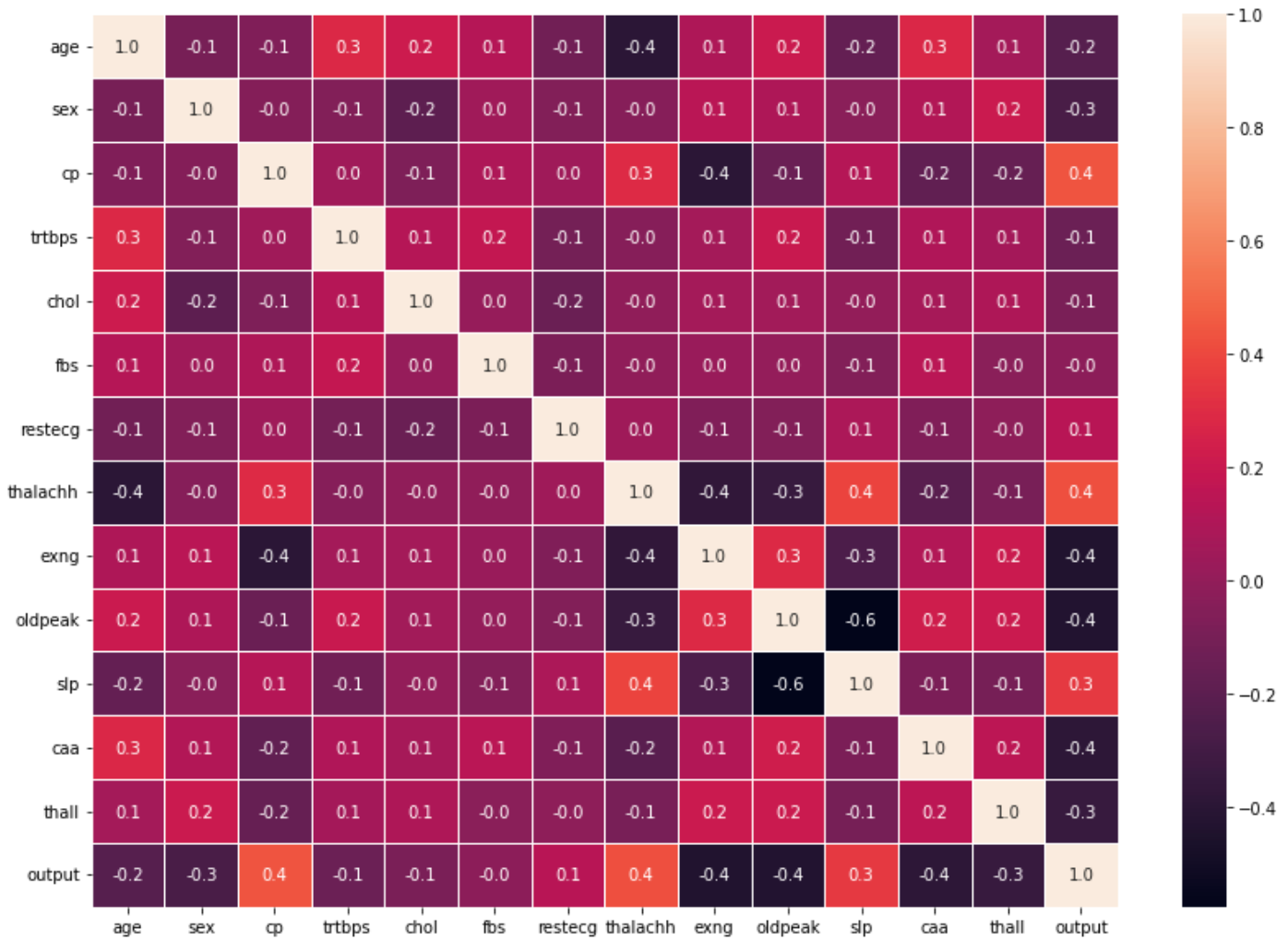
***Group 3***:- *Report 4*

**CSE523 Machine Learning**

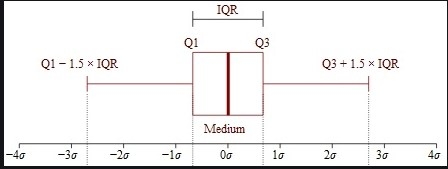
**Heart Attack Prediction**

| **Group Members** | **Roll No** |
| --- | --- |
| **Shivam Thakker** | **AU1940193** |
| **Devarsh Sheth** | **AU1940189** |
| **Pranav Gandhi** | **AU1940313** |
| **Meet Jhaveri** | **AU1940284** |

* We further wanted to check how each variable is correlated to each other. So we carry out correlation analysis by using plt.figure() and sns.heatmap() by passing in df.corr().
* Our output will be graph showing how much correlation is there between each variable with one another as shown below:-



* Further we would like to drop Uncorrelated features. First let’s do this for Categorical features. We will find features which has correlation less than 0.15 using pd.Dataframe() and passing in np.abs(df.loc[:, categoryList].corr()["output"]) < 0.15 and then storing it in a variable and Then dropping them using df.drop().
* We also did this same for Numerical Features. Dropped the features which are correlated less than 0.15.
* Then we also removed the outliers by finding interquartile range. We first found q1 = np.percentile(currentItem, 25) and q2 = np.percentile(currentItem, 75).



* And then finding Iqr by formula Iqr = q3 - q1.
* Now we will find upper limit and lower limit using below mentioned formula:-

Upperlimit = q3 + 2.5 \* Iqr

Lowerlimit = q1 - 2.5 \* Iqr

* Then we removed every sample which was higher then Upperlimit and lower than Lowerlimit using df.drop().
* Hence, In this way we got rid of Outliers.

***Thank you!***