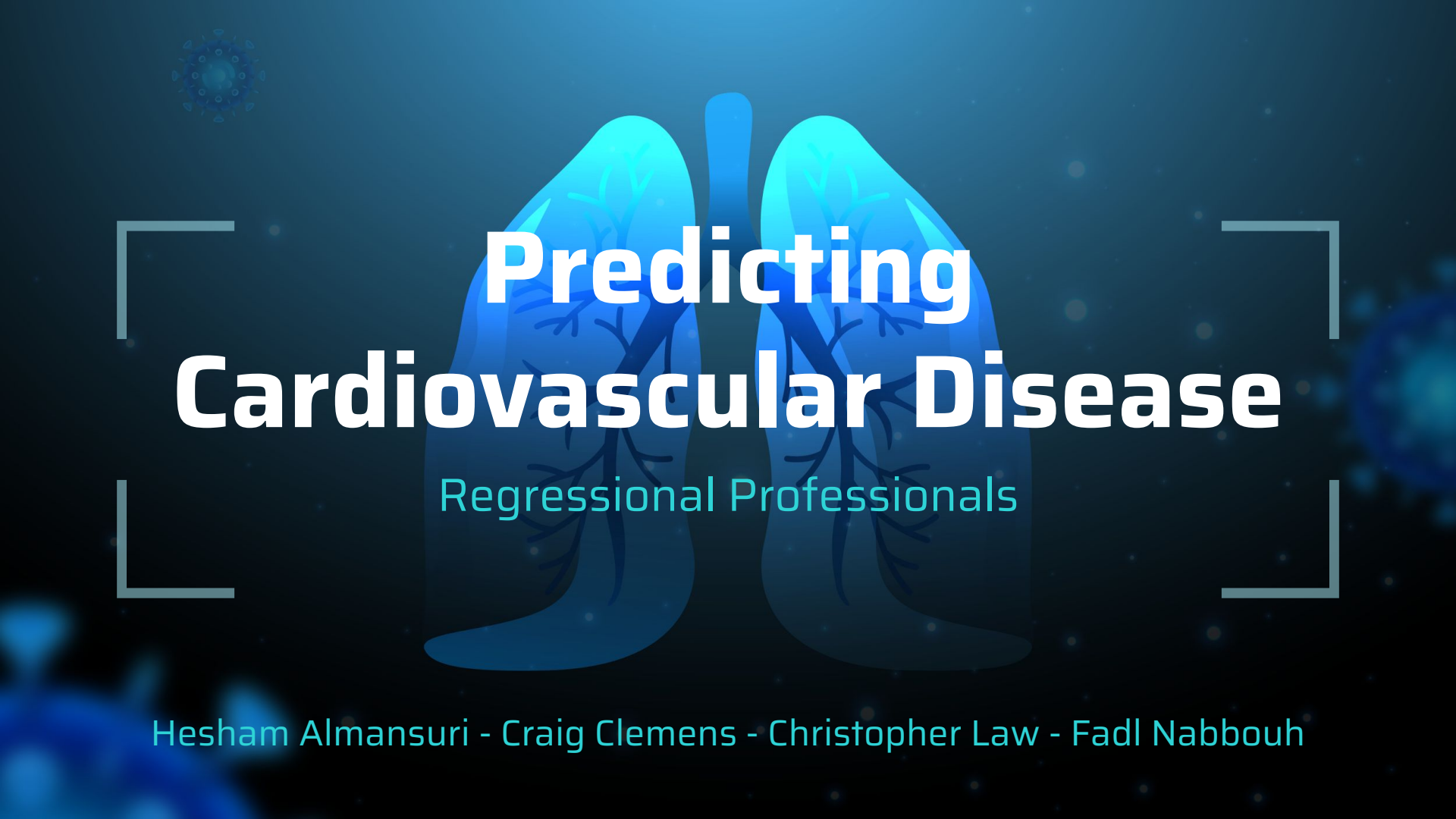


Predicting Cardiovascular Disease

Regressional Professionals

Hesham Almansuri - Craig Clemens - Christopher Law - Fadl Nabbouh

The Problem and our Goal

01

Cardiovascular disease is one of the leading causes of death worldwide.

02

As with any disease, early detection is key to reducing the number of deaths. However, this remains challenging due to asymptomatic nature of most cardiovascular diseases

03

We are attempting to create a model that will accurately predict, to a high accuracy, the prevalence of cardiovascular disease in the general population

Our Primary Dataset

Elements:

- Age
- Height
- Weight
- Gender
- Ap_hi (Systolic Blood Pressure)
- Ap_lo (Diastolic Blood Pressure)
- Cholesterol
- Smoking
- Alcoholism
- Active Lifestyle

Conversions:

- Age / 365 = Age in Years
- BMI (kg/m2) = weight(lbs)/height(in)2

Link: https://www.kaggle.com/sulianova/cardiovascular-disease-dataset

What is *not* a strong predictor?

Where is the noise in our dataset?

Geography

Cardiovascular disease affects people regardless of where they live

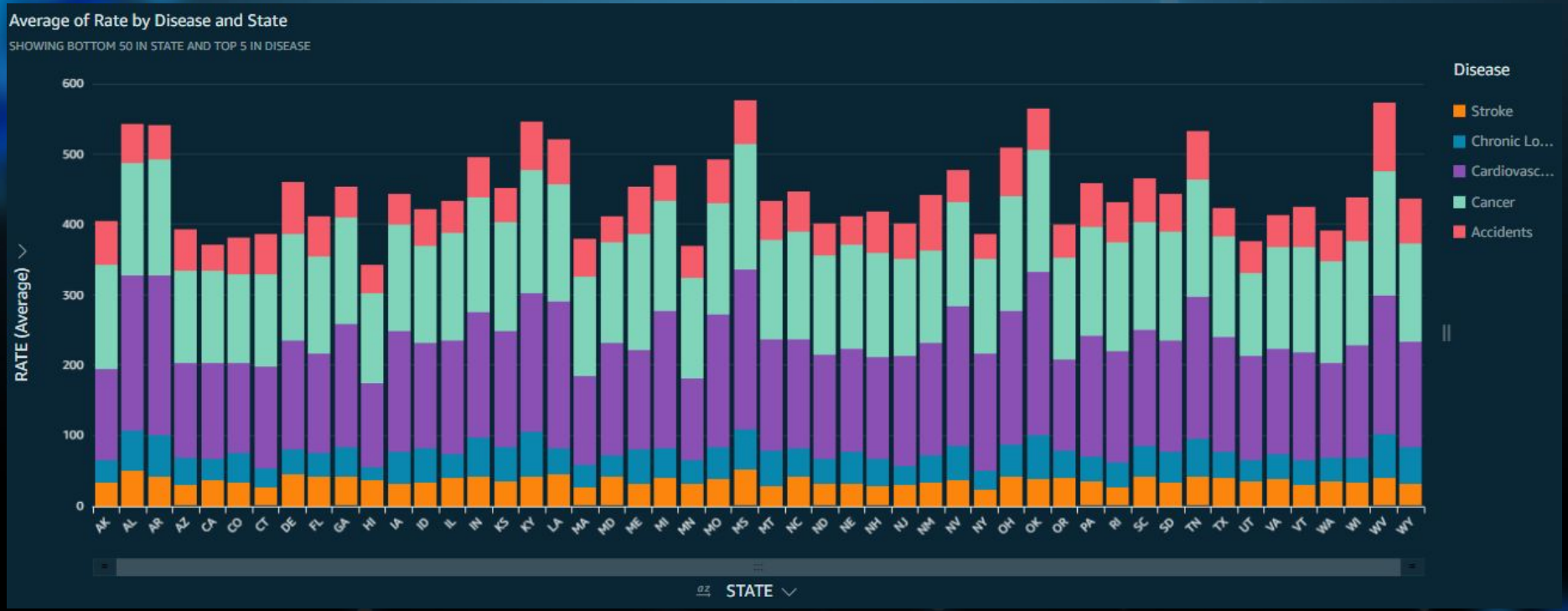
Gender

Although there are more males in our dataset, gender is not a predictor of cardiovascular health

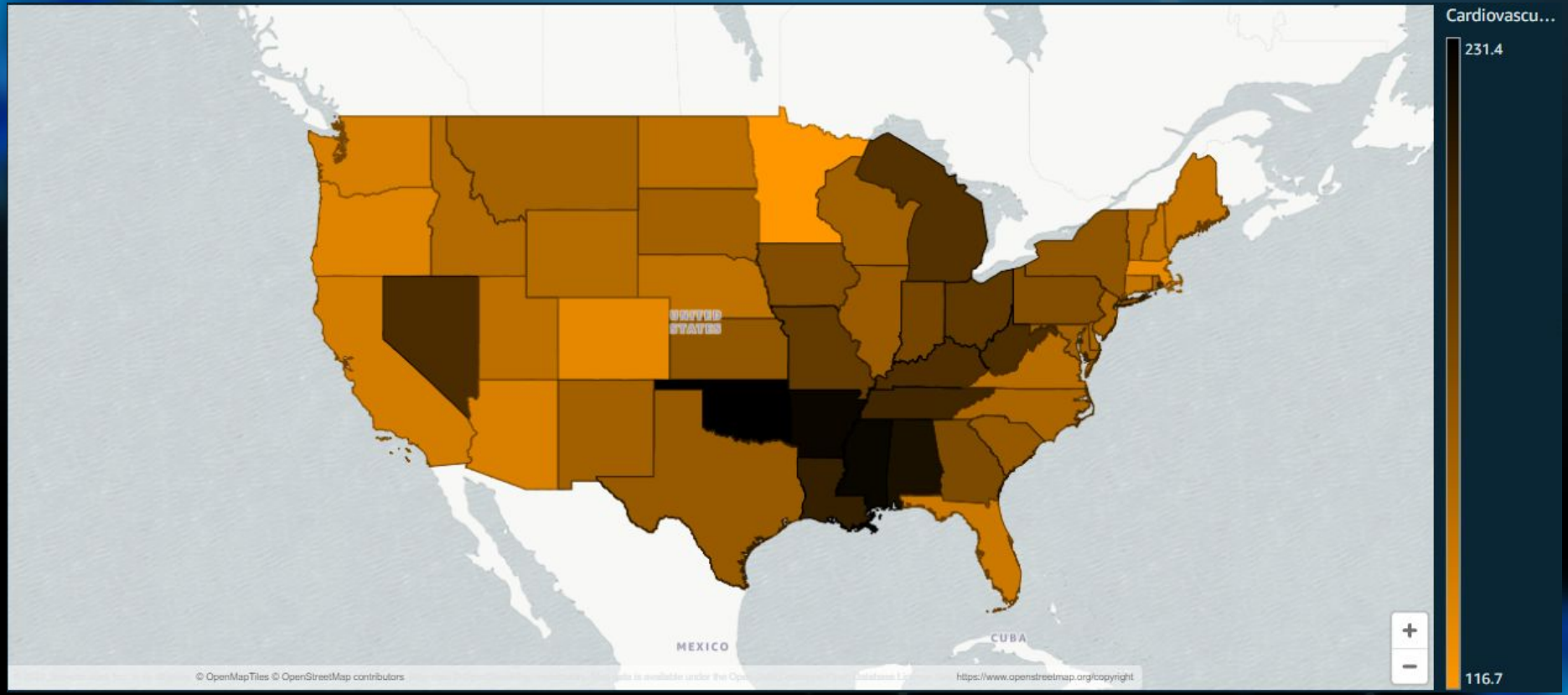
Smoking

Even if it is a predictor of other issues such as cancer, surprisingly it is not a strong predictor

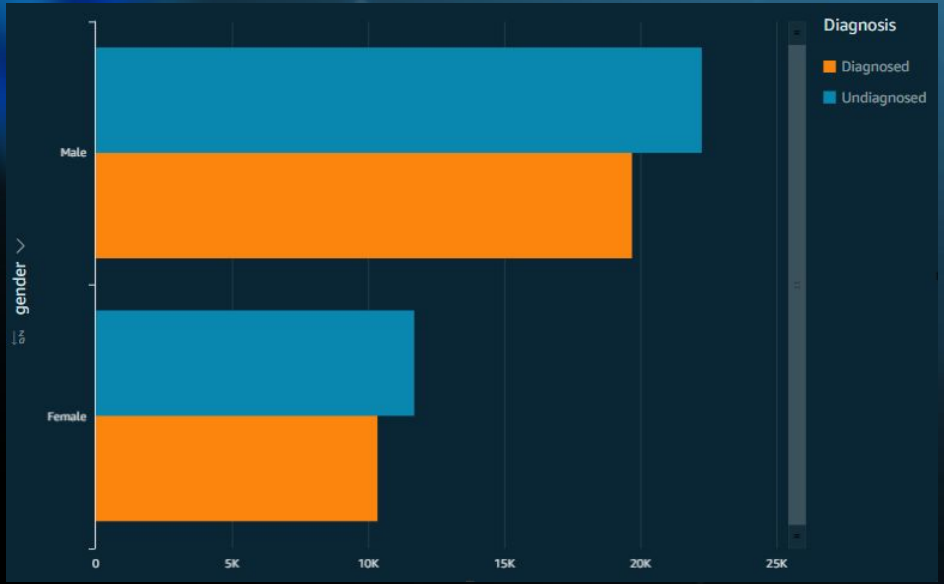
Causes of Death by State

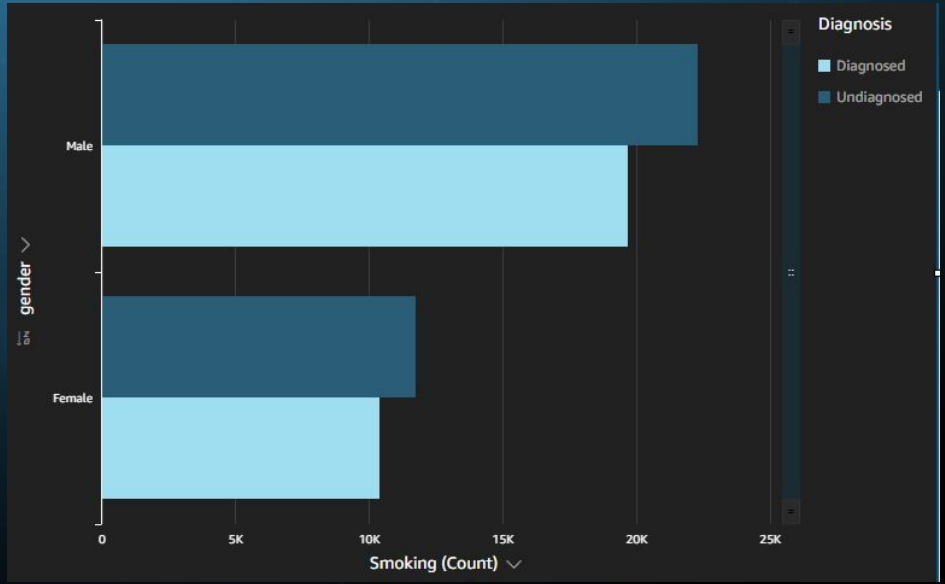


Cases of Cardiovascular Disease by State



Cases by Gender & by Smoking





More “Lifestyle Data” Veracity

Alcoholism

Although damaging to the liver, not an accurate predictor

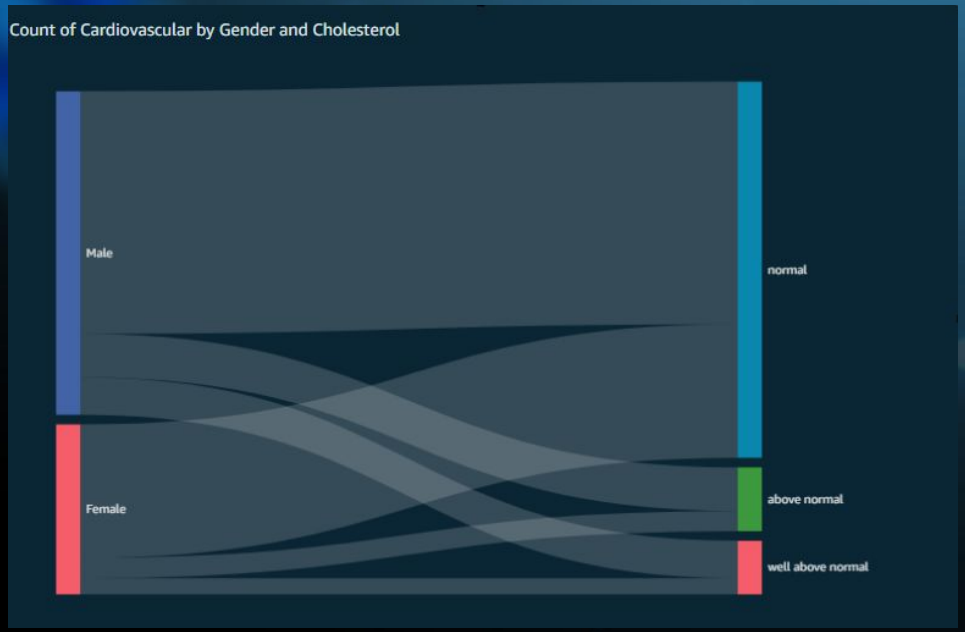
Cholesterol & Glucose levels

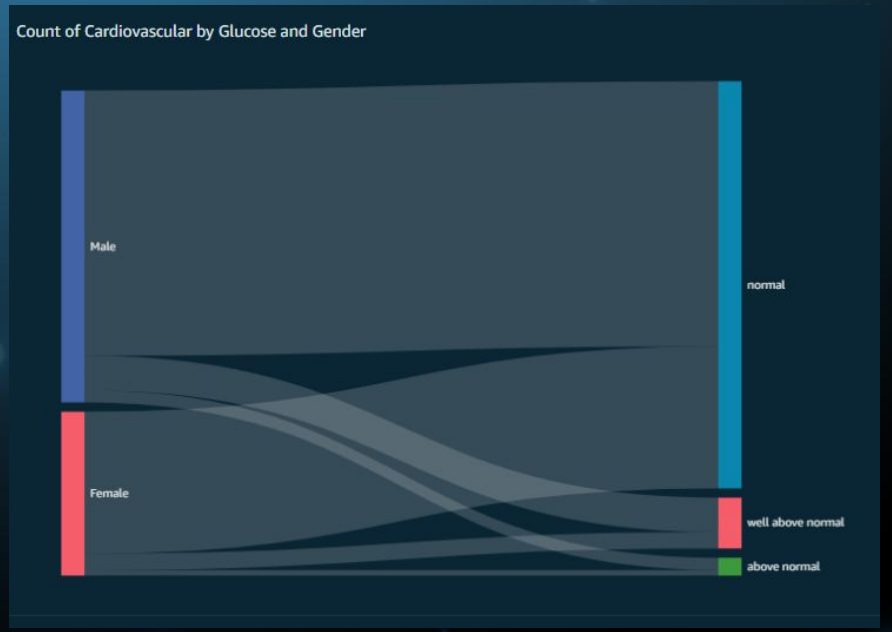
A objective number, but not a direct correlation

Active Lifestyle

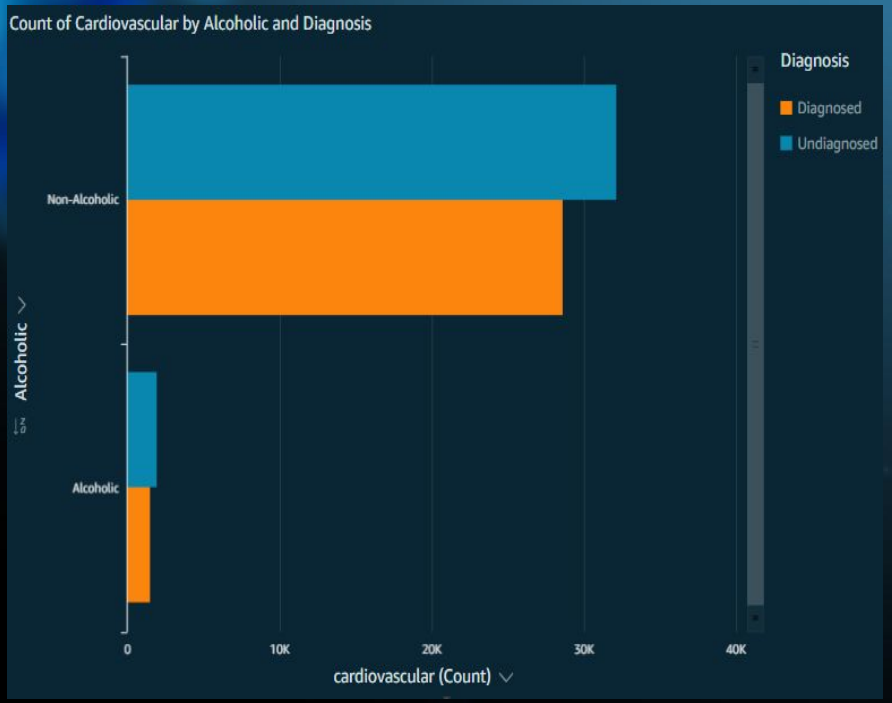
Even those who said they smoked and drank, still said they were “active”

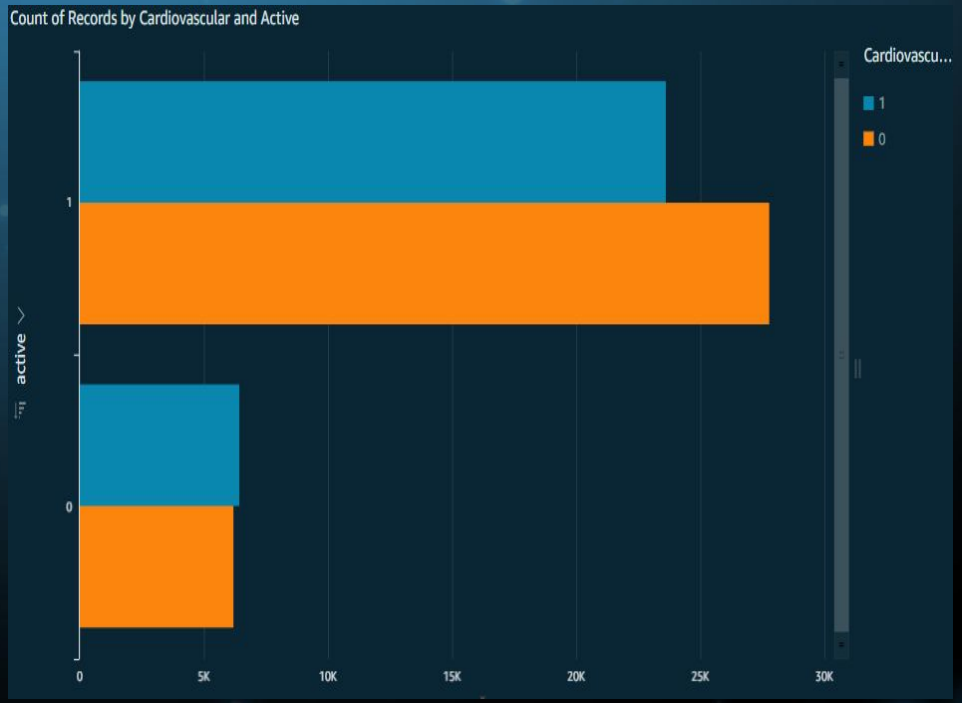
Cases by Gender and Cholesterol





Alcoholism and Active Lifestyle

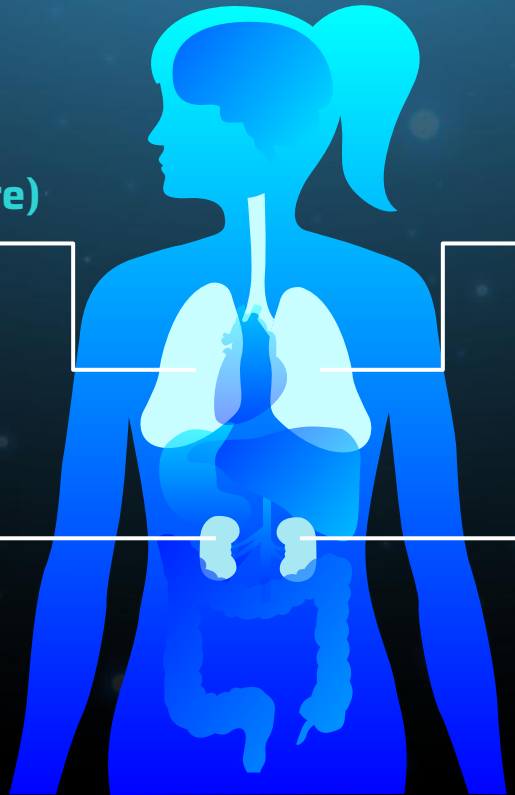




What *is* a strong predictor?

What statistics should we focus on?

Strong Predictors of Cardiovascular Disease



AP_HI (Systolic Blood Pressure)

The force your heart exerts on the walls of your arteries each time it beats

AP_LO (Diastolic Blood Pressure)

The pressure on your arteries when the heart rests between beats

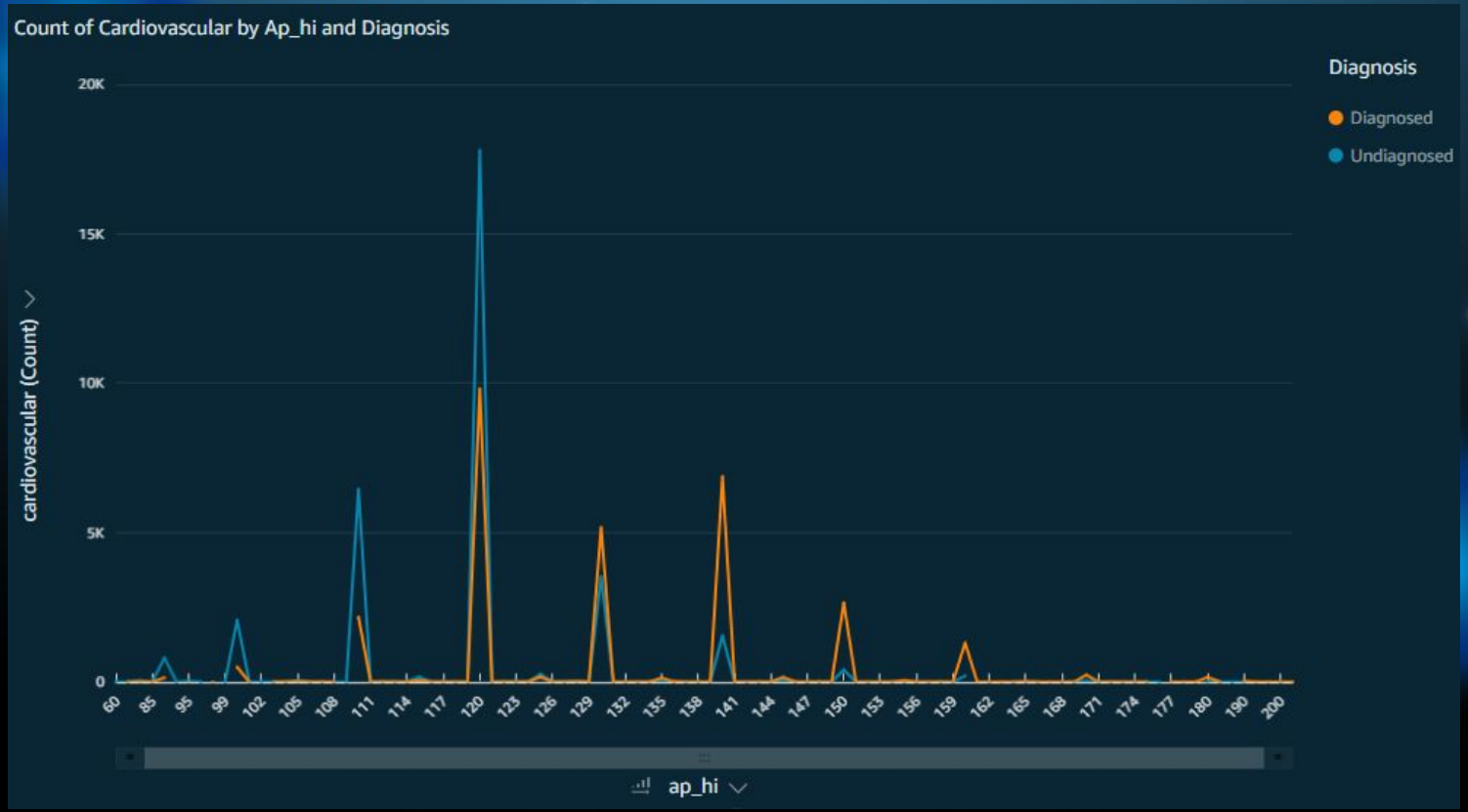
BMI

Body Mass Index which is a calculation of weight and height

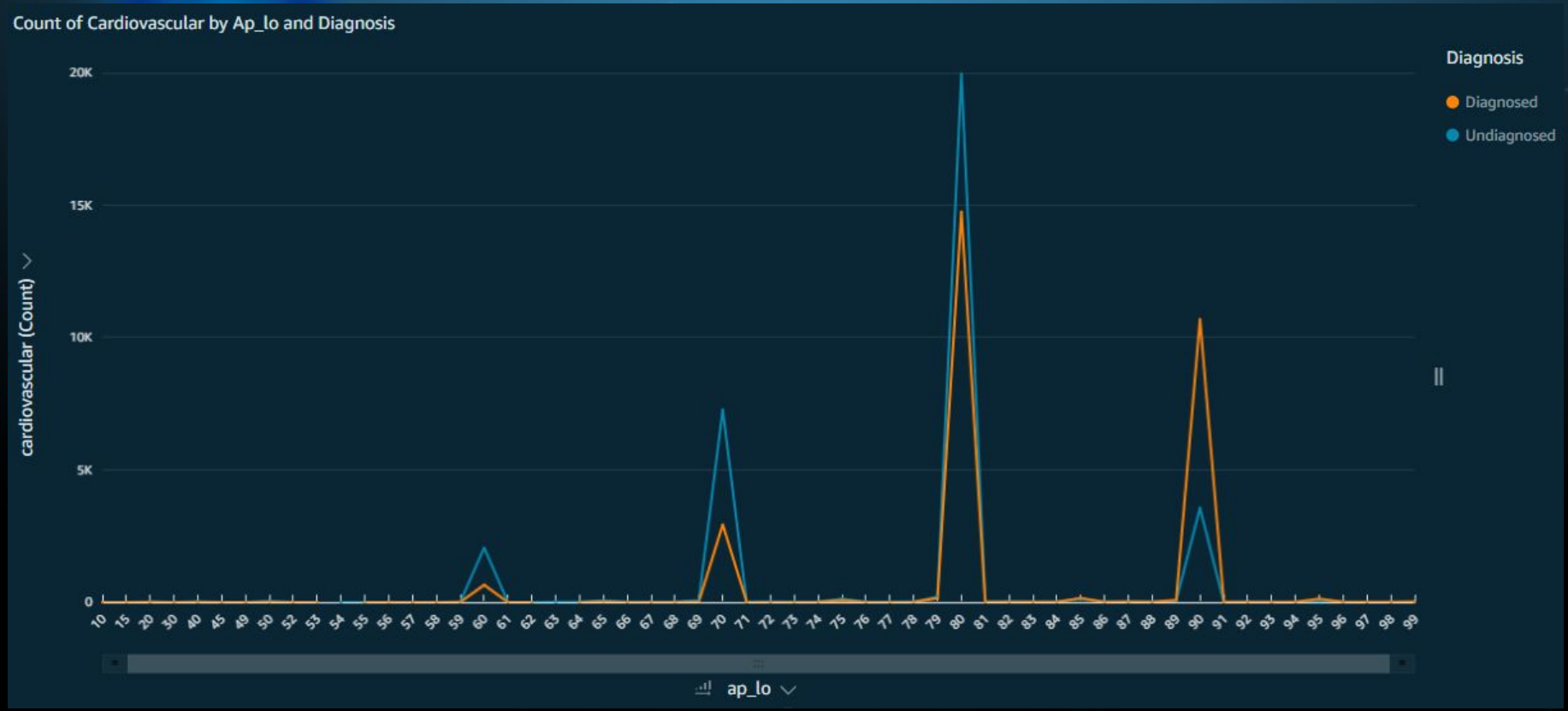
Age

As a person ages they become more susceptible to cardiovascular disease

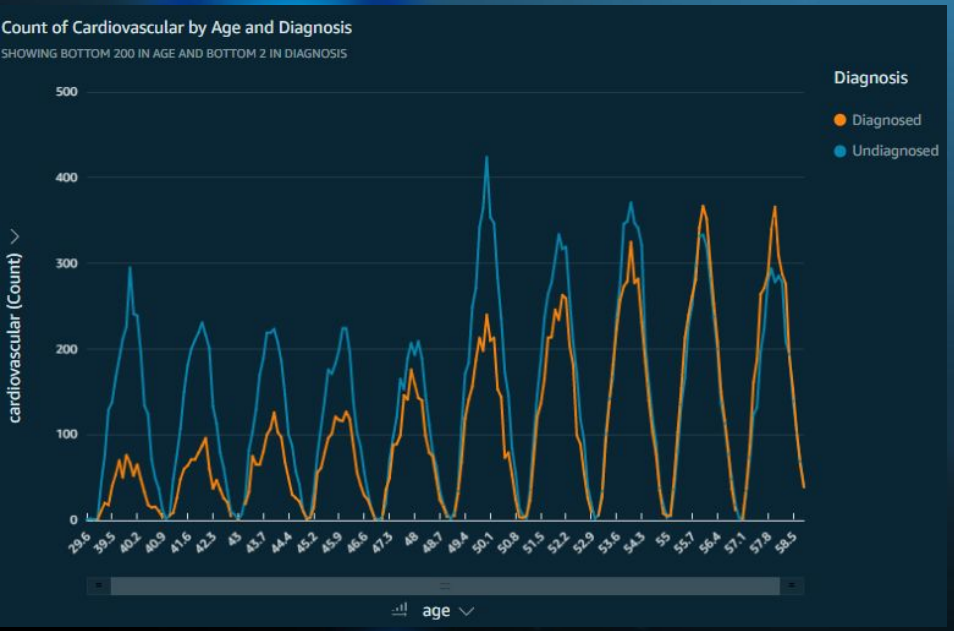
Systolic Blood Pressure vs. Cardio Diagnosis

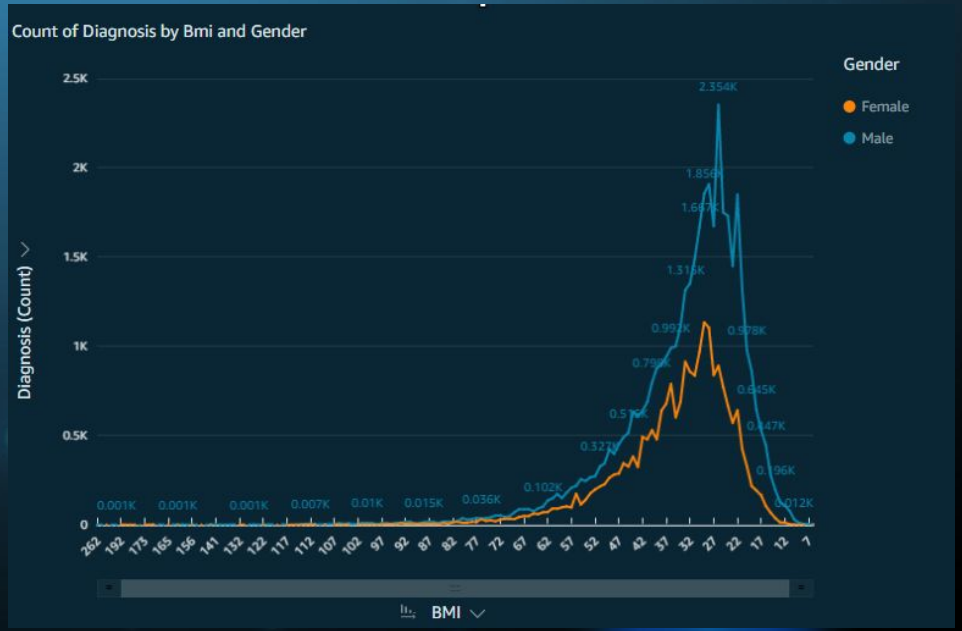


Diastolic Blood Pressure vs. Cardio

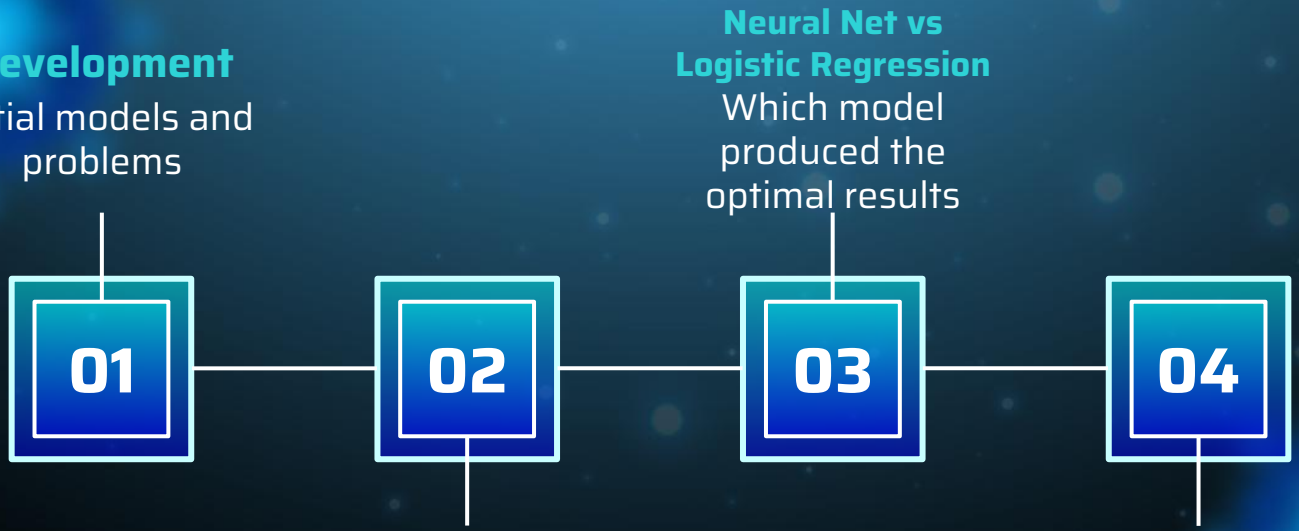


Age and BMI as Predictor





Our Machine Learning Model



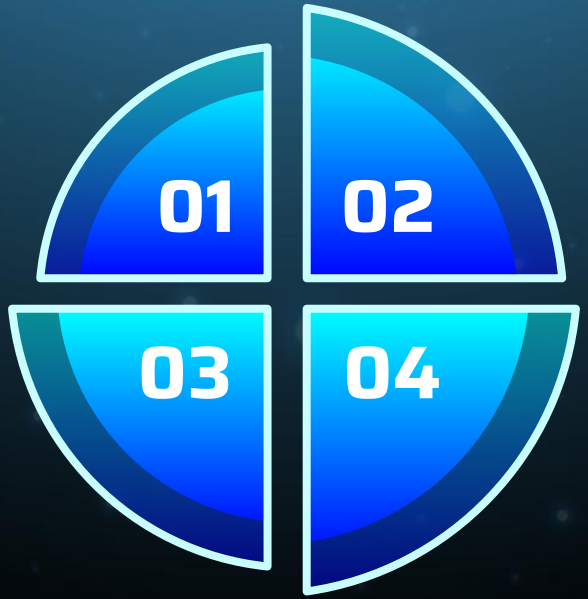
Neural Net Development

Creation

Using our merged, cleaned data we created a fairly standard neural network

Epochs

Although we could theoretically run the model forever, the accuracy peaked at ~77 epochs



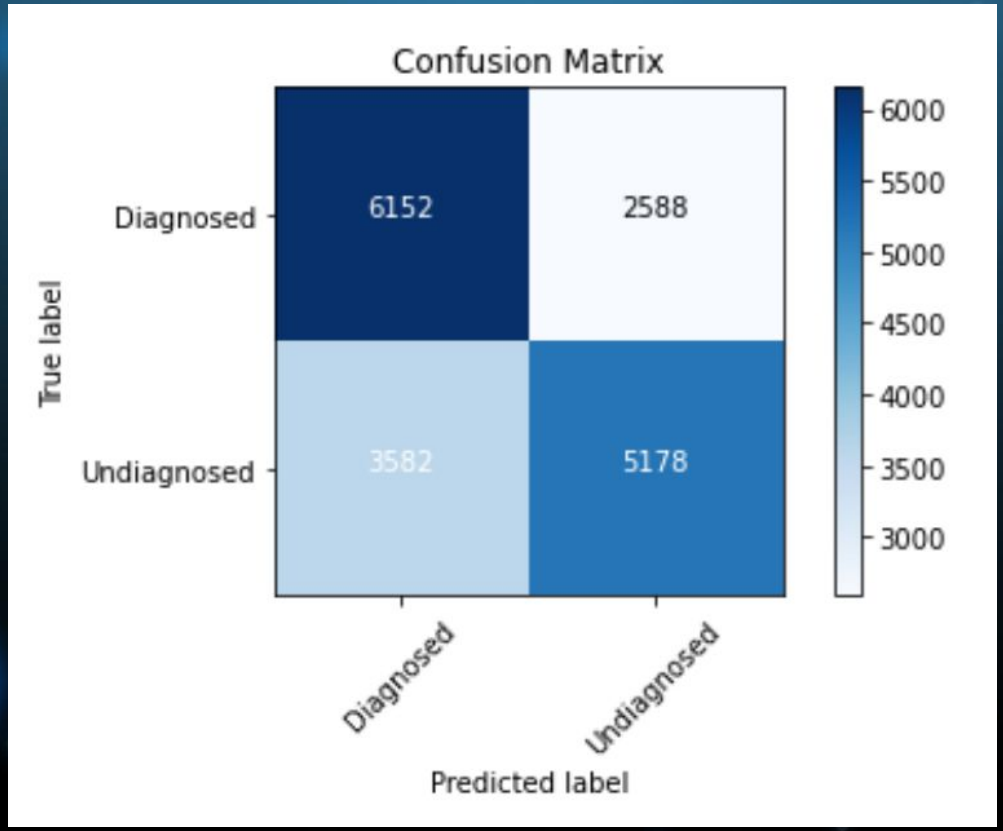
Layers and Activations

After some experimentation we found the ideal activations and layer counts

All about the data

More than anything: having a strong dataset resulted in a satisfactory accuracy

Confusion Matrix



Neural Network vs. Logistic Regression

Neural Network

Has a loss of 54% and an Accuracy of 74%

Overall, good results

Logistic Regression

Had an accuracy as low as 65%

Unsatisfactory for the purposes of dianogsis

Results

Even with an incredibly robust dataset

Our Logistic Regression model lags behind our Neural Net model significantly

73.15%

Using our neural network model we can accurately predict if a patient is at risk of cardiovascular disease more than 7 out of 10 times.

Tools and Technology

AWS

Amazon offered us a one-stop shop for most of the backend services we needed to host our project

Quicksight

AWS also offered us a dashboard builder where we could host our data in a visually appealing way

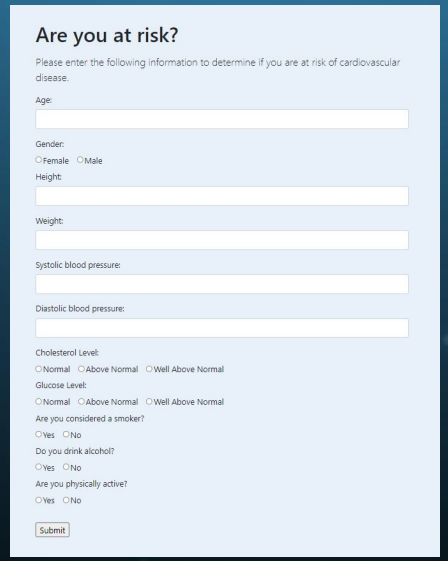
Database Development

Utilizing Postgres we created our database and connected our data to our model using Python

Our application

Our client facing software was developed using Javascript, hosted on electric beanstalk, and connected to our database via our API

Our application



http://cvd-env-v3.eba-ibeaiqtf.ca-central-1.elasticbeanstalk.com/

Thanks

Hopefully people and healthcare professionals who have symptoms of early onset cardiovascular disease can use our tool and seek medical attention and treatment