HEALTHCARE ANALYTICS AND PREDICTIONS ON HEART DISEASE DATA

TECHNOLOGIES: BUSINESS INTELLIGENCE

DOMAIN: HEALTHCARE

ABSTRACT:

- Using the dataset provided, we built a simple model for predicting the Heart Disease in patients which can be used in future. Our model revealed a 86% of Accuracy using past data and few major features which can help predicting the heart disease in future.

INTRODUCTION

Health is real wealth in the pandemic time we all realized the brute effects of covid-19 on all irrespective of any status. Techniques of data science and predictive analytics can be used to predict the Heart Disease. In this project, we are provided with the Heart disease.csv. This database contains 76 attributes, but all published experiments refer to using asubset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The “goal” field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Attribute Information:> 1. age> 2. sex> 3. chest pain type (4 values)> 4. resting blood pressure> 5. serum cholesterol in mg/dl> 6. fasting blood sugar > 120 mg/dl> 7. resting electrocardiographic results (values 0,1,2)> 8. maximum heart rate achieved> 9.exercise induced angina> 10. oldpeak = ST depression induced by exercise relative to rest> 11. the slope of the peak exercise ST segment> 12. number of major vessels (0-3)colored by flourosopy> 13. thal: 3 = normal; 6 = fixed defect; 7 = reversible defect.

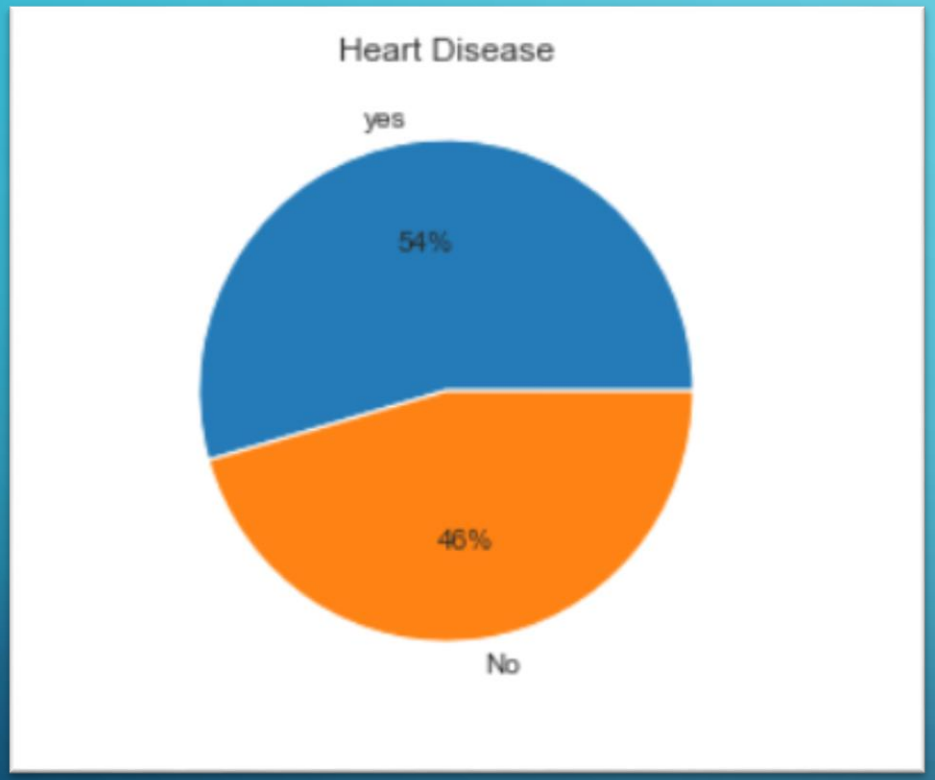
PROJECT OBJECTIVE AND BENEFITS

- The goal of this project is to use techniques of data science to Analyze and Predict the Heart Disease, to Find key metrics and factors and show the meaningful relationships between attributes. We are required to analyze this health and medical data for better future preparation.

- BENEFITS:
- Detection of Heart Disease
- Gives better insights of patients
- Helps in easy flow for managing resources
- Heart disease is predicted

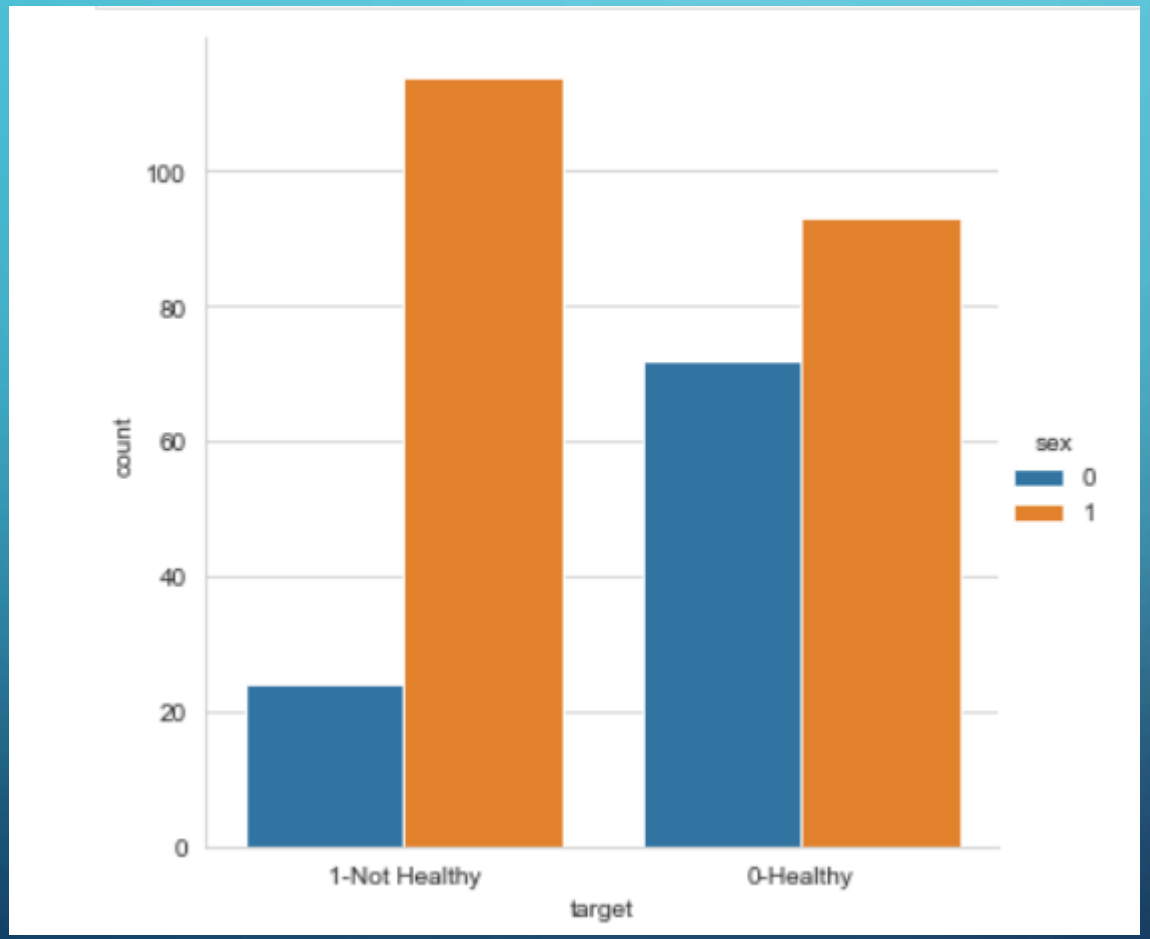
EXPLORATORY DATA ANALYSIS

The dataset was important in Python and calculations were performed using python. We plot the following figures:

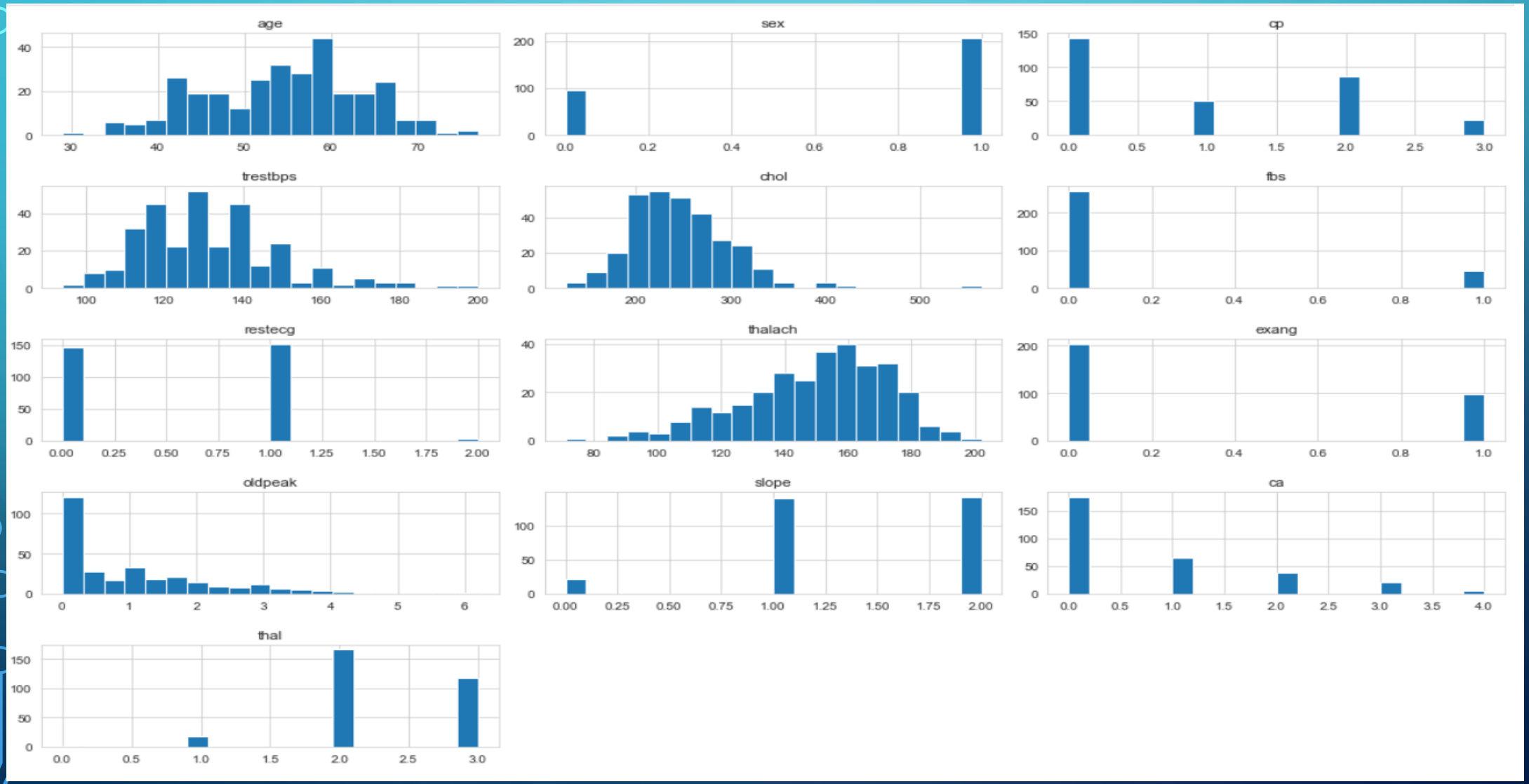


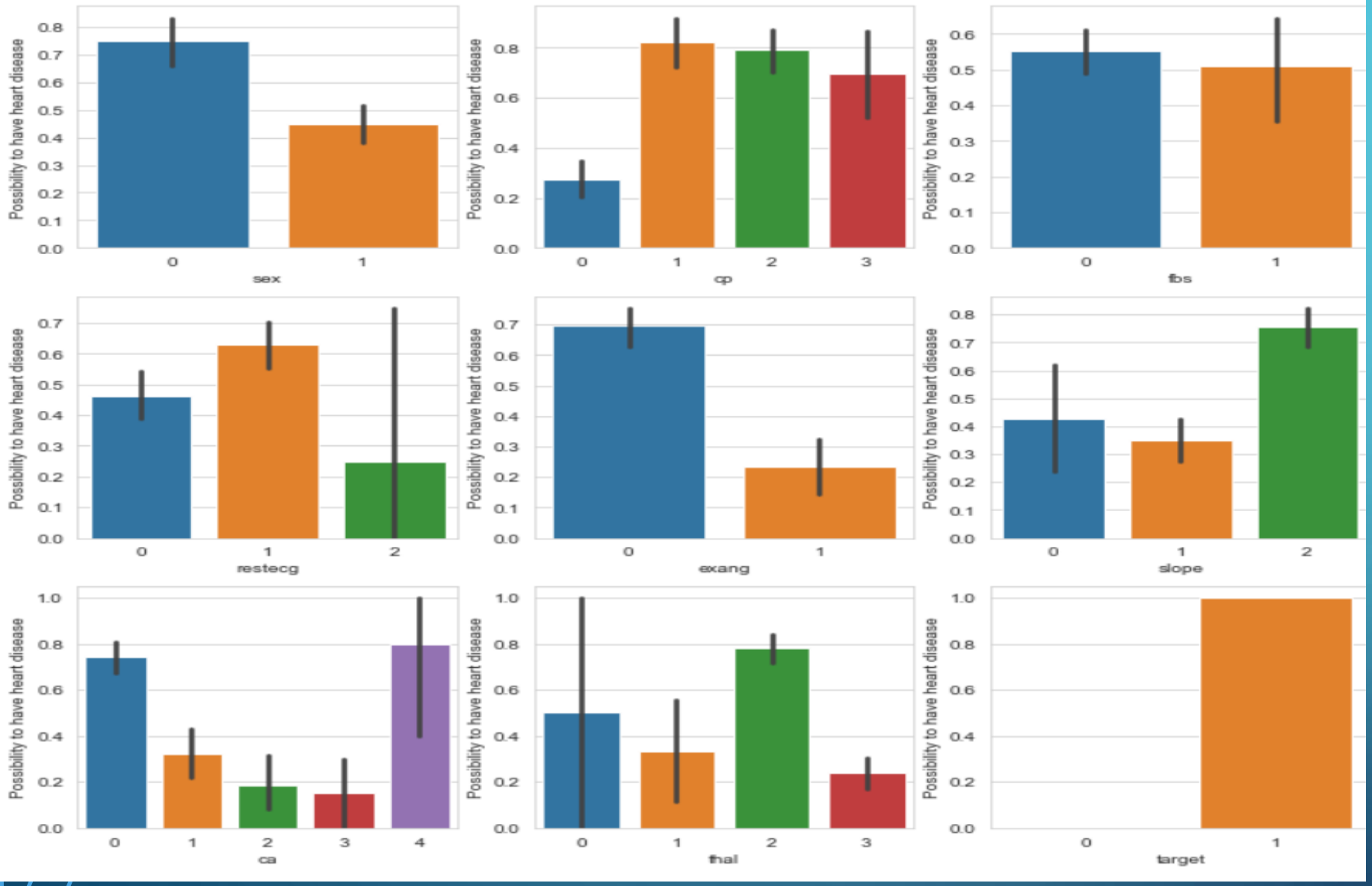
The Pie chart shows the percentage of patients having Heart Disease(yes) and patients who do not have heart disease (no).

The following graph shows the count of male and females who are healthy and not healthy.
It is observed that females are more Healthier than male.



THE HISTOGRAMS CONCLUDES AGE, CHOLESTEROL, RESTING BLOOD PRESSURE, AND MAXIMUM HEART RATE ACHIEVED PLAYS THE MAJOR ROLE IN DETECTION OF HEART DISEASE.





- Here we can observe that females have the higher possibilities in order to get heart disease.
- In chest pain Type 2 (atypical agina(bar1)) has the possibility to have heart disease.
- Exercise induced angina has the higher possibility of having heart disease.

The above Barplot shows the possibilities of having heart disease as per the features.

MODEL SELECTION

- Use of Random Forest for predicting Heart Disease gave 86% of accuracy. A random forest is **a machine learning technique that's used to solve regression and classification problems**. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees.

PREDICTIONS

- We have demonstrated the relationship between the feature responsible for Heart disease, we can use random forest model to predict if the patients have chances of heart disease or not.
- Features i.e. Age, chest pain, resting blood pressure, cholesterol, and Maximum Heart rates achieved are responsible for having heart disease.
- Patients who have Type 2 (atypical agina), high resting blood pressure=136 or more, and high cholesterol(230-257.40) have chances of getting heart disease.