Predicting whether a person is suffering from a heart related disease or not!

Table of contents

Scope of Interest..

As this project basically focuses on applying data science techniques into medical sciences, the scope of this data science project would mostly be limited to medical experts! Doctors would be thrilled to know that there is a model that would predict the well-being level of a person's heart with some mere numeric values.
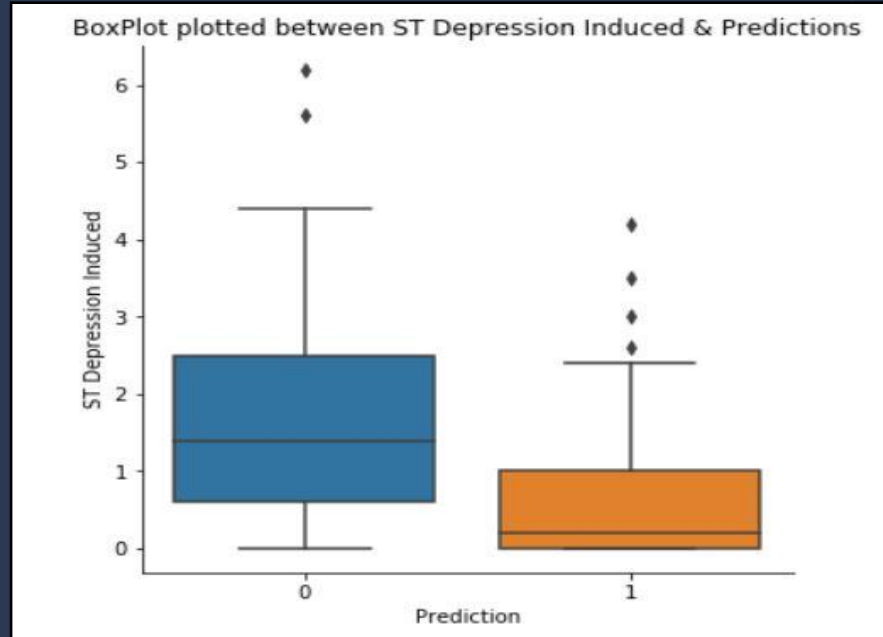
Data Cleaning

The amount of work I had to do as a part of data preparation and cleaning has been very small. The dataset I used had already applied all the techniques related to data preparation. All the categorical value based columns excluding the target were converted to numerical value based columns performing the one hot encoding operation. Many unwanted features that wouldn't help in prediction were removed and the number of features including the target came down to fourteen which is generally the number of features a good dataset has. All the data redundancy issues were taken care of properly. I then had to rename all the features to proper and meaningful names and after importing the .csv file as and into a data frame it looked like the following,

	Age	Sex	Chest Pain Type	Blood Pressure	Cholestrol	Blood Sugar	Electrocardiac Measurement at Rest	Maximum Heart Rate	Exercise indused Angina	ST Depression Induced	Slope of Peak Exercise ST Segment	Number of Major Vessels	Thalassemia	Prediction
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Maximum Heart Rate V/S Target...

BoxPlot plotted between Maximum Heart Rate & Predictions
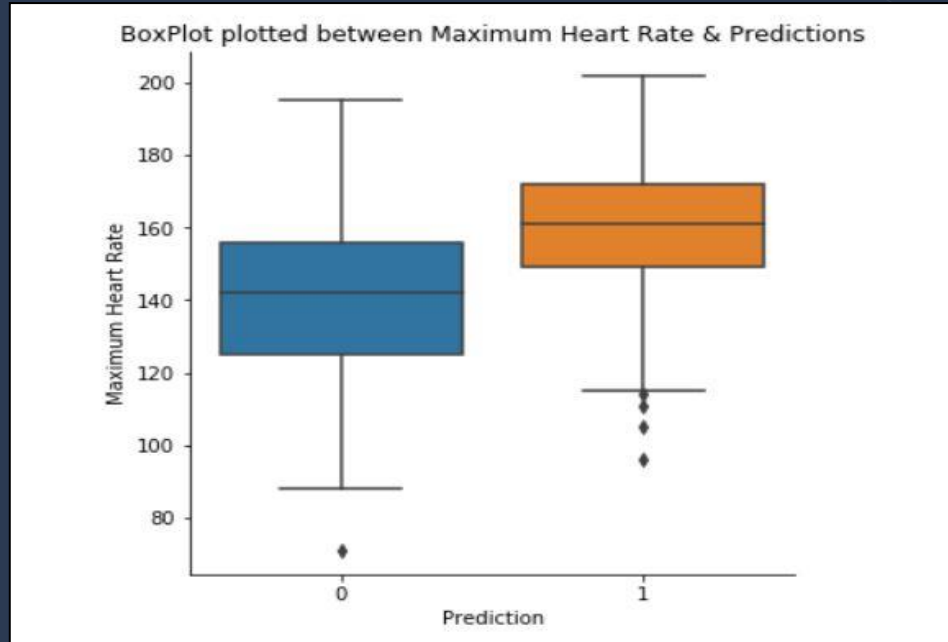
ST Depression Induced V/S Target...

Exercise induced Angina V/S Target…

Relationship between Exercise indused Angina & Predicitons
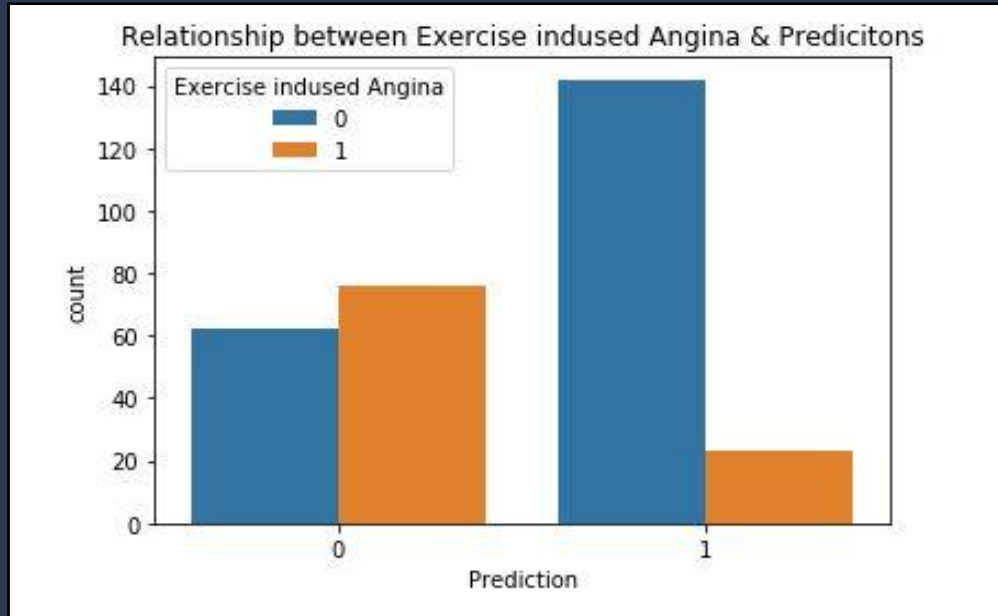
Chest Pain Type V/S Target...

Thalassemia V/S Target...

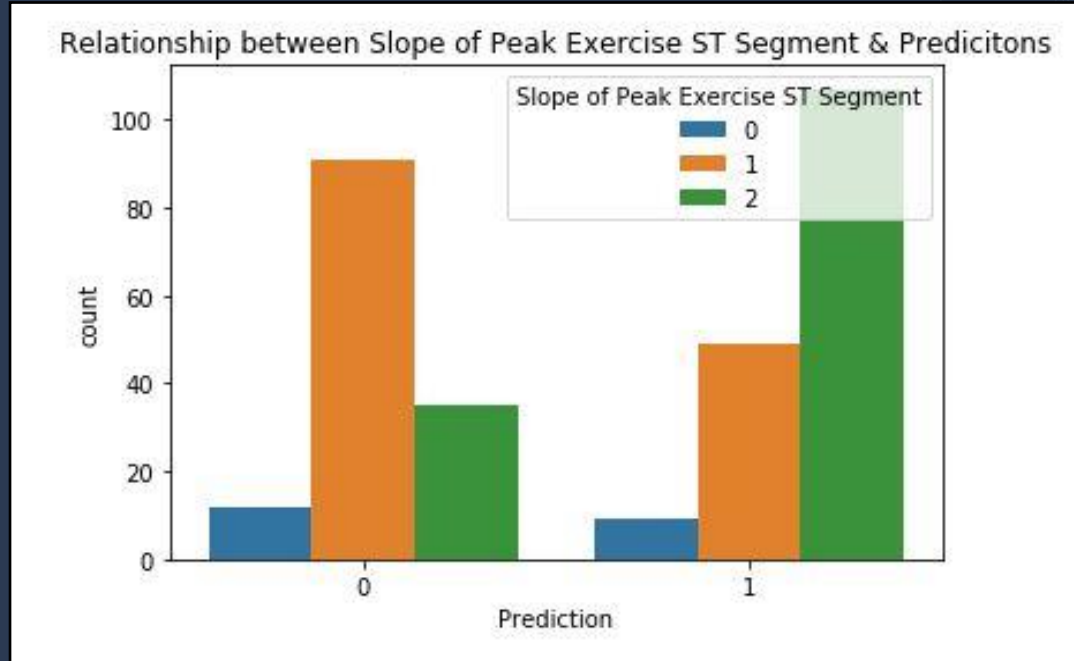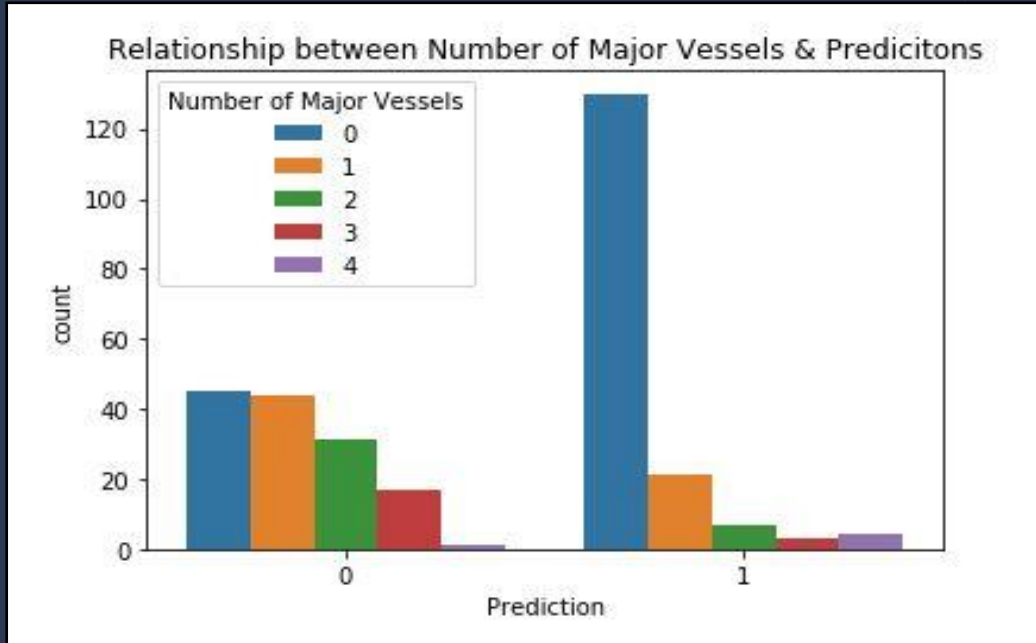Relationship between Thalassemia & Predicitons
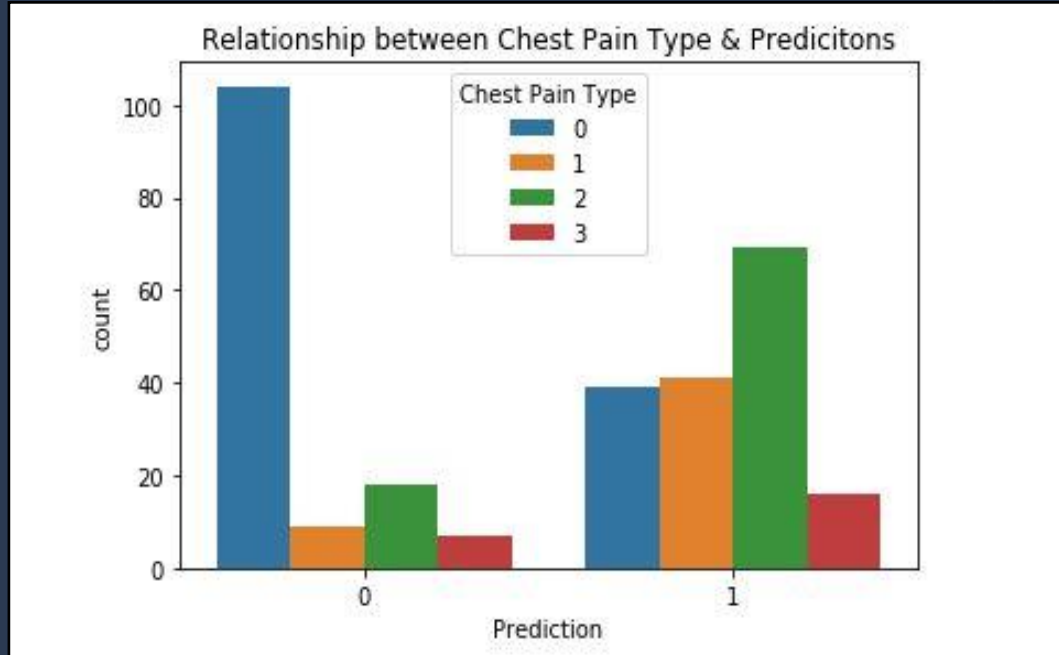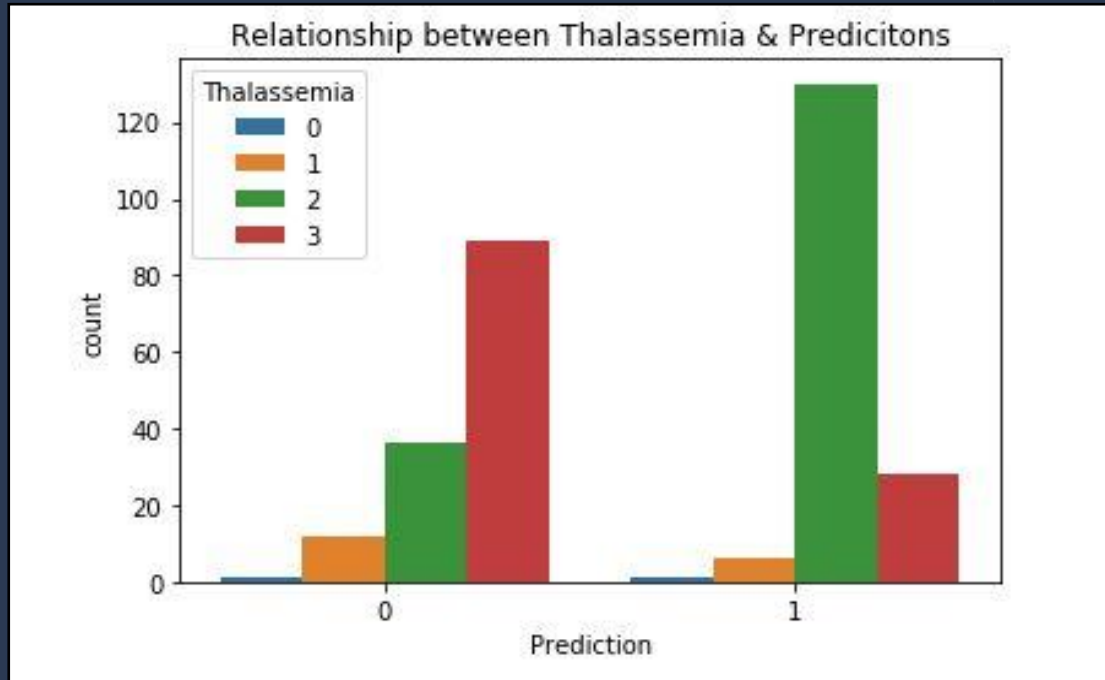
Slope of Peak Exercise ST V/S Target...

Number of Major Vessels V/S Target...

Future Scope & Conclusions

The model was trained and tested on a dataset that had entries that could expand itself as time goes by. Even the columns i.e. the attributes could be even more convincing and could increase in number therefore leading to easy and better predictions. The model was trained and tested on different standardized classification algorithms. The best could be therefore chosen, an API could be designed, connected to a web application, therefore making it easy for patients and users as well. There's a lot left to do beyond imaginations. 😊

This is a report that is being made for the project which is a part of applied data science capstone project. In my study I learnt a lot, about how different features could contribute to predicting the whether a person is suffering from a heart related disease. This project would further be taken up and with certain improvements would definitely become a good, efficient, dependable working model....