Stroke Predictions

Kai Dawson-Fischer
Data Scientist Consultant

Can physicians predict who will be a stroke victim before one even occurs?

Context

- Strokes are the second most leading cause of death worldwide.*
- Annual Mortality Rate of 5.5 Million*
- 11% of all total deaths

Dataset

Using 11 different Clinical Features:

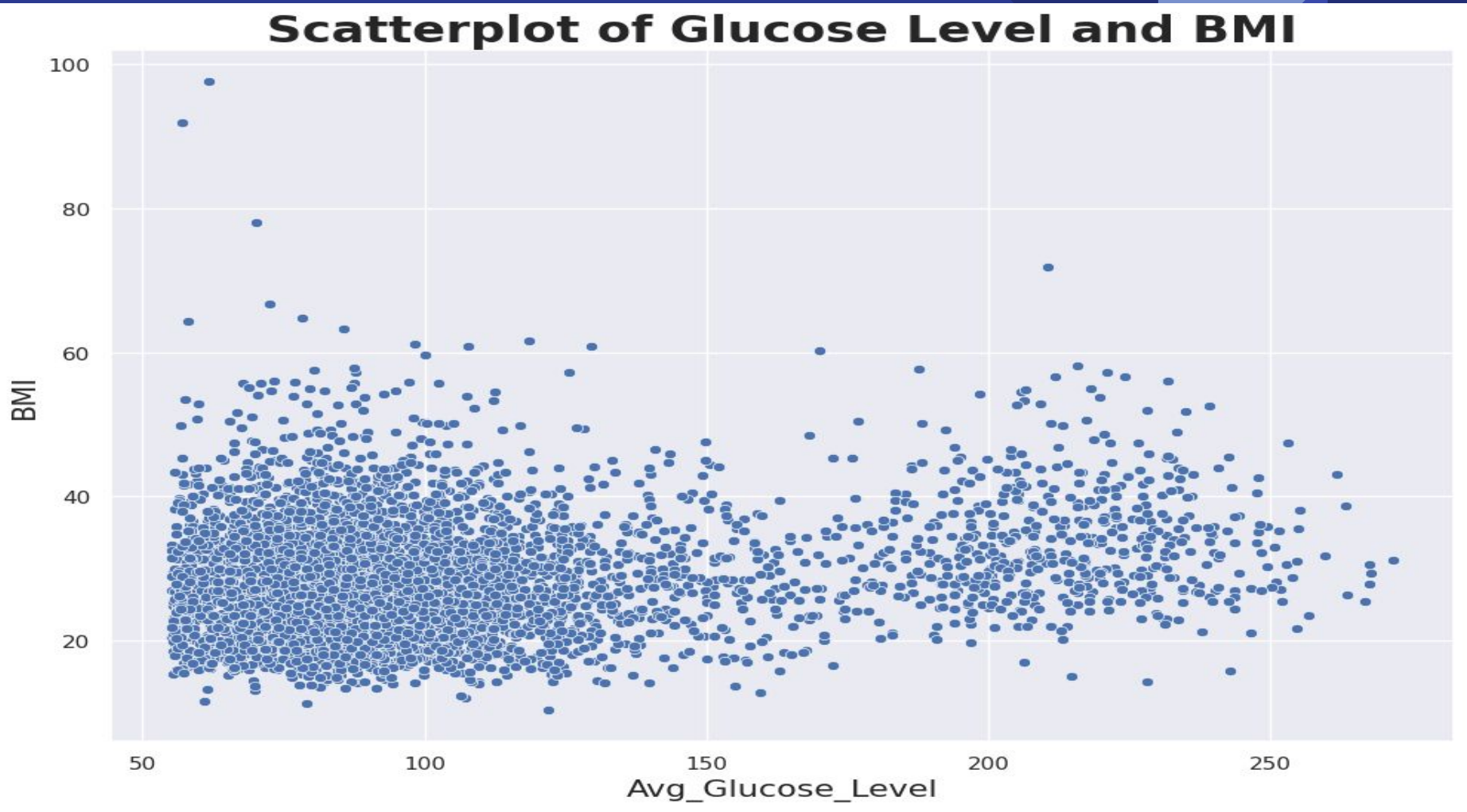
Gender, Age, Hypertension, Heart Disease, Marital Status, Work Type, Residence Type, Average Glucose Level, BMI, and Smoking Status.

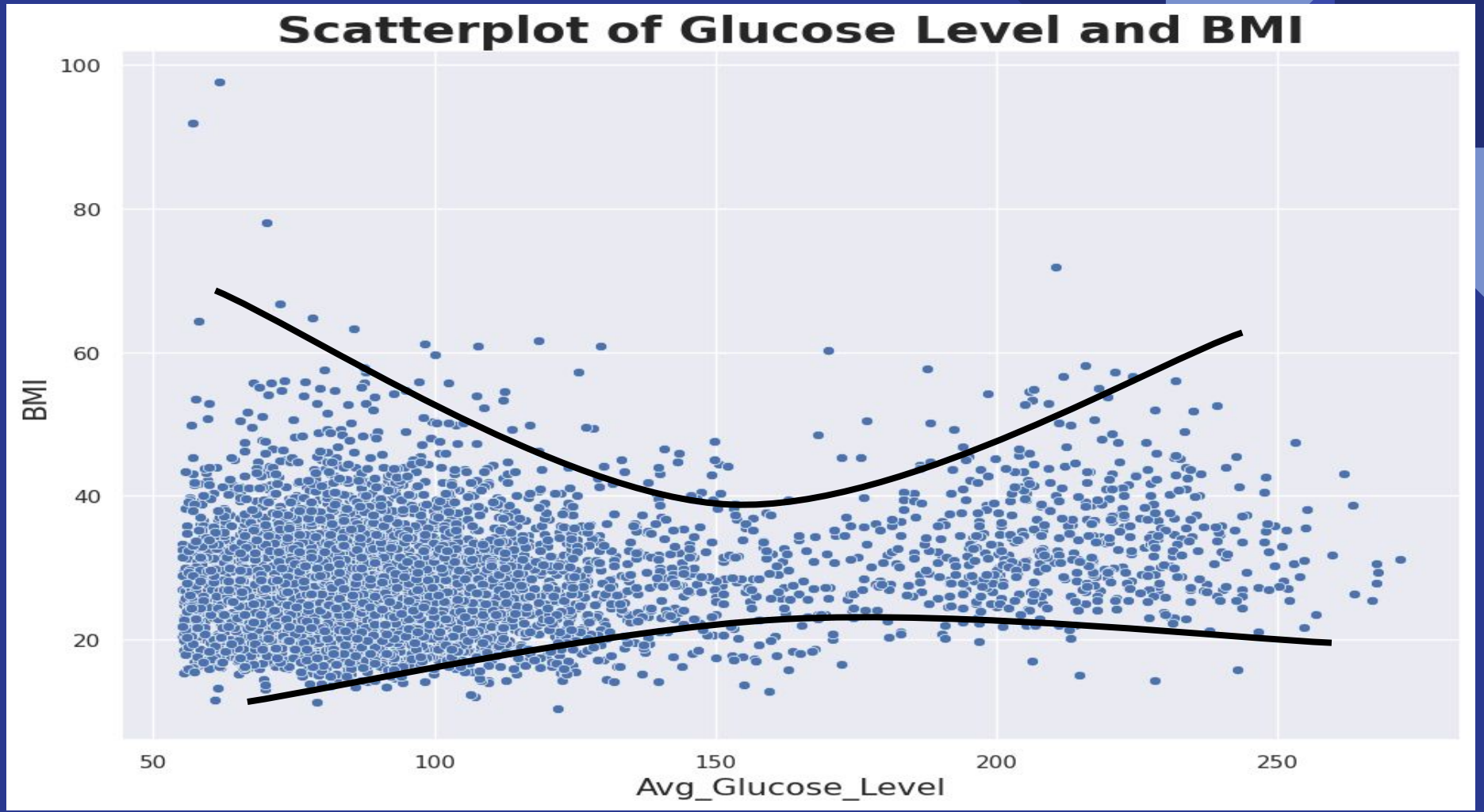
Target:
Stroke Occurrences

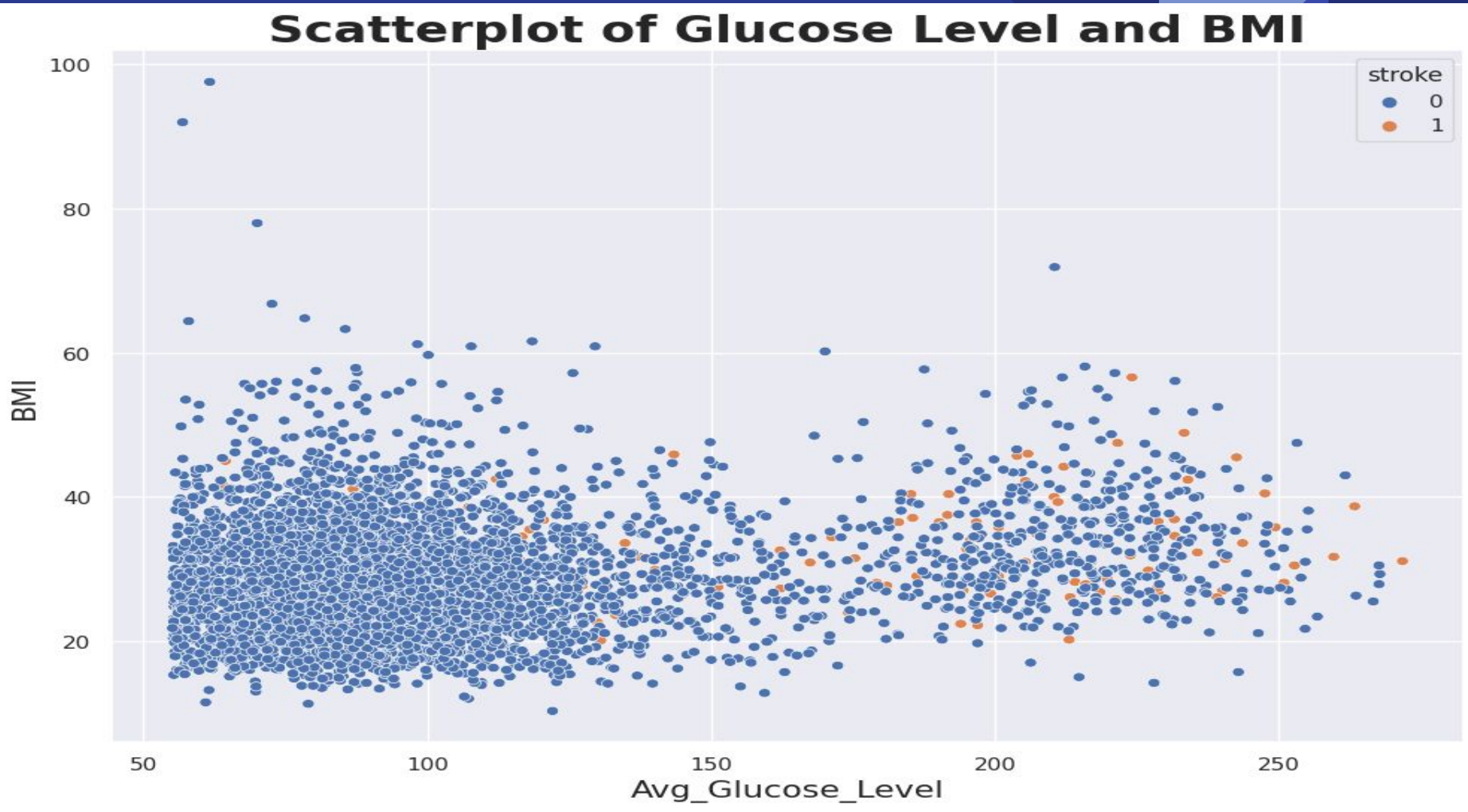
Problem statement

Is there a way to predict Strokes before they happen using Patient data?

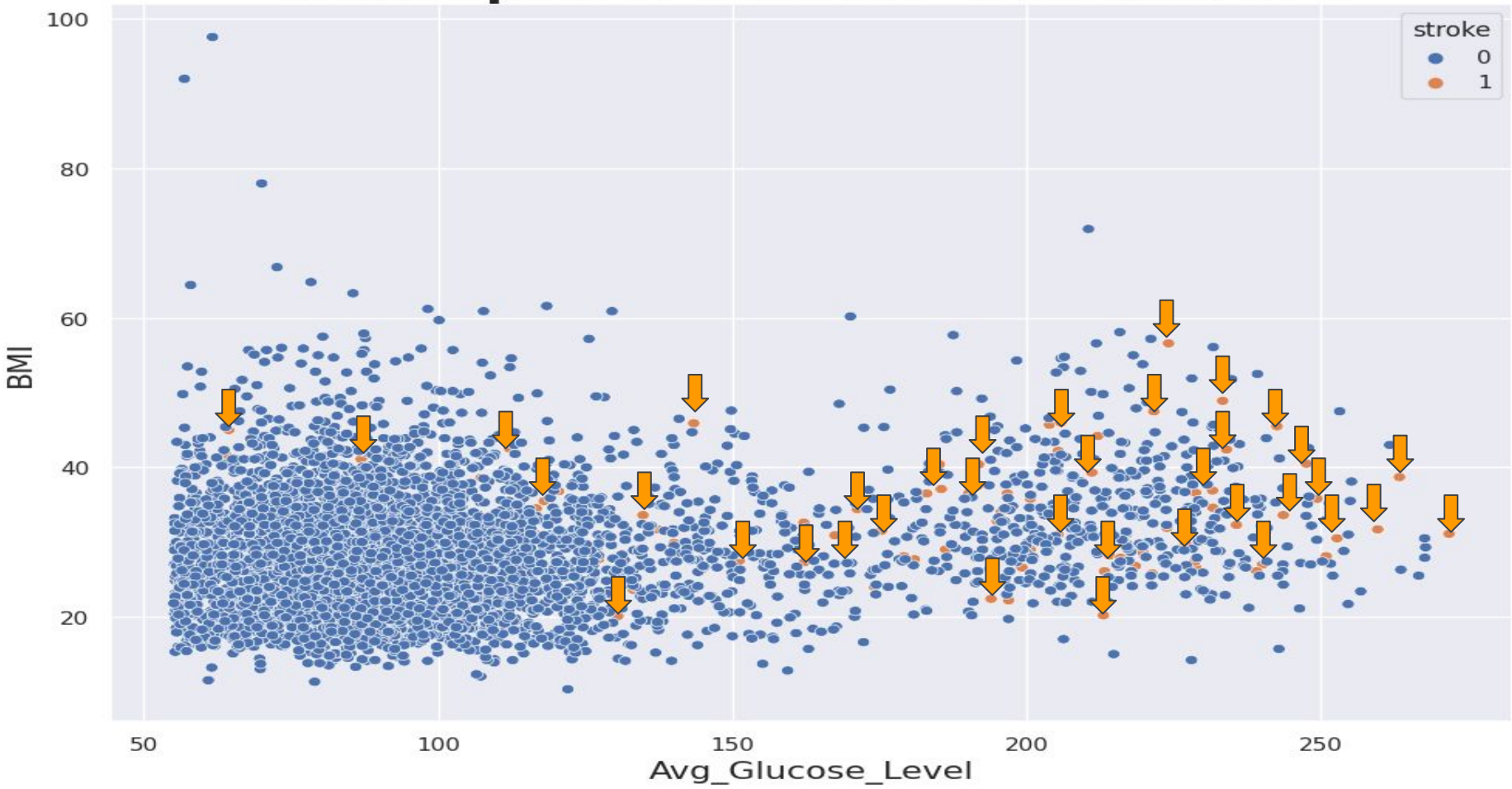
Exploratory Results


Scatterplot of Glucose Level and BMI
BMI
100
80
60
40
20
50
100
150
200
250
Avg_Glucose_Level

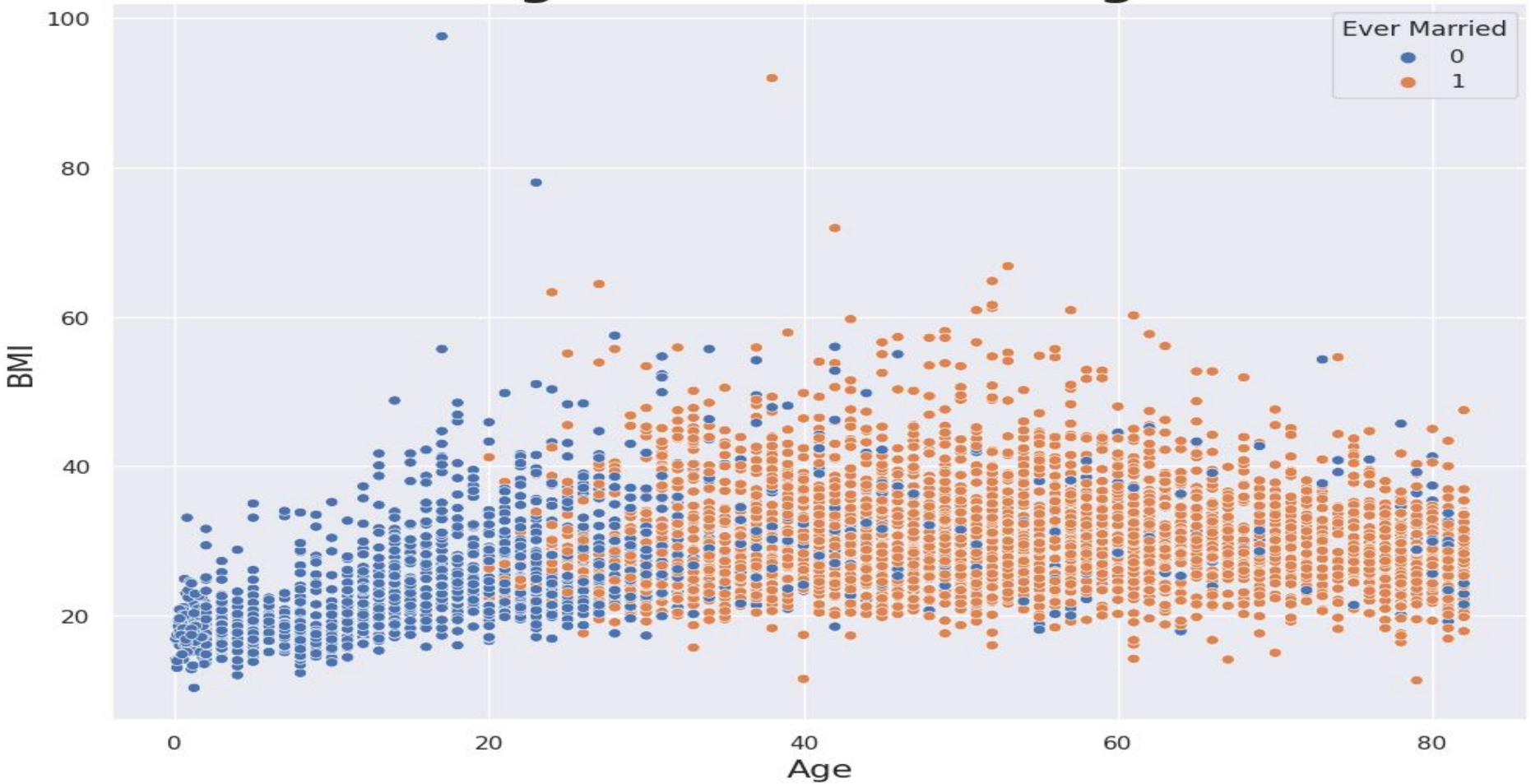

Scatterplot of Glucose Level and BMI
100
80
60
40
20
BMI
50
100
150
200
250
Avg_Glucose_Level

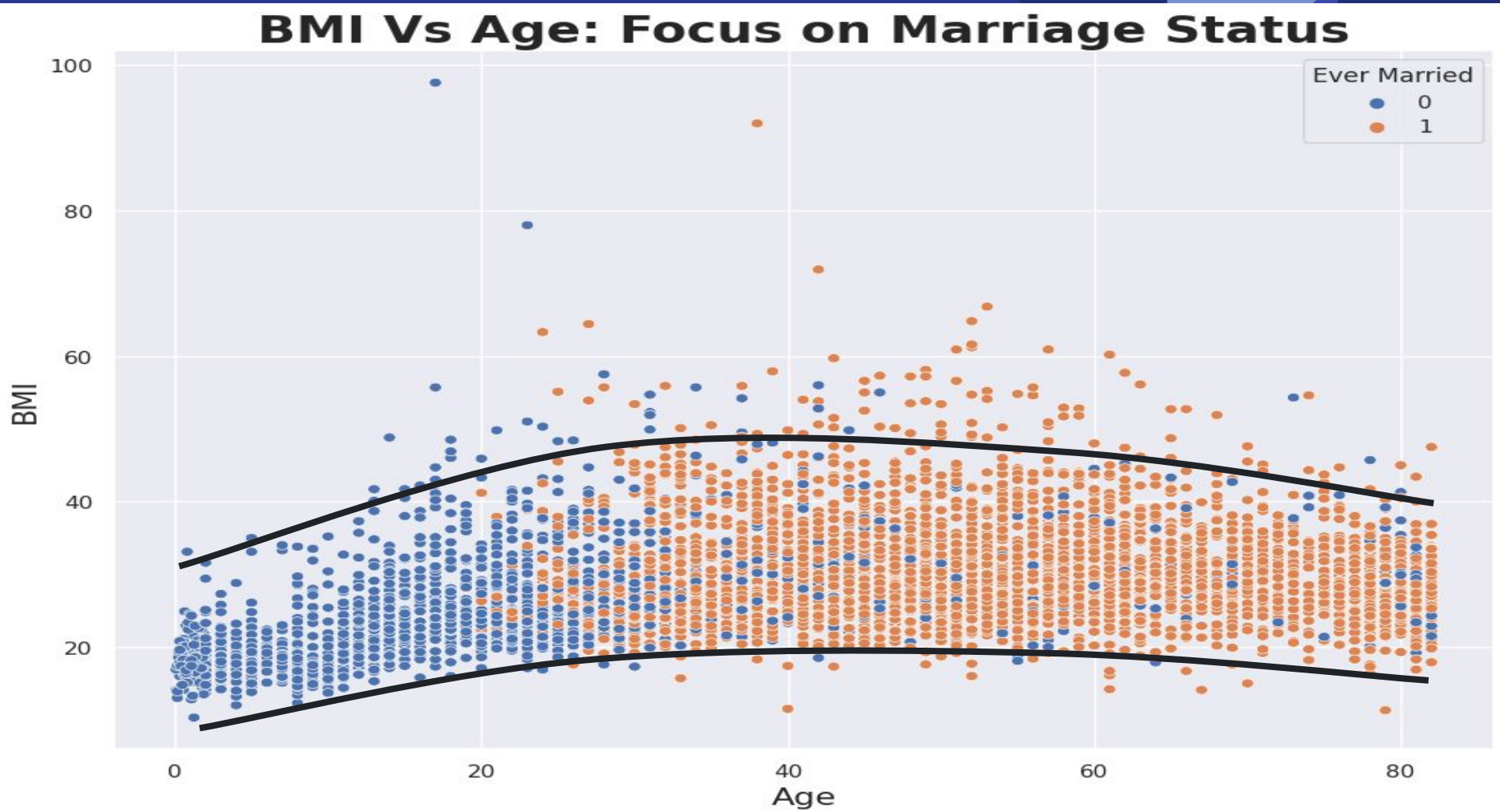

Scatterplot of Glucose Level and BMI
stroke
0
1
BMI
100
80
60
40
20
50
100
150
200
250
Avg_Glucose_Level

Scatterplot of Glucose Level and BMI

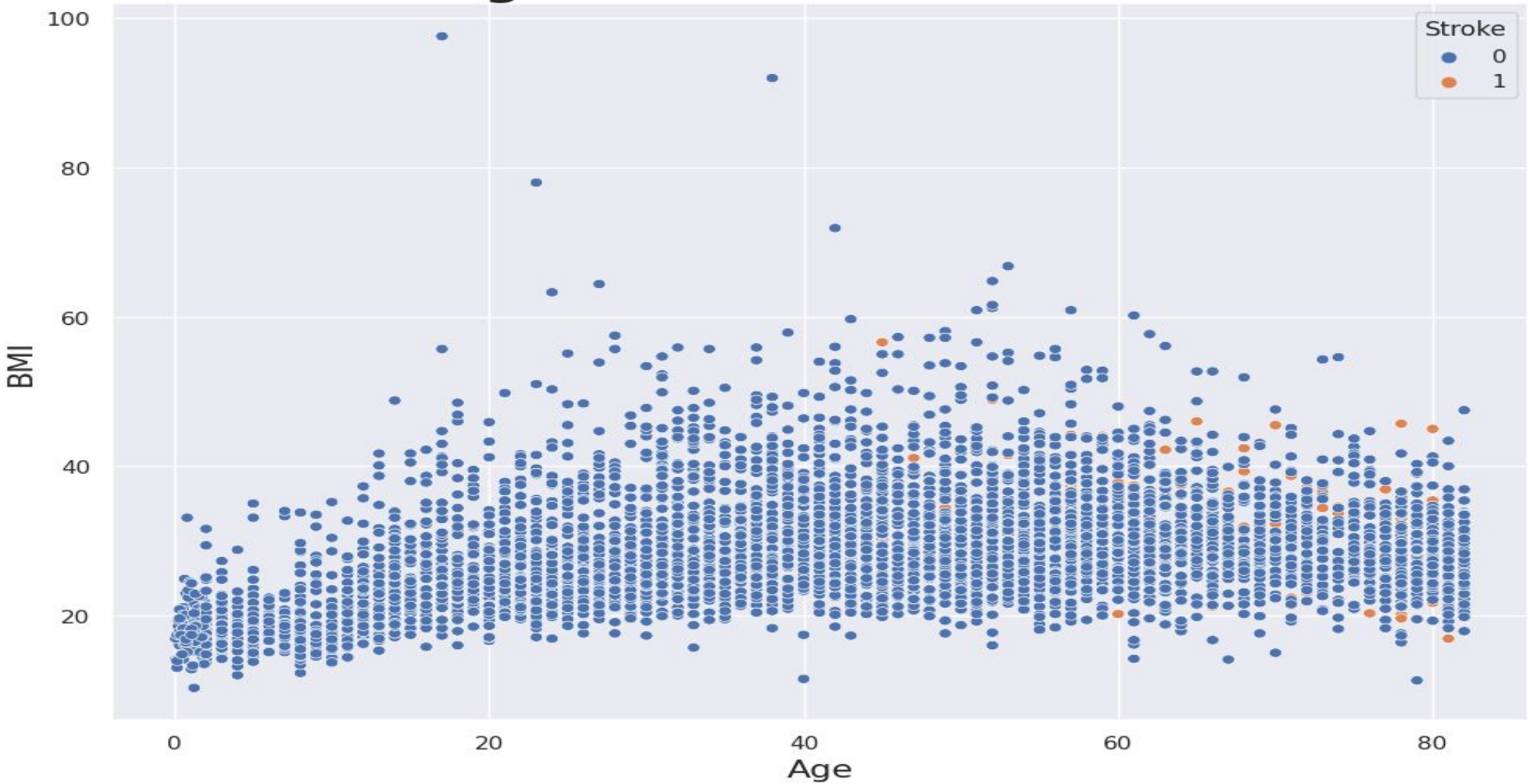


BMI Vs Age: Focus on Marriage Status

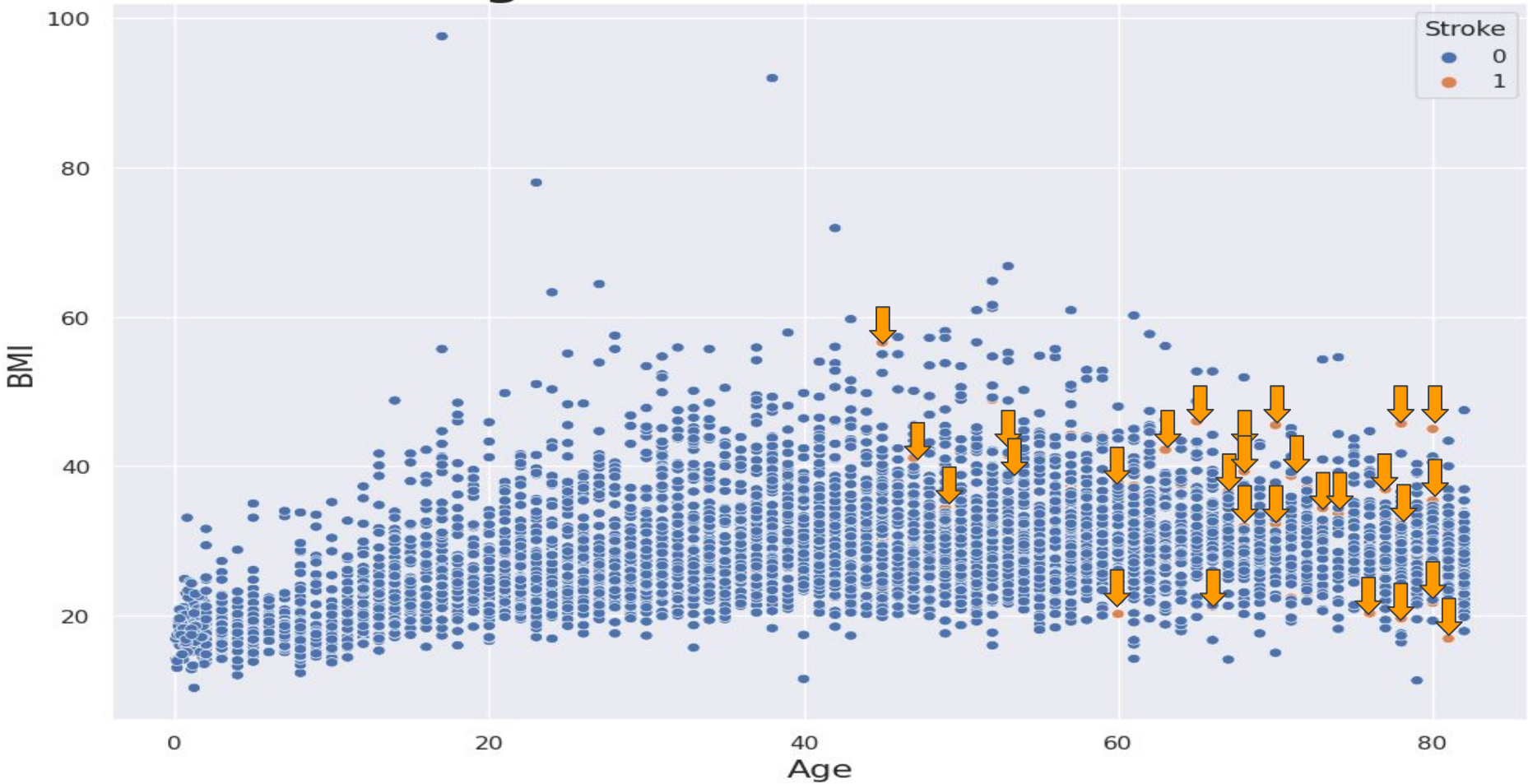



BMI Vs Age: Focus on Marriage Status
Ever Married
0
1
BMI
Age
100
80
60
40
20
0
20
40
60
80

BMI Vs Age: Focus on Stroke Occurrence



BMI Vs Age: Focus on Stroke Occurrence



Challenges Deep-Dive

Challenge 1

Lack of Stroke Victims

Small percentage of actual Stroke Victims to Non-Stroke Victims.

Challenge 2

Other Symptom Interference

High Glucose Level and BMI interfering with results that might otherwise predict other diseases other than Strokes.

Challenge 3

Increase conversion

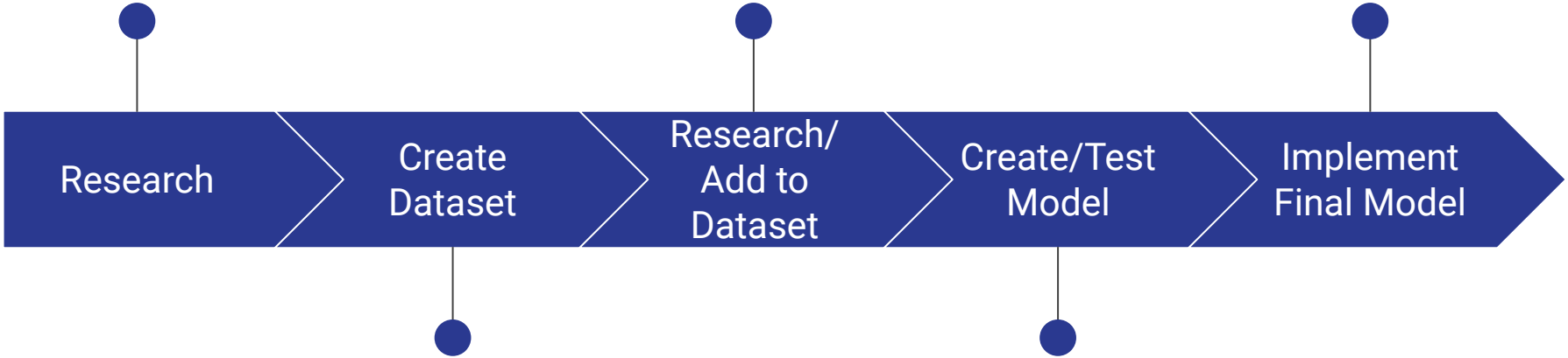
Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Final Results Recommendations

Research into patients who are Stroke Victims

Research Further into adding a control group of patients

Out of the New Dataset, Implement the Optimum Model



Create a Dataset with just Stroke Victims

Create a new Model/Test the KNN Model on new Dataset

Thank you For Attending!

This is the Time For Questions!

*Resources:

Donkor, Eric S. “Stroke in the 21st Century.” National Library of Medicine, Published: 2018 Nov, 27, (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6288566/#:~:text=Stroke%20is%20ranked%20as%20the,rate%20of%20about%205.5%20million.)

Dataset Source:

Fedesoriano. “Stroke Prediction Dataset”, Kaggle: fedesoriano, 2021. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset. [Accessed: 04/14/2023]