Stakeholder

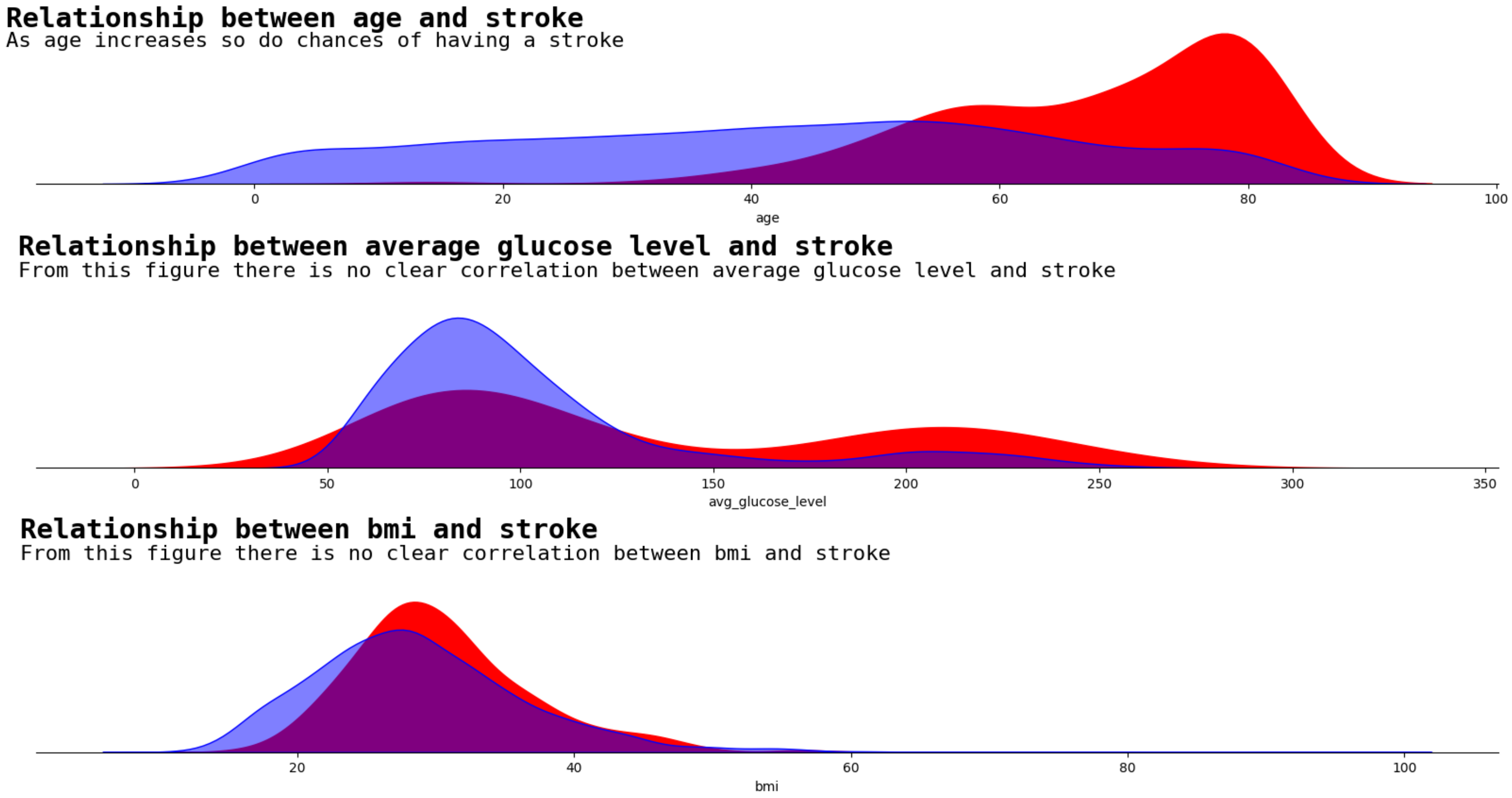
Primary Healthcare Office Staff

Doctors, Nurses, PAs and NPs

How can we predict stroke using general health data

- It is important to know who is at risk for stroke
- That being said there are many complicated health factors that can be at play when a patient has a stroke
- Is it possible to predict what patients are more at risk of having a stroke based on general healthcare data and not heavy duty labs that can take days to produce results

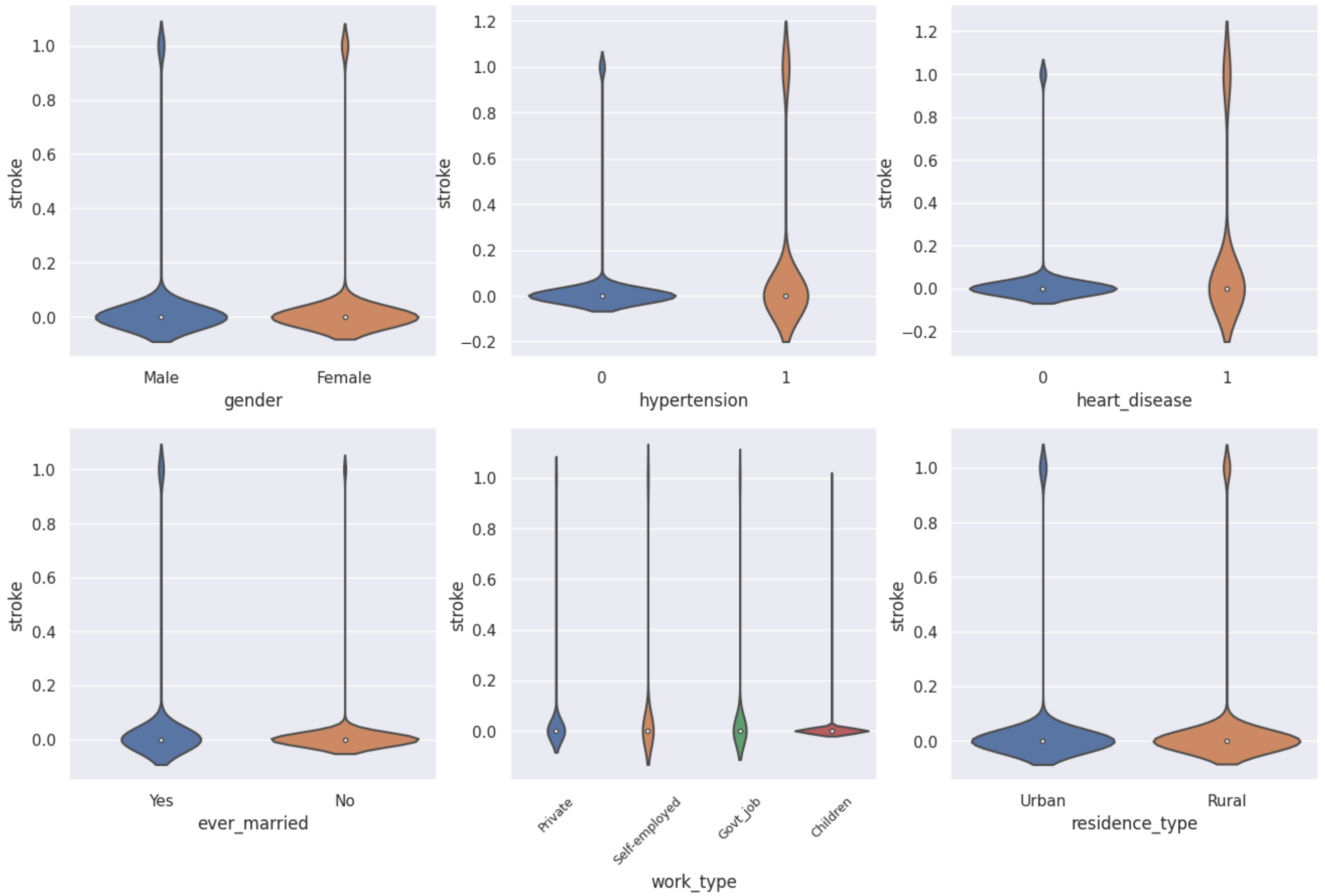
This shows the relationship between patient age, avg glucose level and BMI



What to take out of the graph

- We can look at this data and see that age when compared to average glucose level or bmi has the largest correlation with patients who are having strokes.
- As patients get older their chance of stroke increases.
- While average glucose level and bmi do have a correlation it would only help our predictions to keep receiving this information about patients as time goes on. This will help make our final model even more accurate in the end

These features are easily accessible by any health care worker.



General Features in Predicting Stroke

- Gender is not a huge player
- Residence is not a huge player
- Hypertension, Heart Disease and ever being married do have a large distribution of patients who have had stroke
- People in every walks of life can have a stroke so if we could predict it happening. While anyone can have a stroke it does seem that younger people are less at risk however that does that not preclude them from having one which is why prediction is important

We tried three main models

- This prediction is classification which means there are 2 outcomes. Having a stroke and not having a stroke.
- The data is unbalanced when it comes to our target. This is due to the fact that stroke is an uncommon issue over the population.
- This makes it so when predicting strokes we need to be able to figure out what can make our model better.
- This will allow us to further improve our model in predicting with better accuracy

Logistic Regression Model

- The logistic regression model that was used to make predictions was a unique balance of samples and feature selection.
- The model ending up performing well with both pick up true and false instances of stroke as well as false negatives which are much more costly when it comes to patient health.
- While false positives may be an issue with this model, being able to correctly predict strokes while possible assuming someone may have had a stroke when they didn’t is not a bad trade off.

Model Statistics

- Our model had a true positive rate of 87% which represents the instance of stroke recall score.
- The model had a 71% true negative rate which represents the non instance of stroke recall score.
- This means that it caught 83% of strokes and 71% of non strokes
- The model had a false positive rate of 29% but impressively a false negative rate of only 13%
- This means that the model using only general patient data can predict stroke with an accuracy of 74%