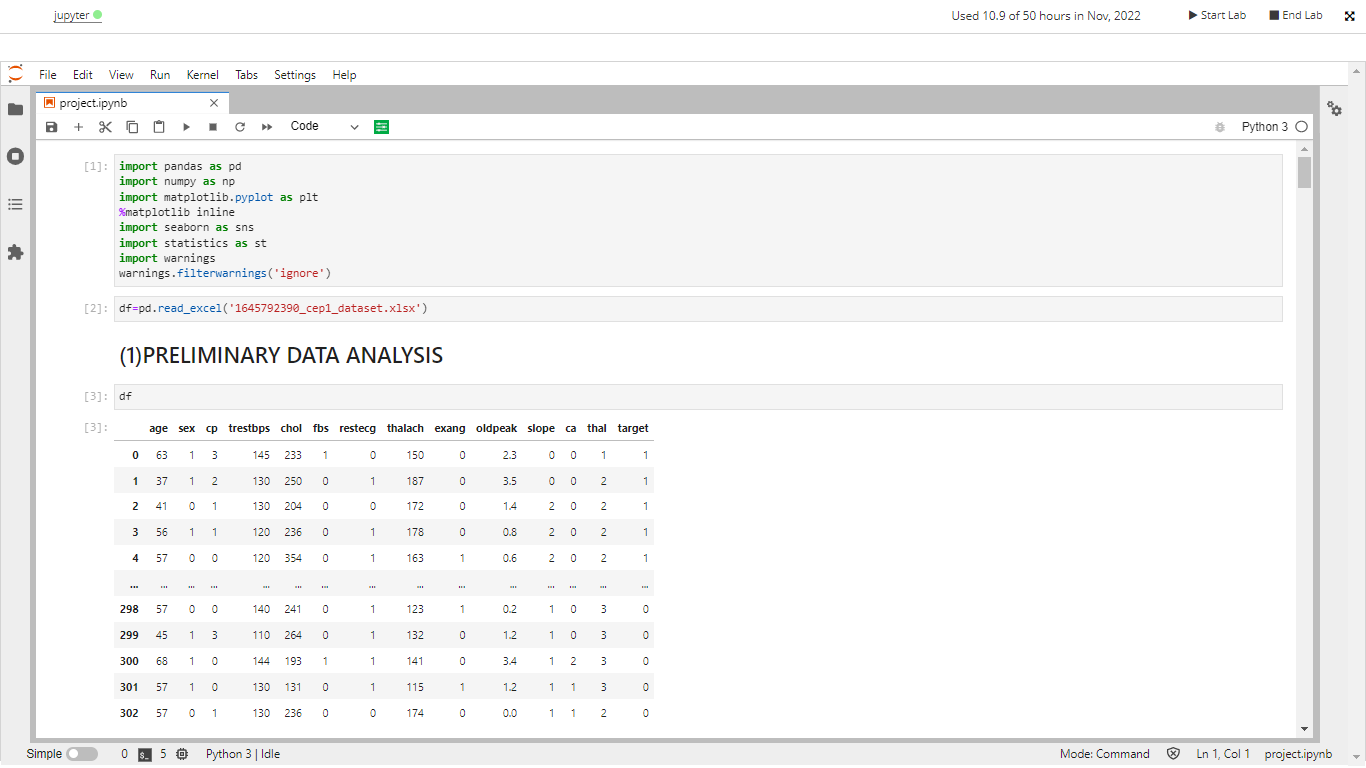
**MACHINE LEARNING PROJECT ON HEALTH CARE AND CARDIOVASCULAR DISEASES**

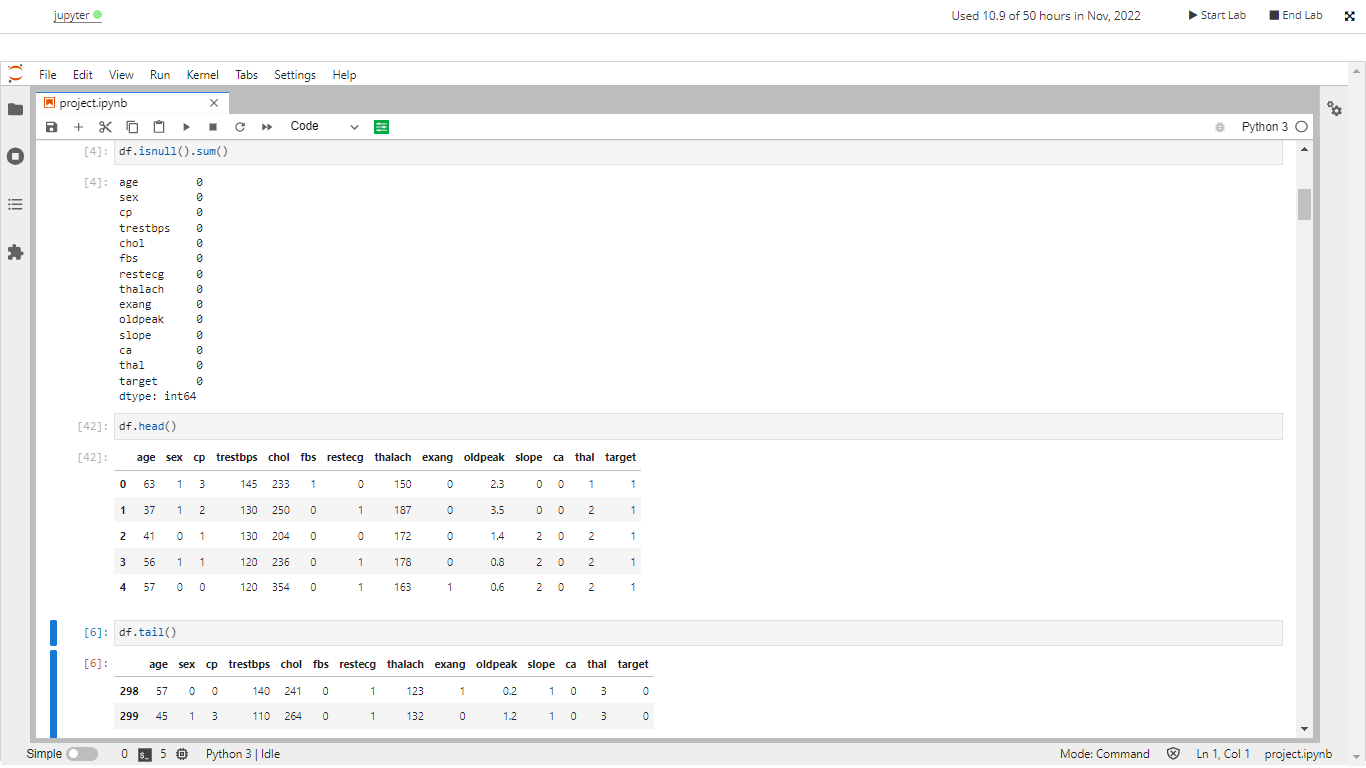
* B.MANISHA SINGH

The project consists of detailed study of the given data and two models built using logistic regression and random forest methods and the comparison of their efficiencies. The code has been marked according to the question numbers and the inferences are included. The code and question wise screenshots are as follows



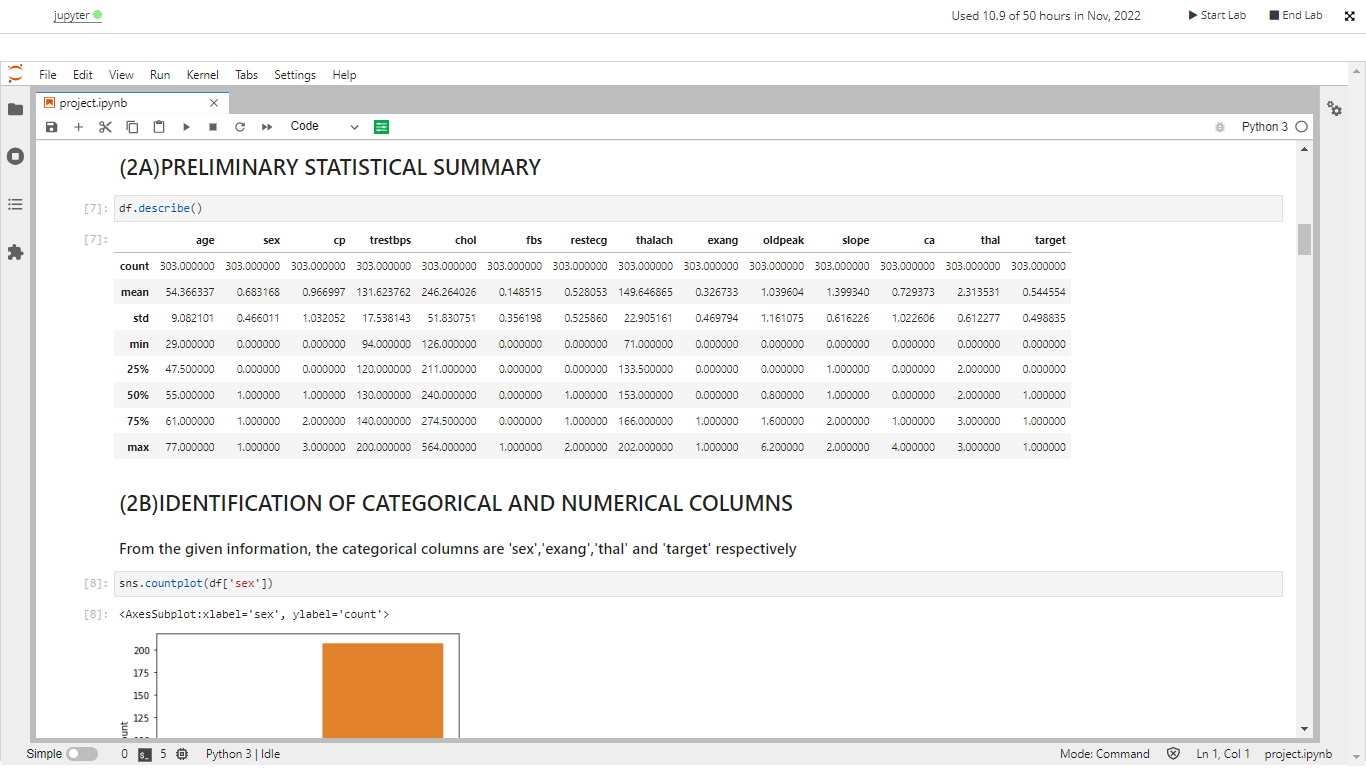
**1. PRELIMINARY ANALYSIS**

****

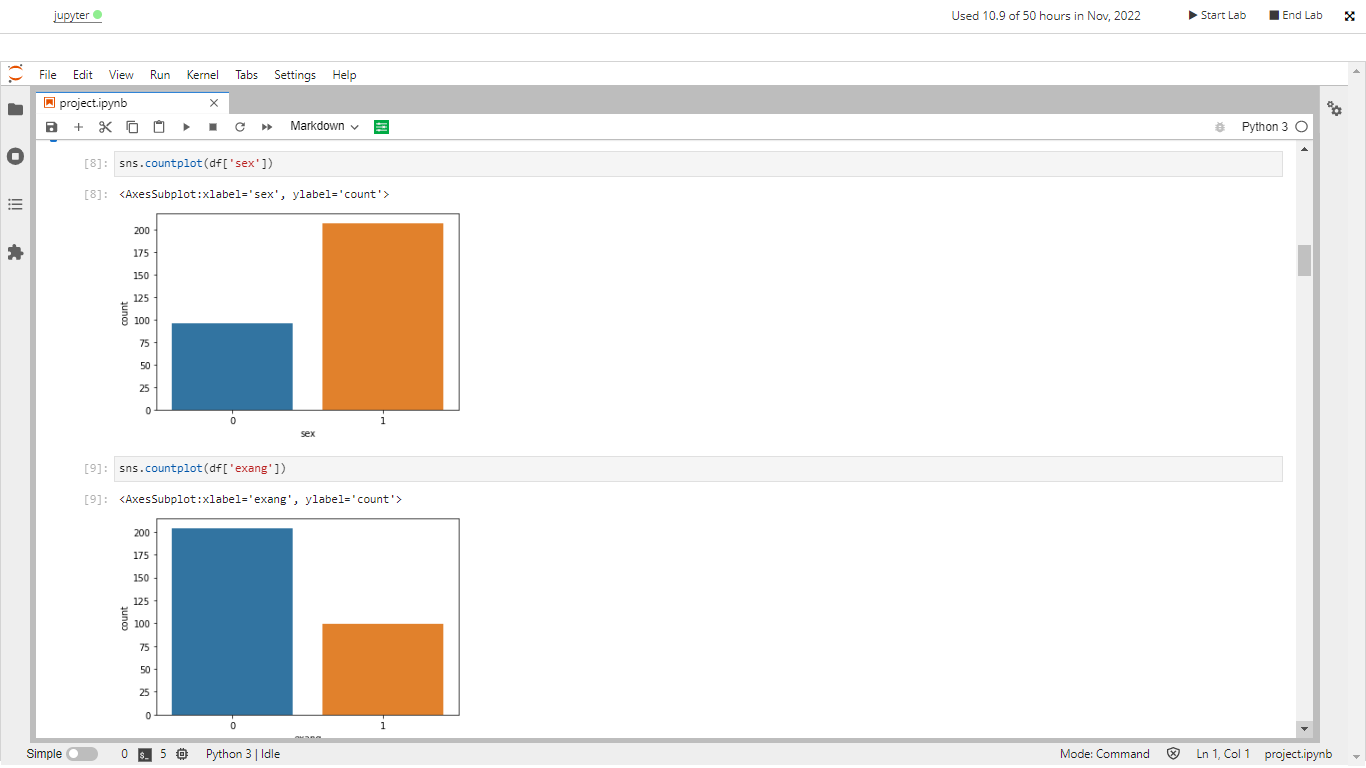
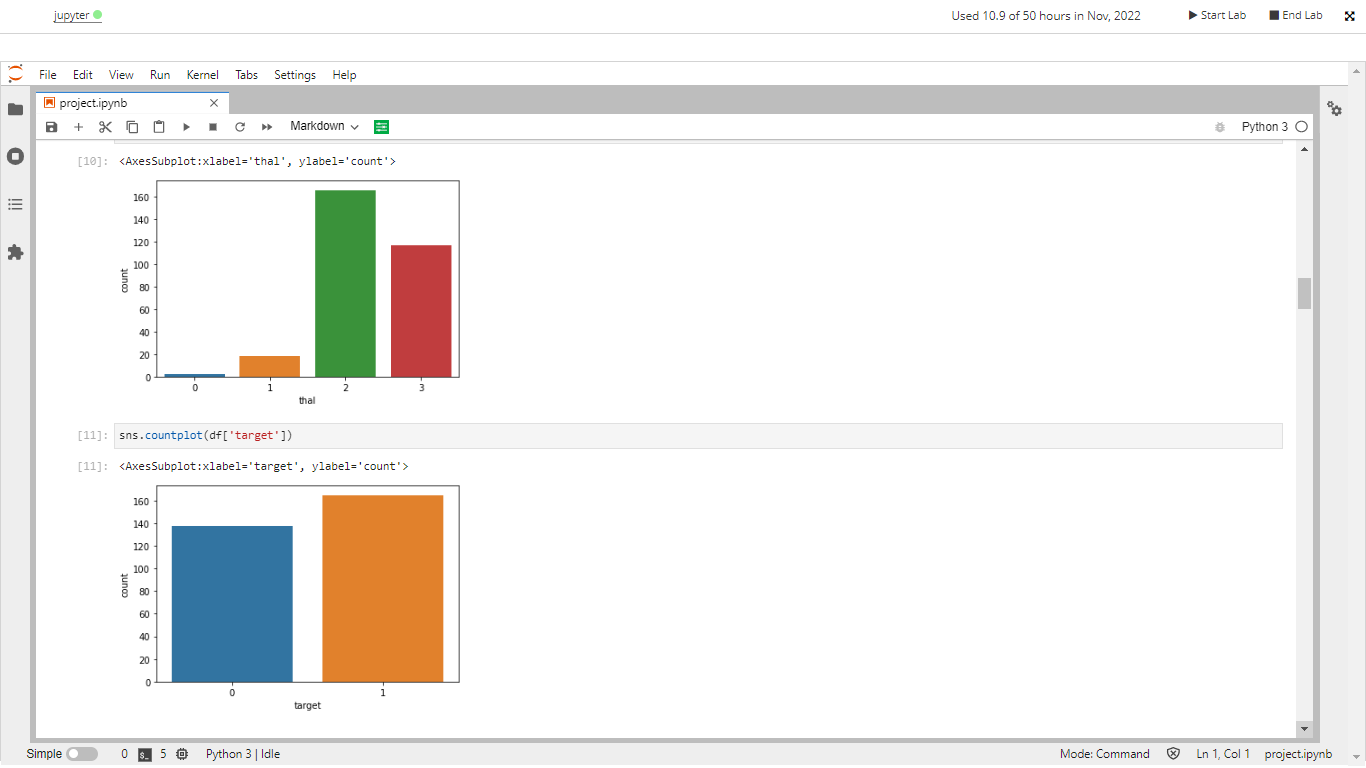
****

**2. DETAILED REPORT**

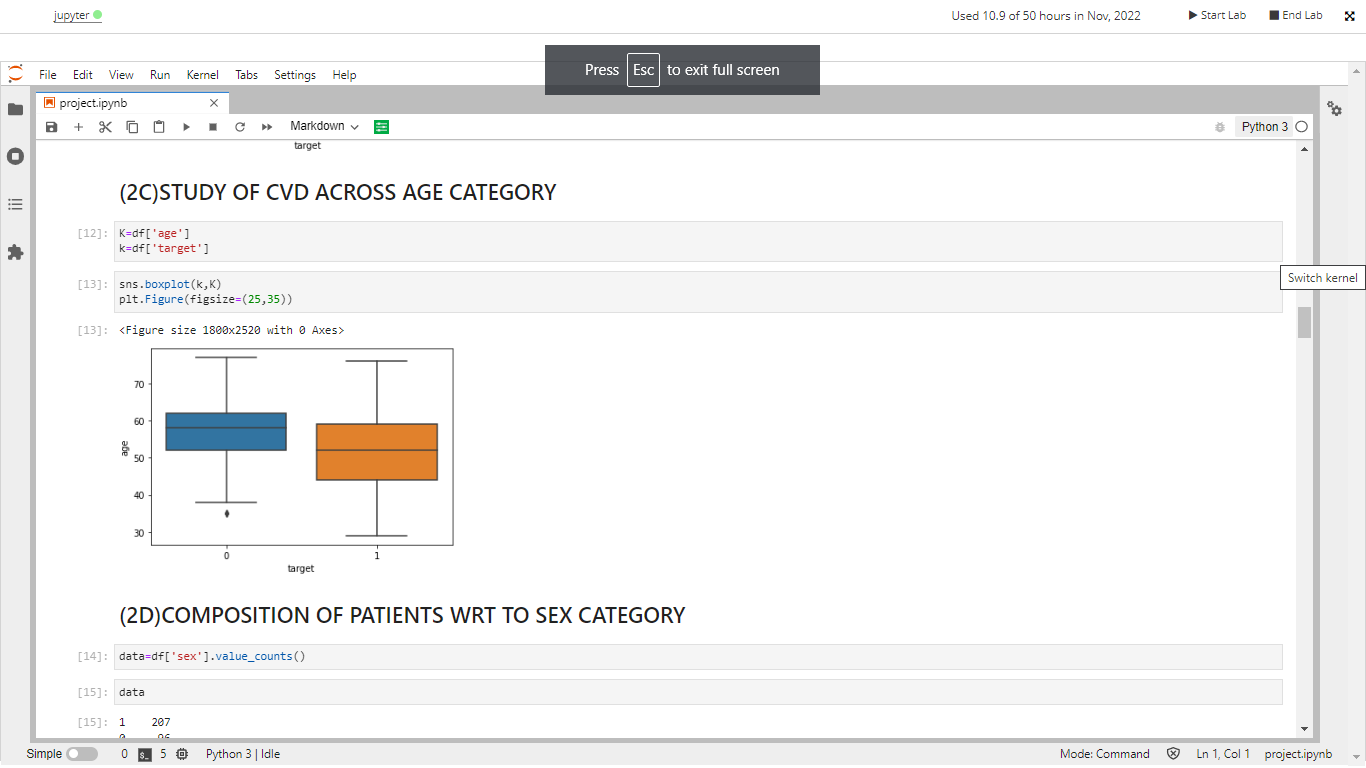
**A. PRELIMINARY STATISTICAL ANALYSIS**

****

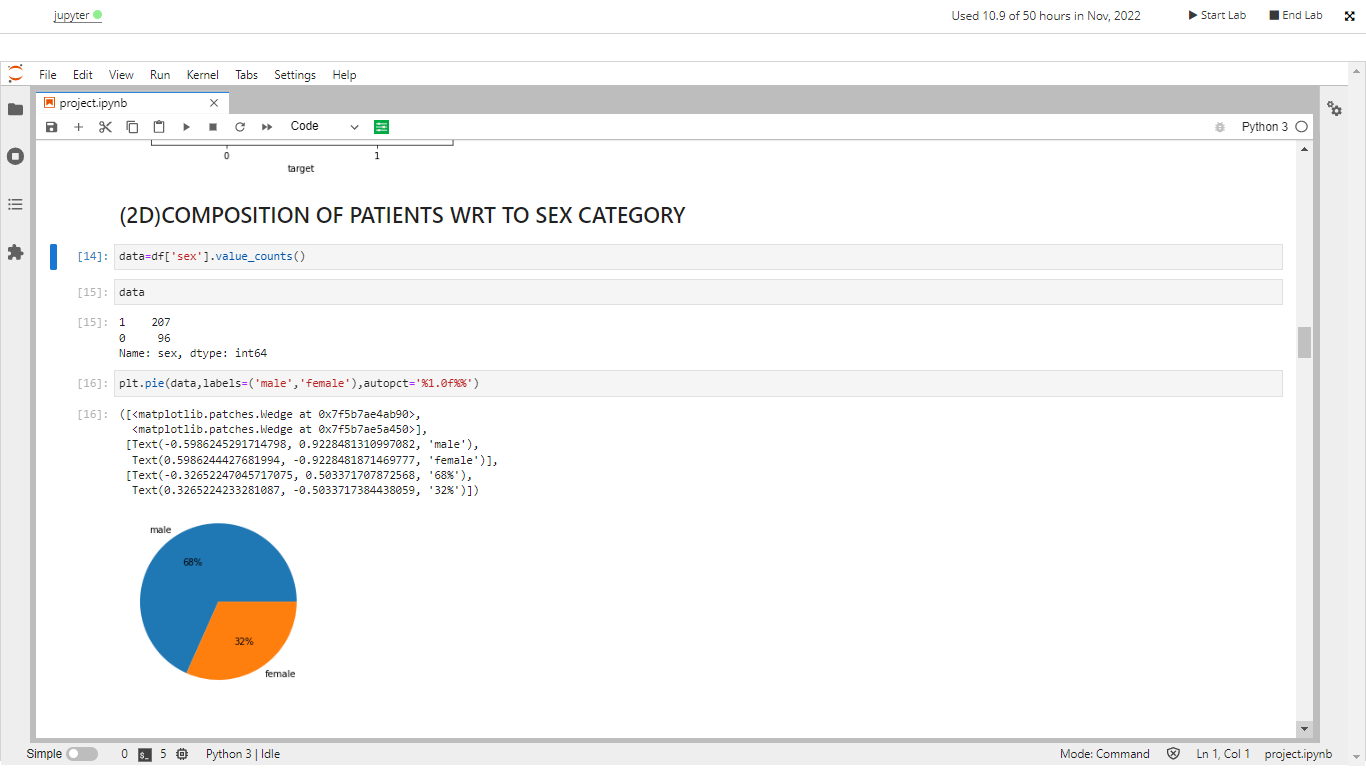
**B.** **IDENTIFICATION OF CATEGORICAL AND NUMERICAL COLUMNS AND COUNT PLOTS**

** **

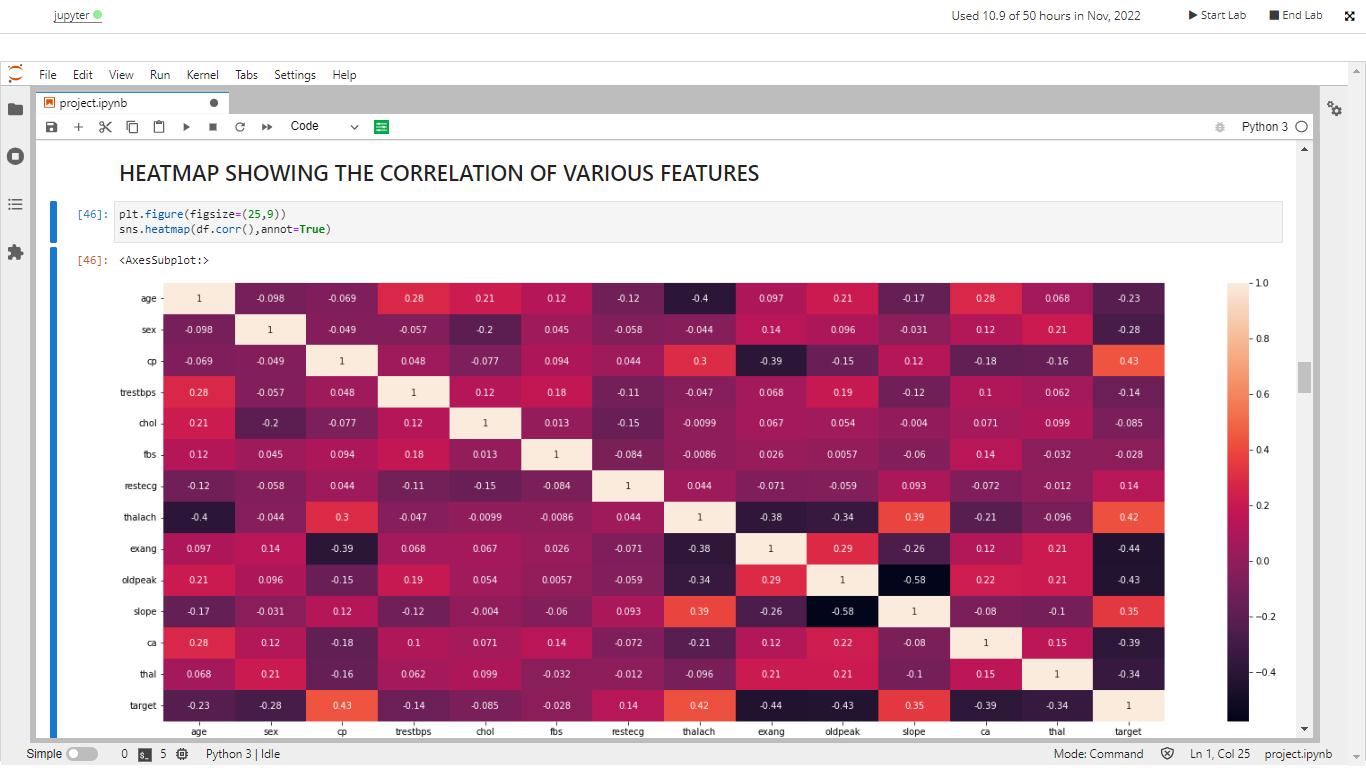
**C.STUDY OF CVD ACROSS AGE CATEGORY**

****

**D.** **COMPOSITION OF PATIENTS WRT TO SEX CATEGORY**

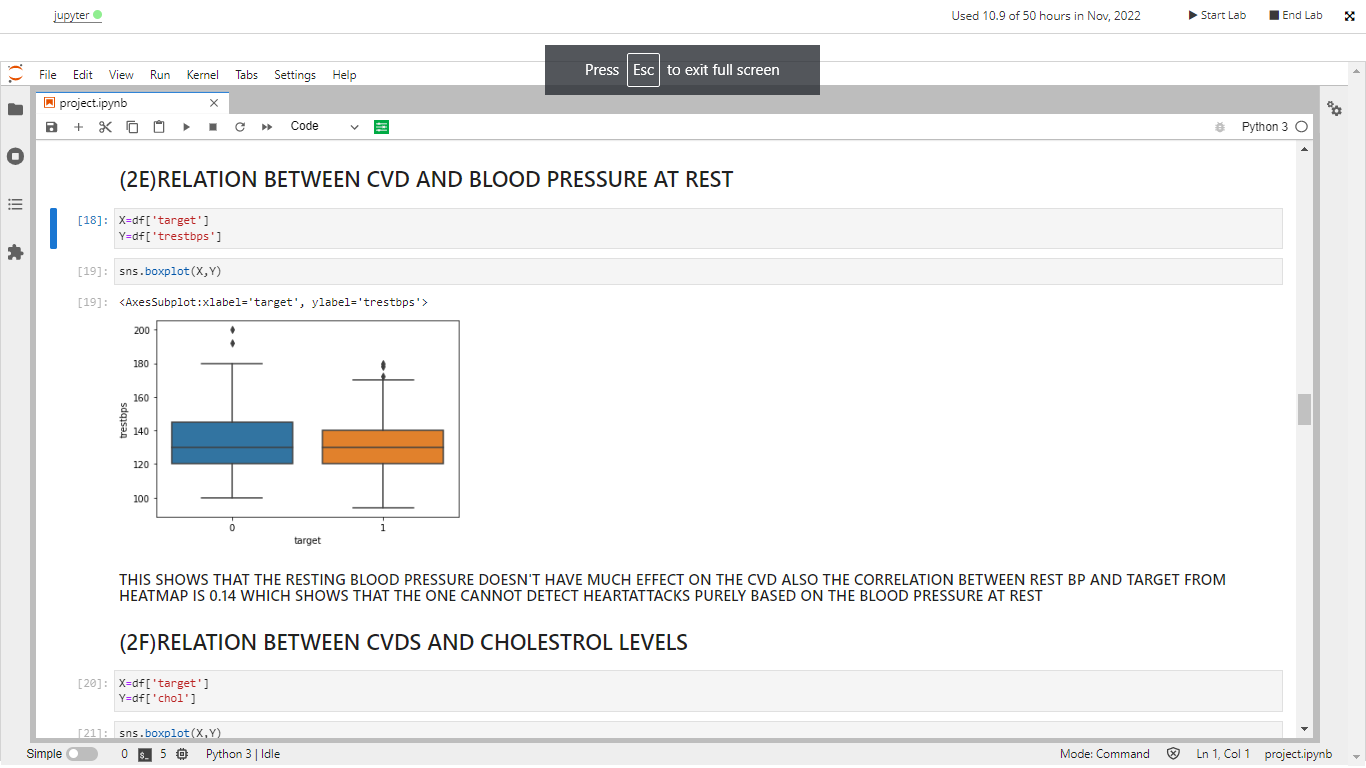
****

**HEATMAP SHOWING CORRELATION BETWEEN ALL FEATURES**

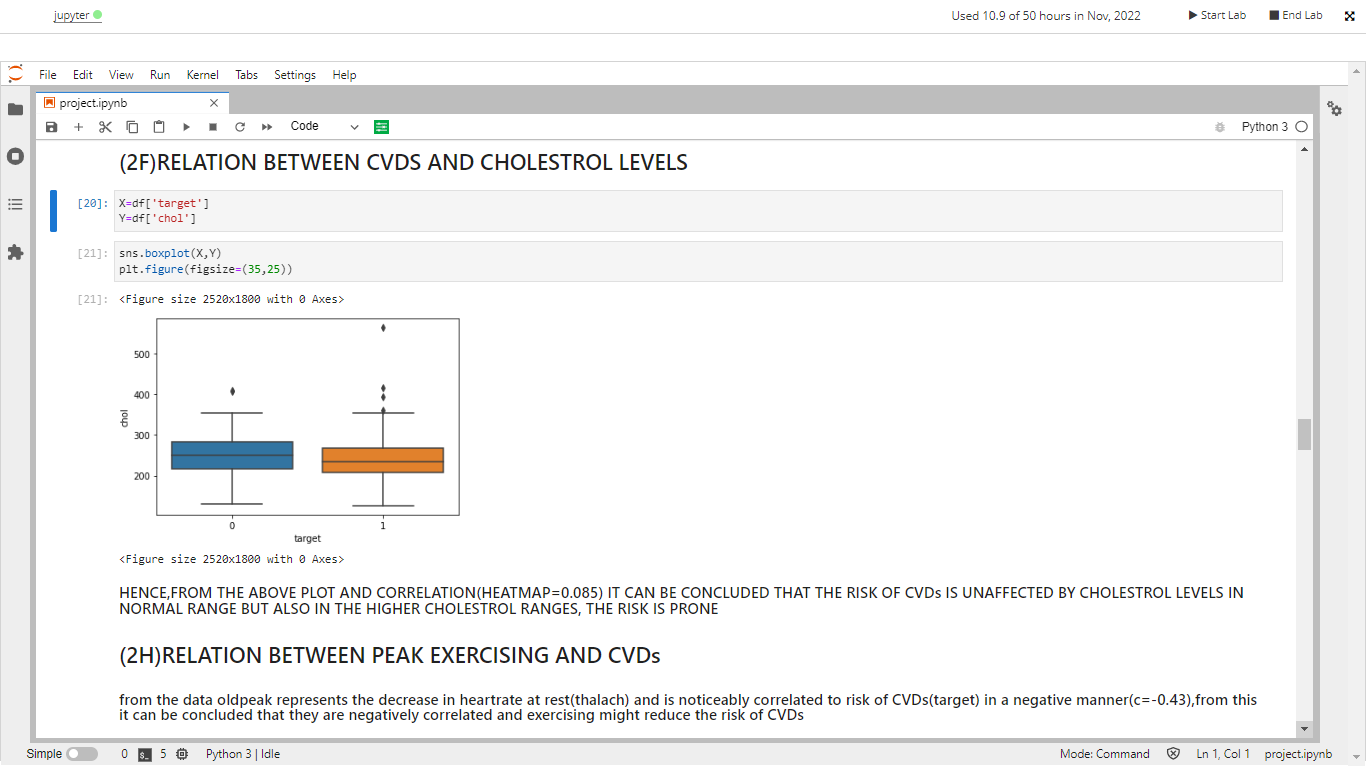
****

The above heat map helps us understand the correlation and helps us draw the conclusions for the upcoming questions easily.

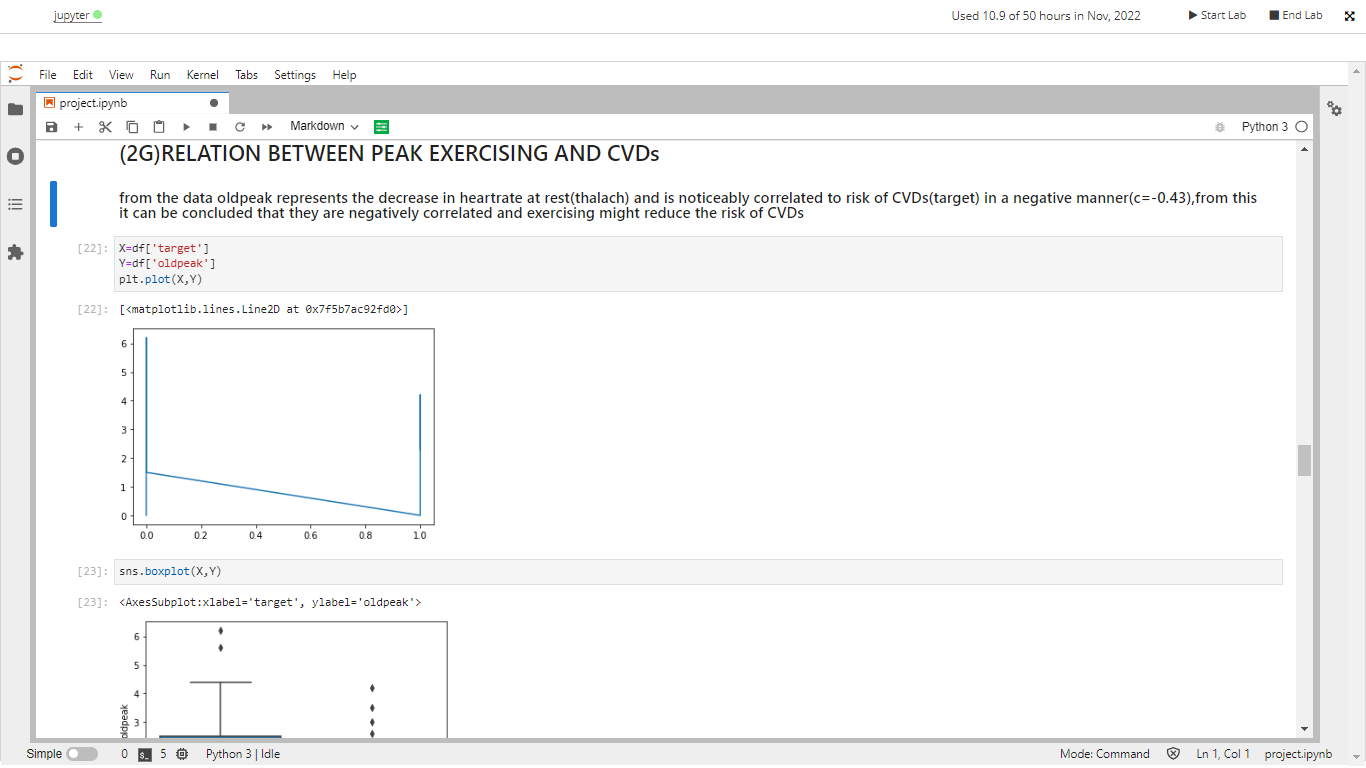
**E.** **RELATION BETWEEN CVD AND BLOOD PRESSURE AT REST**

****

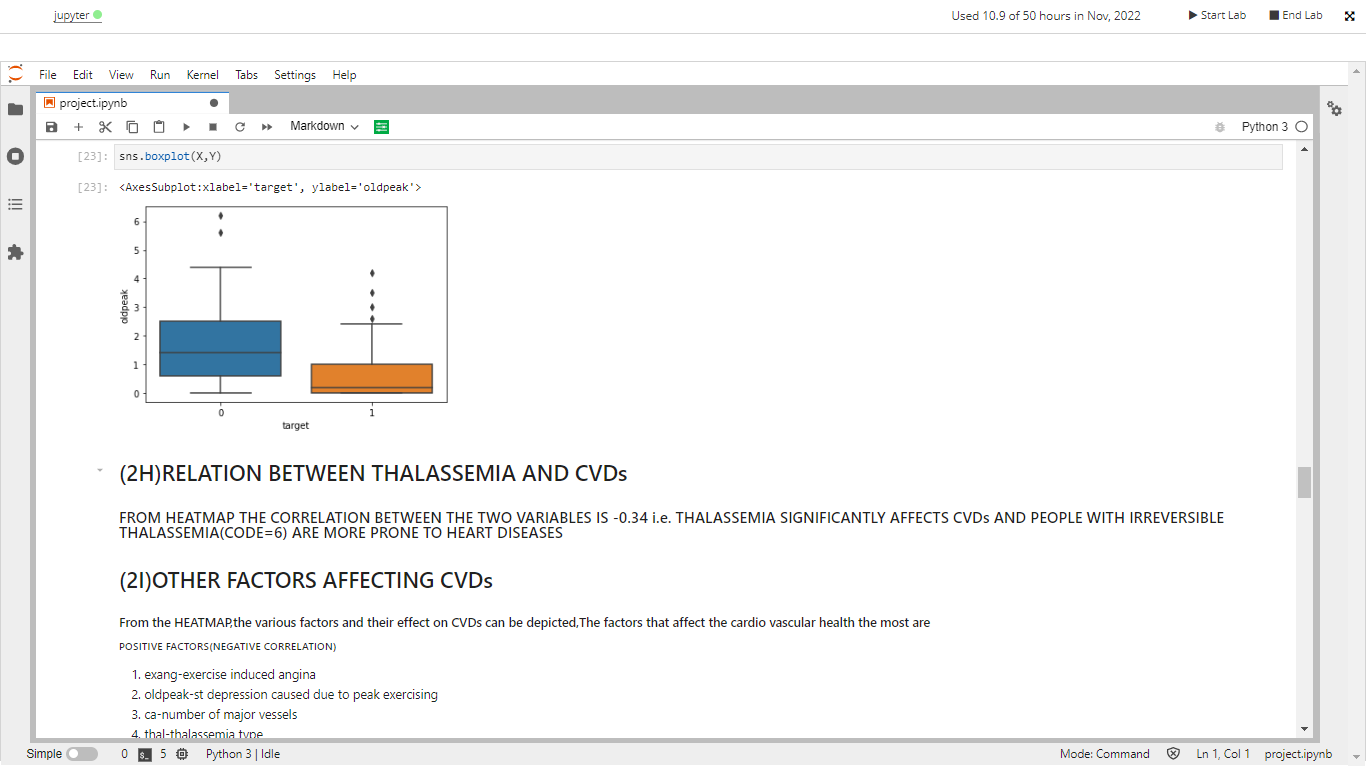
**F.** **RELATION BETWEEN CVDS AND CHOLESTROL LEVELS**

****

**G.RELATION BETWEEN PEAK EXERCISING AND CVDs**

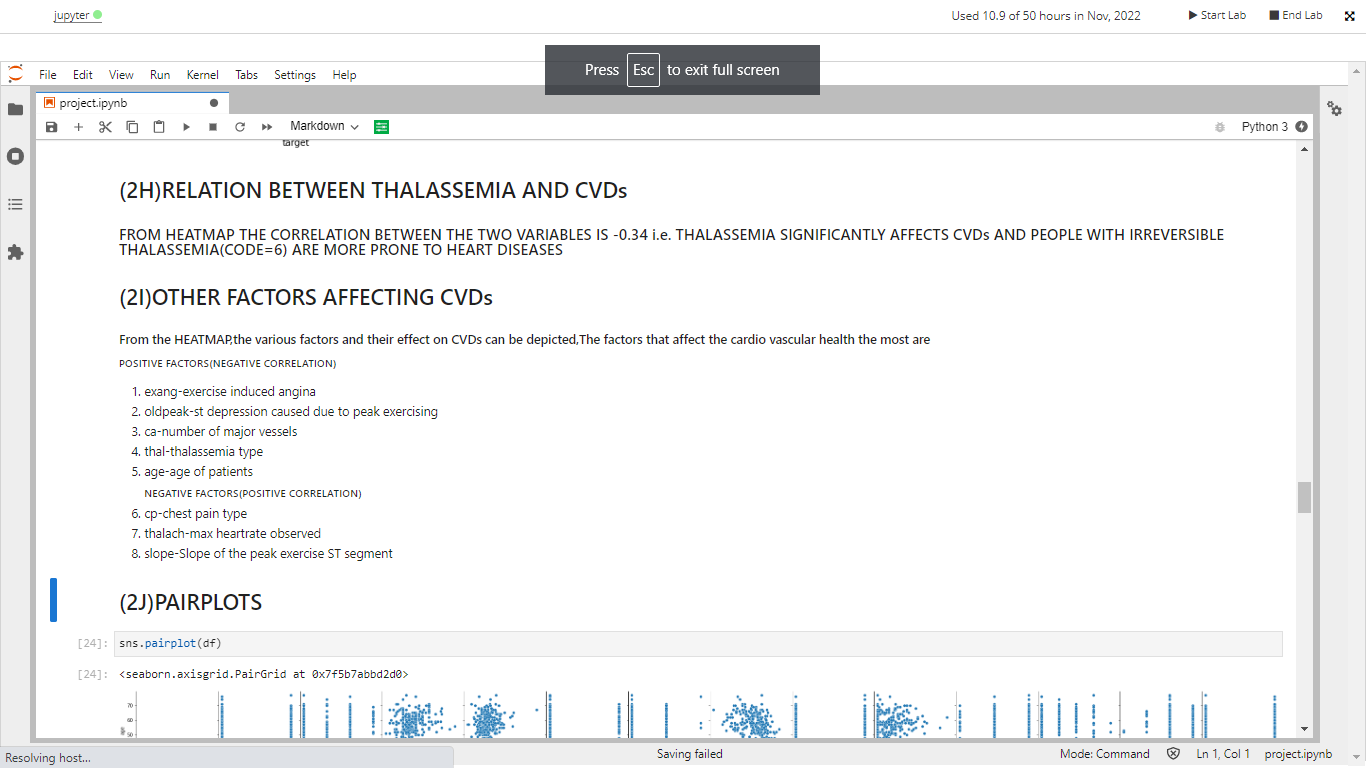
****

**H.RELATION BETWEEN THALASSEMIA AND CVDs**

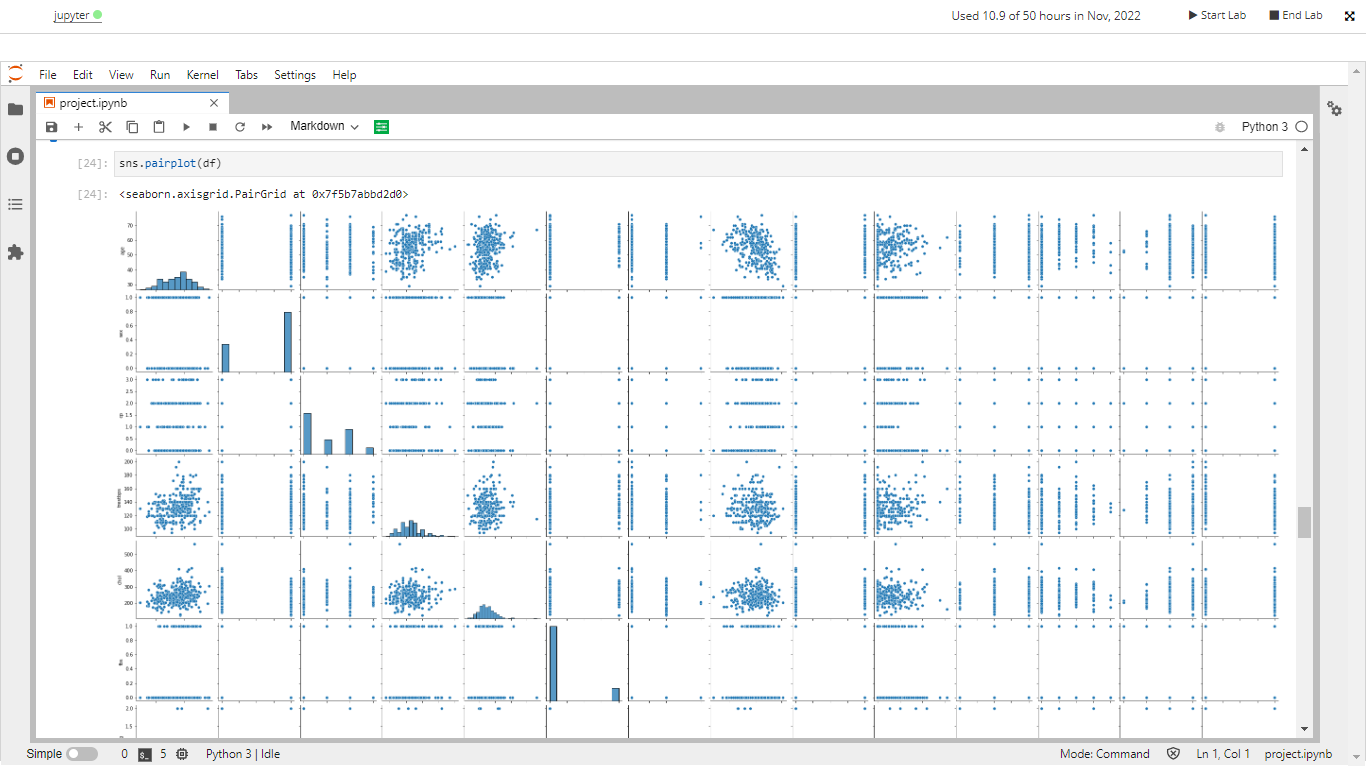
****

**I.OTHER FACTORS AFFECTING CVDs**

These inferences have been made based on Heat map

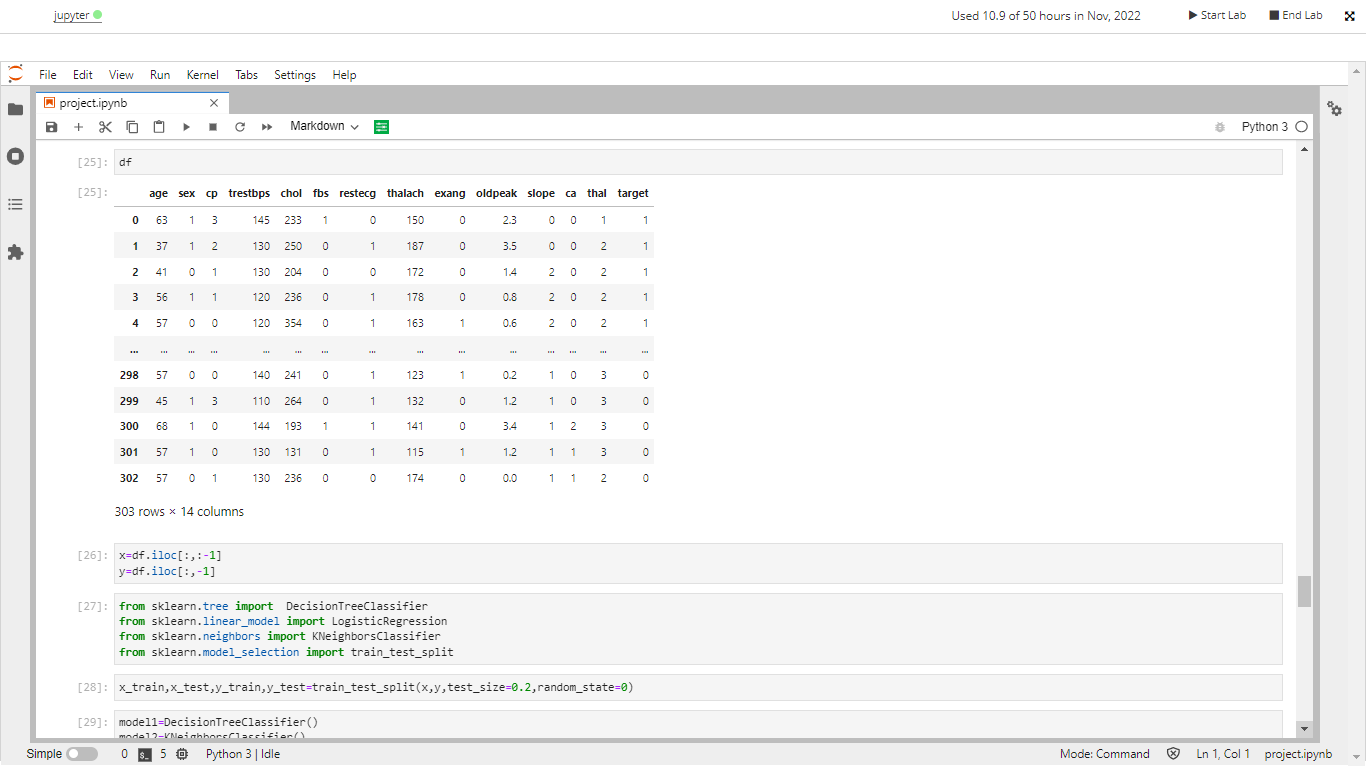


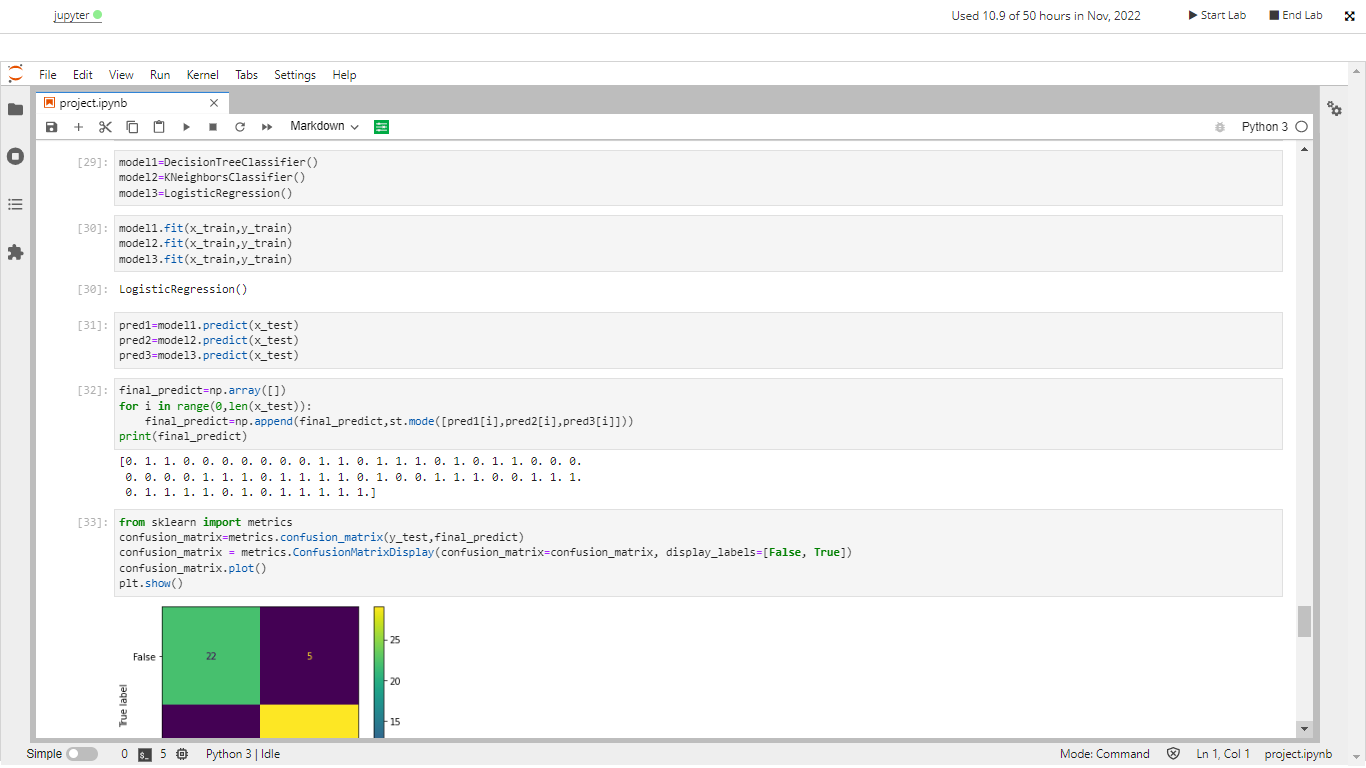
**J.PAIRPLOTS OF ALL FEATURES**

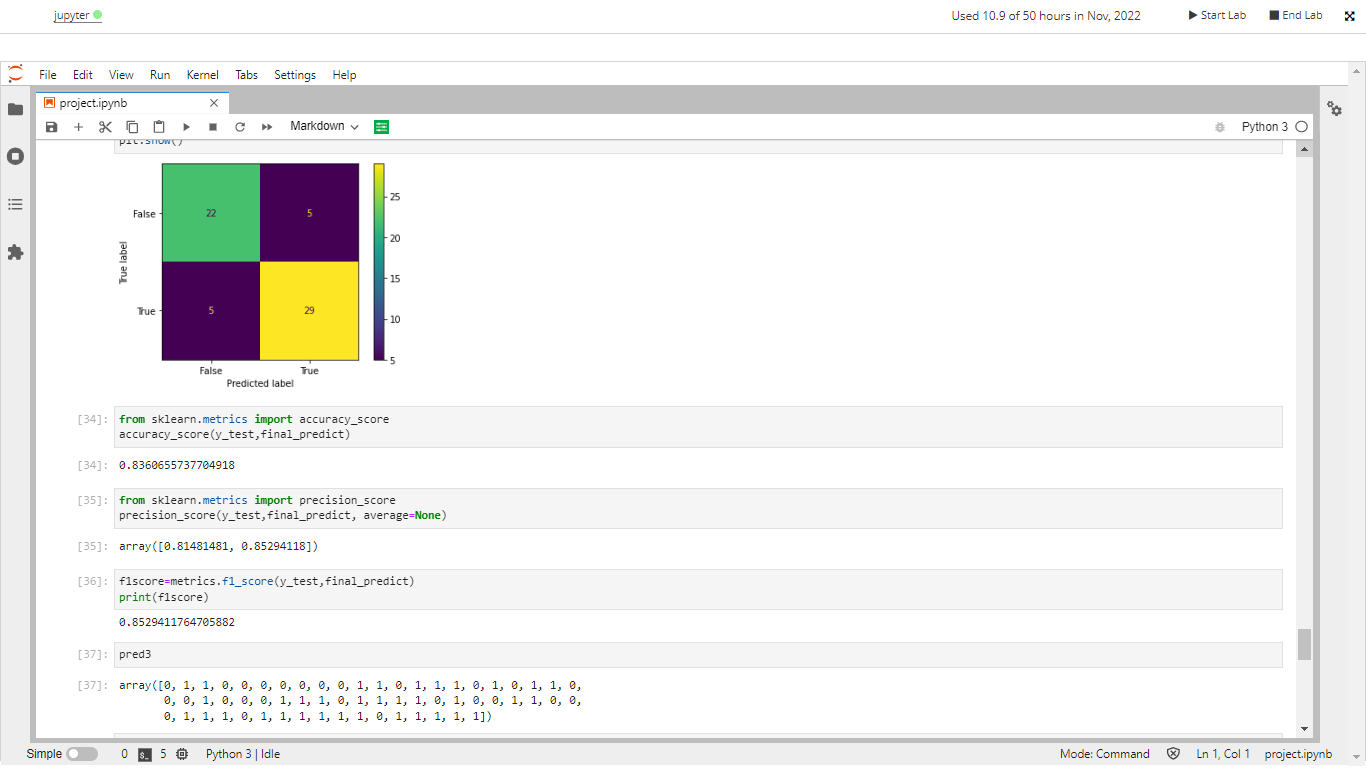
****

****

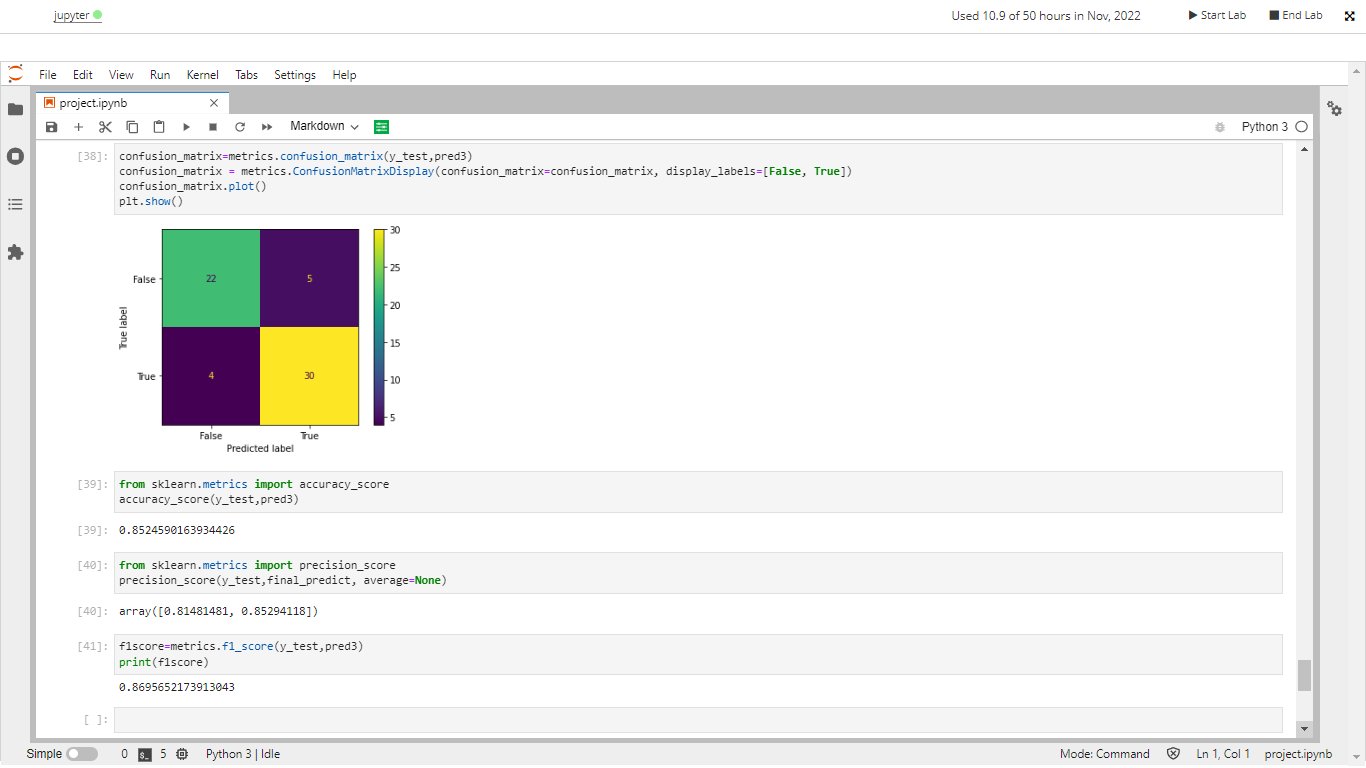
**MODEL**

****

****

****

**RANDOM FOREST MODEL STATS**

****

**LOGISTIC REGRESSION MODEL STATS**