

An introduction to our data and goals.	Some analysis of the medical features of o..	Some statistics from the lifestyles feature..
--	--	---

Stroke Prediction

Project Introduction

For this project we used a dataset comprised of some key health and lifestyle data from over 5,000 patients including if each patient suffered a stroke. With this dataset we set out with two tasks and two questions. The two tasks were to visually analyze the dataset itself for trends, and secondly to create a predictive machine learning model which would analyze a patients health and lifestyle variables to predict if they were likely to suffer a stroke. We tested many different types of models, comparing performance, but did settle on the results from a Logistic Regression model.

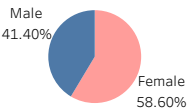
A small disclaimer that the none of the team members who assembled this project are medically trained, and our dataset was of course limited in scope somewhat. We were however, able to find some interesting correlations from the data, and create an accurate predictive model from ..

Total Cases

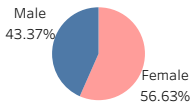
Stroke Status	
Non-Stroke	4,861
Stroke	249

Data Source..

Patient Gender Breakdown



Stroke Cases Gender Breakdown



Model Results

Logistic Regression Prediction Results

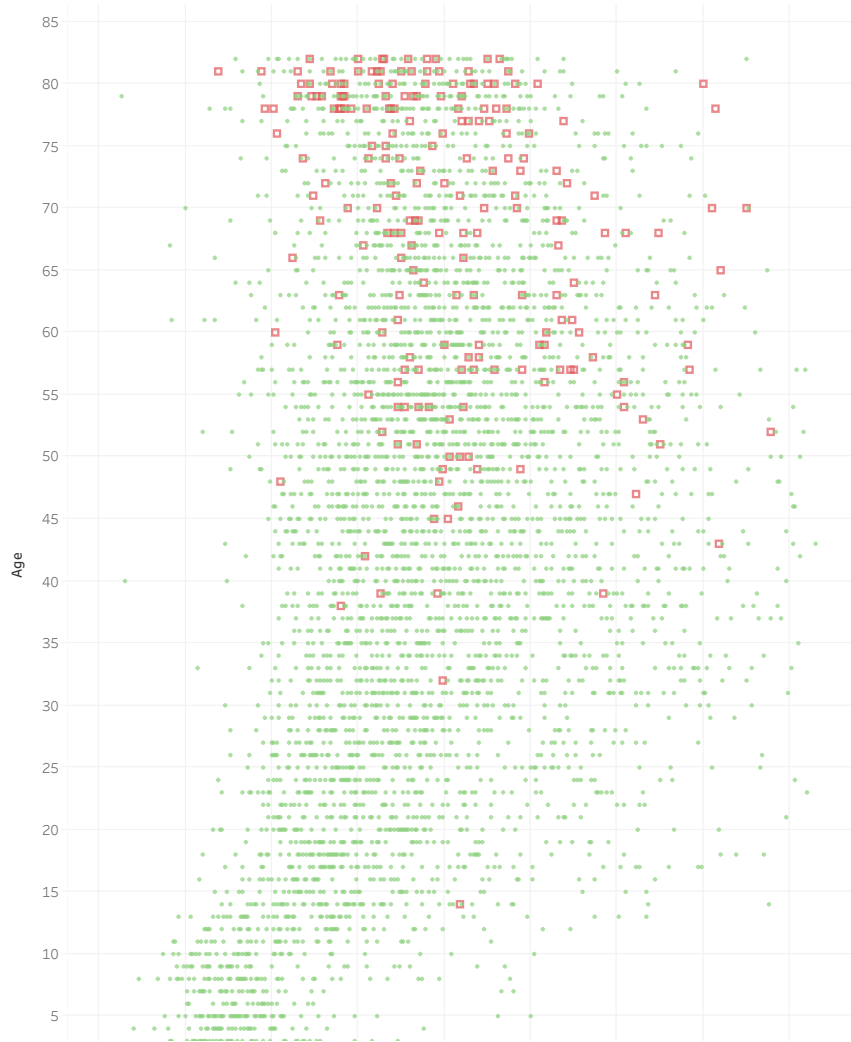
Model Prediction Accuracy: 84.32324%

Model F1 Score: 82.0

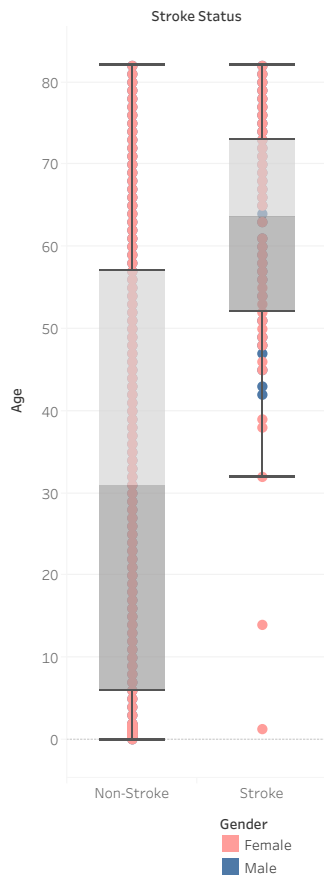
Model Description

Medical Data

Stroke Cases Relative to BMI and Age



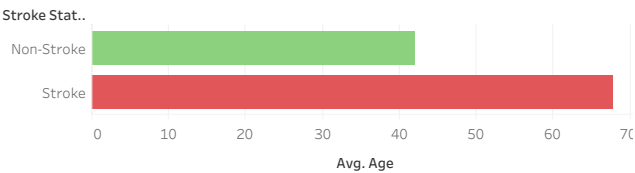
Stroke Status Box Plot



Age and BMI Correlate Strongly

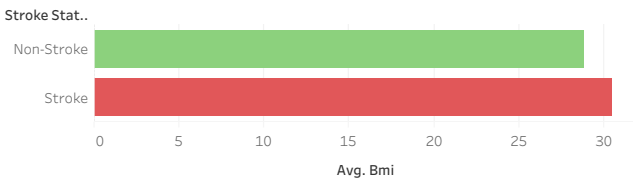
The scatter plot to the left shows the strong correlation we discovered from age and BMI with cases of stroke, you can clearly see the very strong

Average Age



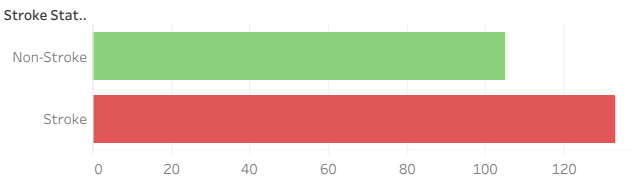
When examing the data around patient age in our dataset, it stands out from both our analysis and our predictive models that age was the strongest correlation, and thus risk factor, for having a stroke. See above the average age of stroke cases was near 70, and non-stroke cases down just above 40. The box and whisker chart digs into this a bit deeper with some more descriptive statistics and confirms this.

Average BMI



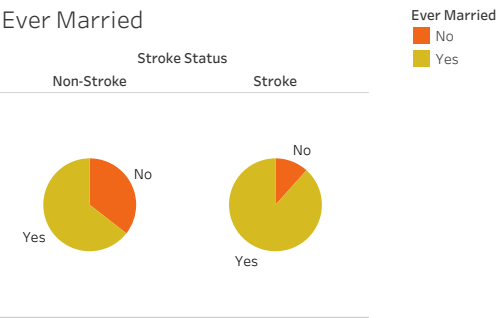
In contrast to age, we found BMI did not correlate strongly as a stroke predictor. The difference in average BMI between stroke and non-stroke outcomes among our data was only just under 2. We found BMI in conjunction we age did appear to correlate.

Average Glucose Level



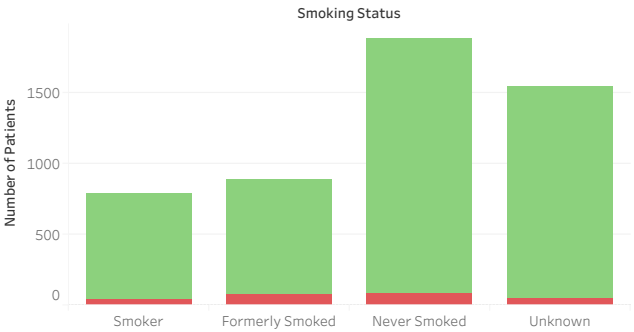
Lifestyle Data

Ever Married



Included with our patient data were not only medical features, but some key features for each patient about different lifestyle factors. While we did try to create predictive models that included these factors we found they seemed to reduce the accuracy of the model and eventually did not include them in the final model. We did analyze these lifestyle factors to see if there were any interesting correlations or trends to stroke risk. Smoking Status, Type of Work, and Residence Type all seemed to have absolutely no correlation to stroke risk.

We found that a noticeably higher percentage of the patients who suffered strokes were married than the patients who did not have a stroke. However, this was found to not be predictive and likely just a characteristic of the dataset used. The data included some patients who were children which likely skewed these marriage statistics this way.



Closing Summary

