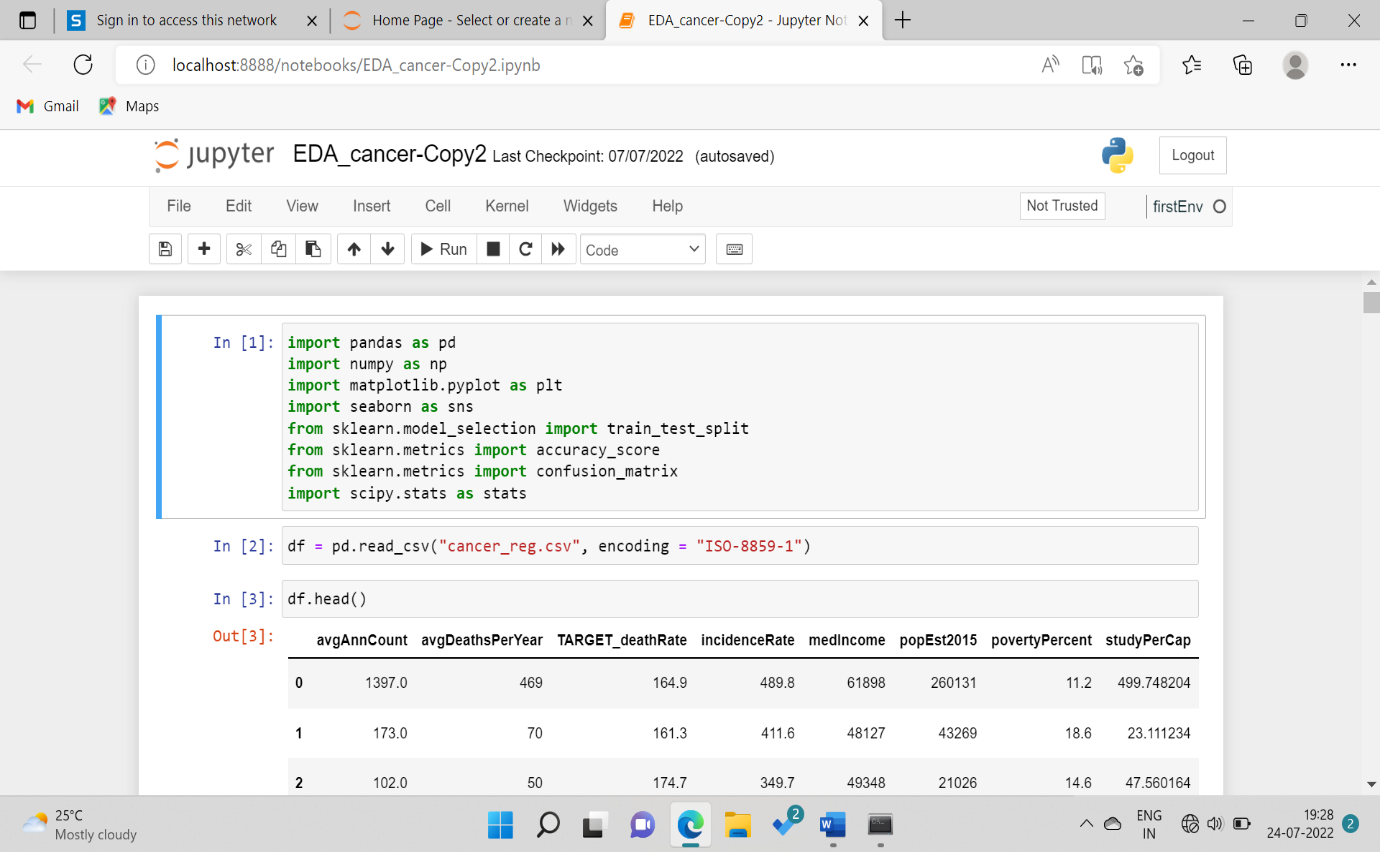
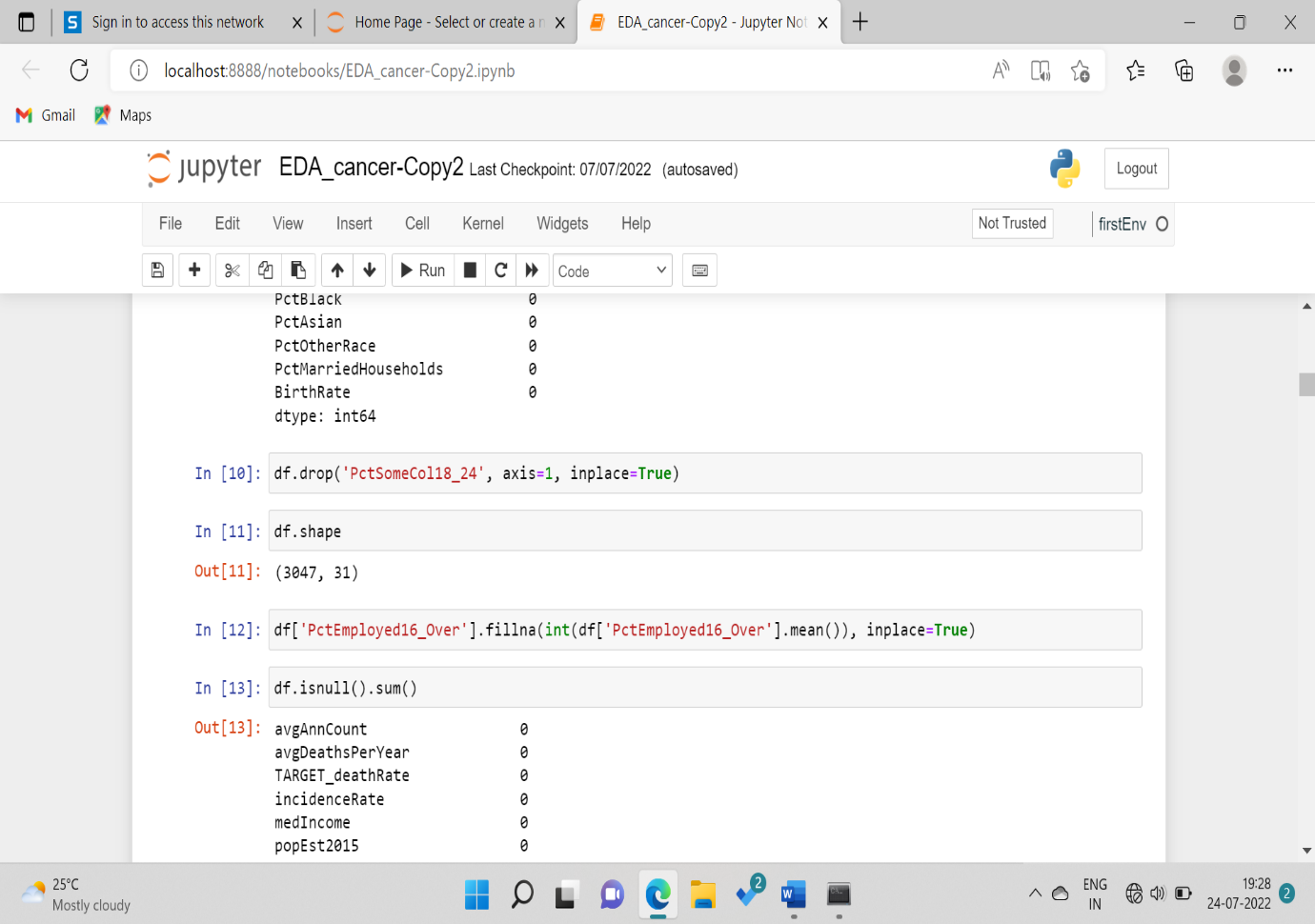
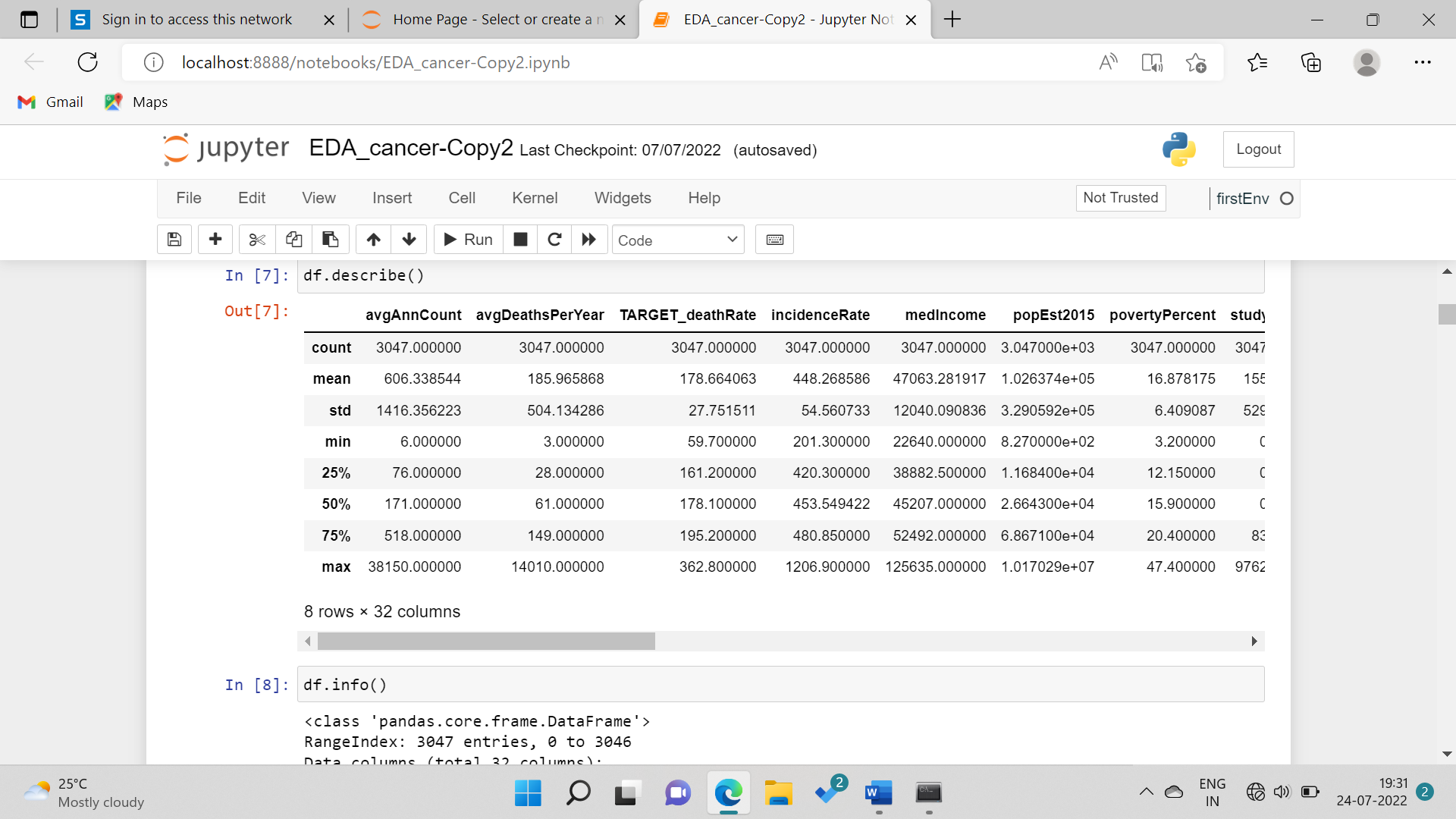
**Documentation Flow**

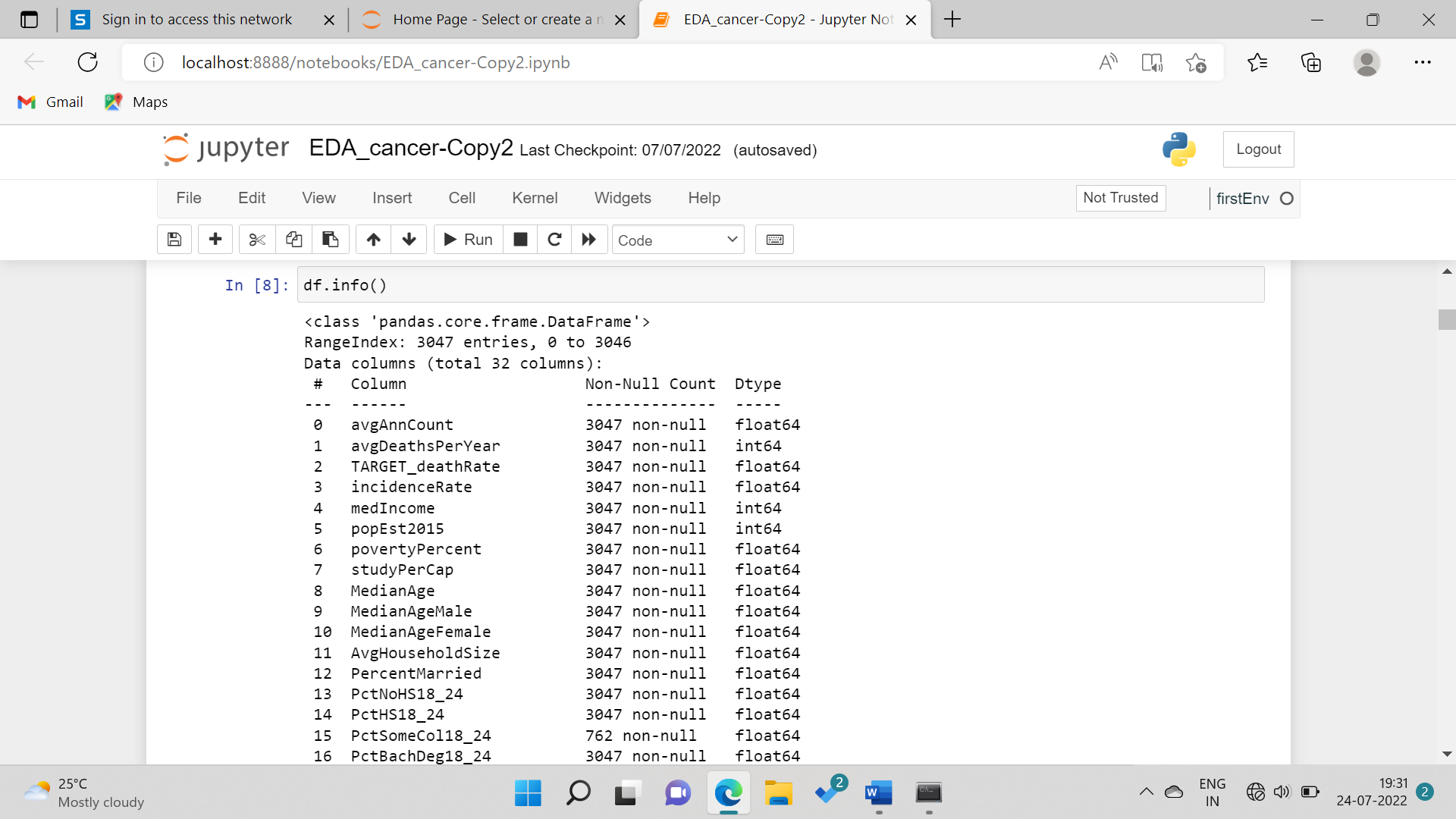
**Project - 3 {Regression Modelling Exercise}**

1. **Problem Statement->** We are provided with real time regression dataset and we have to perform EDA on that to find out and compare various machine learning algorithms and find which has the better rms value and accuracy and perform graphical visualization for better understanding of the data.
2. **EDA->** After the data analysis part i.e. importing the dataset and removing null values. We used three machine learning algorithms and did comparative study namely- Logistic regression, Random Forest and XGBoost. We found out the accuracy, root mean squared error and mean squared error for each of them.

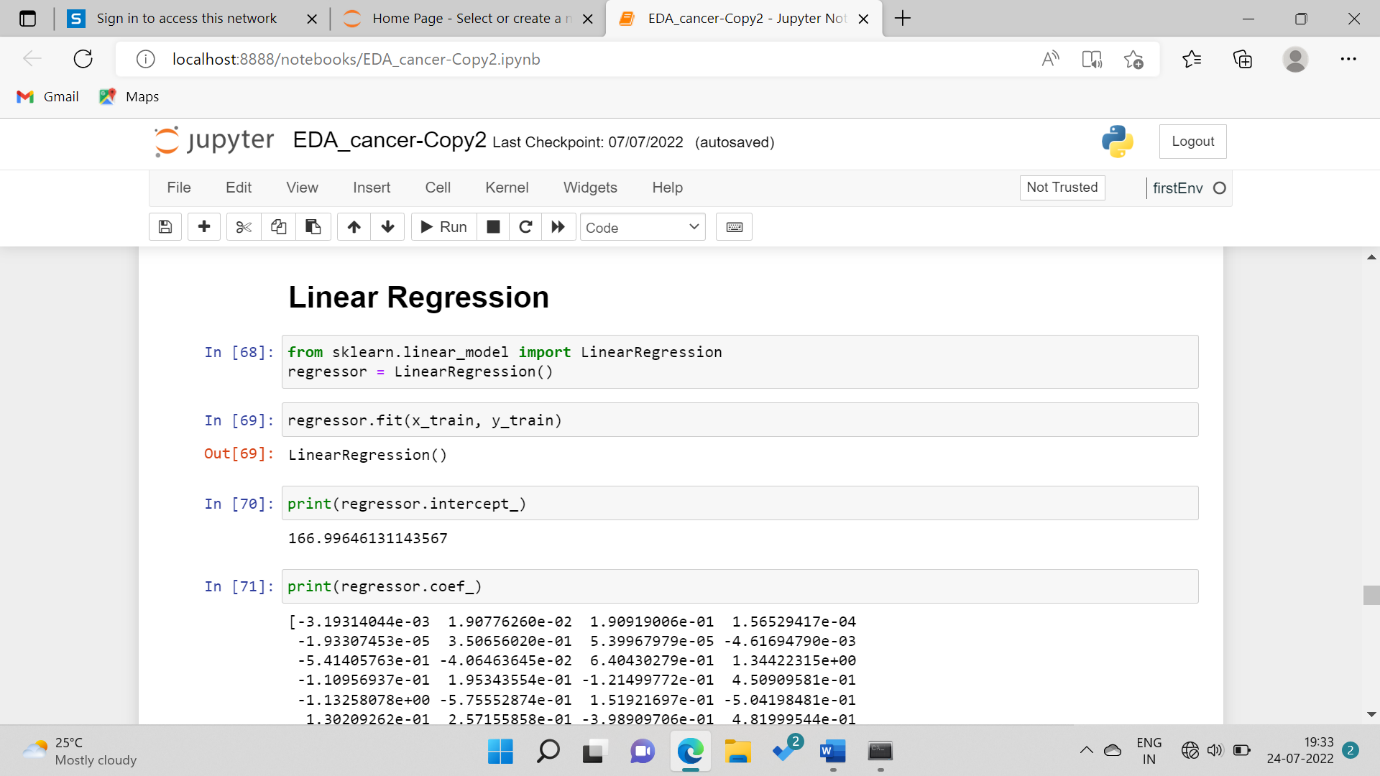


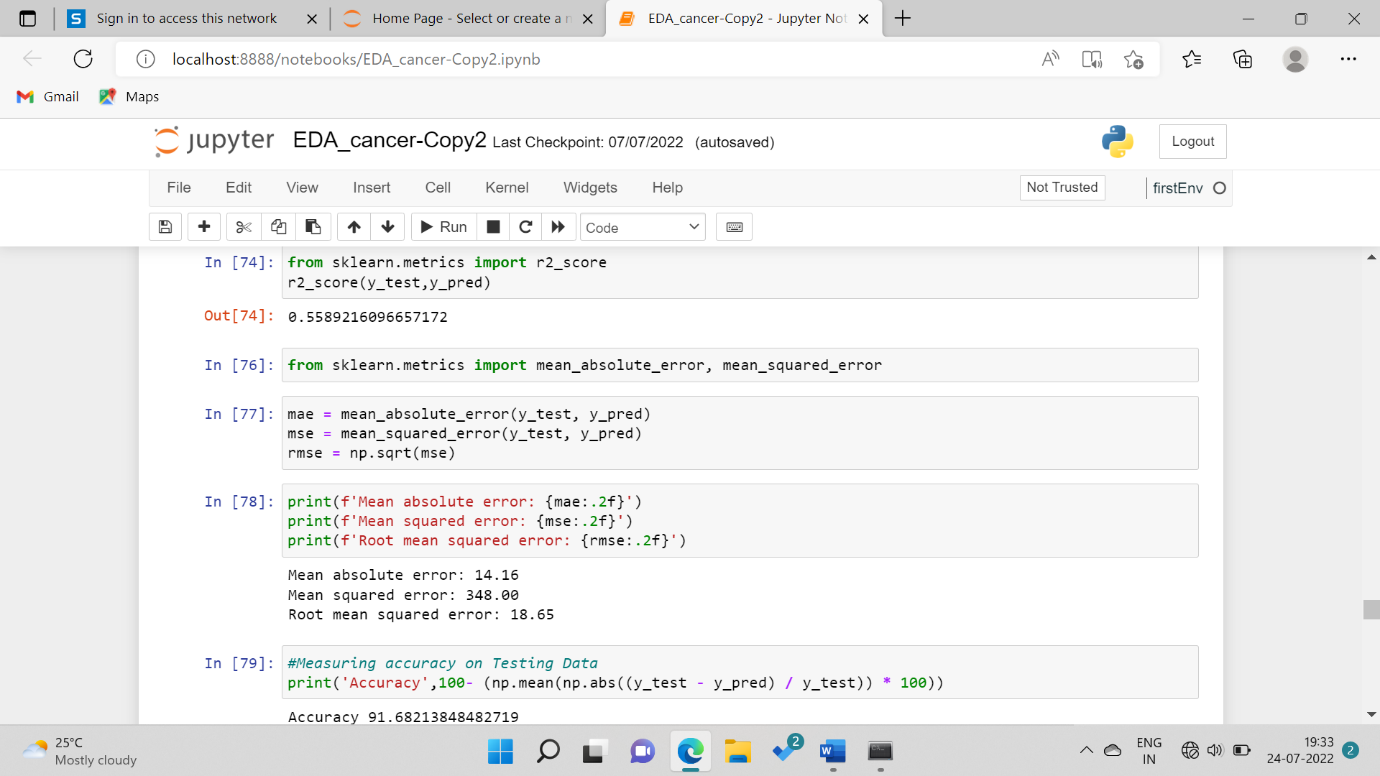


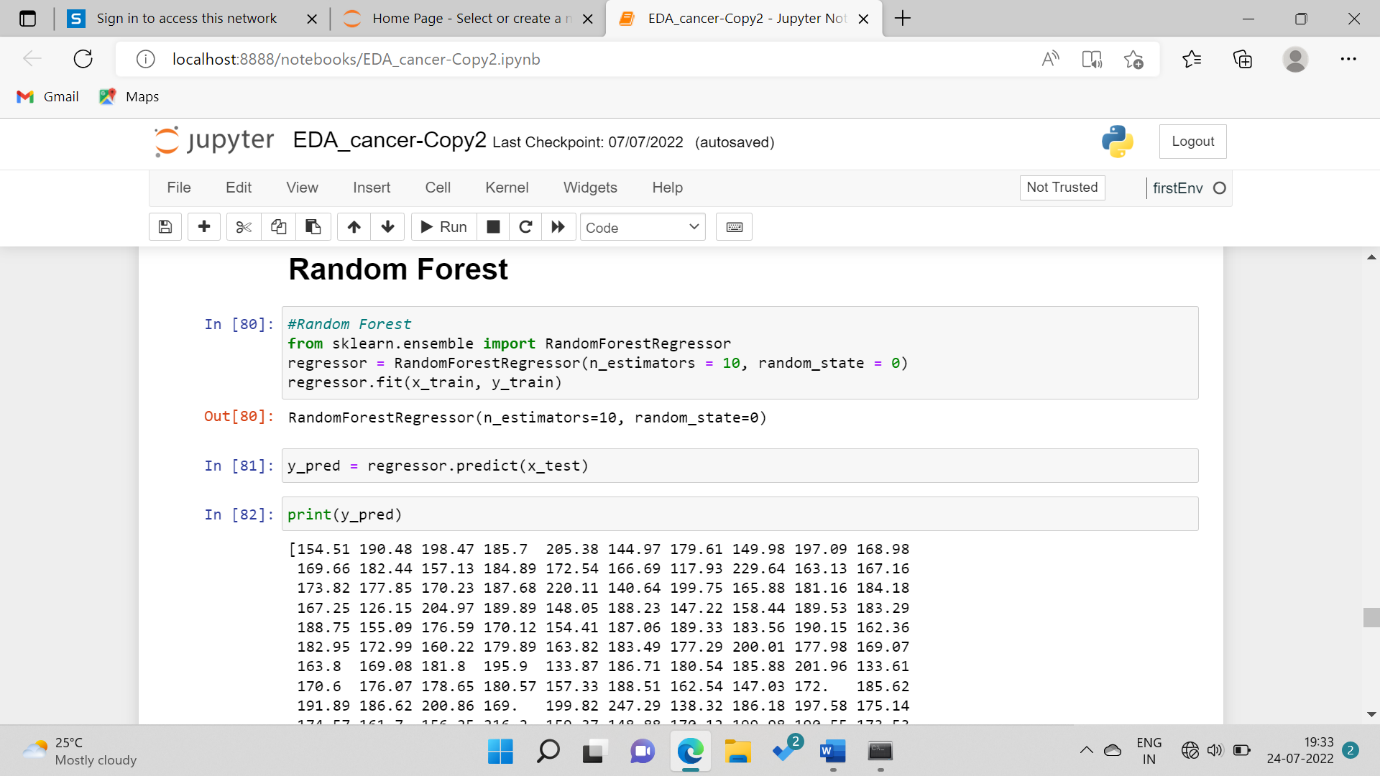


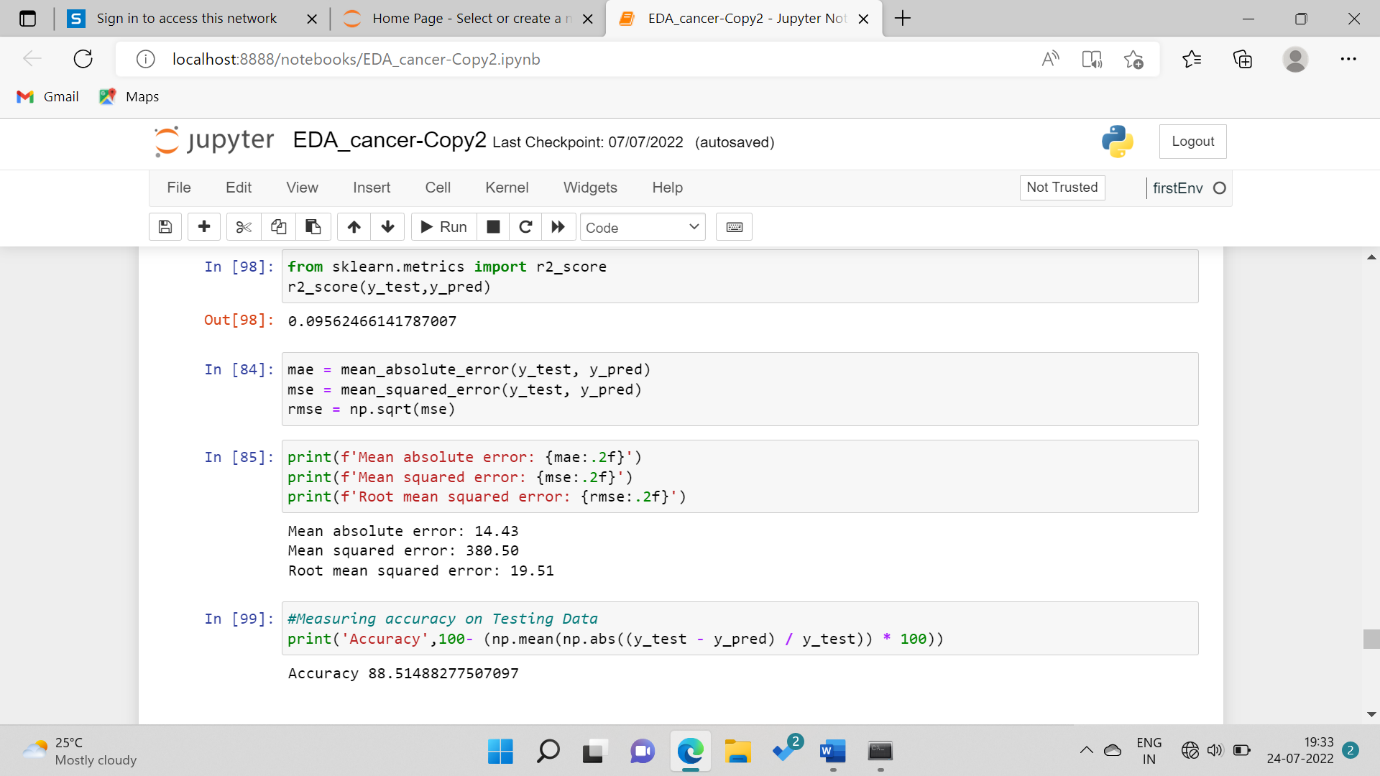


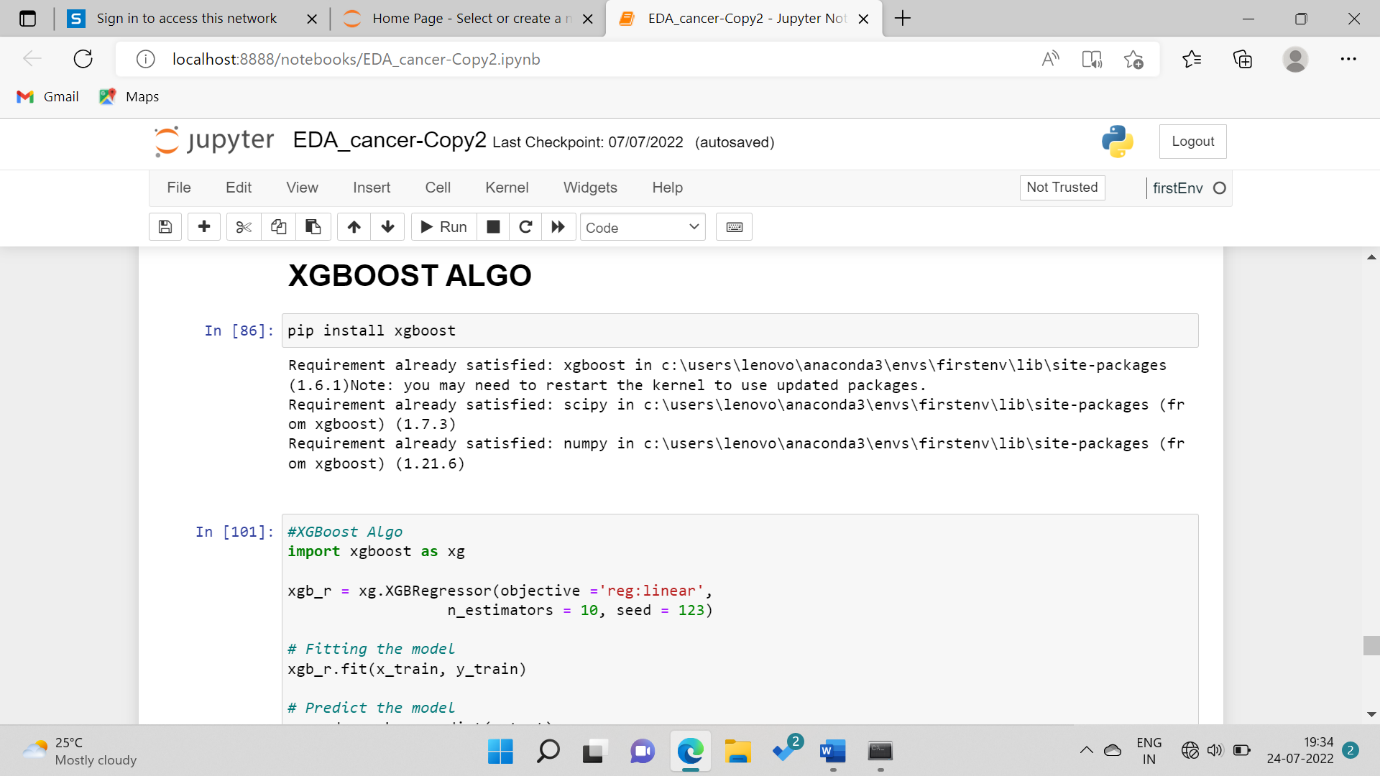
1. **REGRESSION MODELING->** A regression model provides a function that describes the relationship between one or more independent variables and a response, dependent, or target variable.
2. **COMPARATIVE STUDY->** We have used four machine learning algorithms for comparative study – Random Forest, XgBoost, Linear regression where we have observed that the root mean squared error value for Linear regression is the least and highest for XgBoost and on calculating the accuracy for each of them, we found that the linear regression was 91.68% accurate, whereas random forest was 88.51% accurate and XgBoost was 91.84% accurate which shows that XgBoost is the most accurate model.

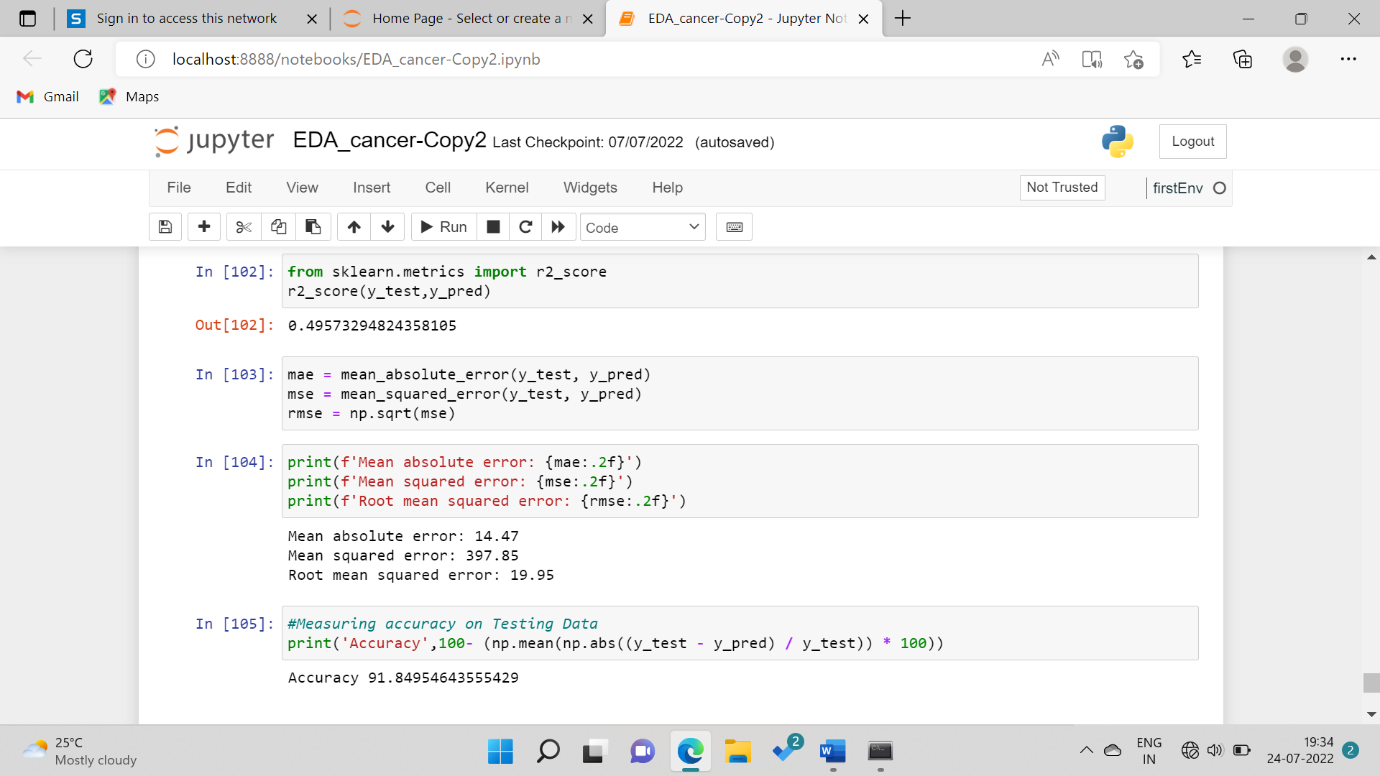




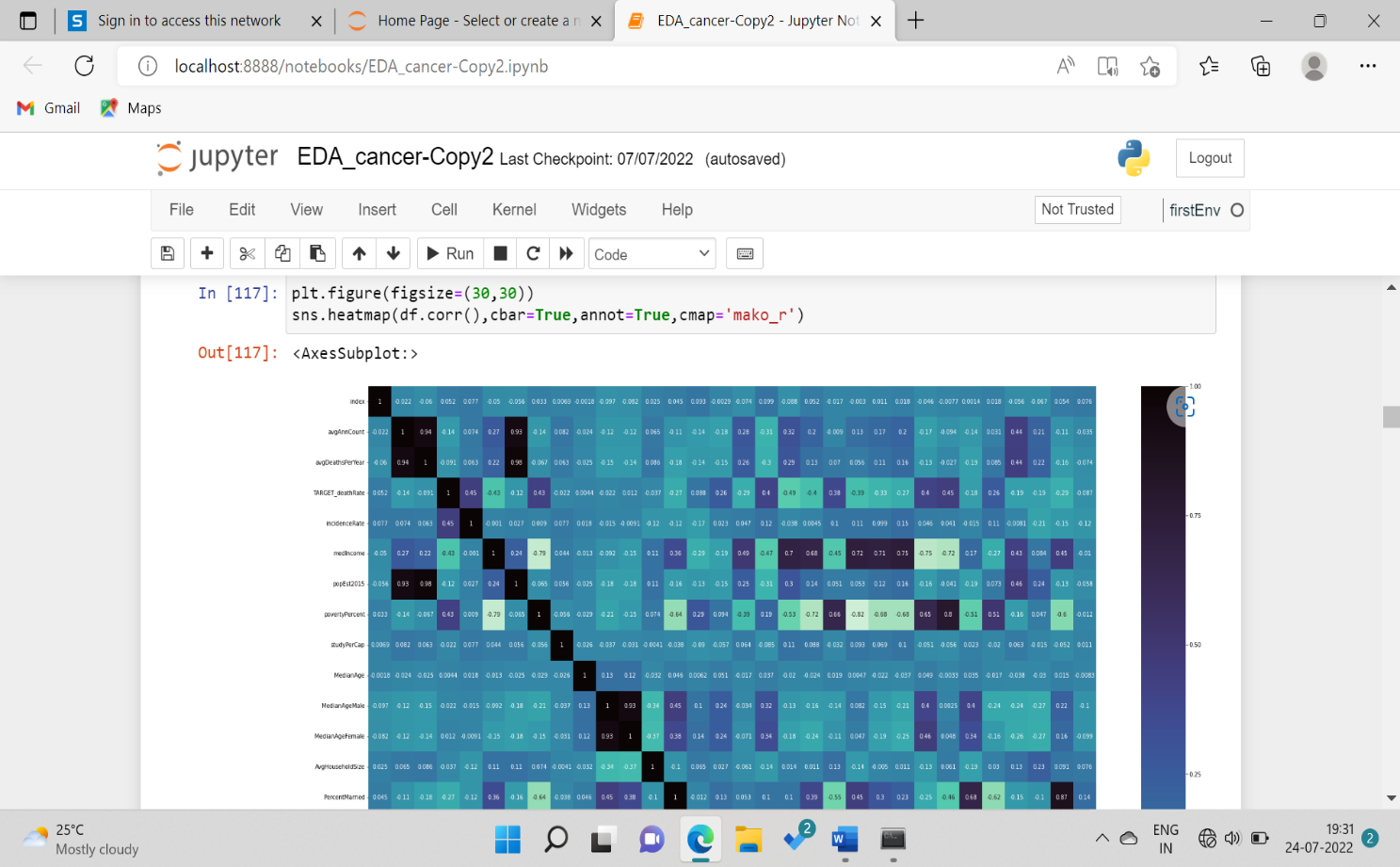




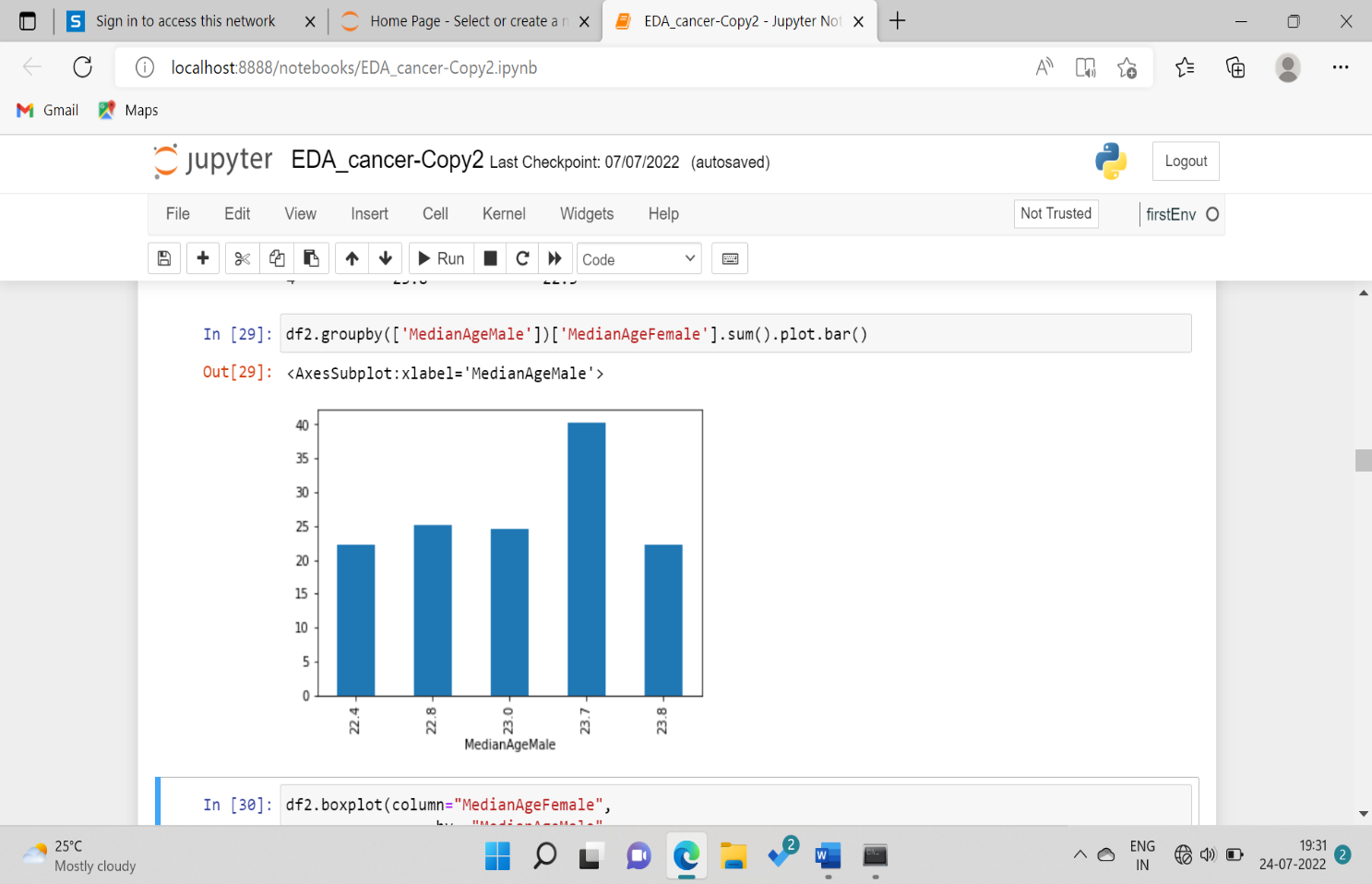




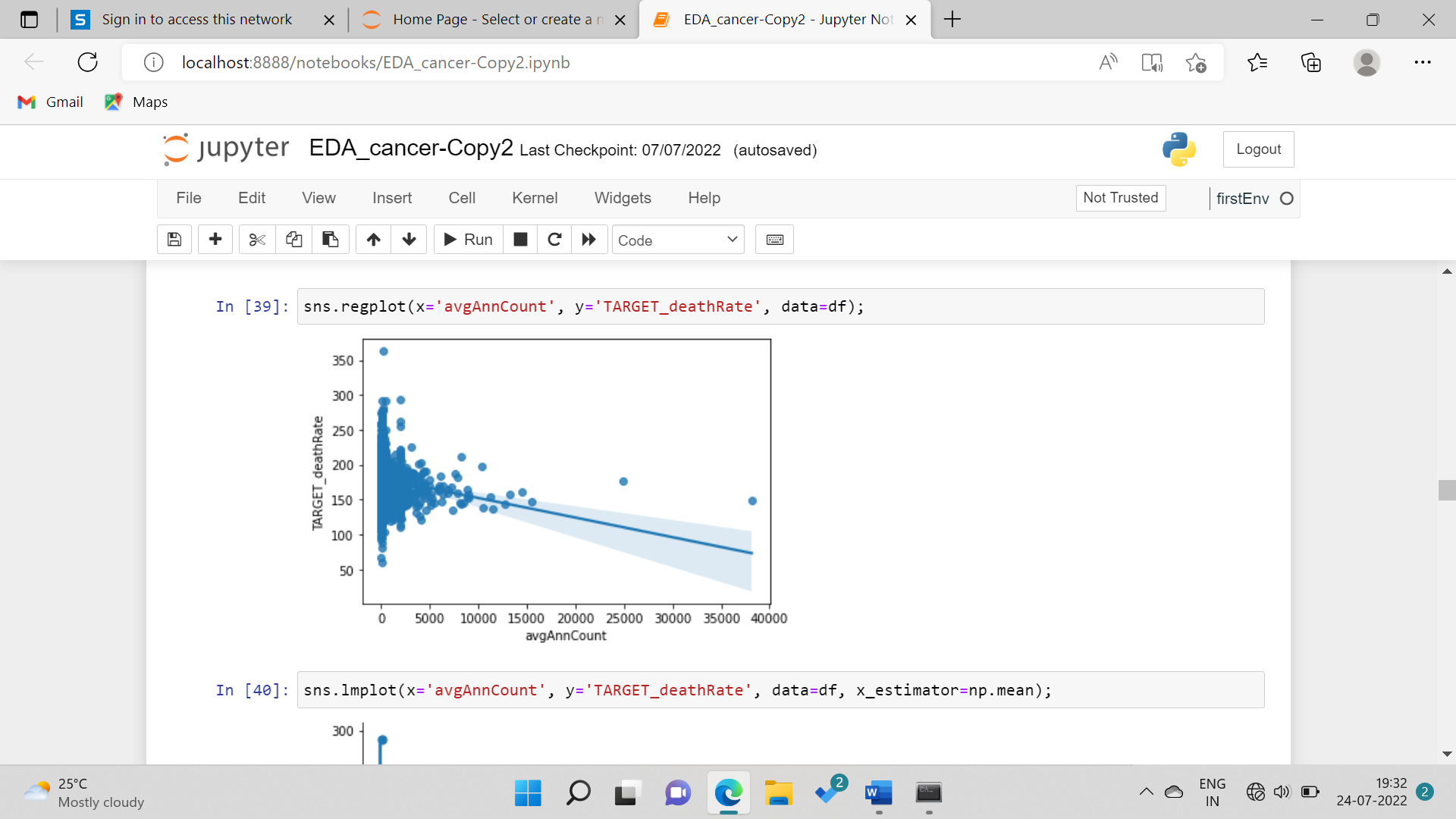
1. **INFERENCE->** The graphical visualization of the dataset and their inference.



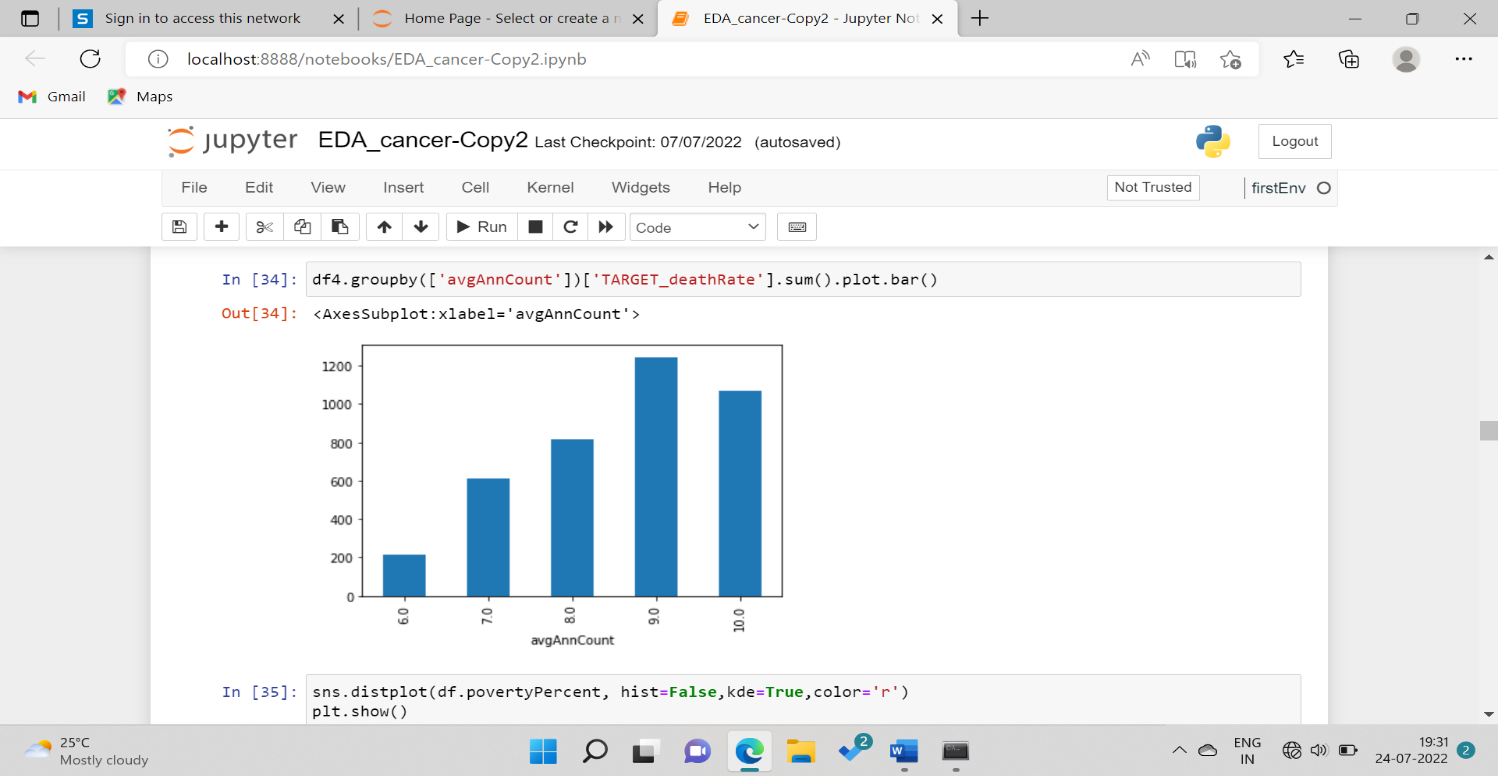
**Inference->** This graph shows the correlation values between the variables**.**



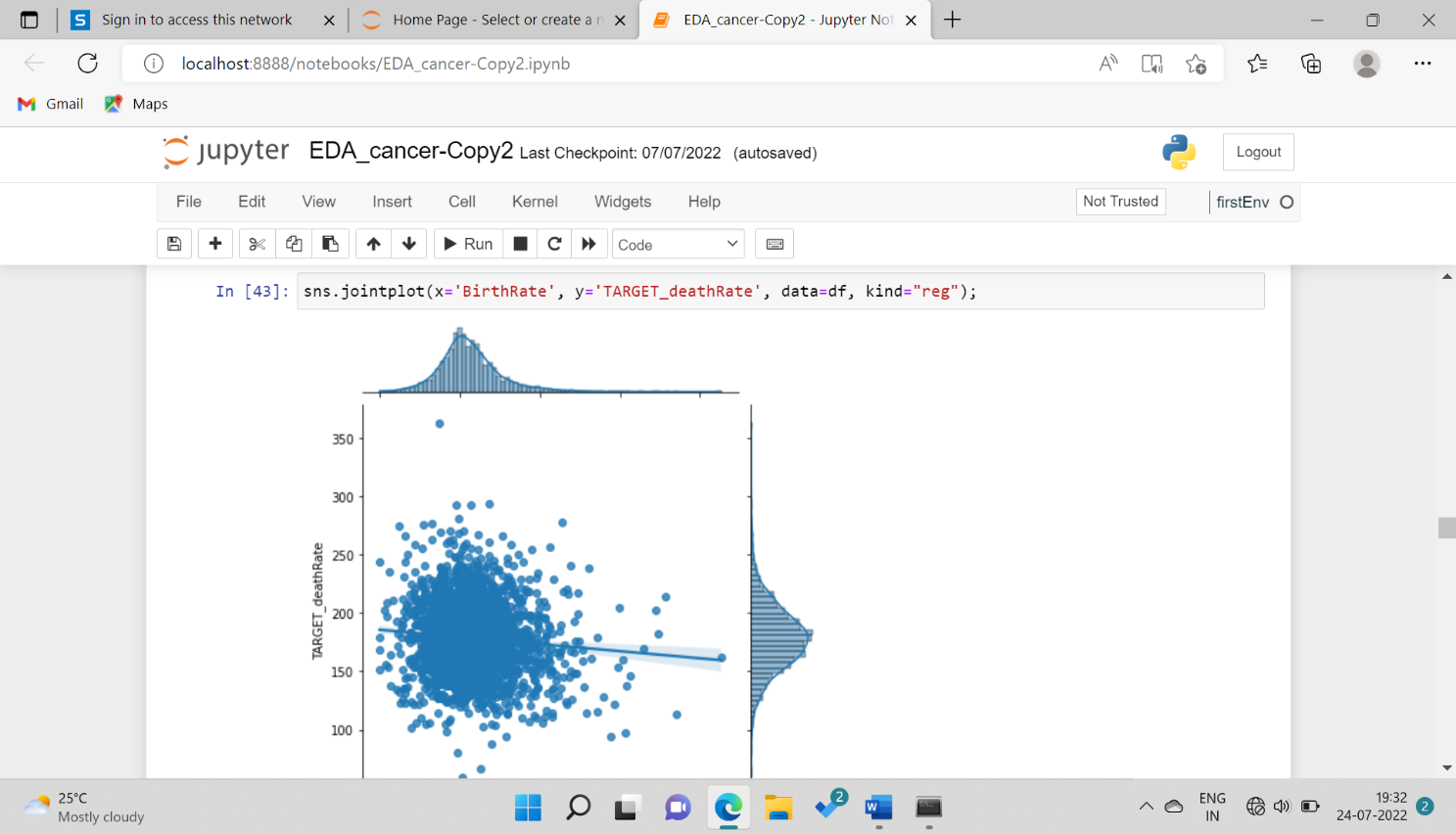
**Inference->** This graph shows the relation between the MedianAgeMale and MedianAgeFemale



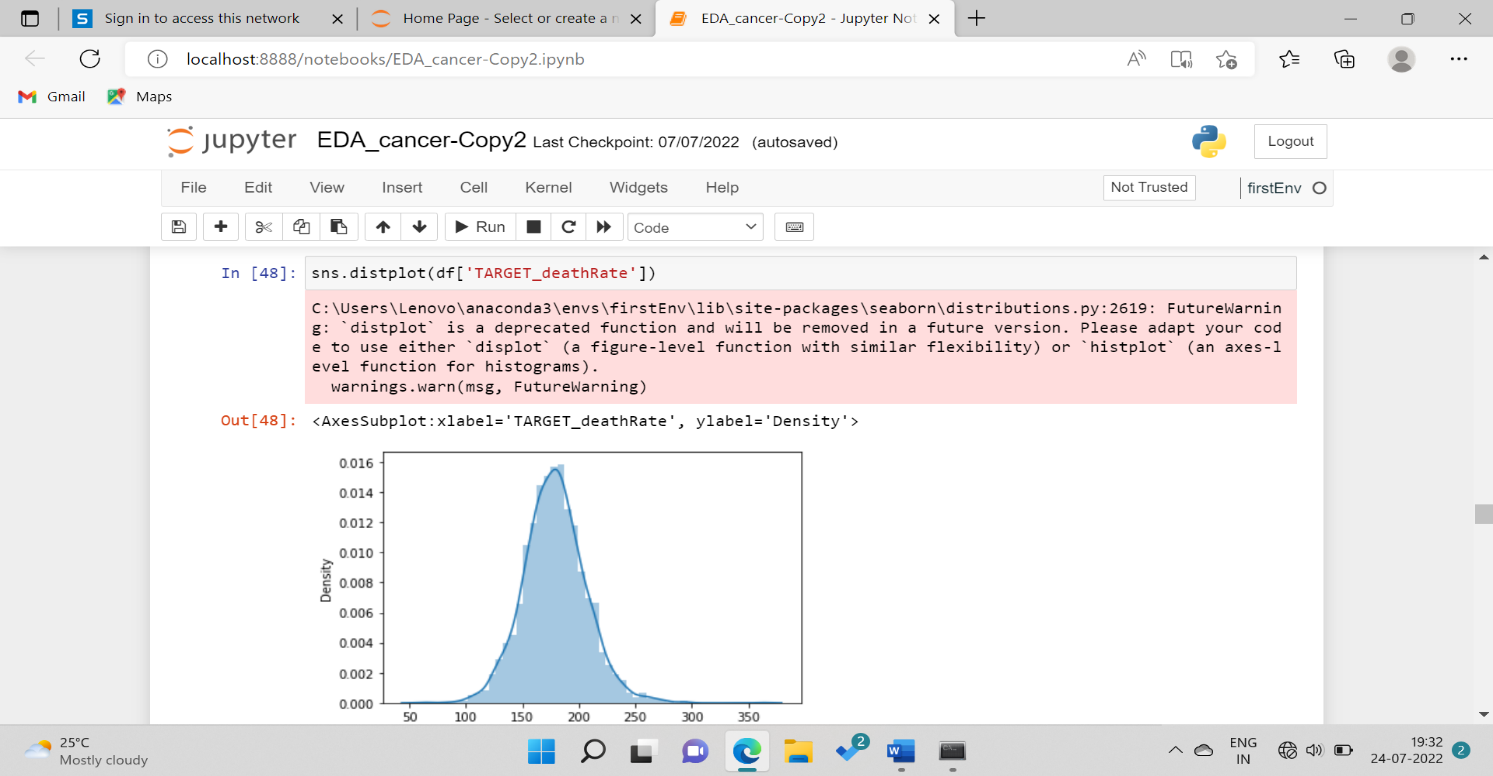
**Inference->** This graph shows the best fit of the data (best fit line).



**Inference->** This graph shows the relation between the average count and target death rate.



**Inference->** This graph shows the relation between the x and y i.e., how the dependent variable(y) varies with the independent variable(x).



**Inference->** This graph shows the distribution plot of target death rate and tells us where the target values fall in a distribution.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_