Project 2

Predicting Stroke

Problem: Predicting Stroke

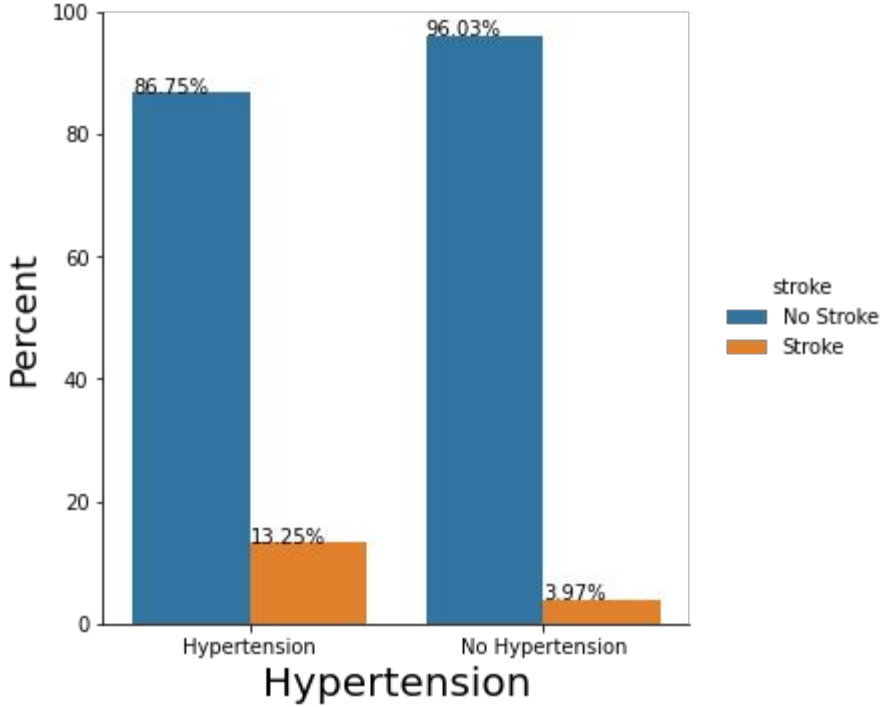
The problem explored in this project is to discover which of several features increase the likelihood of stroke.

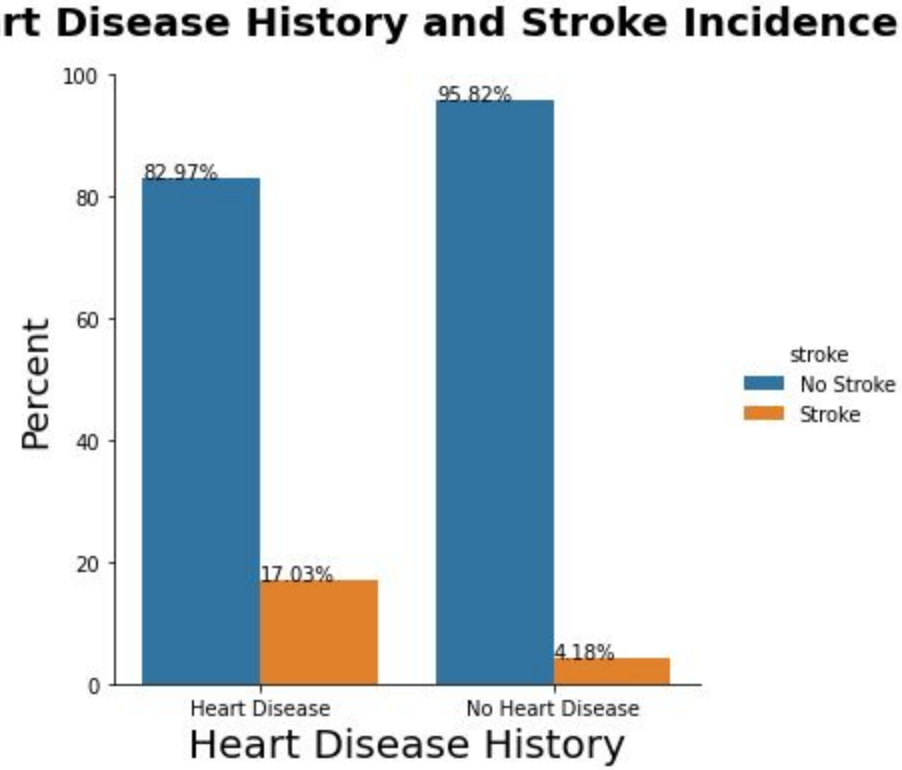
Data Set:

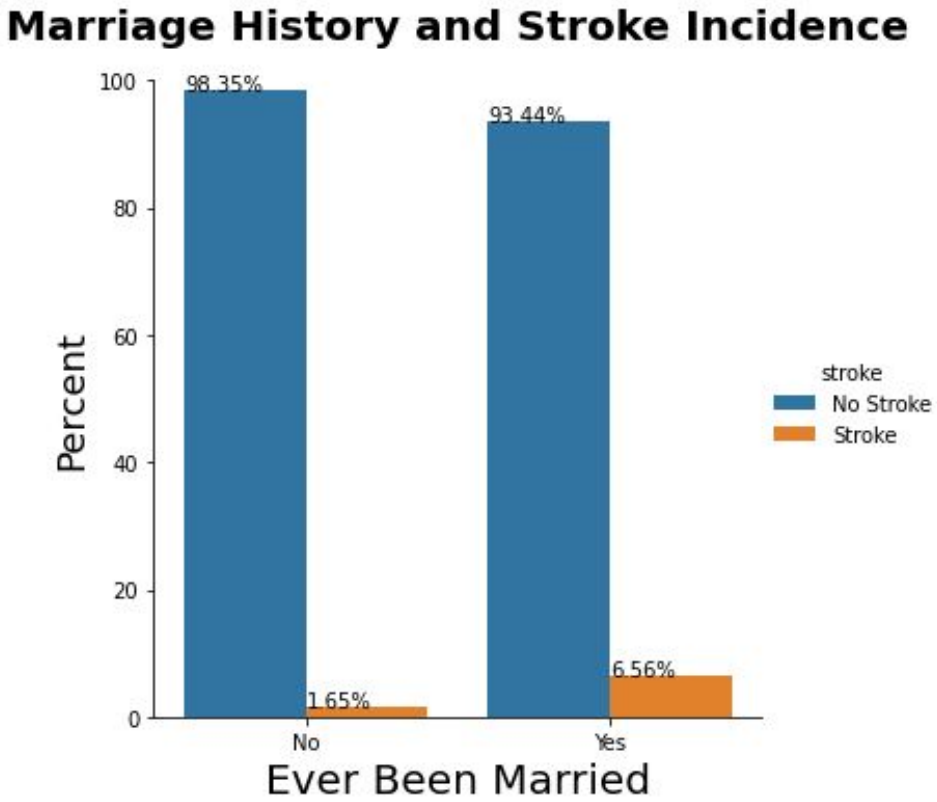
The data used for this project originates from:

(https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset)

Hypertension History and Stroke Incidence




Heart Disease History and Stroke Incidence
100
80
60
40
20
0
Percent
82.97%
17.03%
95.82%
4.18%
Heart Disease
No Heart Disease
Heart Disease History
stroke
No Stroke
Stroke


Marriage History and Stroke Incidence
100
80
60
40
20
0
Percent
98.35%
1.65%
93.44%
6.56%
No
Yes
Ever Been Married
stroke
No Stroke
Stroke

Predictive Models Explored:

Bagged Tree Classifier

Logistic Regression

LightGBM Classifier

Model 1: Bagged Tree Classifier

- Great at overall predictions

- Where did it fall short?

Model 2: Logistic Regression

- Great at overall predictions initially, but where did it fall short?
- How did it do after some tuning?

Model 3: Light GBM

- Great at overall predictions initially.
- Where did it fall short?
- How did it improve after some adjustments?

Model Chosen for Production - LightGBM !!!

Why?

- Reduce False Negatives
- Maintain overall predicting

Final suggestions:

- Predicting stroke