Group 7: Stroke Risk EDA

Emily Park, Emerson Noble, Kaining Wan, Ikmal Mahadi, Nancy Yu

Motivation and Objective

What is a Stroke? - A stroke is damage to the brain from interruption to its blood supply

According to Mayo Clinic, There are more than 200,000 cases a year in the United States.

Objective: We are going to examine the data from our data set and determine what factors lead to higher risk of stroke.

Motivation: By completing our objective, we educate individuals who may fall into some of these categories so that they may take necessary precautions to prevent a stroke happening to them or their loved ones.

Dataset

The dataset we are using is Stroke Prediction Dataset from Kaggle. This dataset contains 5110 observations and 12 attributes. The attributes for this dataset are as follows:

1) id
2) gender
3) age
4) hypertension

5) heart_disease
6) ever_married
7) work_type
8) Residence_type

9) avg_glucose_level
10) bmi
11) smoking_status
12) stroke

Methodology

Prep the dataset →

Check correlation →

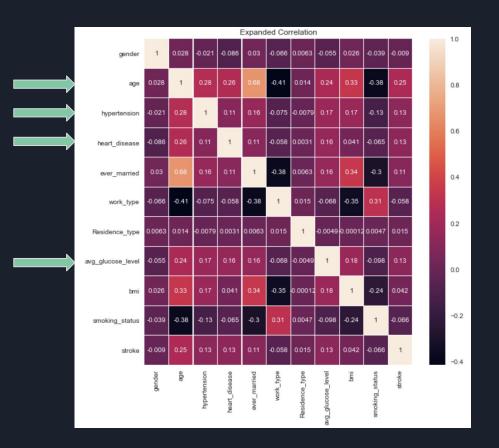Analyze relationship of features

Plot each Feature that has high correlation vs. Stroke

→ Check the remaining features to determine if there is strong correlation to higher risk of Stroke
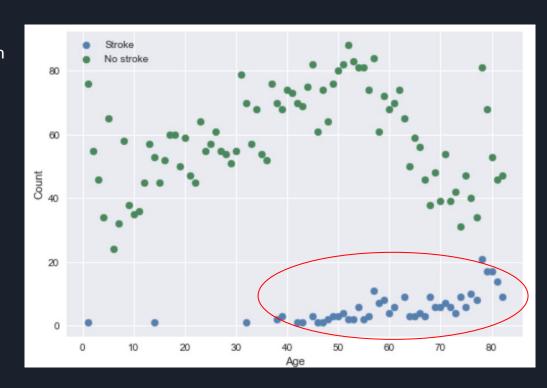
Let's look at Correlation

- From this heatmap, top 4 features that are most correlated to stroke are:
  - **Age**
  - **Hypertension**
  - **Heart disease**
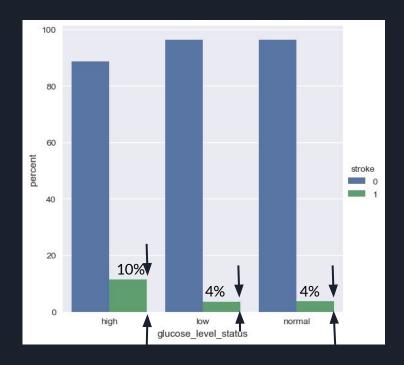 - **Average glucose level**

Expanded Correlation

How does age contribute to the risk of stroke?

- People who had strokes were generally older in age
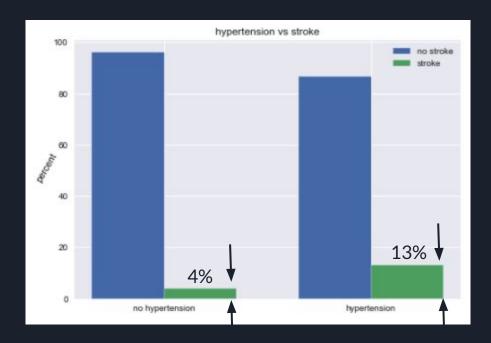- Can conclude that the older someone is, their chance of getting a stroke is higher

How does Avg Glucose Level contribute to Stroke risk

- People with higher glucose level (>140 milligrams/deciliter) more likely to get stroke compared to the people who have low(<70 mg/dl) and normal glucose level (>70 and <140 mg/dl)
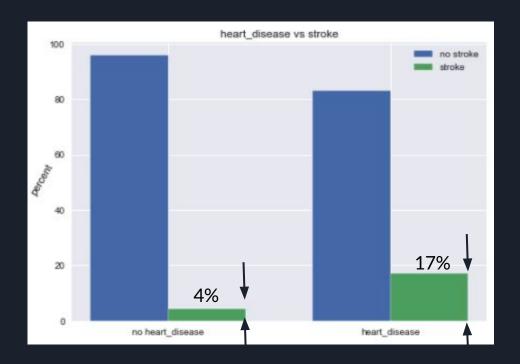
How does Hypertension contribute to Stroke risk?

- The proportion of stroke in hypertension group is 13%, three times higher than 4%.
- Hypertension is an important risk factor of stroke
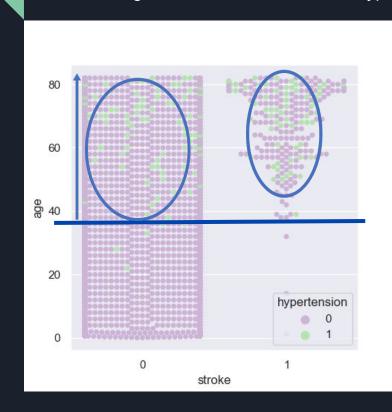

hypertension vs stroke

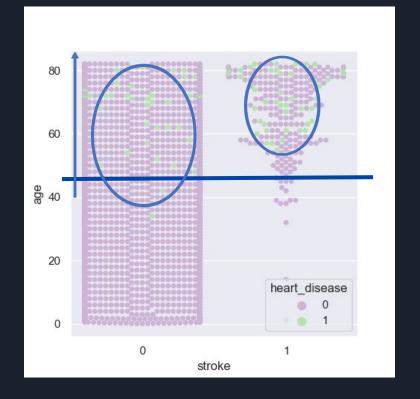# How does Heart Disease contribute to Stroke risk?

- The proportion of stroke in heart disease group is 17%, four times higher than 4%.
- Heart disease is an important risk factor of stroke
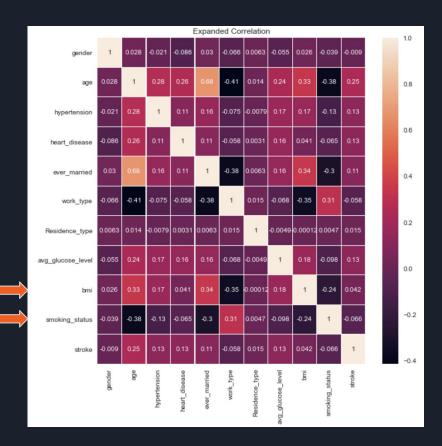
Why the Aged is more likely to have a stroke?

- The Aged account for most of the hypertension and heart disease

Let's look back at Correlation

- For BMI and Smoking Status we found that from our analysis it actually leads to stroke
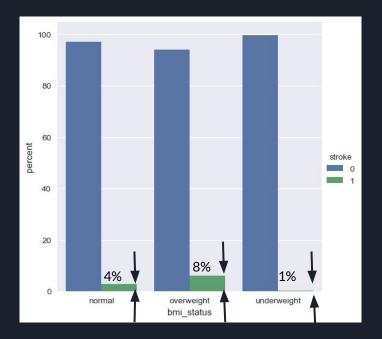
How does Higher BMI contribute to Stroke risk?

A person who is overweight (BMI > 25)

is more likely to get stroke compared to a person
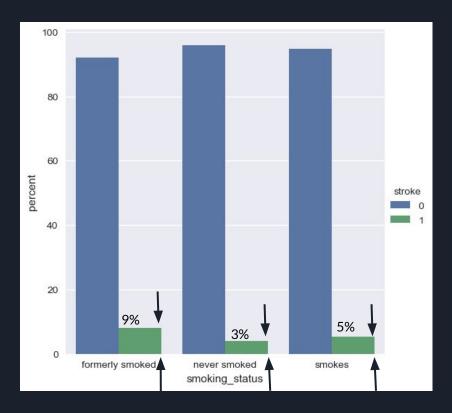
Who have normal and underweight BMI.

Is a person that does not smoke at a lower risk of stroke?

A person who smokes and has smoked in the past,

has a higher chance of getting a stroke

than someone who never smokes.

Conclusion and Future Work

From our analysis, factors that put someone at higher risk of stroke are:

- **Old age**
- **Hypertension**
- **Heart disease**
- **High BMI**
- **Smoking**
- **High Glucose Level**

Conclusion:

Based on our data analysis, we can now identify the people who are at risk of getting stroke and help them manage their risk factors.

Future work:

1) Use machine learning to predict someone getting stroke more accurately.
2) Use a wider range of data from other countries regarding factors that lead to stroke globally.