**Name of the project:-**

**Sales Channel Prediction Case Study**

Submitted by:

Sujay Anil Bharmal.

**ACKNOWLEDGMENT**

This includes mentioning all the references, research papers, data sources, professionals, and other resources that helped you and guided you in the completion of the project.

Introduction:- Problem Definition

In this Sales Channel Prediction Case Study,  this is the problem had in lots of companies. When a company enters a market, the distribution strategy and channel it uses are keys to its success in the market, as well as market know-how and customer knowledge and understanding. Because an effective distribution strategy under efficient supply-chain management opens doors for attaining competitive advantage and strong brand equity in the market, it is a component of the marketing mix that cannot be ignored. The distribution strategy and the channel design have to be right the first time. The case study of the Sales channels includes a detailed study of TV, radio, and newspaper channels. They predict the total sales generated from all the sales channels.

**Data Analysis:-**

In this project, we have the 3 inputs and 1 output that are available according to my information I see that this case is a supervised learning case.

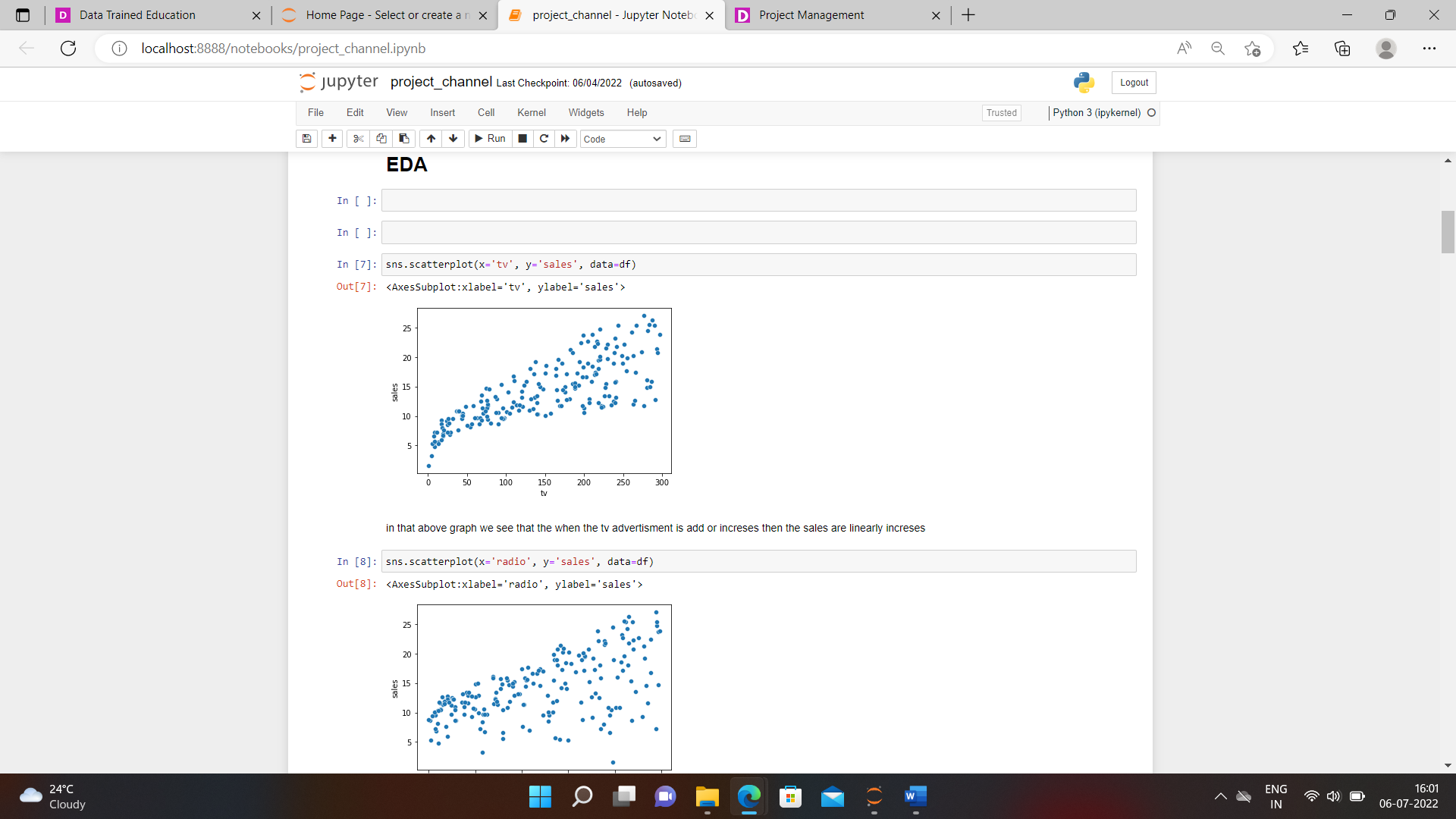
First, we import all the necessary libraries. Like pandas, NumPy seaborn, and matplotlib, and then run that cell. After that, we read the all data and save it into the df folder. There are 200 rows and 4 columns. TV, Radio, and newspapers are the input and sales are the output variable.

First, we see that there are in this dataset’s null values are present or not, we give df. IsNull().sum() code to run and we see that there is no null value present.

Then we check all column datatypes by using the df.info() code.

EDA process:-

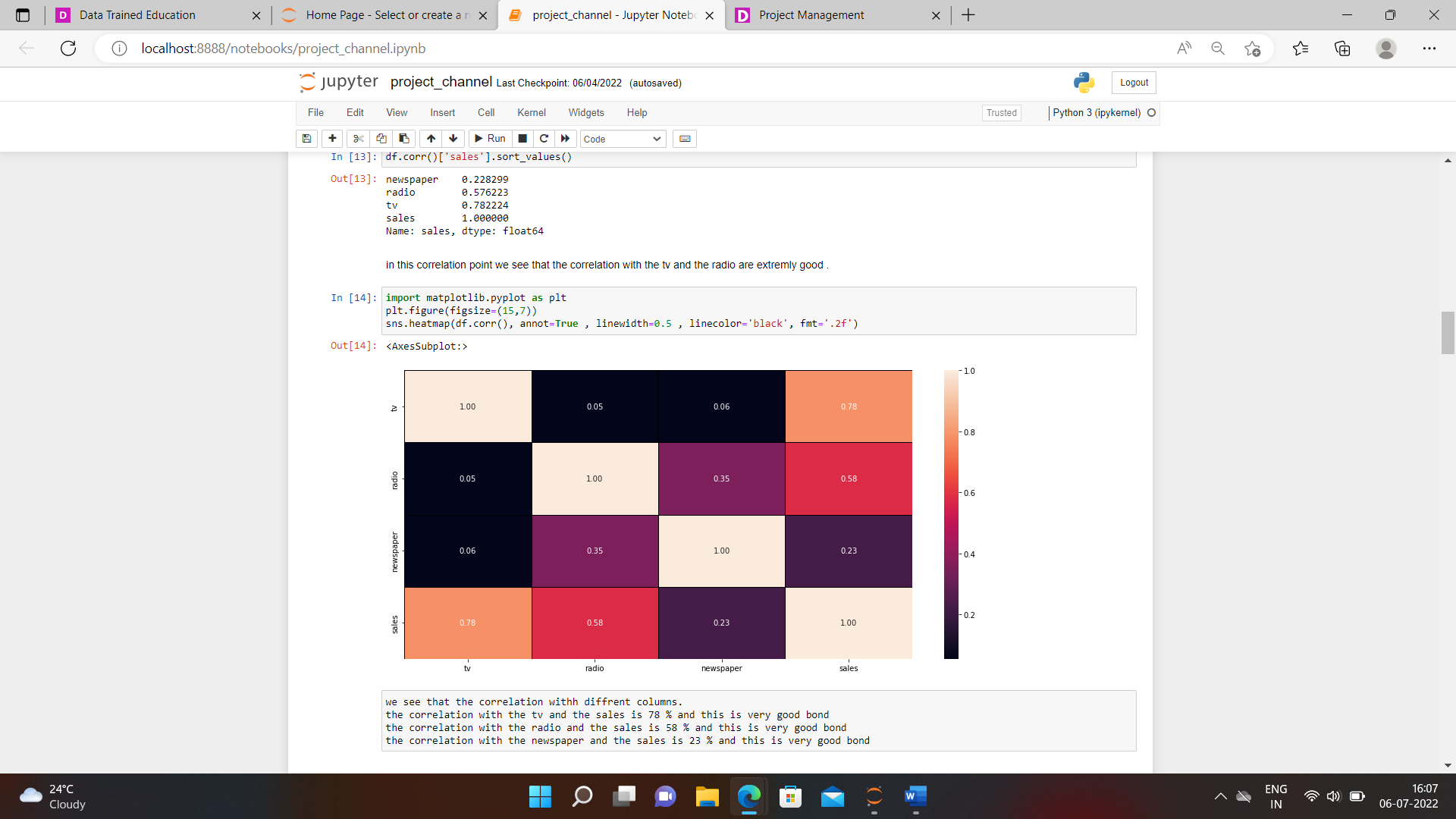
For the EDA process, we use the seaborn and matplotlib libraries,



We see that in that photo we use the scatterplot and relation between the TV and the sales.

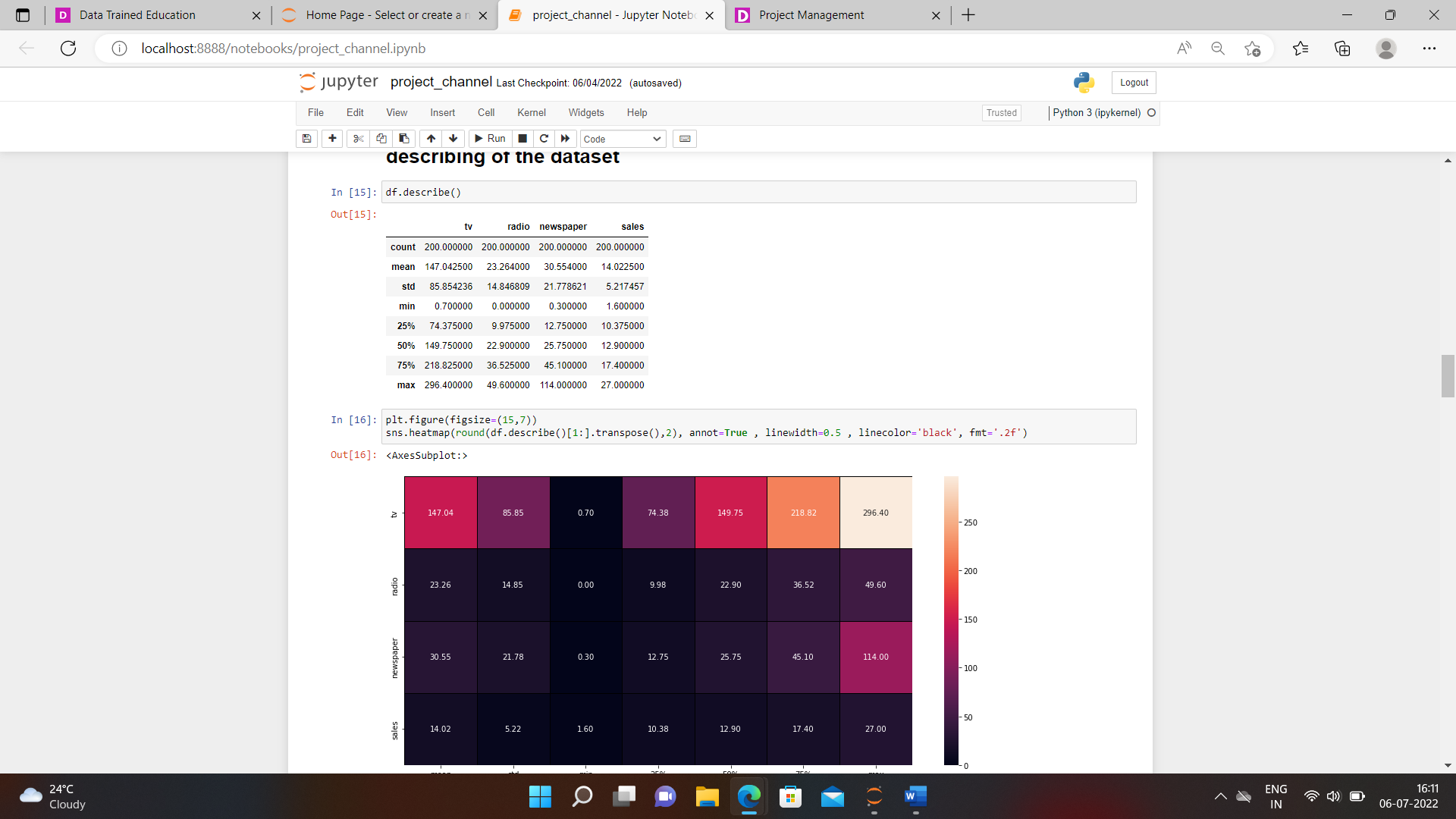
Also, we see the relation between Radio to sales and newspapers to sales.

After that, we use the correlation function and see the co-relation of columns with each other. Use code df.corr(). And plot that correlation using the matplotlib.



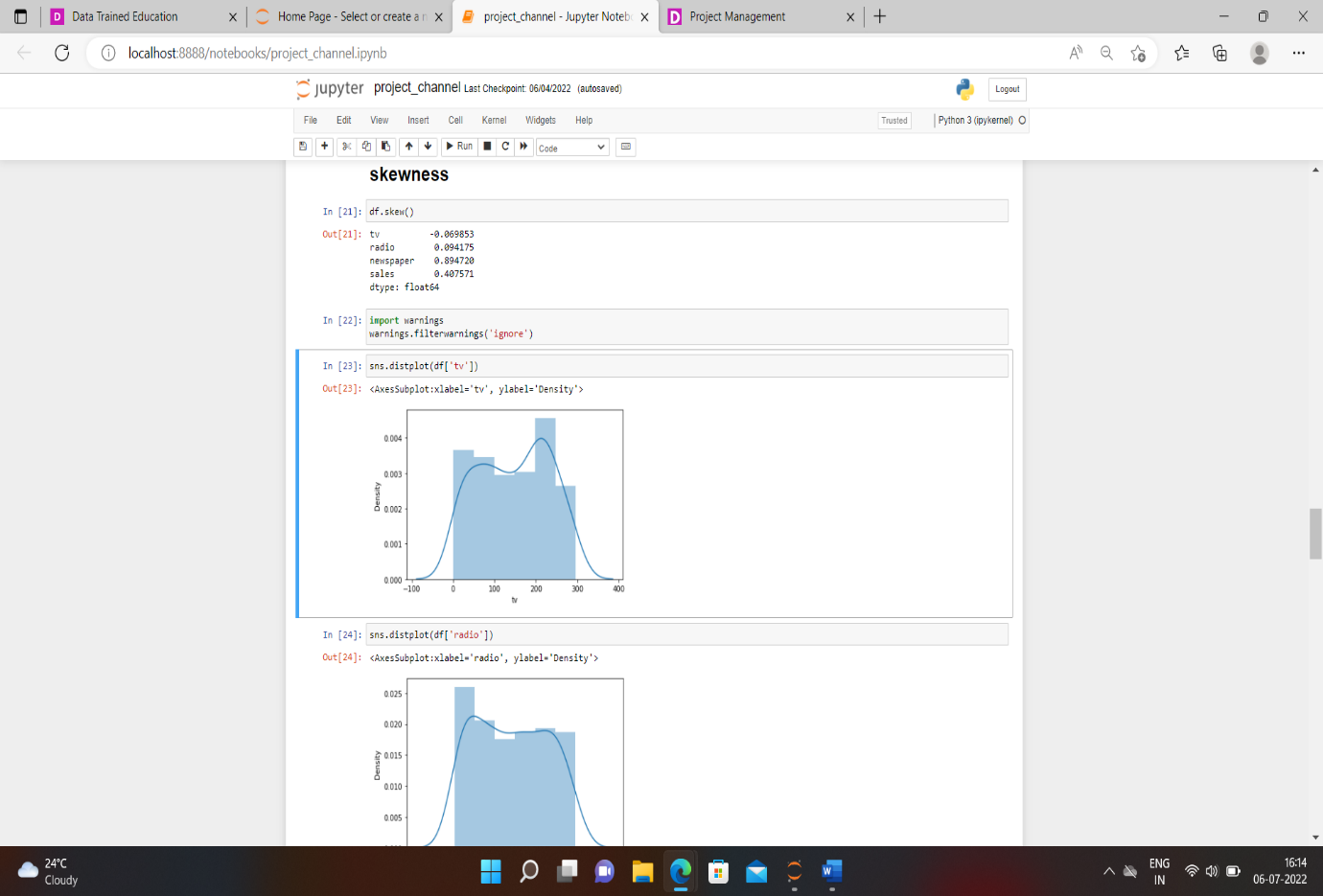
We see the correlation with each other.

Then we come to the next step called describing of data and which we see using the df. describe() method. And also see the there count, mean, standard deviation, minimum value, 25 percentile, 50 percentile as well as 75 percentile, and maximum values of the data set.

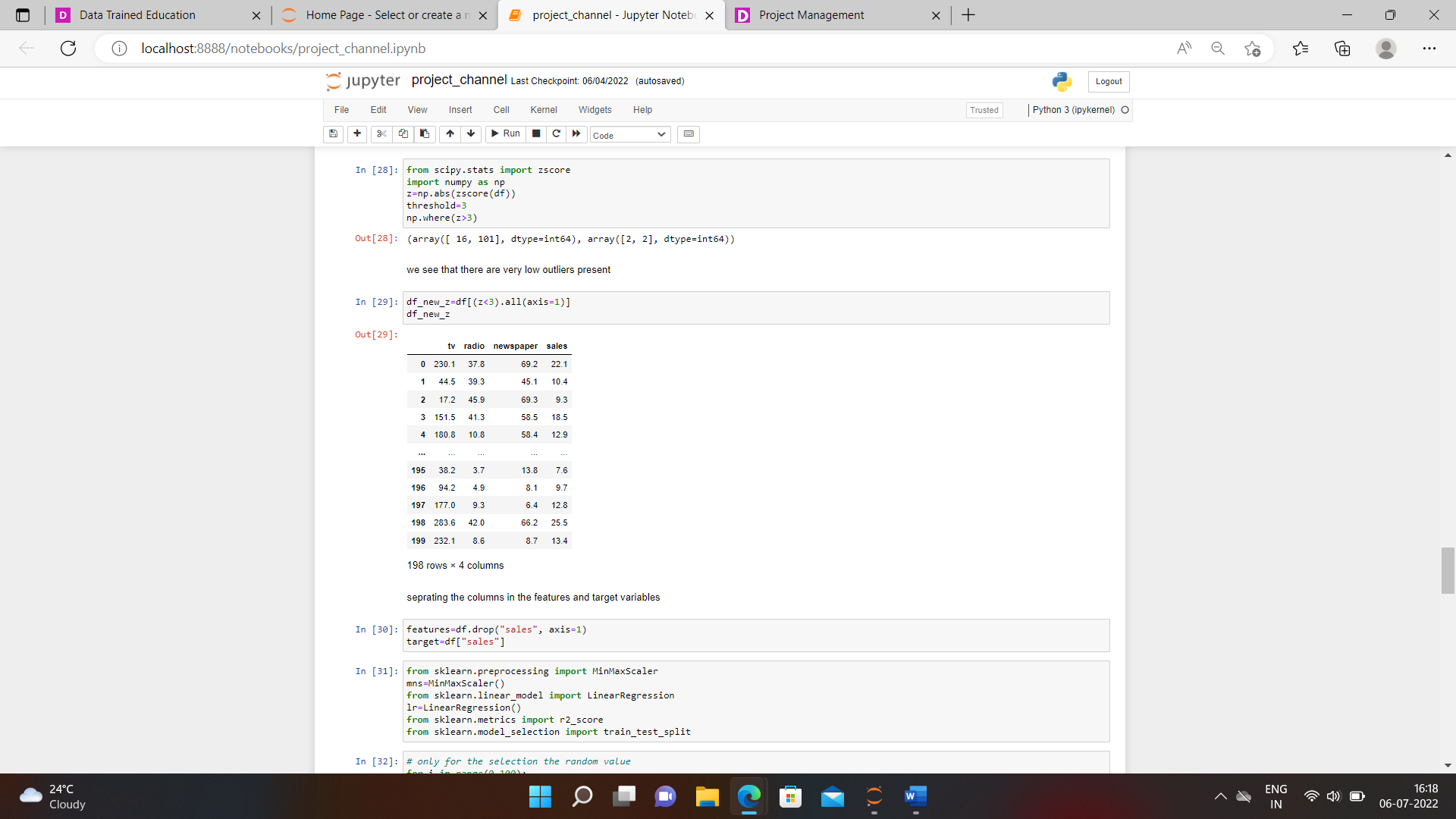


Using the describe method we see the outliers are present so use the box. pl ot we see the outliers.

After that, we go further skewness of the data so we use the df. skew() method to see the skewness of the dataset. And also see

 the distplot of all the columns.

After that, we remove the outlier so we use another library score and remove the outliers.



We see that we remove the outliers and our EDA process will be completed.

**Building Machine Learning Models:-**

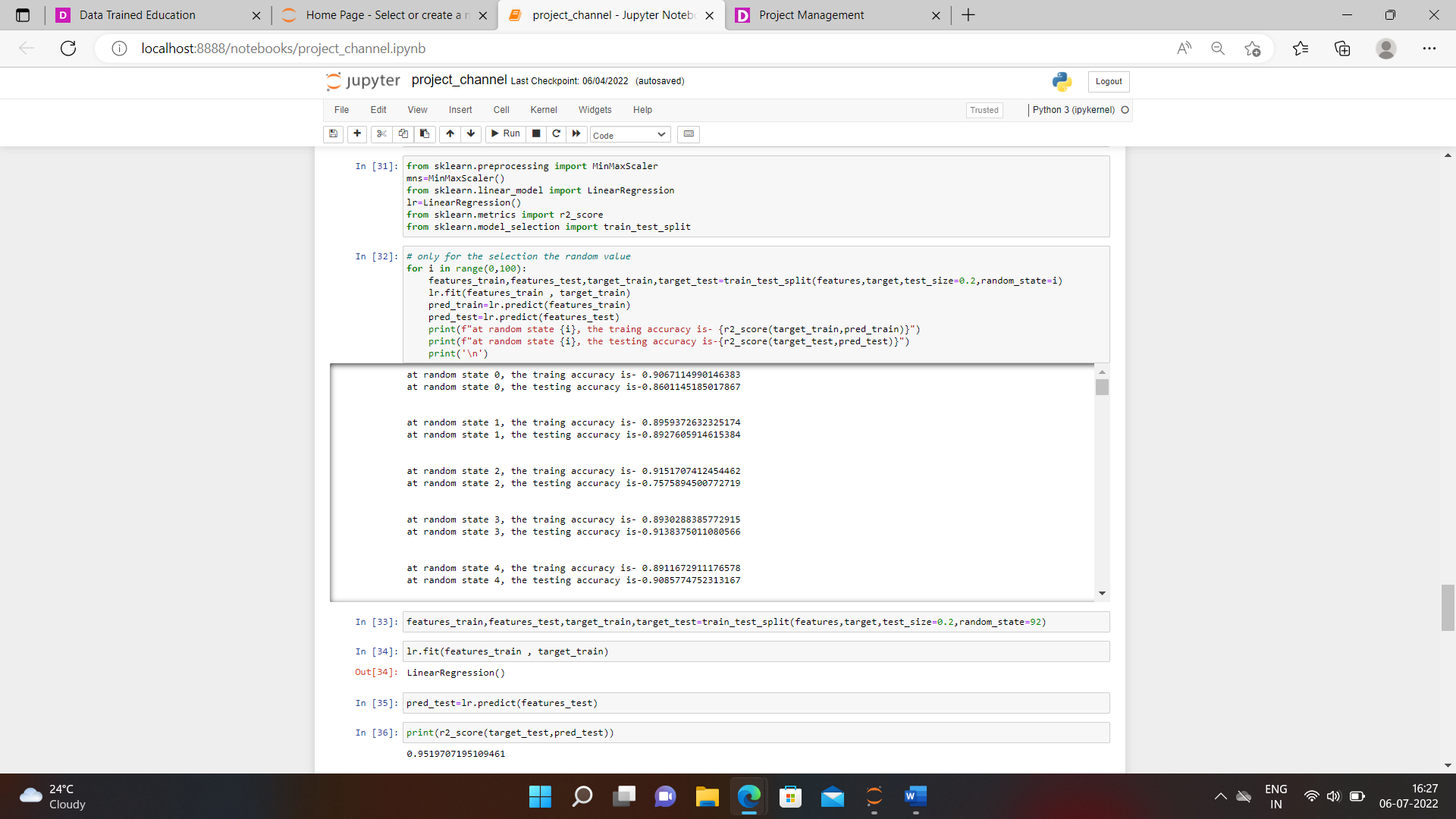
For building the model we split the data into 2 parts feature is called input data and which contains the TV radio and newspaper and the second part is the target and contains the sales column.

Firstly we import all libraries are MinMaxScaler and this is used for all values are in one shape, then we import the LinearRegression which is used for the regression purpose.

For accuracy score we import r2\_score

And splitting the data we use the train \_test\_split method.

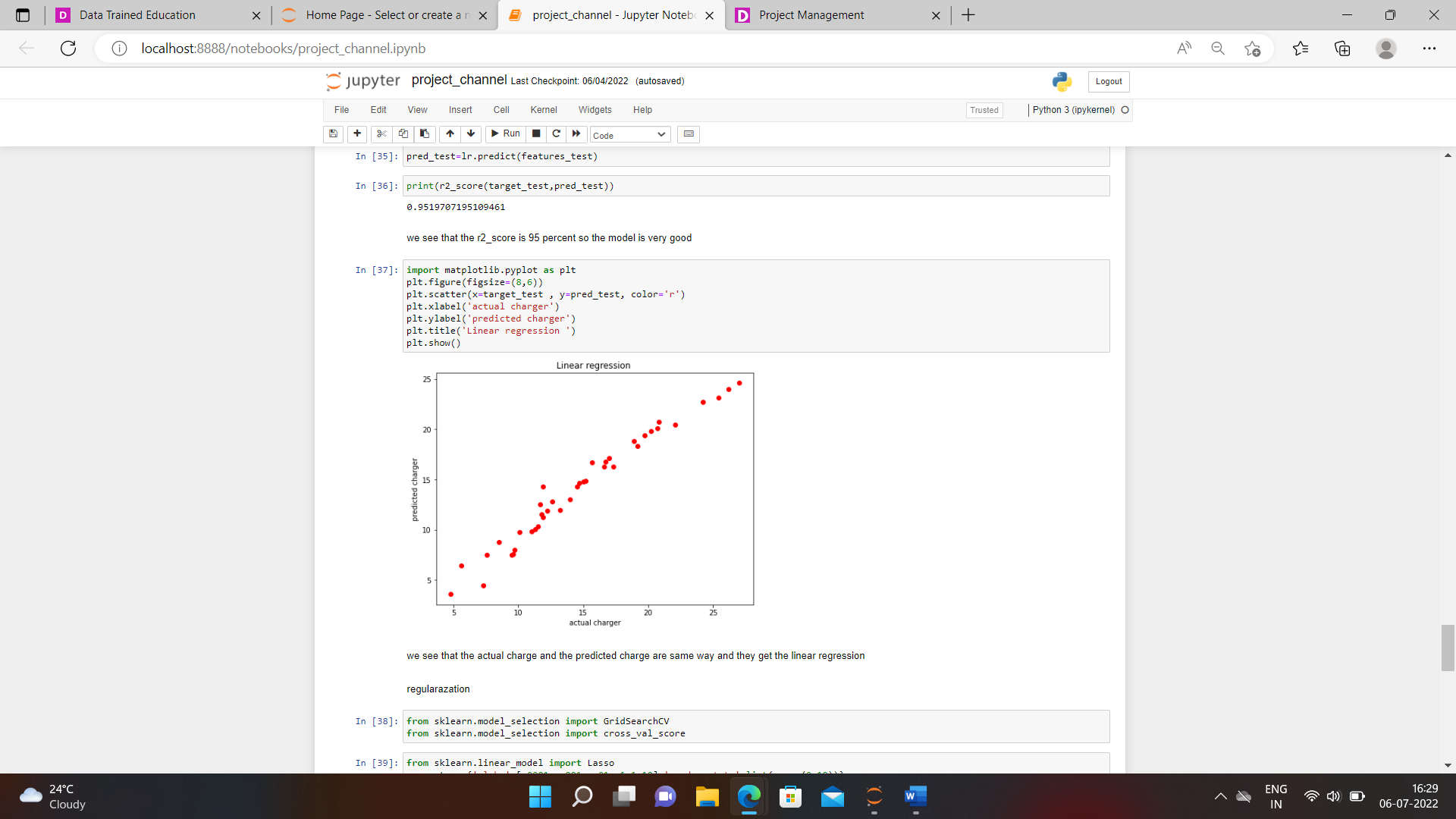
After that, we take the data in one loop and see the best accuracy of training and testing data.



After that, we fulfill the data according to the linear regression method.

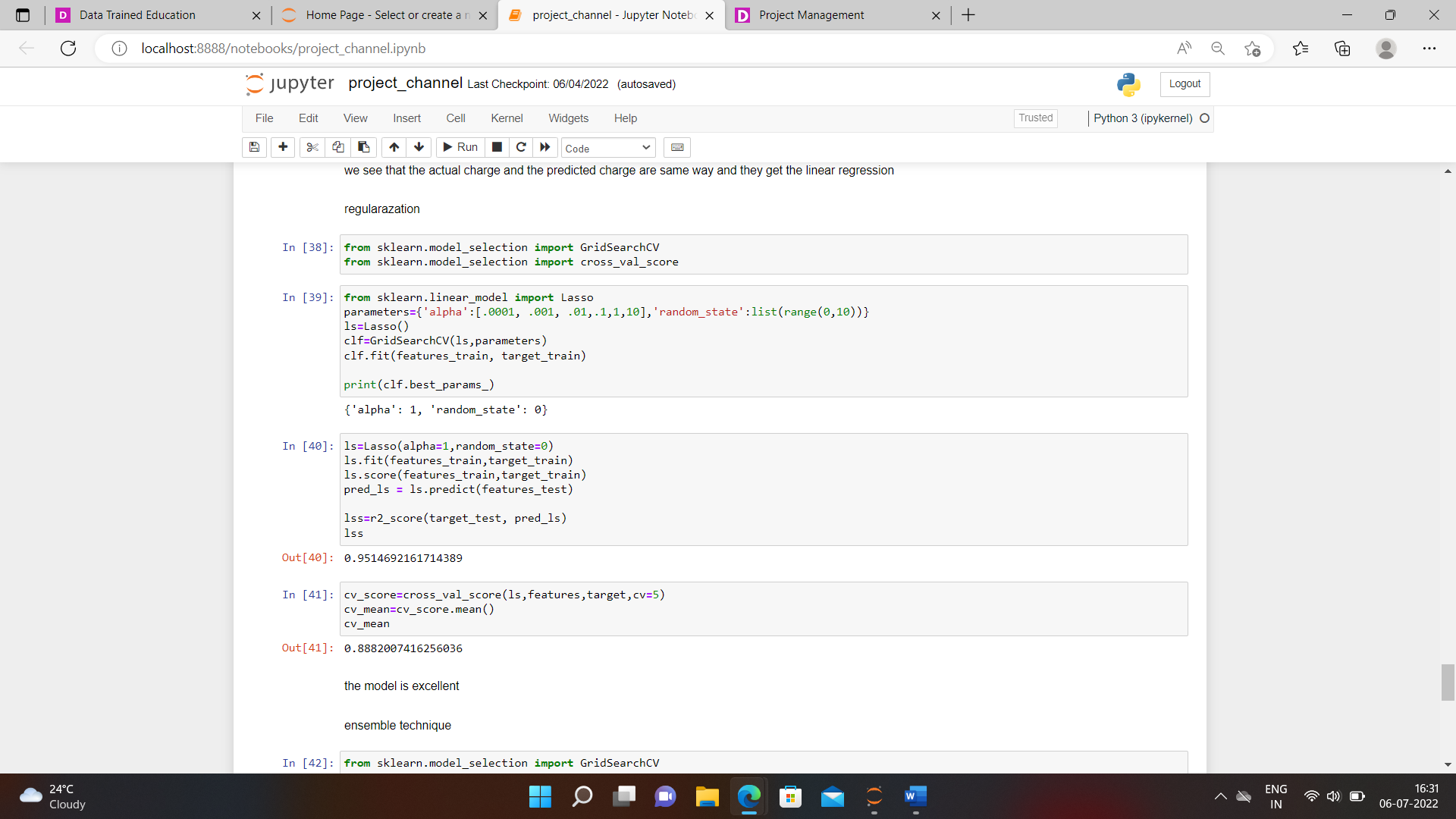
We see that our r2\_ score is 95%.

Then we plot that data using the matplotlib method.



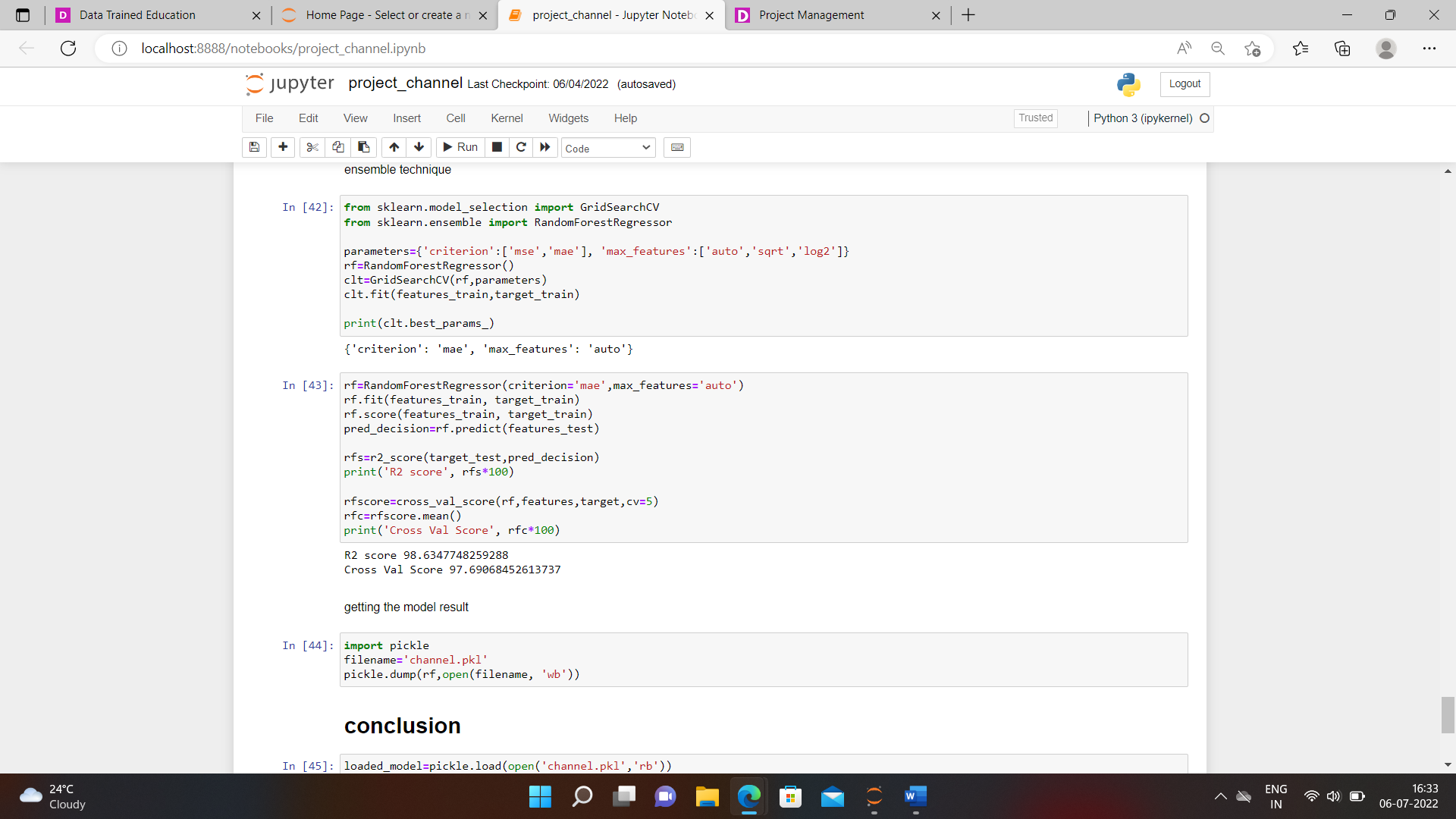
We see the actual charge, as well as the predicted charge, are in the linear form.

Then the next process is the Regulation method and cross\_validation method.



After that, the last process is the assembly technique.

There we see the r2\_score and the cross Validation score.



Our model will be ready with 98 % accuracy.

**Conclusion**:-

We save the best result using the pickle and see the predicted and the actual values of the sales according to our input.

