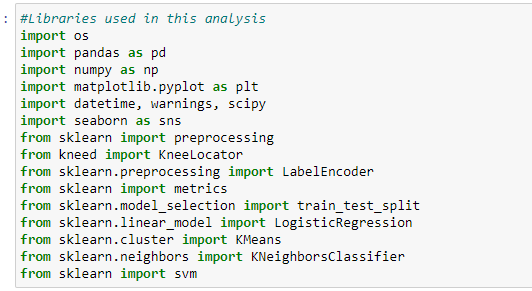
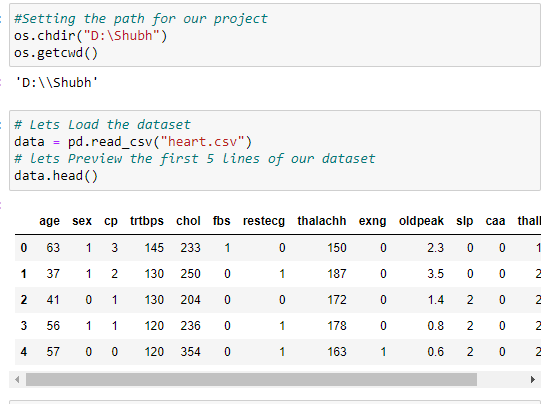
**HEART DISEASE PREDICTION**

We aim to find the model that best predicts heart disease risk so that it can be used to screen for the disease in high-risk populations.

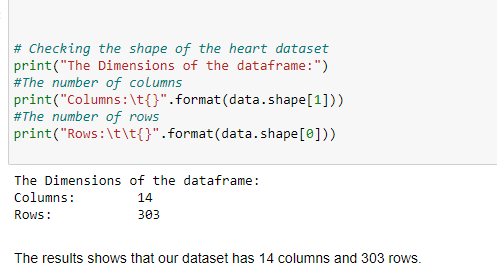
First, we loaded the most important libraries in this analysis, as shown below



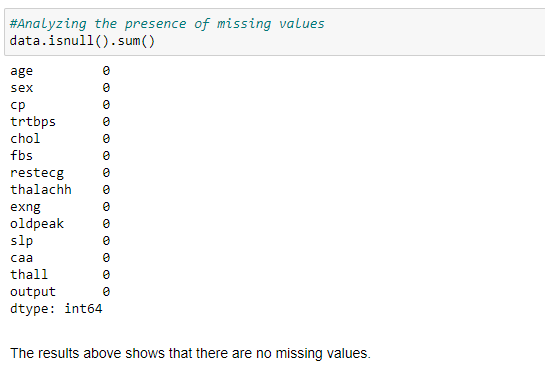
The second step was setting up the path of my project, loading the dataset, and displaying the first lines of the dataset.



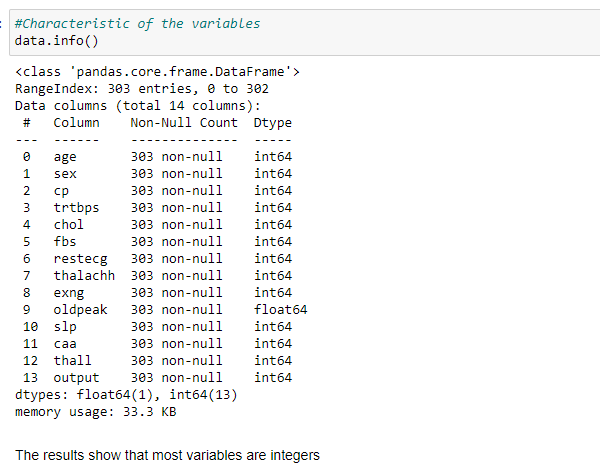
We also checked the dimension of the dataset



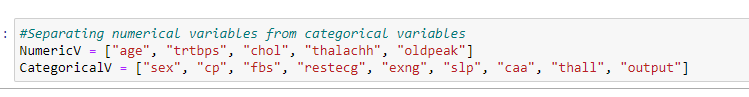
The step was to analyze whether the dataset had missing values. Fortunately, the results revealed that there were no missing values.



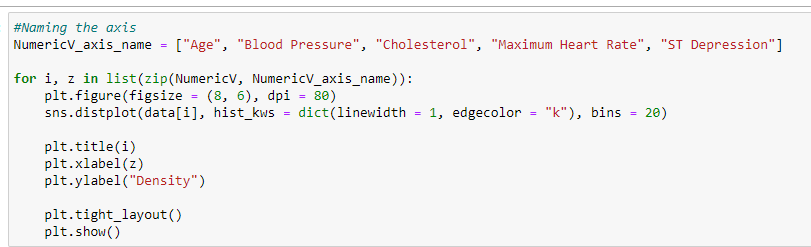
I also analyzed the characteristics of the variables as shown below



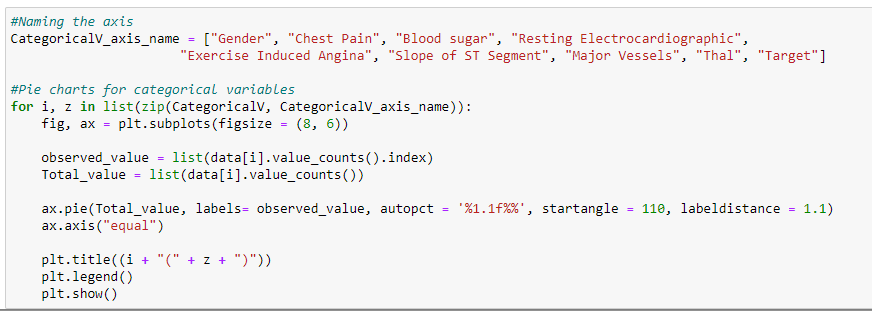
To analyze the variables better, I separated the numerical variables and categorical variables as shown below



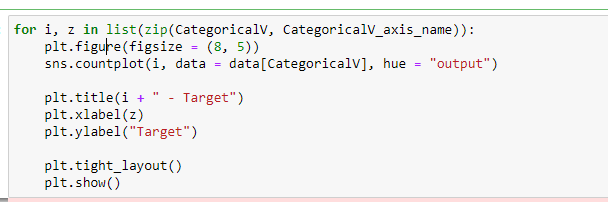
I first analyzed the numeric data using density plots. The code produced all the graphs for the numerical variables only.



Second, I analyzed the categorical variable using pie charts where the following code produced all the charts.

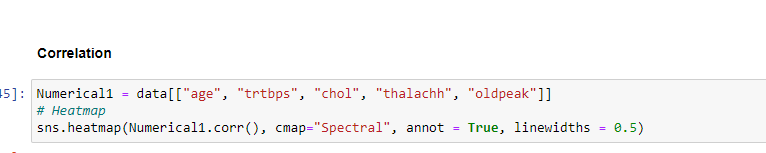


The next visualizations included visualizing the categorical independent variables and the dependent variables using bar charts as shown below;

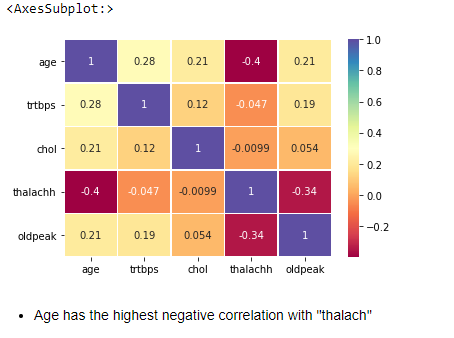


Correlation

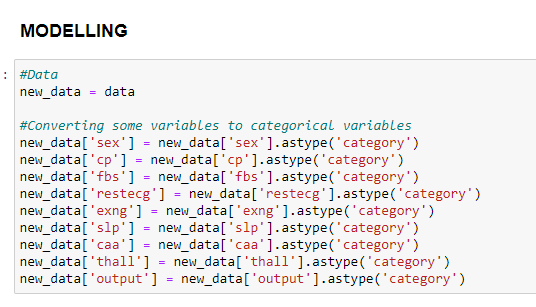
Correlation analysis was done to check whether variables are related to each other. This was done using a heatmap as shown in the code below



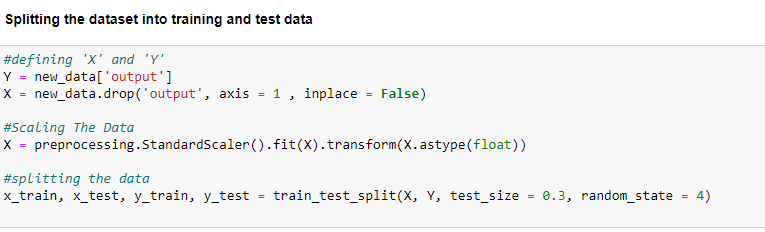
The results of the correlation are as shown below



Before modeling the data, we converted some variables to categorical variables, as shown below



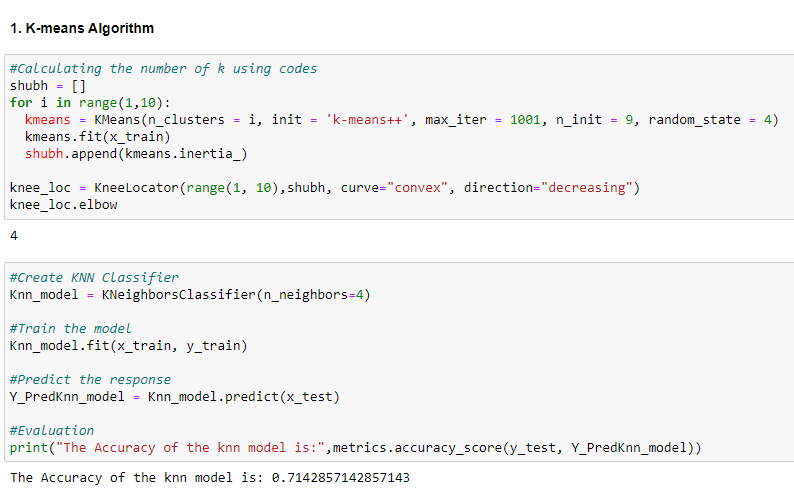
The data was then split into training and testing data, as shown below



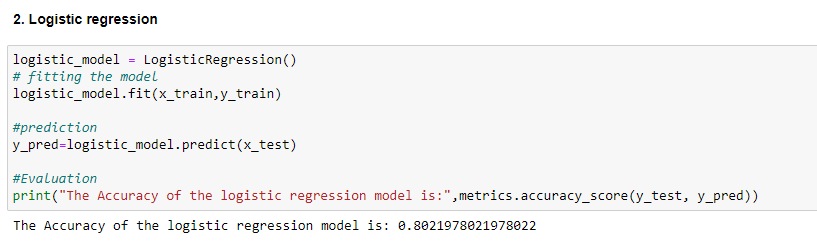
**CLASSIFICATION**

In this section, I will use three classification methods: K-means Algorithm, Logistic regression, and Support vector machines (SVM).

K-means algorithm was integrated and evaluated using accuracy score as shown below



The second algorithm was logistic regression, where it was evaluated using an accuracy score as shown below



The final algorithm was the support vector machine which was also evaluated using an accuracy score as shown below

