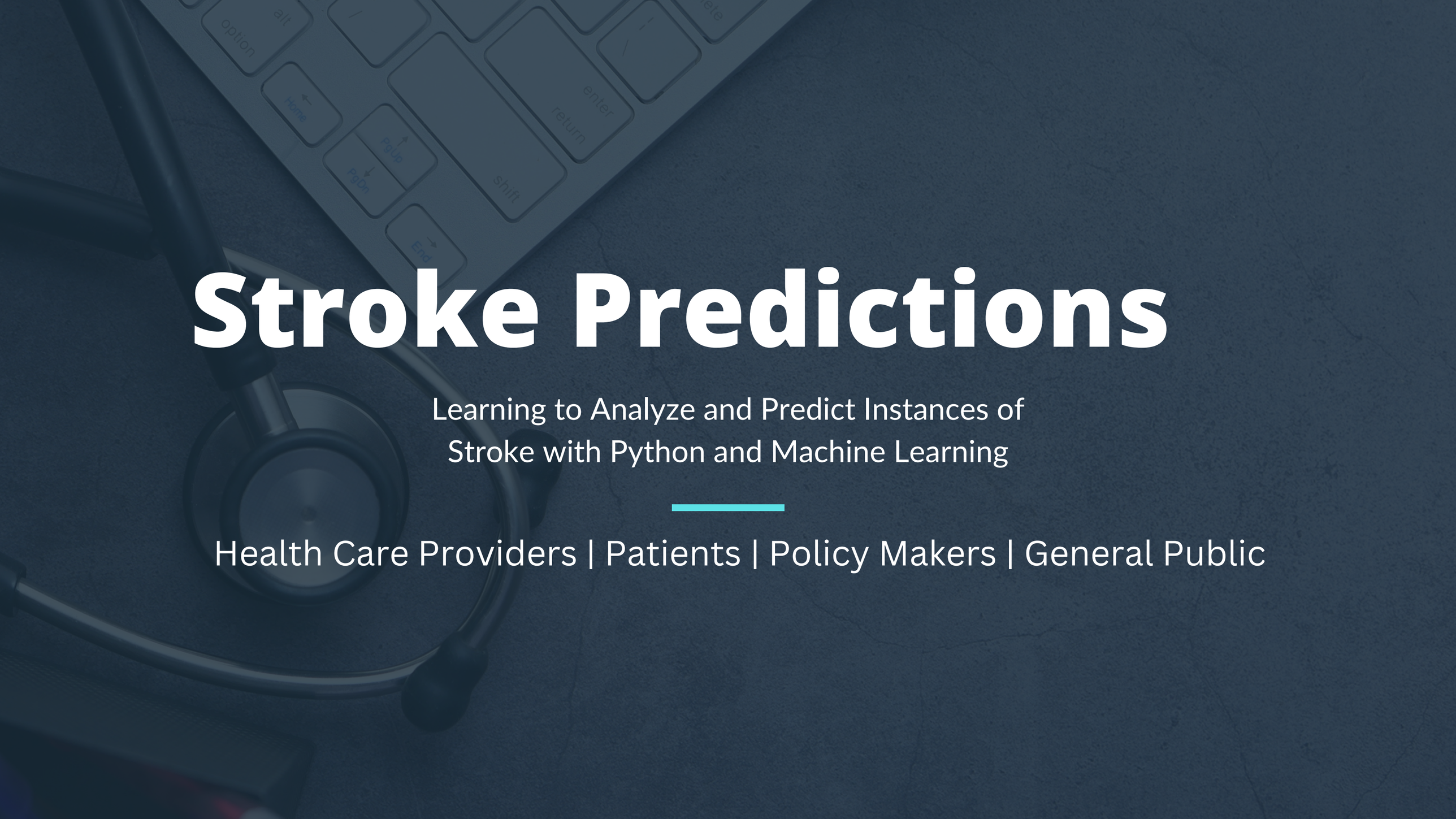


Stroke Predictions

Learning to Analyze and Predict Instances of
Stroke with Python and Machine Learning

Health Care Providers | Patients | Policy Makers | General Public

How well can we predict future instance of stroke, and what data is relevant?

A stroke, also known as transient ischemic attack or cerebrovascular accident, happens when blood flow to the brain is blocked This prevents the brain from getting oxygen and nutrients from the blood. Without oxygen and nutrients, brain cells begin to die within minutes. Sudden bleeding in the brain can also cause a stroke if it damages brain cells.

NIH National Heart, Lung, and Blood Institute

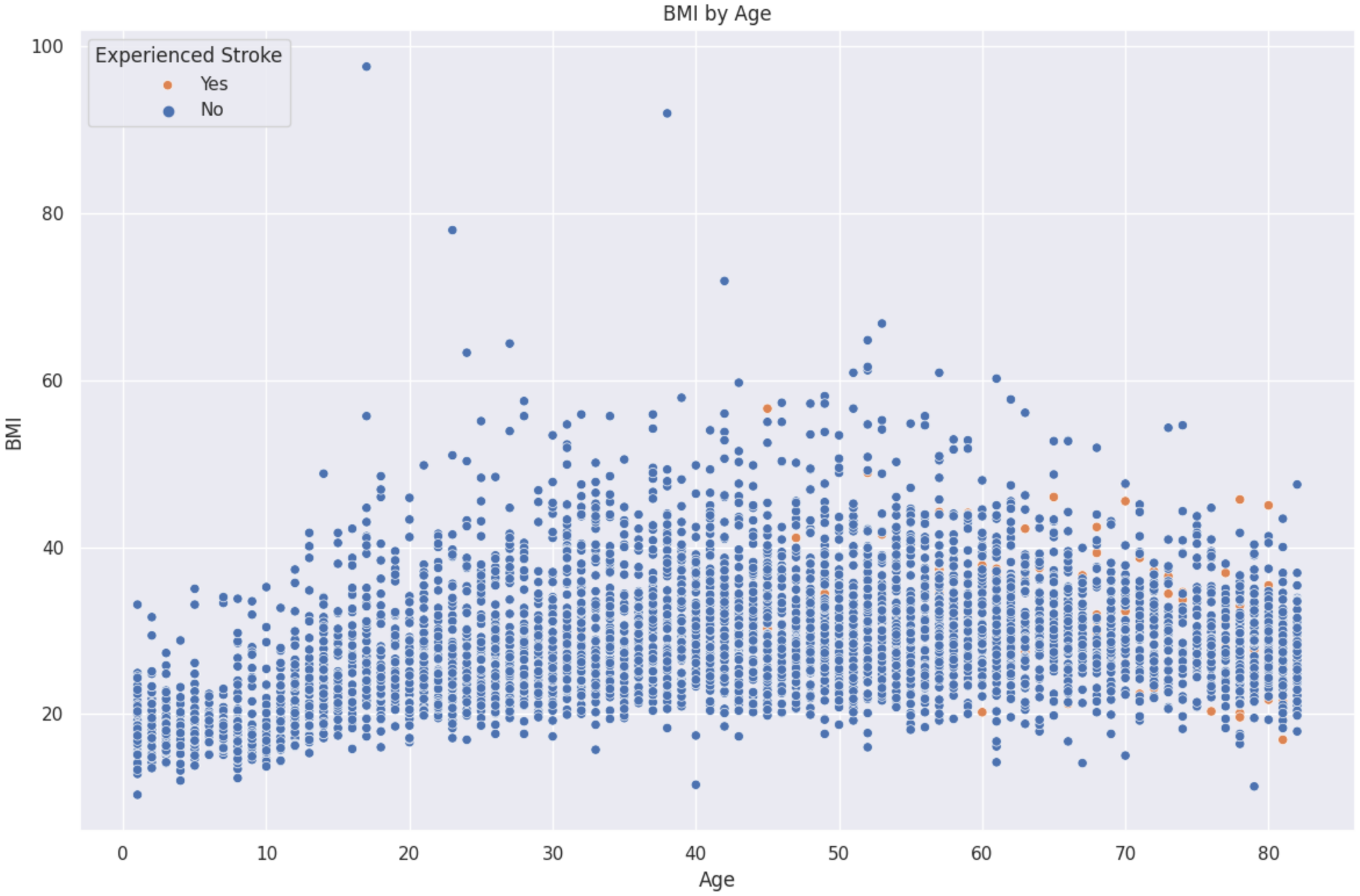
Stroke Prediction Data Set

Initial Data Set

- **"ID"**
  - **unique identifier**
- **"Gender":**
  - **"Male", "Female" or "Other"**
- "Age":
  - Age of the patient
- "Hypertension":
  - 0 if the patient doesn't have hypertension
  - 1 if the patient has hypertension
- "Heart Disease":
  - 0 if the patient doesn't have any heart diseases
  - 1 if the patient has a heart disease
- **"Ever Married":**
  - **"No" or "Yes"**
- **"Work Type":**
  - **"Children", "Govt job", "Never worked", "Private" or "Self-employed"**
- "Residence Type":
  - "Rural" or "Urban"
- "Avg Glucose Level":
  - Average glucose level in blood
- "BMI":
  - Body mass index
- "Smoking Status":
  - "Formerly smoked", "Never smoked", "Smokes" or "Unknown"*
- "Stroke":
  - 1 if the patient had a stroke
  - 0 if not

Prepared Data Set

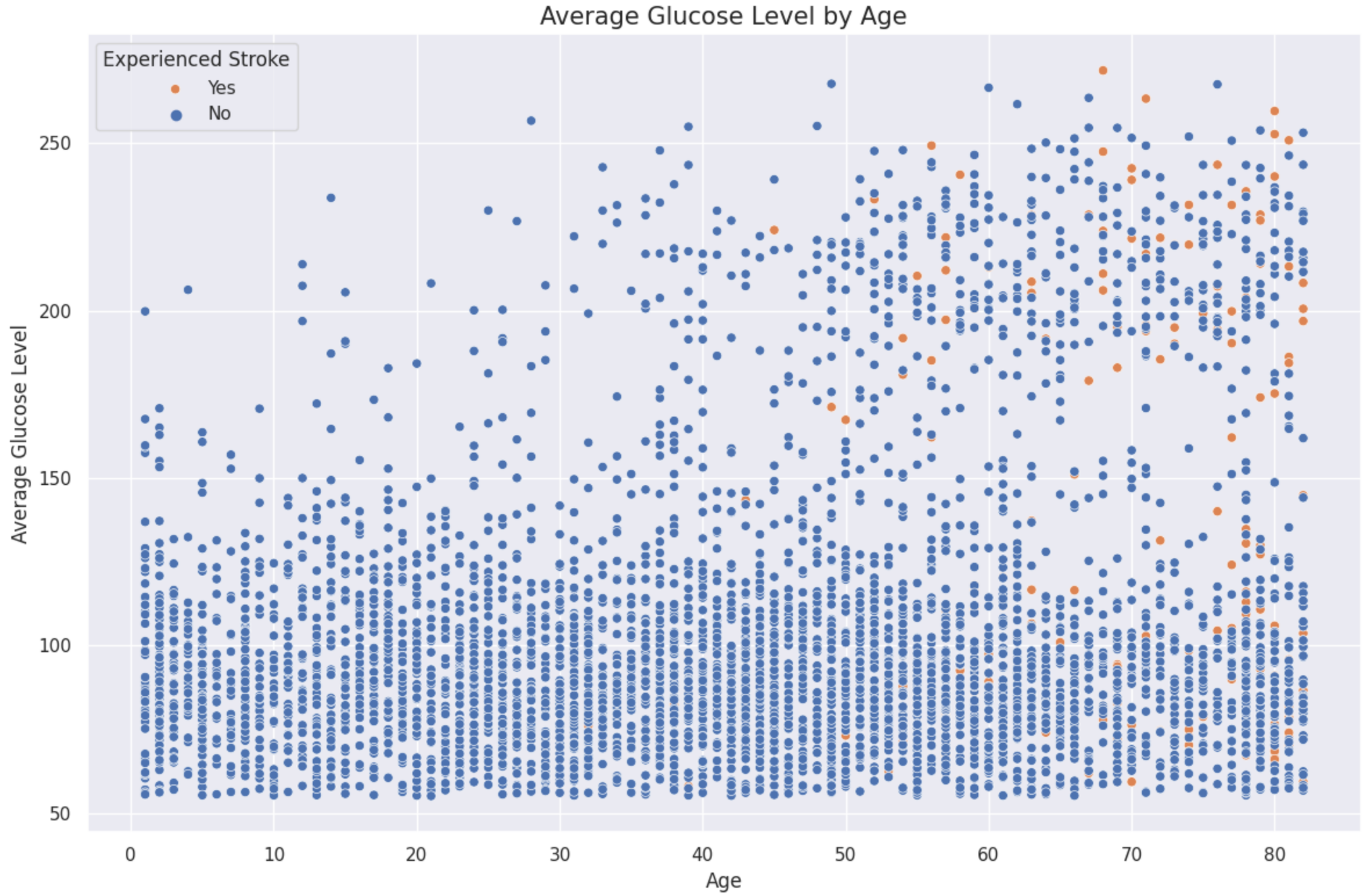
- "Gender":
  - "Male" : 1988
  - "Female" : 2879
- "Age":
  - Age of the patient
- "Hypertension":
  - 0: 4416
  - 1: 451
- "Heart Disease":
  - 0: 4624
  - 1: 243
- "Stroke":
  - 1: 4658
  - 0: 209
- "Work Type":
  - "Children", "Govt job", "Private" or "Self-employed"
- "Residence Type":
  - "Rural" or "Urban"
- "Avg Glucose Level":
  - Average glucose level in blood
- "BMI":
  - Body mass index
- "Smoking Status":
  - "Formerly smoked", "Never smoked", "Smokes" or "Unknown"



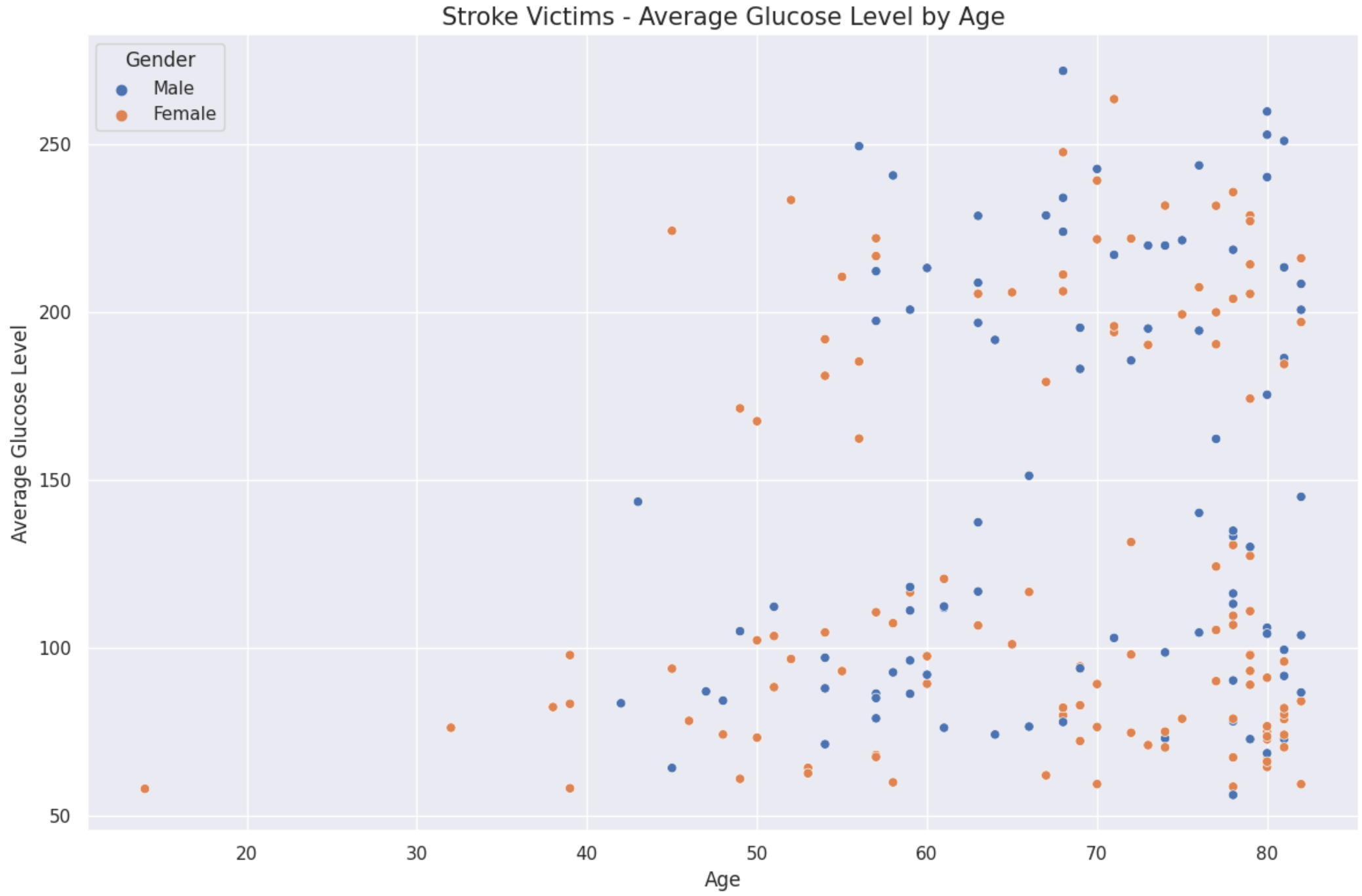
- According to the Centers for Disease Control and Prevention (CDC), a healthy BMI ranges from 18.5 to 24.9

- BMI seems to rise slightly starting at age 30, levels out, then begins to fall slightly around age 65.

- When we isolate the stroke victims, we can see that stroke occurences are more frequent as age increases.
- Most stroke victims have a BMI of 25 or higher - approximately 82.8%
- We see stroke occurrences start earlier in females. This could be due to the higher amount of data we have received from females (Almost 1000 more data points)
- The highest density of stroke occurences are from ages 70 - 85.

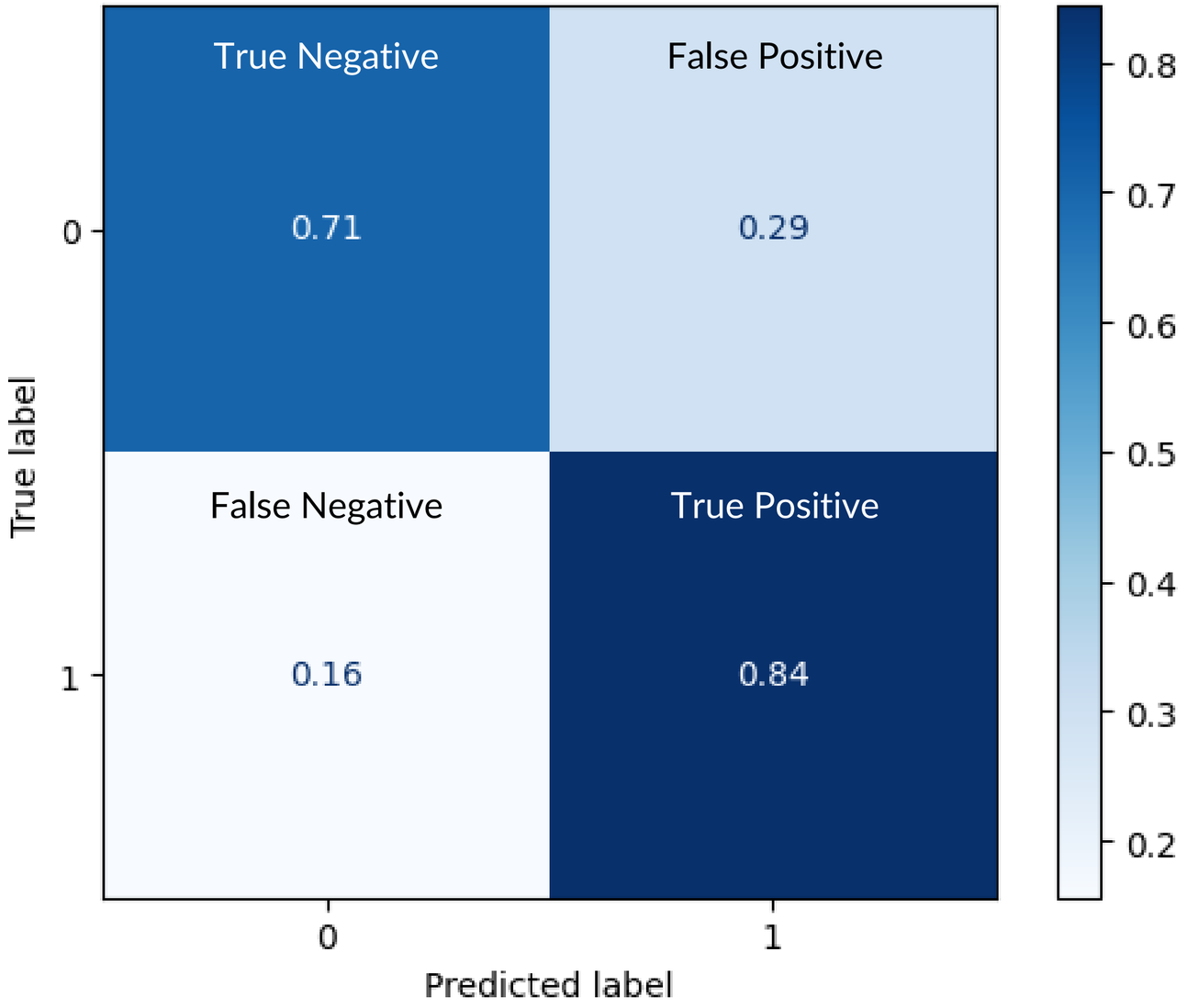


- 570 of 4,867, or approx. 11.71%, of data points are outliers
- 73 of 570, or approximately 12.81%, of avg. glucose level outliers were stroke victims
- 47% of outliers are male, while 53% are female



- 73 of 209, or approximately 34.93% of stroke victims had outlying avg. glucose levels
- Of all stroke victims without outlying average glucose levels, 45% male and 55% female

Production Model



Accuracy: 71% **Precision:** 94% **Recall:** 71%

- 71% accuracy means that we had 71% correct predictions in our model
- Precision and recall deal with false positives and false negatives, respectively
- In predicting a stroke, while important to accurately predict the possibility of a stroke, it is also very important to lower the chances of any false negatives.
- A false negative in this case is not predicting a stroke, but having one occur.
- While we have a relatively low amount of false positives, the heavily imbalanced distribution of data limits our predictions accuracy and overall production

Final Reccomendations

- Monitor older populations with higher than average BMI and Avg. Glucose Level
  - Particularly around the age of 60 and beyond
- Gathering more data on stroke victims
- Gathering more balanced data in gender