BRAIN STROKE PREDICTION

AGENDA

01
ABOUT THE DISEASE

02
DATA UNDERSTANDING

03
EXPLORATORY DATA ANALYSIS

04
DATA PREPARATION

05
MODELING

06
CONCLUSIONS

AGENDA

01.

ABOUT THE DISEASE

ABOUT THE DISEASE

DEFINATION

A stroke is a medical condition in which poor blood flow to the brain or bleeding causes cell death.

MAIN FACTORS

The main risk factor for stroke is high blood pressure. Other risk factors include high blood cholesterol, smoking, obesity, diabetes and end-stage kidney disease

ABOUT

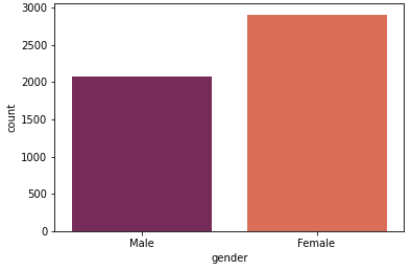
02.

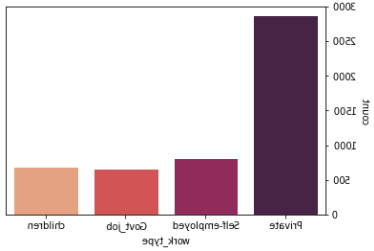
DATA UNDERSTANDING

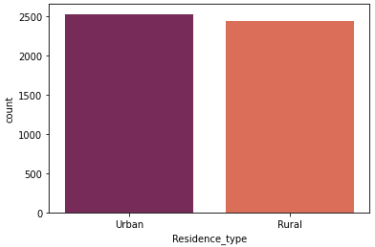
DATA UNDERSTANDING

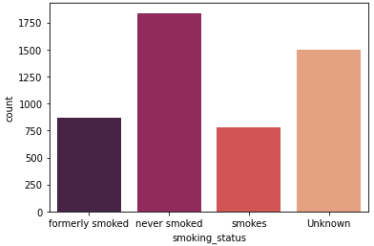
- This dataset is used to predict whether a patient is likely to get stroke or not.
- Each row in the data provides relevant information about the patient.
- Dataset has 4981 rows and 11 columns
- No Null values in this data
- No duplicates in the data

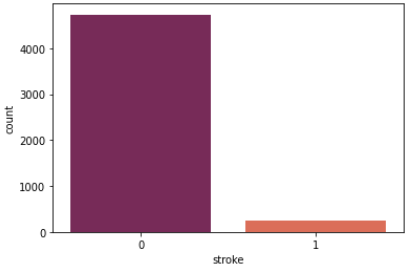
DISTRIBUTION OF DATA

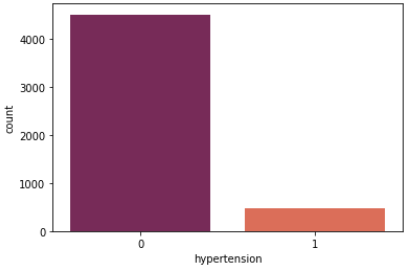


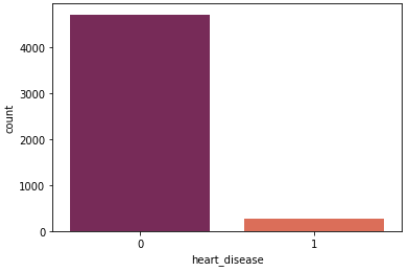


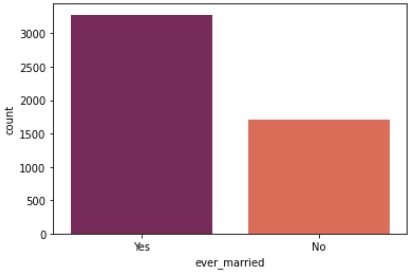












- **We can infer from these graphs that the data is imbalanced**

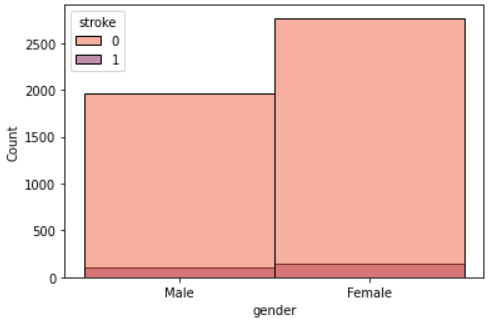
03.

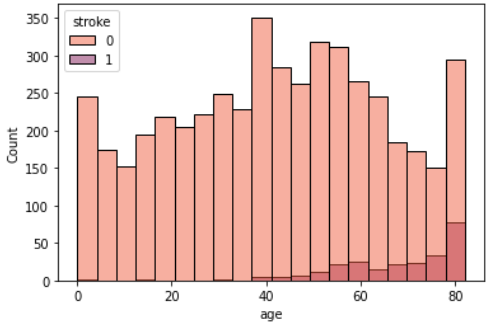
EXPLORATORY DATA ANALYSIS

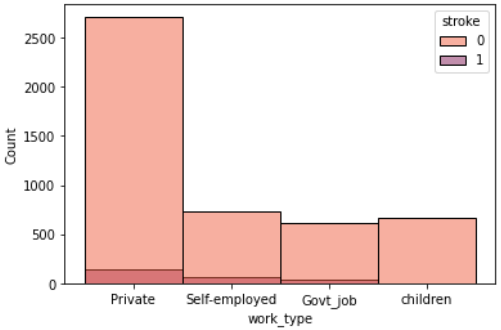
EXPLORATORY DATA ANALYSIS

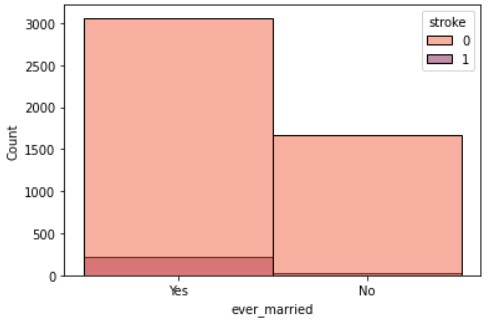
FEATURES AGAINST STROKE

- Females are more likely to have a stroke
- Marriage people are more likely to have a stroke
- Private work type category are more likely to have a stroke
- People who have aged more than 40 are more likely to have a stroke





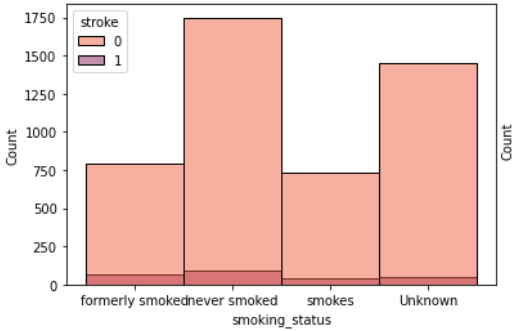


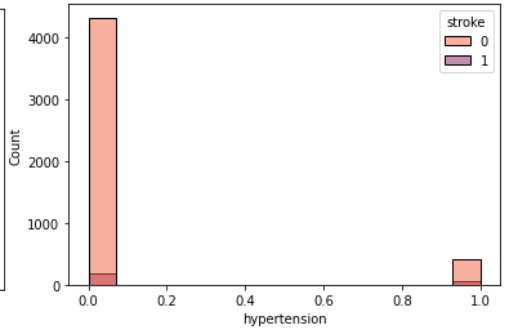


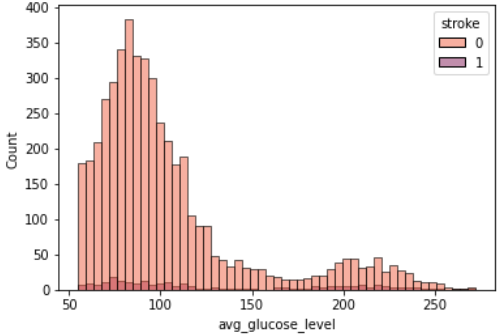
EXPLORATORY DATA ANALYSIS

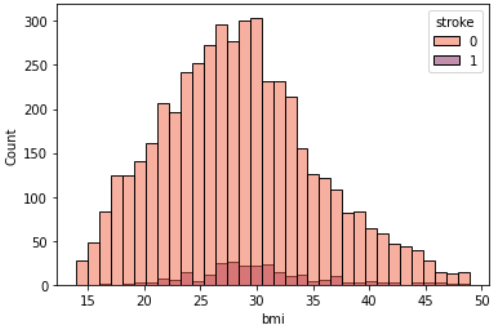
FEATURES AGAINST STROKE (CONT.)

- "never smoked" category is more likely to have a stroke also the category that does not have hypertension and heart disease is more likely to have a stroke
- People that have low levels of glucose are more likely to have a stroke
- High levels of BMI can lead to a stroke





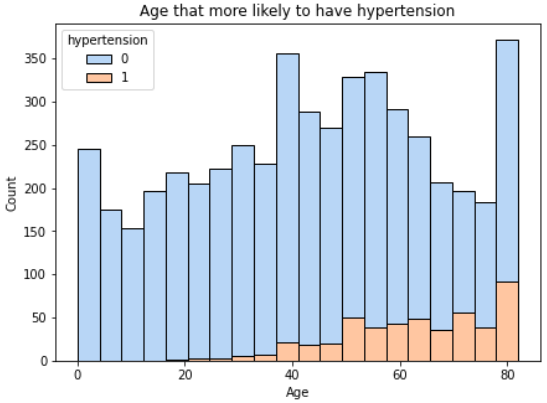


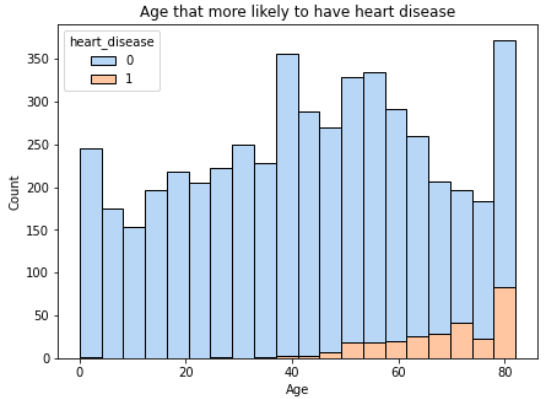


EXPLORATORY DATA ANALYSIS

WHICH AGE CATEGORY IS MORE LIKELY TO HAVE HYPERTENSION AND HEART DISEASE

- People who have aged more than 40 are more likely to have a hypertension and heart disease

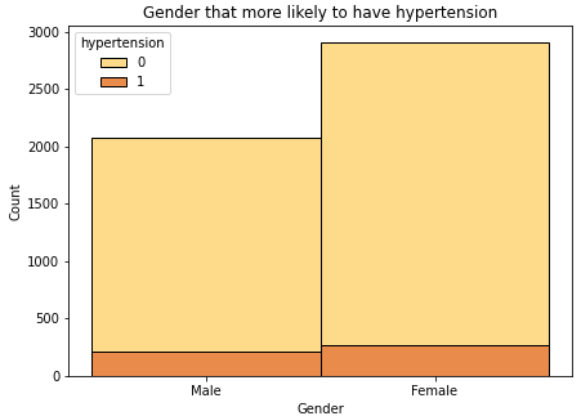


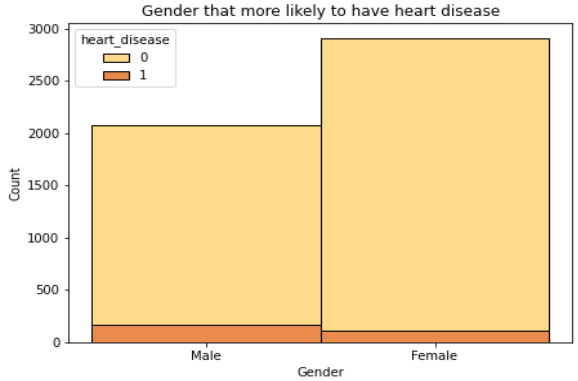


EXPLORATORY DATA ANALYSIS

WHICH GENDER IS MORE LIKELY TO HAVE HYPERTENSION AND HEART DISEASE

- Females are more likely to have hypertension and stroke but Males are more likely to have heart disease

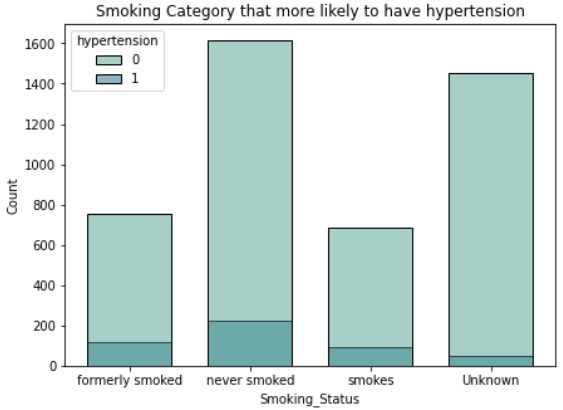


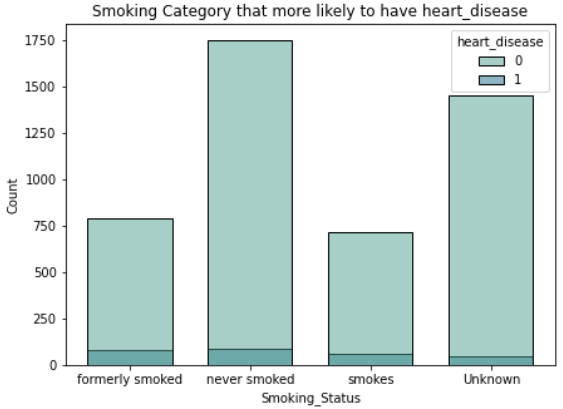


EXPLORATORY DATA ANALYSIS

WHICH SMOKING CATEGORY IS MORE LIKELY TO HAVE HYPERTENSION AND HEART DISEASE

- An unexpected result of smoking is that the "never smoked" category is more likely to have hypertension and heart disease

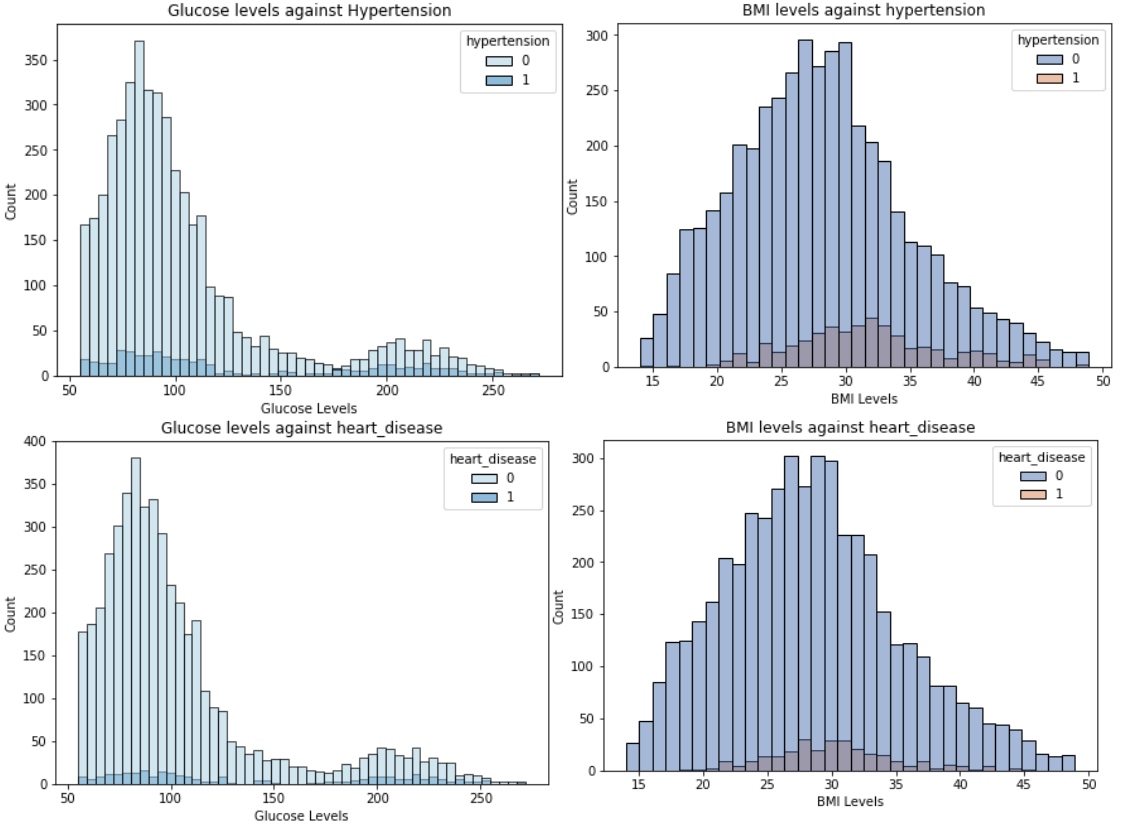




EXPLORATORY DATA ANALYSIS

DOES HIGH BMI LEVELS AND AVG GLUCOSE LEVELS CAN BE A FACTOR FOR HYPERTENSION AND HEART DISEASE?

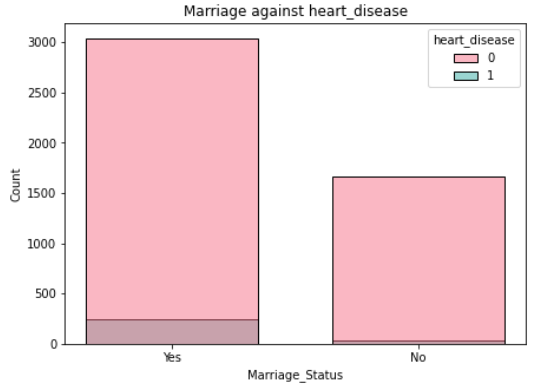
- High levels of BMI can be a factor in hypertension and heart disease
- People with low and high levels of glucose both can get hypertension and heart disease

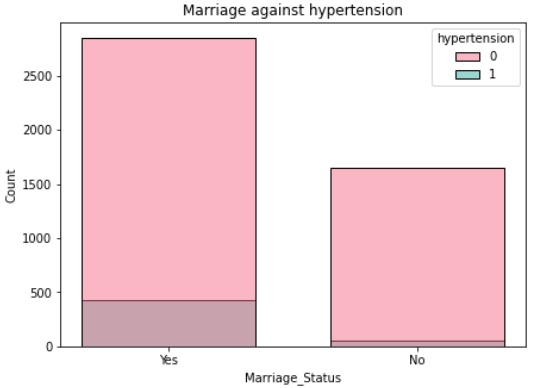


EXPLORATORY DATA ANALYSIS

IS MARRIAGE A FACTOR FOR HYPERTENSION AND HEART DISEASE?

- Marriage people are more likely to have hypertension and heart disease :-)

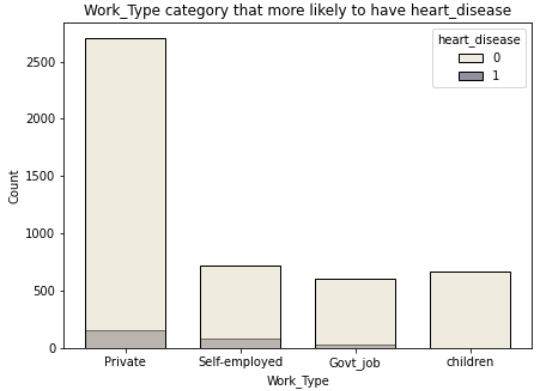


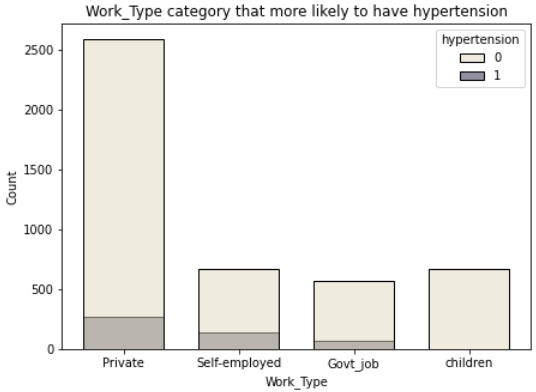


EXPLORATORY DATA ANALYSIS

DOES WORK TYPE CAN BE A FACTOR FOR HYPERTENSION AND HEART DISEASE

- Private work type category are more likely to have hypertension and heart disease

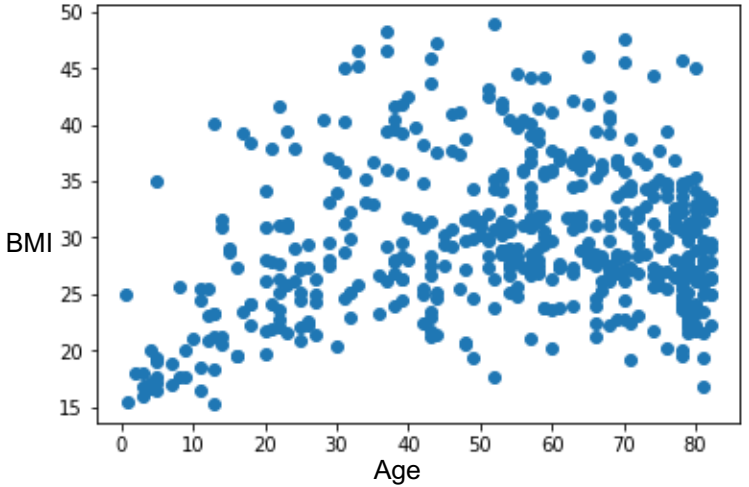




EXPLORATORY DATA ANALYSIS

IS THERE A RELATION BETWEEN BMI AND AGE?

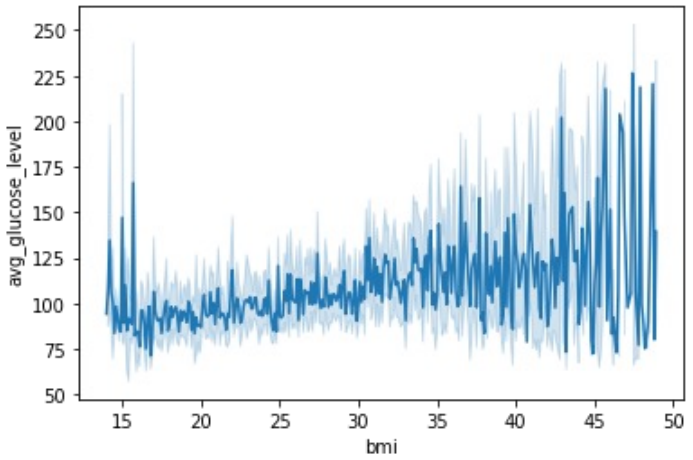
- getting old can cause an increase in BMI levels



EXPLORATORY DATA ANALYSIS

IS THERE A RELATION BETWEEN BMI AND AVG GLUCOSE LEVEL?

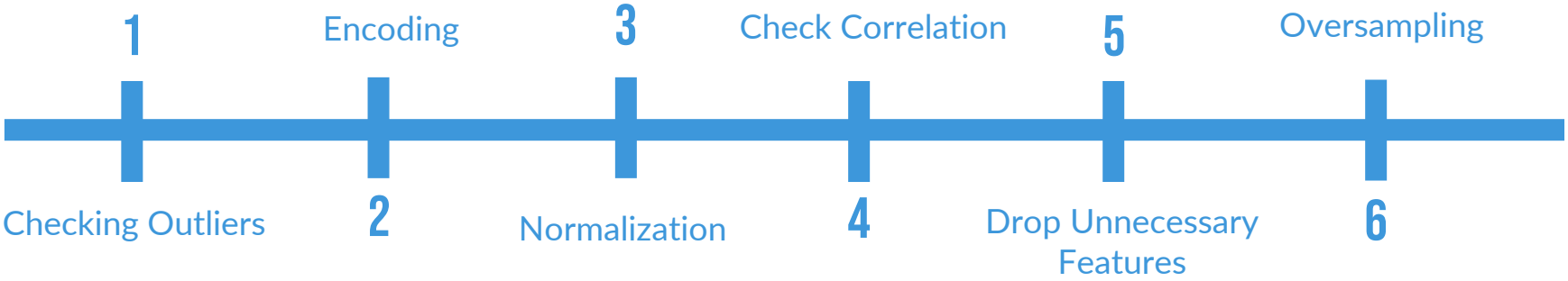
- We can say that Increasing BMI levels can cause an increase of avg glucose levels



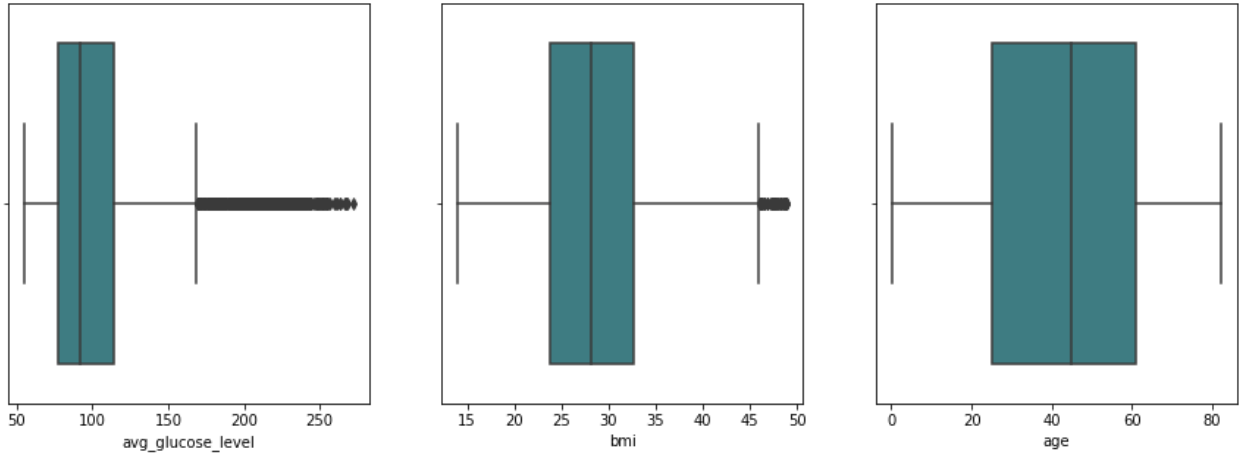
04.

DATA PREPARATION

DATA PREPARATION



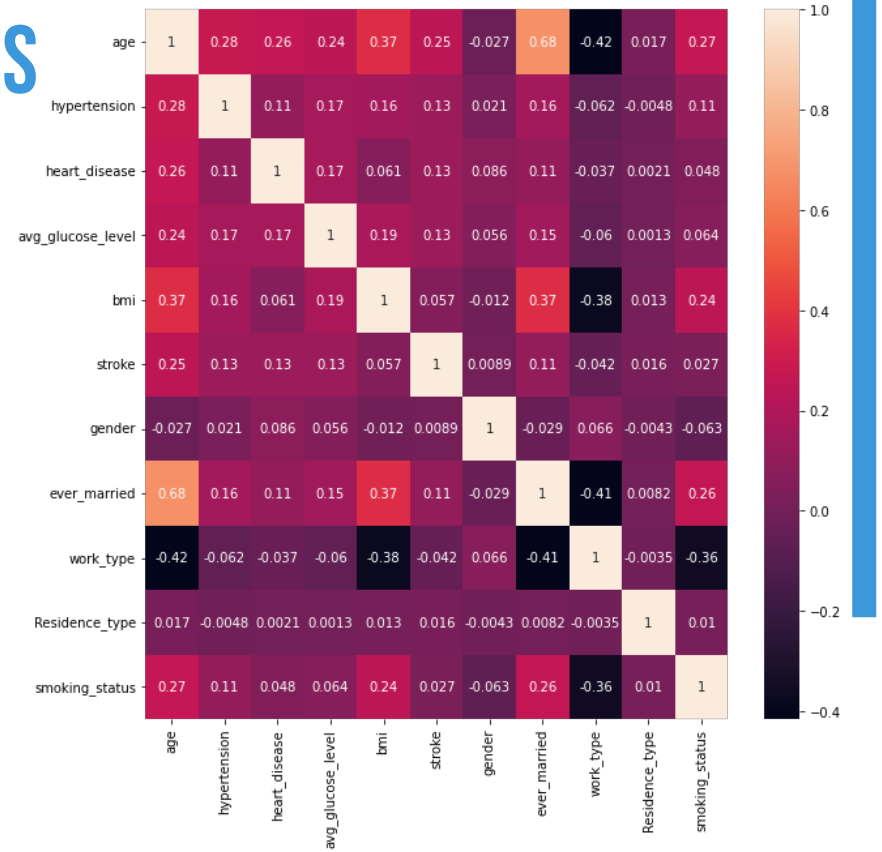
CHECKING OUTLIERS



- The box plot shows that there are outliers for avg glucose levels and BMI, but these values are real data and we must take them into accounts

CHECK CORRELATION
DROP UNNECESSARY FEATURES

- we will take only the features that correlated with stroke like:
  - age
  - hypertension
  - heart disease
  - avg glucose level
  - ever married



05.

MODELING

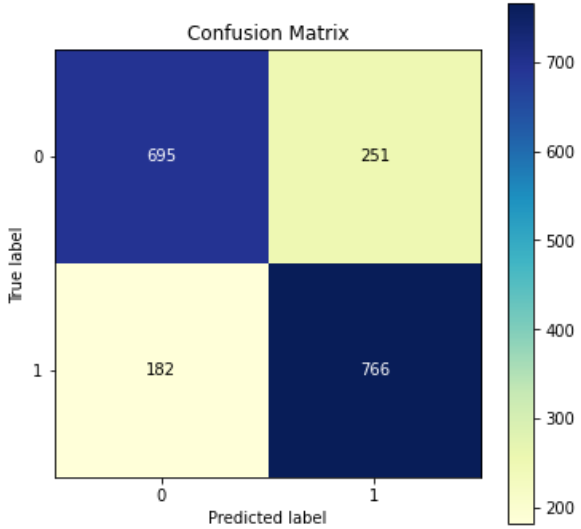
MODELING

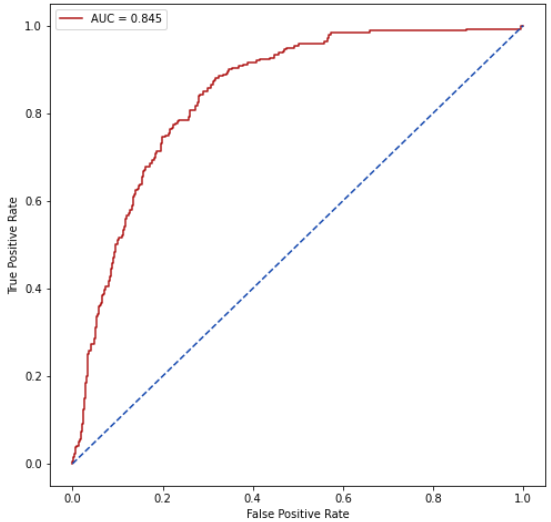
ALGORITHMS USED:

- Logistic Regression
- XGB Classifier
- Decision Tree Classifier
- SVC
- Extra Trees Classifier

LOGISTIC REGRESSION

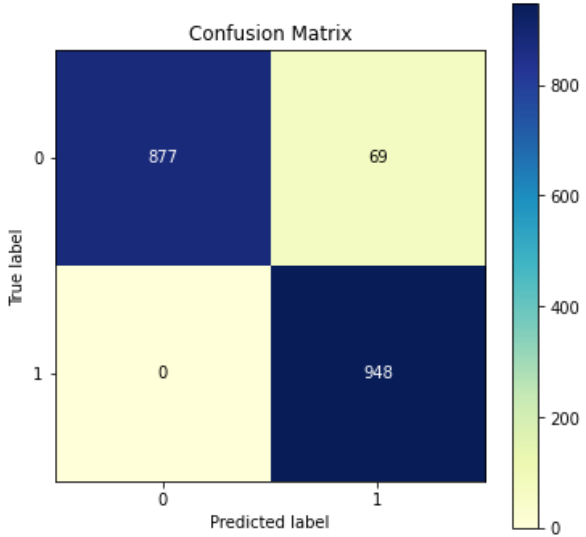
- ACCURACY: 77.1%
- PERCISION: 75.3%
- RECALL: 80%
- F1-SCORE: 77%
- AUC-ROC SCORE: 84.5%

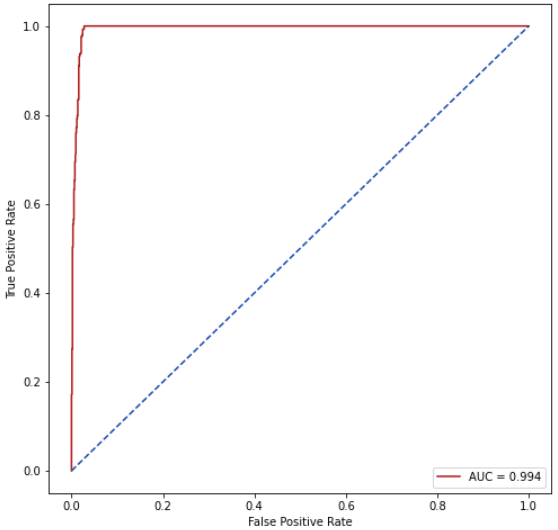




XGB CLASSIFIER

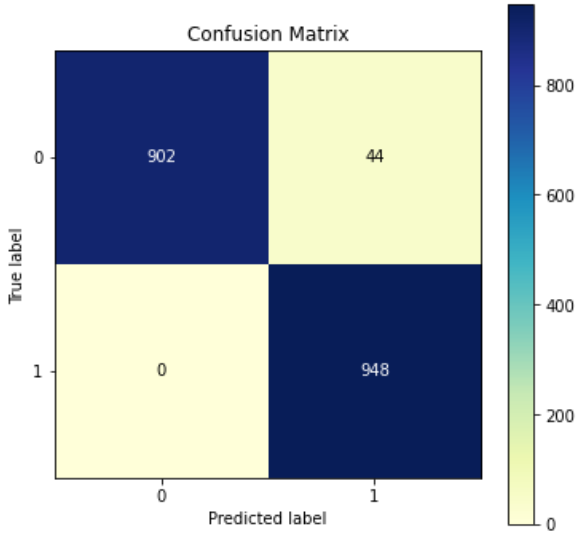
- ACCURACY: 96.3%
- PERCISION: 93.2%
- RECALL: 100%
- F1-SCORE: 96.4%
- AUC-ROC SCORE: 99.3%

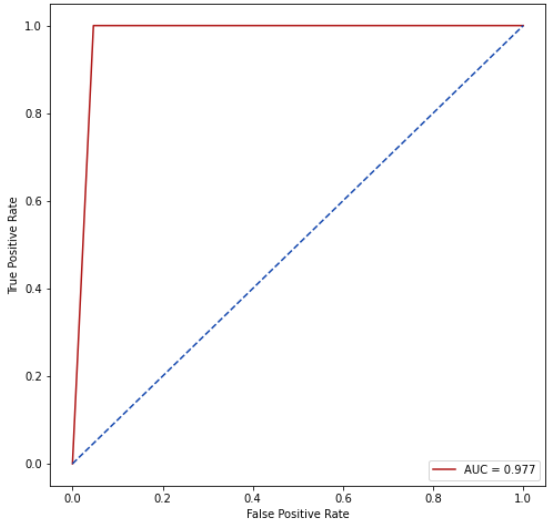




DECISION TREE CLASSIFER

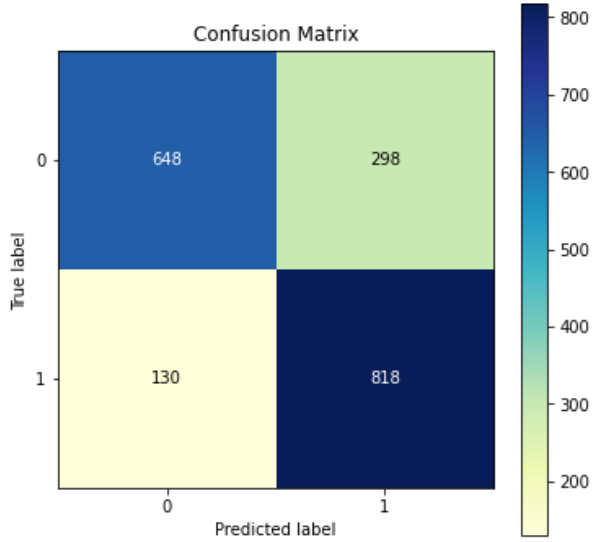
- ACCURACY: 97.6%
- PERCISION: 95.5%
- RECALL: 100%
- F1-SCORE: 97.7%
- AUC-ROC SCORE: 97.6%

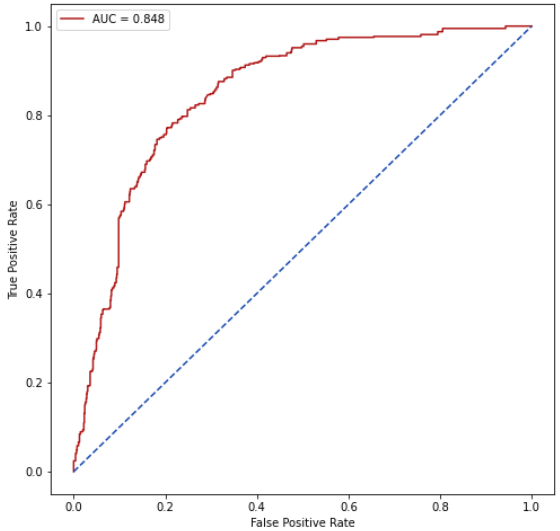




SVC

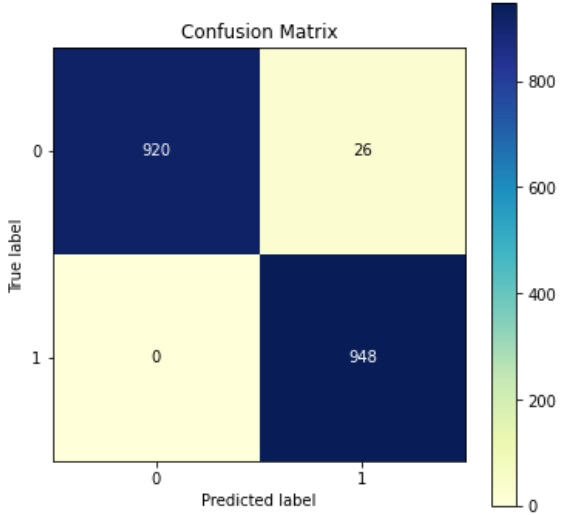
- ACCURACY: 77.4%
- PERCISION: 73.2%
- RECALL: 86.2%
- F1-SCORE: 79.2%
- AUC-ROC SCORE: 84.8%

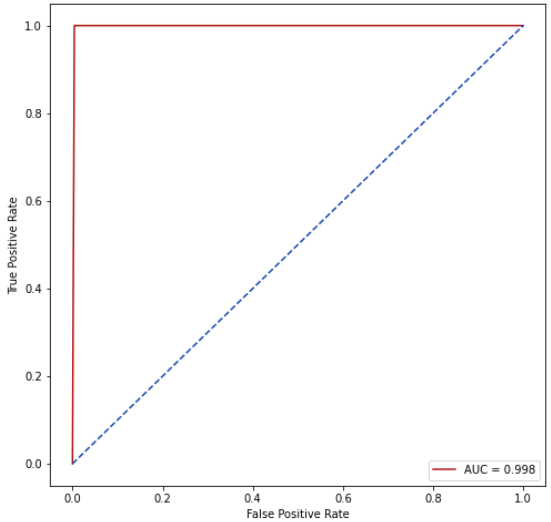




EXTRA TREES CLASSIFIER

- ACCURACY:98.6%
- PERCISION: 97.3%
- RECALL: 100%
- F1-SCORE: 98.6%
- AUC-ROC SCORE: 99.8%

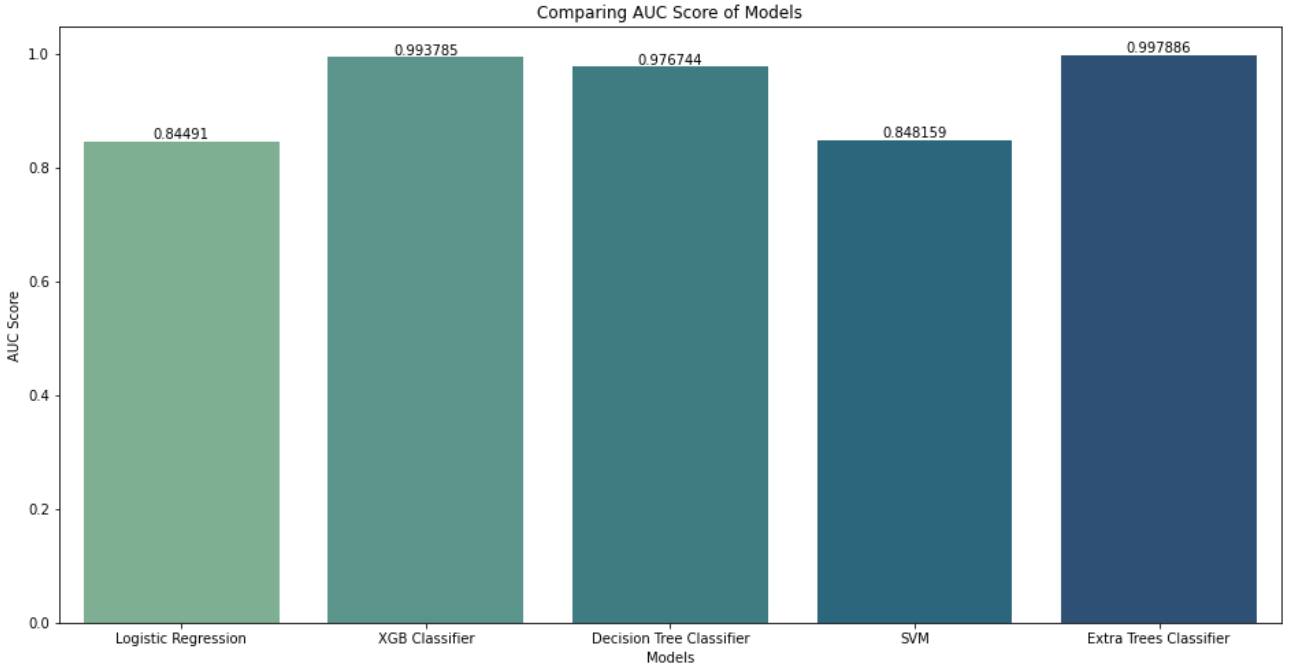




06.

CONCLUSIONS

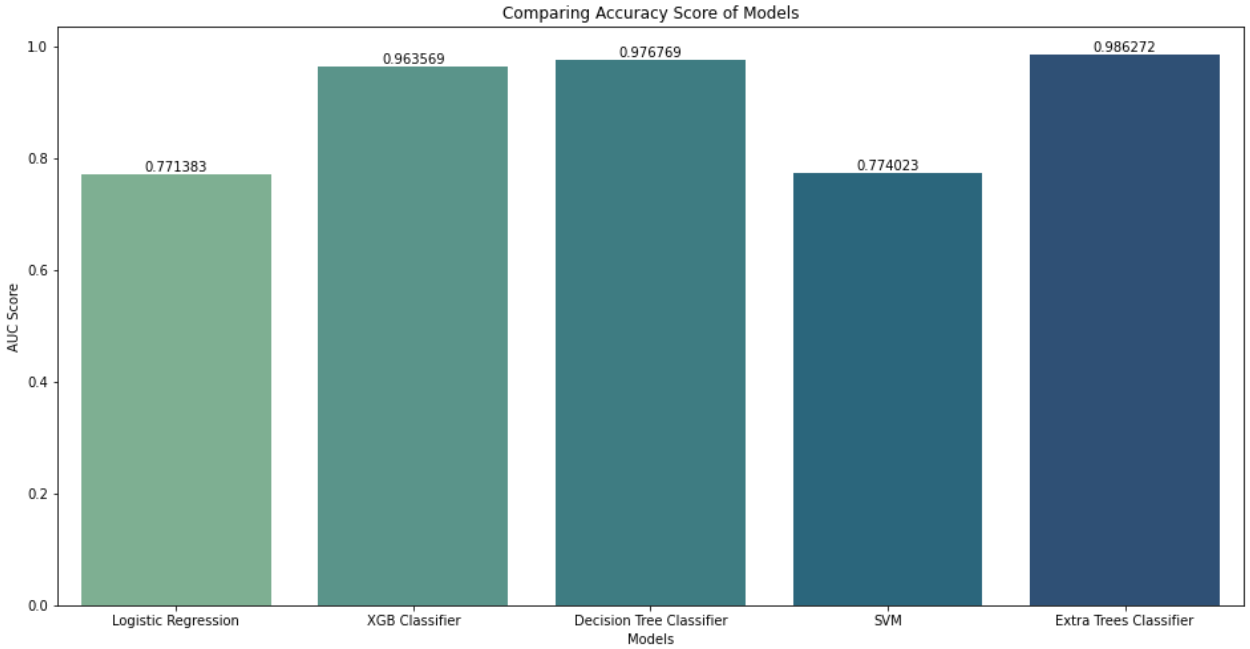
COMPARING AUC SCORE OF MODELS



CONCLUSIONS

- Extra Trees Classifier Model performs the best with 99.8% AUC score

COMPARING ACCURACY SCORE OF MODELS



- Also we can see that Extra Trees Classifier Model performs the best with 98.6% Accuracy score

CONCLUSIONS

THE END

‘’ The model does not misclassify any stroke patient as a non-stroke patient, which is fascinating. We don't want any patient who is suffering from a stroke to be categorized as having a non-stroke and so not receive the necessary medical care. ’’

THANKS THANKS

THANKS!

Do you have any questions?