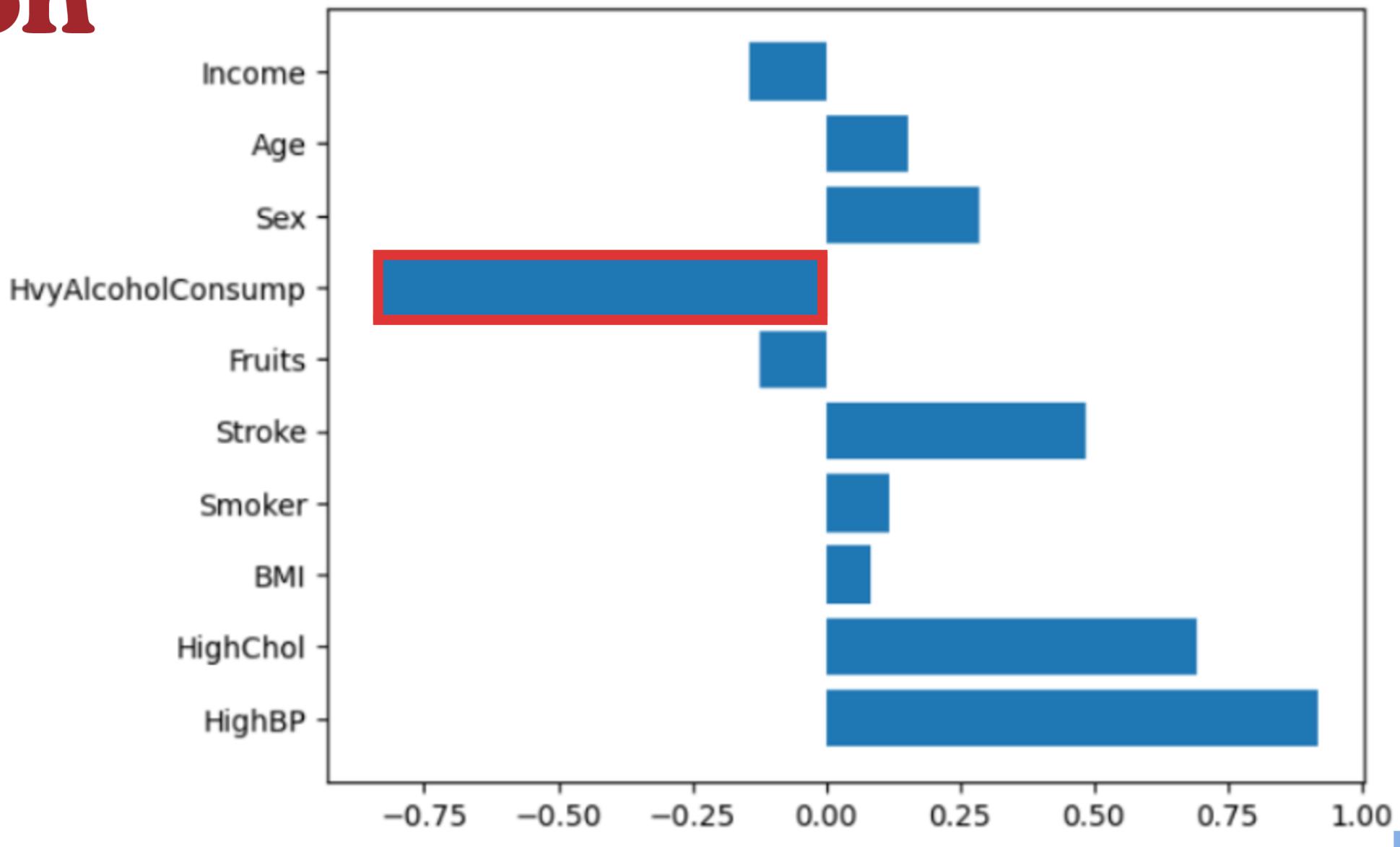
Heavy Alcohol Consumption stands out

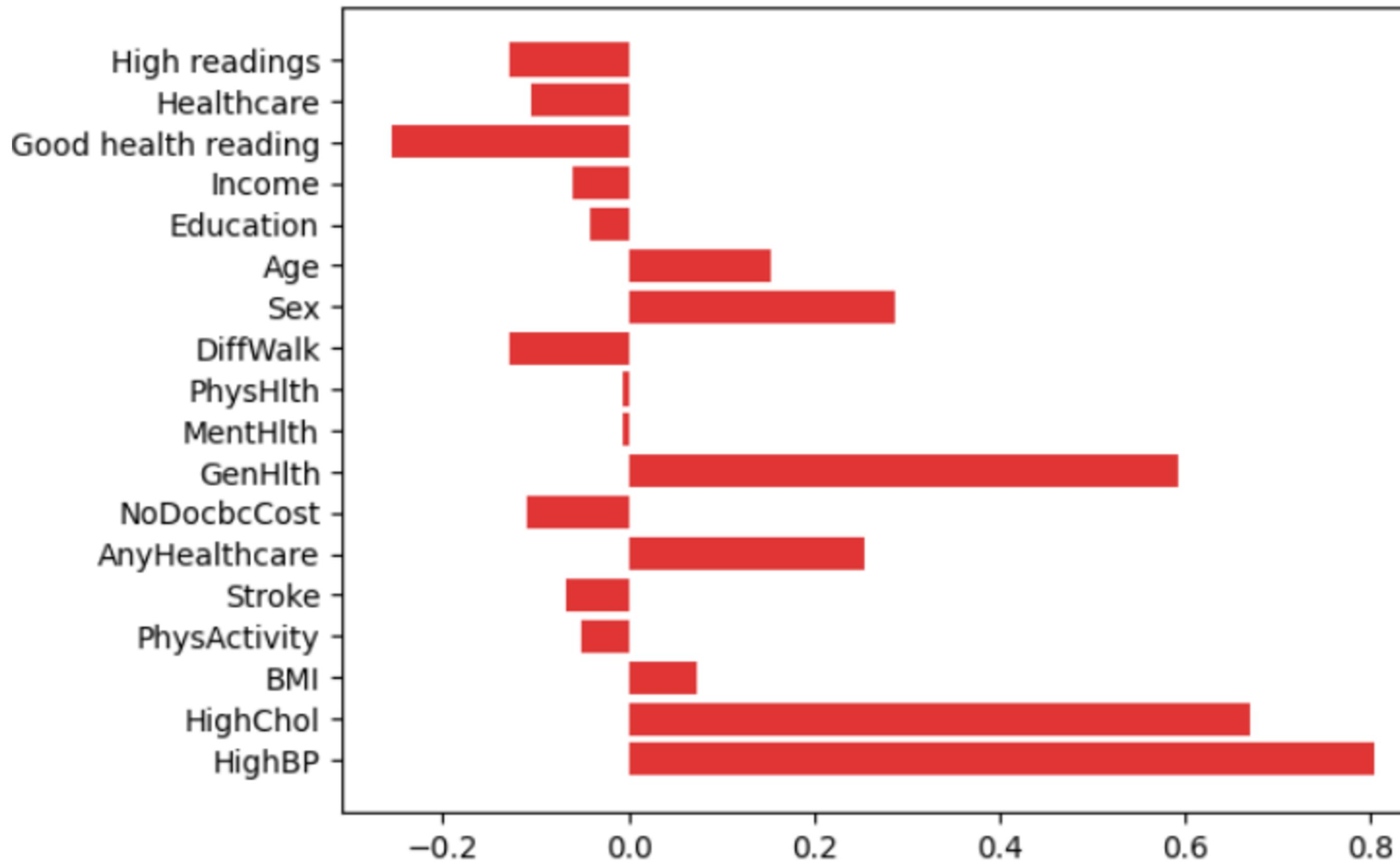


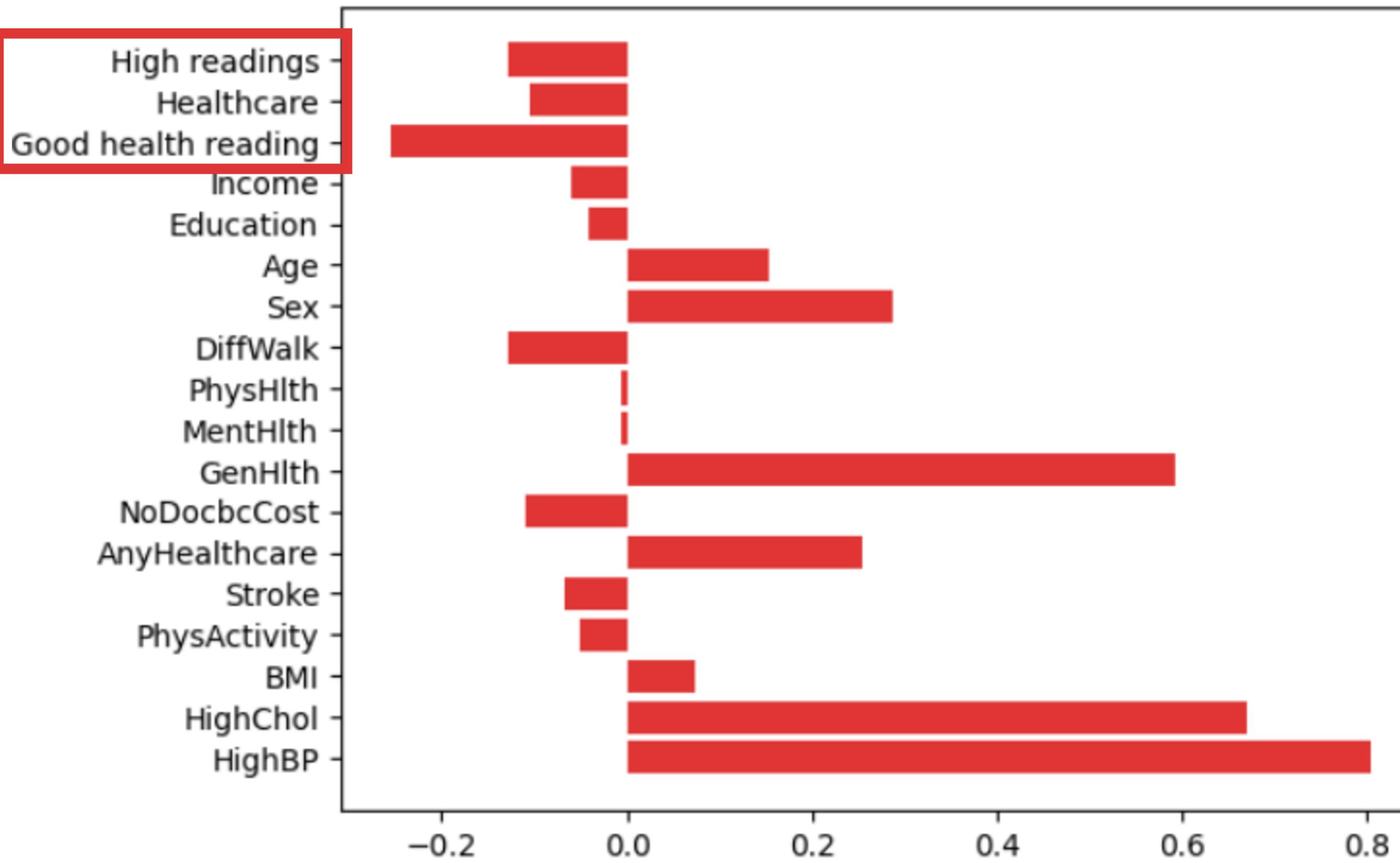


Coefficient of Logistic Regression Model

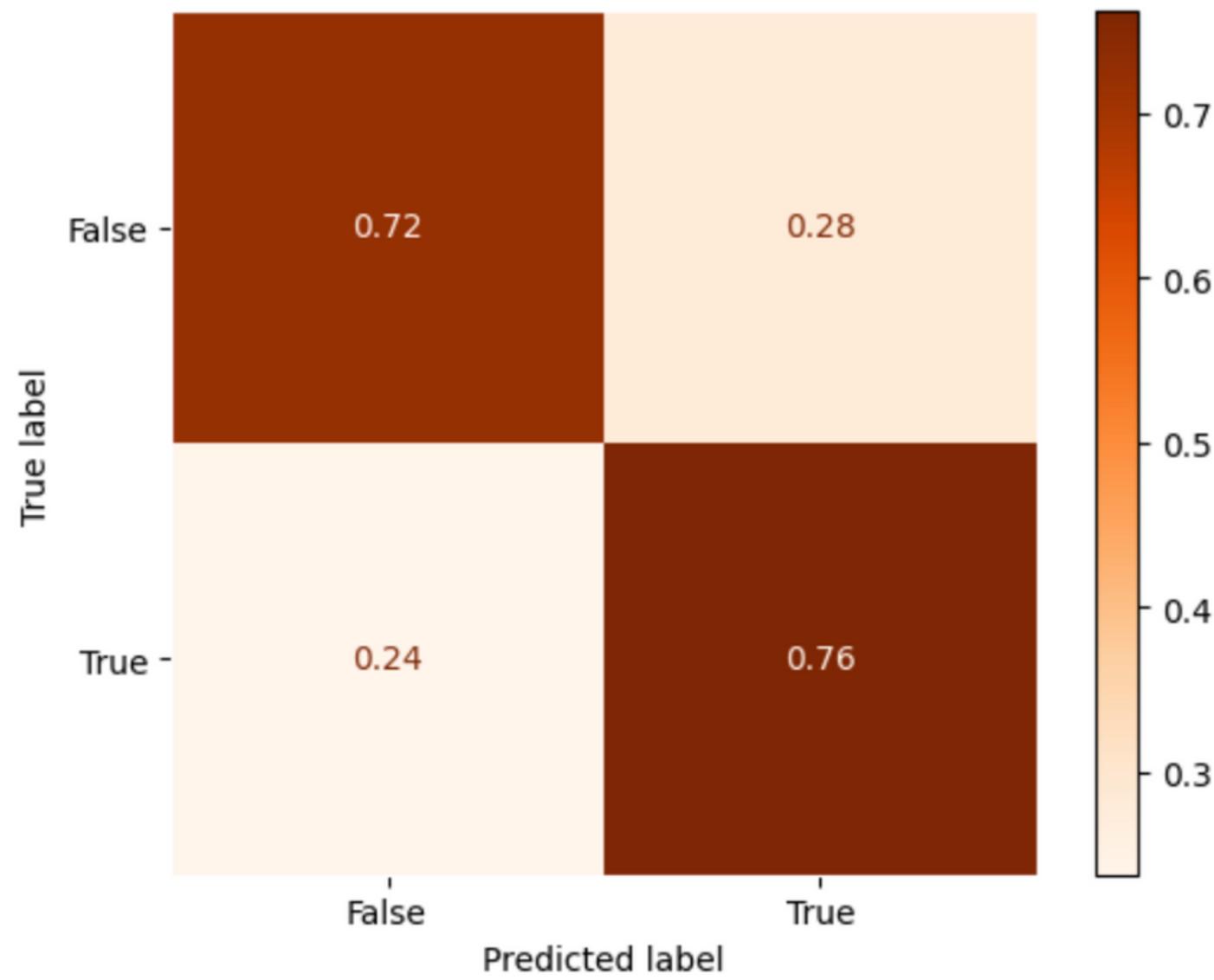
Alcohol prevents your liver from creating glucose, dropping the blood sugar.





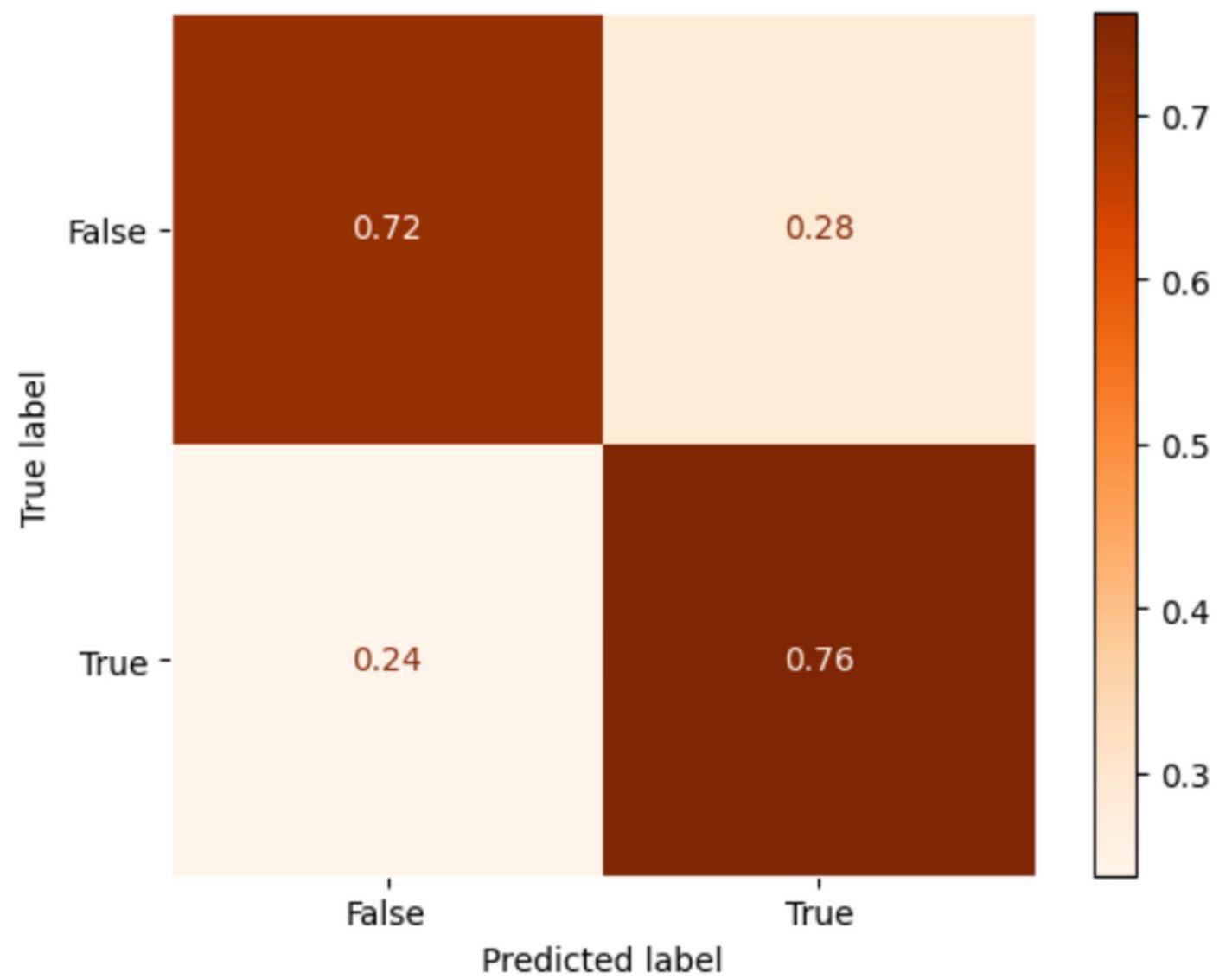


Increasing prediction



With our changes we manage to bring up our accuracy score:

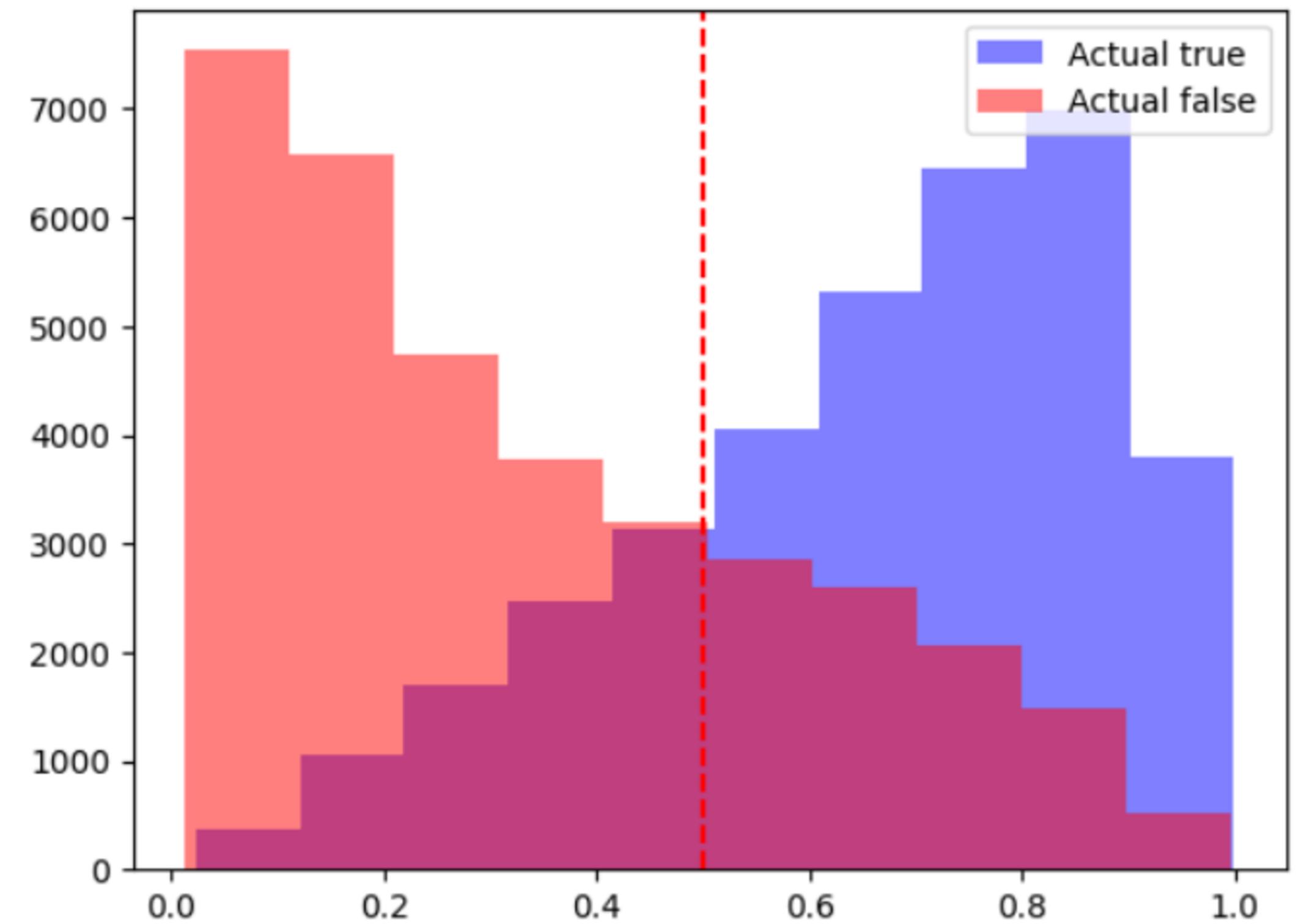
Increasing prediction



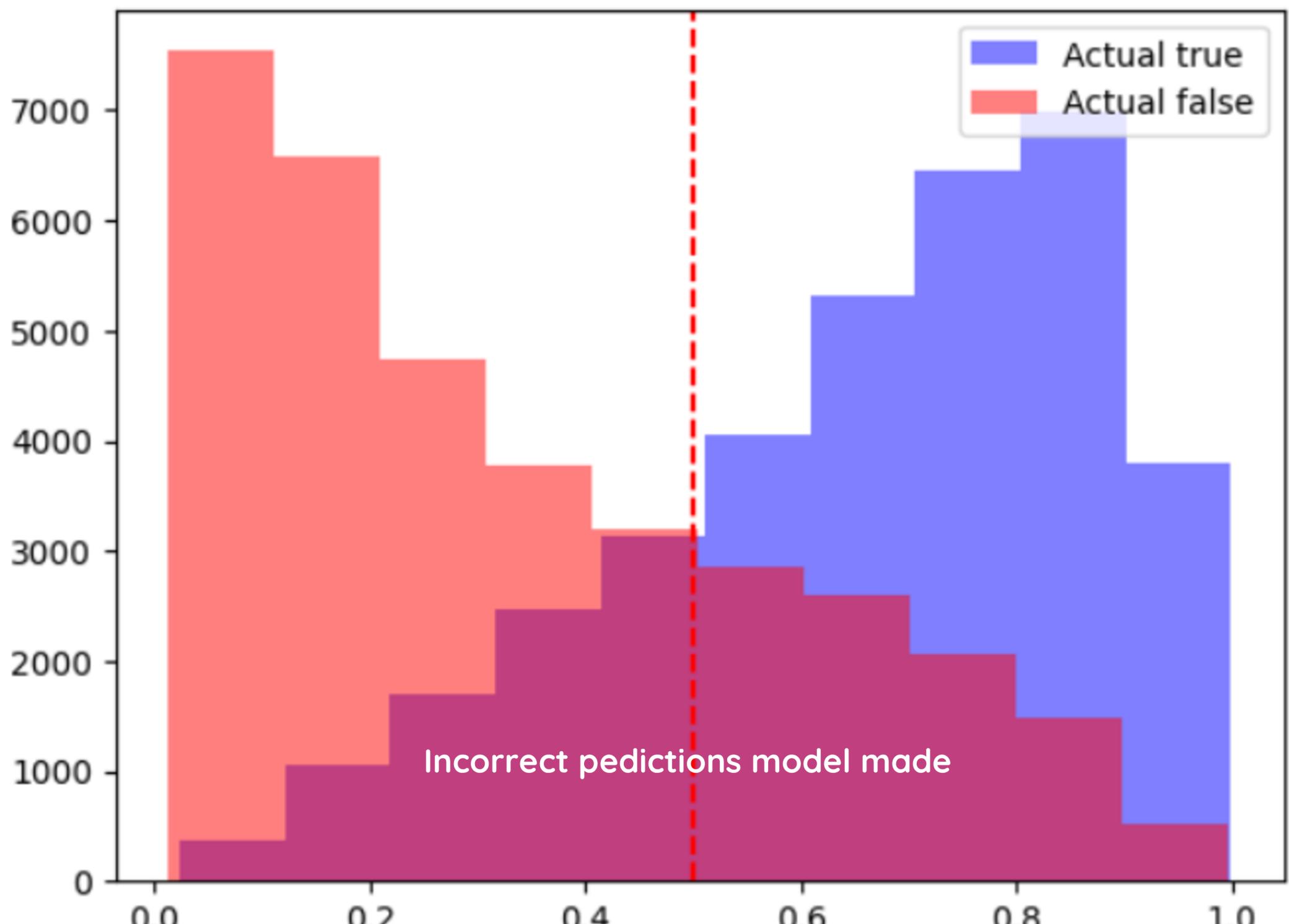
With our changes we manage to bring up our accuracy score:

0.02

Class Separation Histogram



Class Separation Histogram



Demographic Information

Income

Significant?

Gender

Significant?

Age

Significant?

Demographic Information

Income

NO

Gender

NO

Age

NO

Conclusion



Can the data collected here provide **accurate predictions** of whether an individual has diabetes?

What **risk factors** are most **impactful/predictive** of an individual having diabetes?



Conclusion



Can the data collected here provide **accurate predictions** of whether an individual has diabetes?

Yes, up to **74%** of predictions are accurate



What **risk factors** are most **impactful/predictive** of an individual having diabetes?

Conclusion



Can the data collected here provide **accurate predictions** of whether an individual has diabetes?

Yes, up to **74%** of predictions are accurate



What **risk factors** are most **impactful/predictive** of an individual having diabetes?

High BP
High Cholesterol
General Health

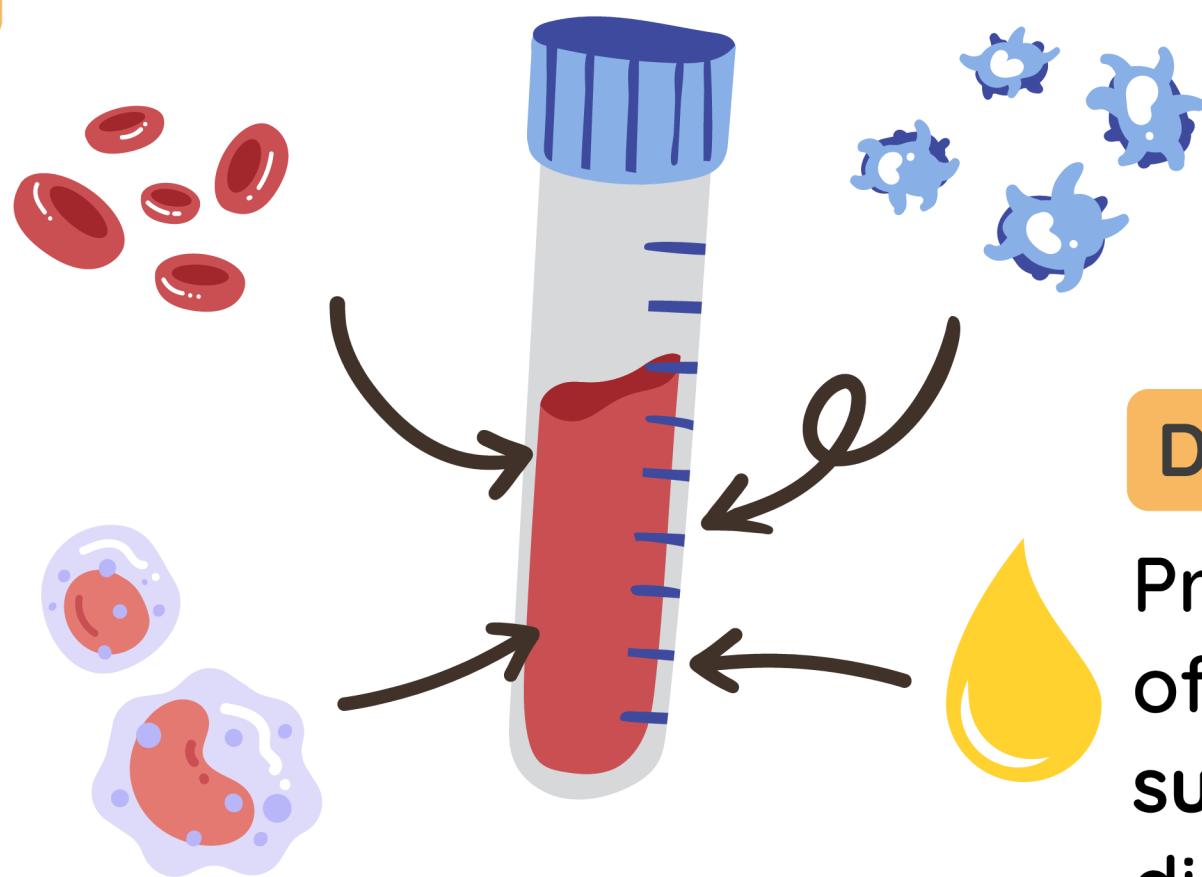
— why 74%? —

Genetics & Family History

Genetics play a major role in diabetes risk because of inherited factors.

Diet Details

Sugar and carb intake, directly affect blood glucose levels.



Medication Use

Certain medications affect blood sugar and cholesterol.

Direct Glucose Measurements

Provides a direct assessment of how the body processes sugar. Directly indicates diabetes, while other factors suggest risk.



Questions?