Predicting Strokes

By: Jordan Buddenhagen

Background Information

- According to the CDC, in 2020 1 in 6 deaths were from cardiovascular disease was due to stroke.
- Roughly 25% of strokes are in people that have already had a stroke.
- Stroke-related costs in the US came to nearly $53 billion: health care services, medicines to treat stroke and missed days of work.
- Stroke is a leading cause of serious long-term disability

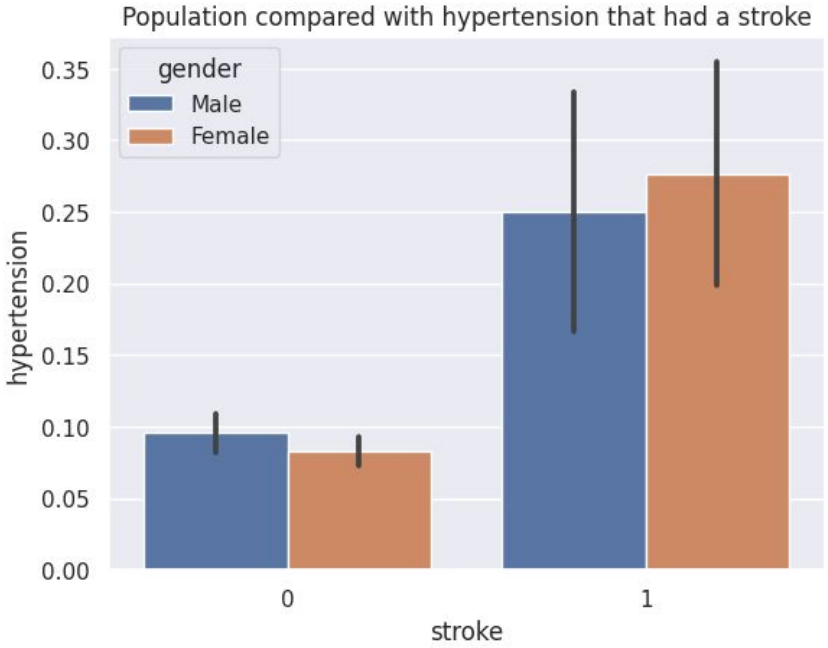
- According to the World Health Organization- stroke is the second leading cause of death globally

About the Dataset

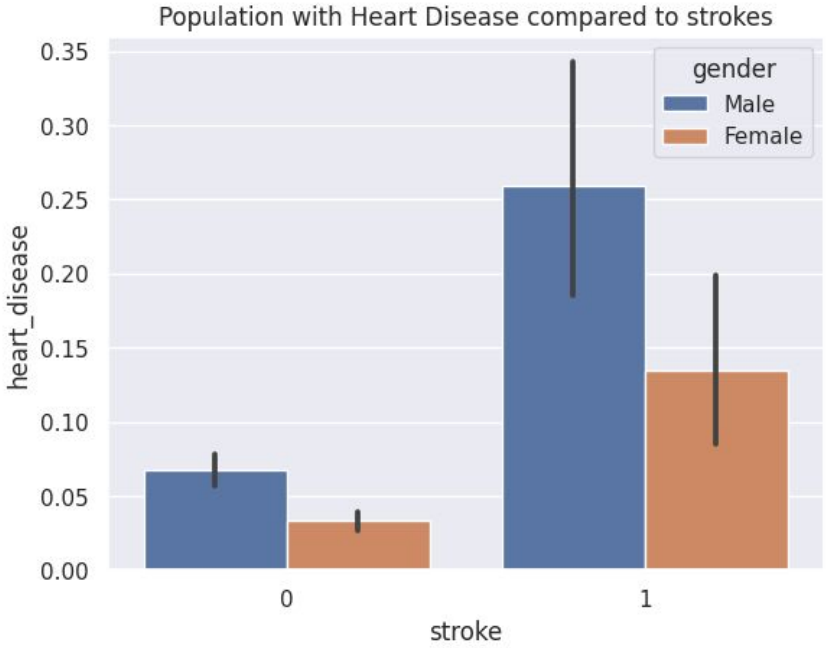
- Stakeholders are medical professionals that want to be able to predict if patients or persons would have stroke due to certain factors.

- Purpose was to create a model to predict whether a patient is likely to get stroke based certain parameters: gender, age, diseases and smoking status.

Visualizations and Trends

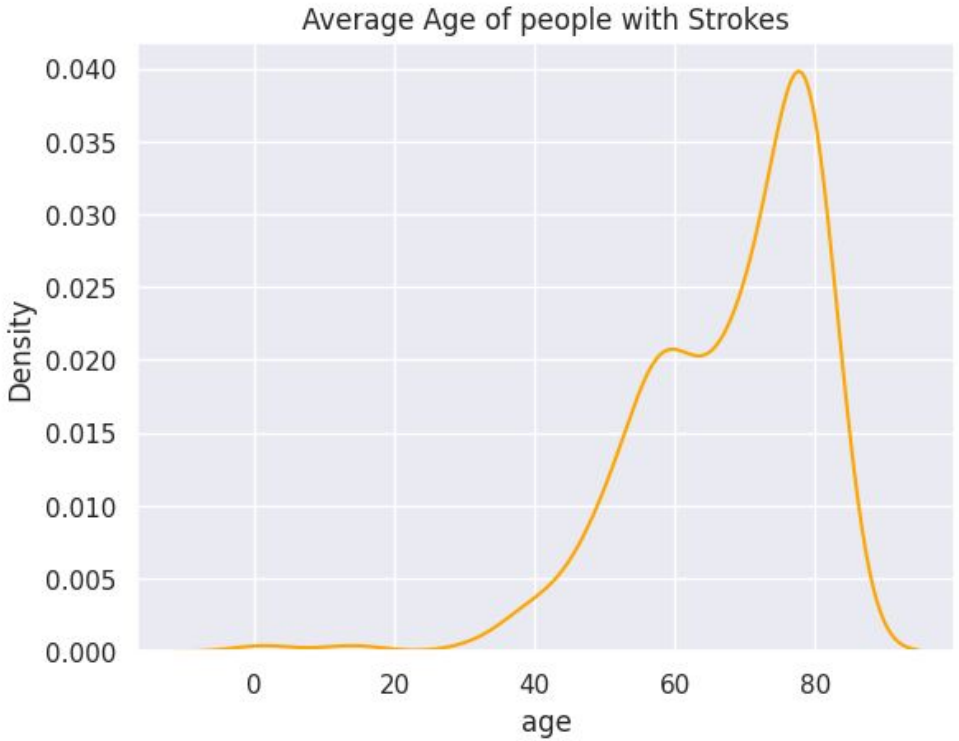


Hypertension leads to a higher possibility of a stroke.



Heart Disease has a higher correlation.

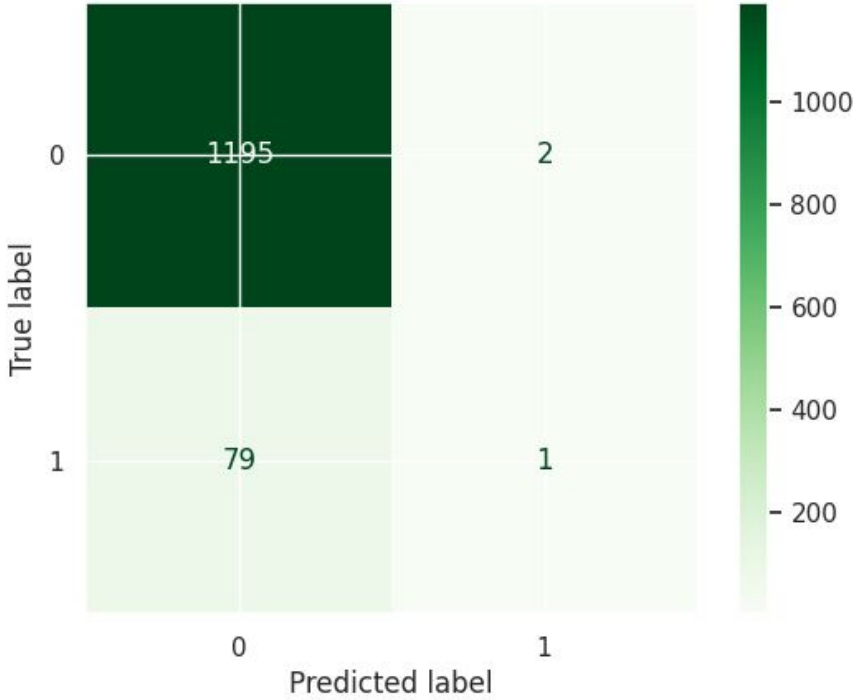
Biggest Factor



- Biggest factor was age.
- Even smoking was not as big as a factor as I thought it would be.

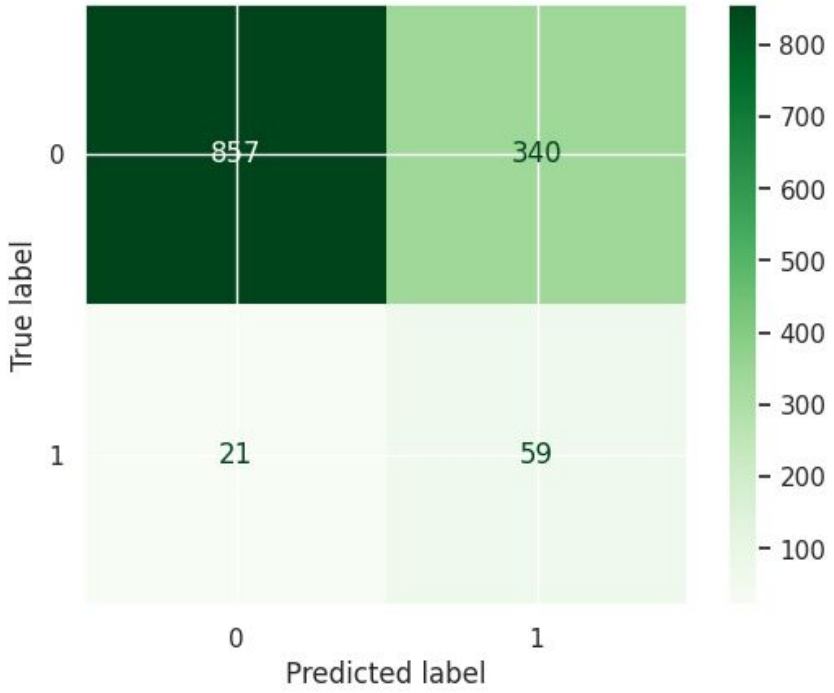
Unbalanced Model

- It was 93.6% accurate!
- However, for a medical data set, it is unable to predict if someone will have a stroke.
- The data set was very unbalanced to begin with and would need more patients that have had a stroke.



Balanced Model

- Model had a lower accuracy of 71.7%.
- However, when we oversampled our data- we were able to better predict patients that may have a stroke.
- The model did run into an issue were it was predicting more false positives.
- False negatives still too high.



Recommendations

Issues and Statements:

- I would not recommend either model.
- Unfortunately, the model needs more statistics on patients that have had strokes
- Easy to clean! That was nice.

For the models to improve:

- More data
- Tune the oversampling models
- Feature select patients over a certain age.

Sources

https://www.cdc.gov/stroke/facts.htm#:~:text=Every%20year%2C%20more%20than%20795%2C000,are%20first%20or%20new%20strokes.&text=About%20185%2C000%20strokes%E2%80%94nearly%201,have%20had%20a%20previous%20stroke.&text=About%2087%25%20of%20all%20strokes,to%20the%20brain%20is%20blocked.

https://www.who.int/data/gho/publications/world-health-statistics

Thank You!