**Name: Abu Hurera, Hamza Iqbal**

**ID: F2018266218, F2018266270**

**Project Data:**

We will be working on dataset of multiple heart patients which contain many values and observation of those patients we will discuss more about that in EDA

**Project Goal:**

After observing and analysis of data we will perform different kind of machine learning models on this data and predict heart attack in other patients

**Exploratory Data Analysis(EDA):**

Definition of data

#### **Age : Age of the patient**

#### **Sex : Sex of the patient**

#### **exang: exercise induced angina (1 = yes; 0 = no)**

#### **cp : Chest Pain type chest pain type**[**¶**](http://localhost:8889/notebooks/AIproject/AI%20Project.ipynb#cp-:-Chest-Pain-type-chest-pain-type)

#### **ca: number of major vessels (0-3)**

#### **trtbps : resting blood pressure (in mm Hg)**

#### **chol : cholestoral in mg/dl fetched via BMI sensor**

#### **fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)**

#### **rest\_ecg : resting electrocardiographic results**

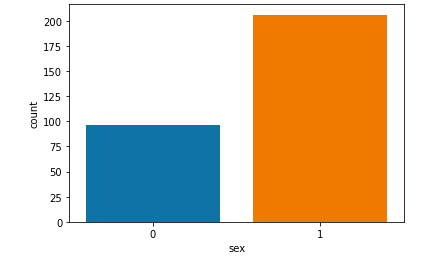
#### **thalach : maximum heart rate achieved**

#### **thalach : maximum heart rate achieved**

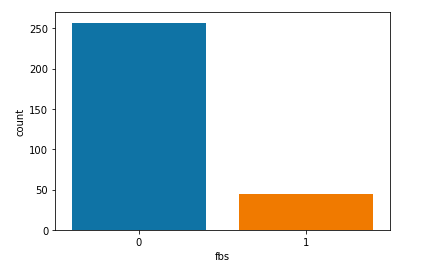
#### **target : 0= less chance of heart attack 1= more chance of heart attack**

**Now we will plot graphs of output on the basis of different features of data**

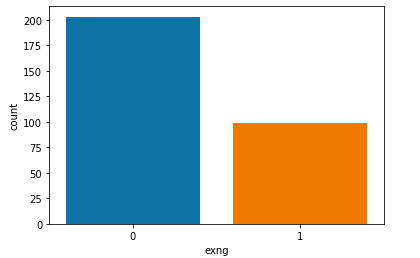
**Output vs Gender:**

****

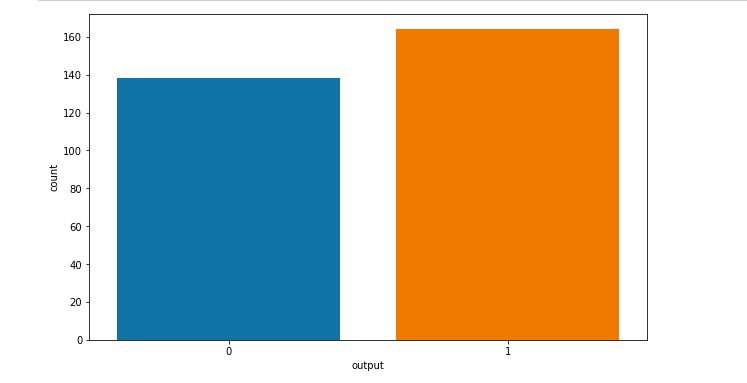
**Output vs fbs**

****

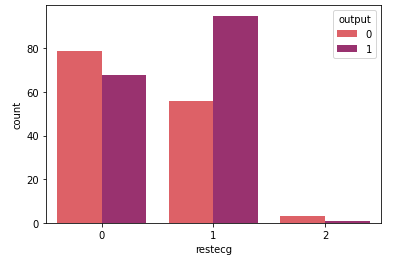
**Output vs exng**

****

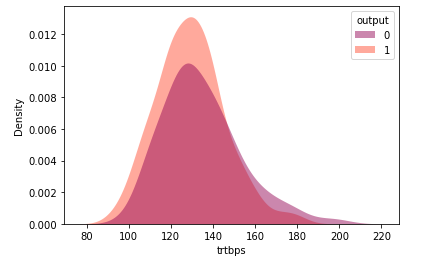
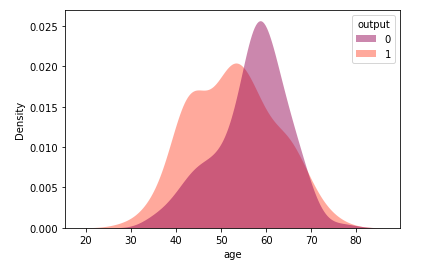
**Countplot of output on the base of gender:**

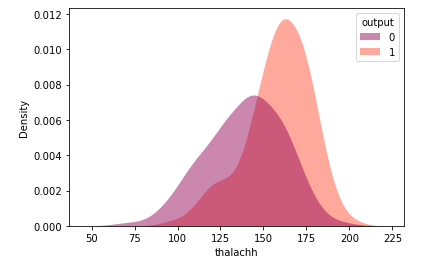
****

**Countplot of output on the base of restecg:**

****

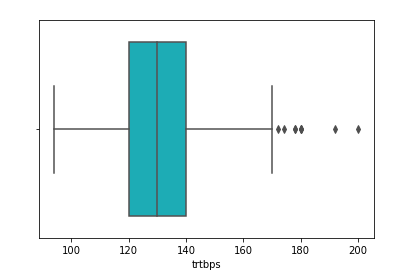
# **Distribution of Non-Categorical Data (Continous Data)**

****

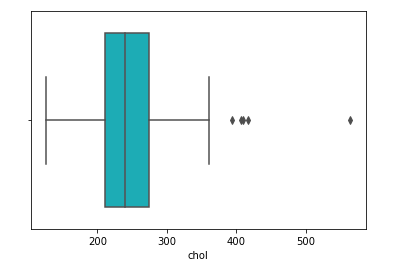
****

**Looking for outliers using boxplot**

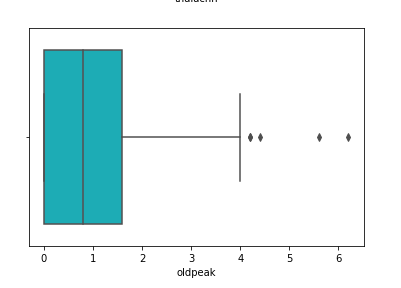
**Boxplot of trtbps**

****

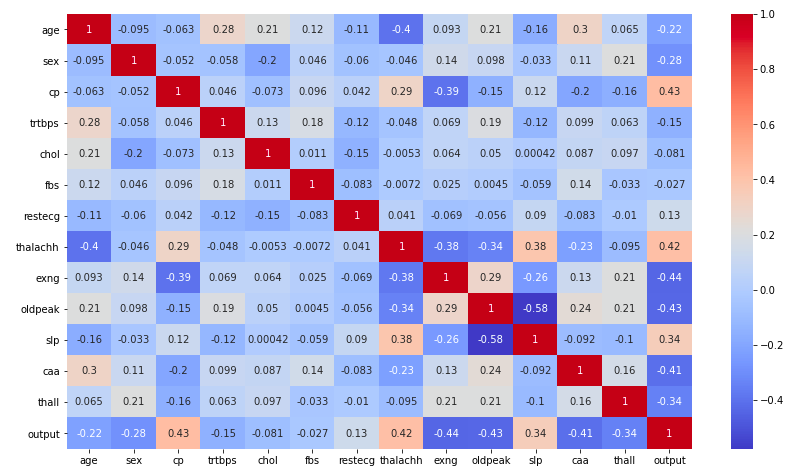
**Boxplot of cholesterol:**

****

**Boxplot of oldpeak**

****

**To Analyze the correlation and outliers we used a Heat Map**

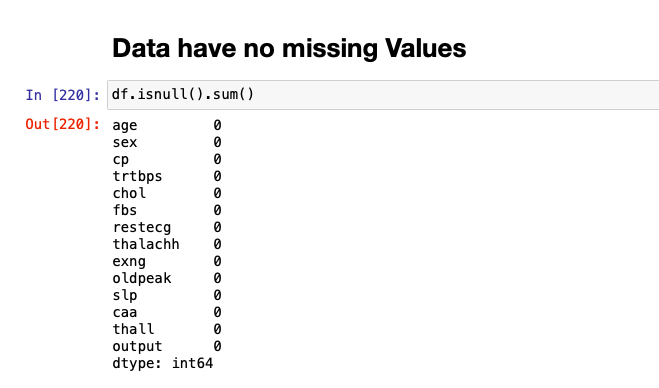
****

**Feature And Label Selection**

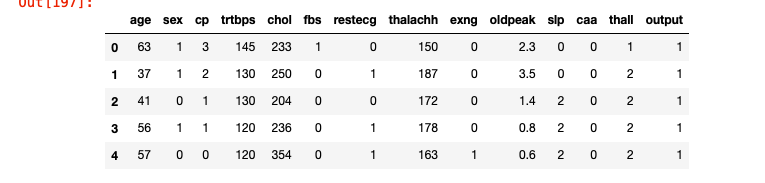
Since every column is numerical value and everything is important in terms of medical domain knowledge so every column of the data will be taken as feature and we also categories data on the basis of output

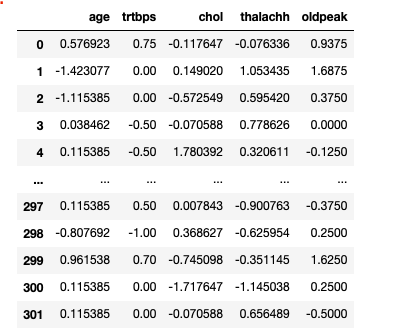
**Here we we will do Feature scaling**

**Handling null Values:**

****

**Data Before Scaling**

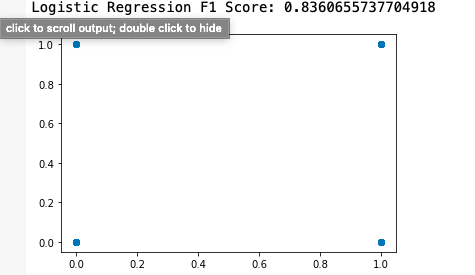
****

**Data After using Robust Scaler**

**Modeling :**

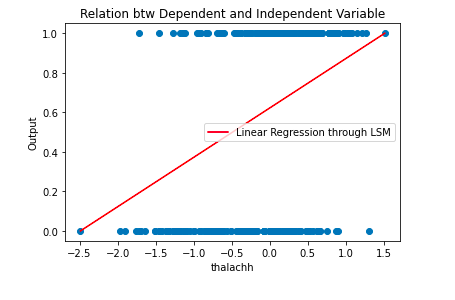
**Logistic Regression:**

After Analysis of data we have learned that our data is have two categories on base of output which is 0 and 1 so here logistic regression is the best solution to make the prediction



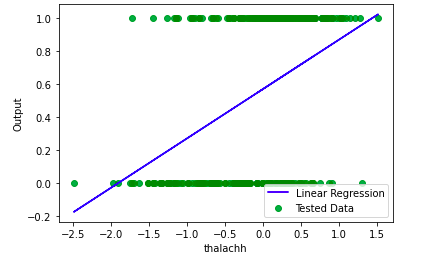
As you can see that Logistic gave some very good results in terms of accuracy score

**Now we will do Linear Regression using least square method based on a important feature**

****

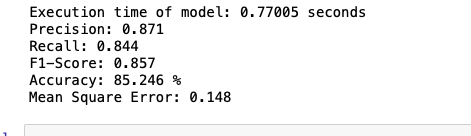
AS you can see it is also pretty accurate

**Now we will do Linear Regression method based on an important feature**

****

**Finally we have used Random Forest Classifier :**

As we have discussed earlier that our data is categorical so it will also give us good result



**Conclusion:**

**After comparing the results of all the model i feel that logistic regression and Random Forest Classifier gives us the best result**