A Deep Dive on Business Intelligence Systems and Infrastructure using Cloud Environment

In this paper author is giving brief reviews on various technologies to designed Business Intelligence Systems. Business intelligence is a computer technology used to identify, obtain and analyse business data for example revenue, products, costs, sales revenue etc. This system generates huge amount of business data and analysing such huge data to extract meaningful information may consume lots of computer processing time. This analysis will helps in any business in identifying which product is in more demand and which is in less demand and by using this analysis they will maintain sufficient stock to boost their revenues.

In propose paper author is suggesting to used Cloud Environment system to manage business data as this cloud server may contains heavy computation systems and can process any size of data efficiently. After extracting meaningful information from processed data then we can apply data mining (machine learning) algorithms to predict future sales and revenues.

To train data mining algorithm author suggesting to use past history and then visualize such history business data to know sales and to know products in more and less demands.

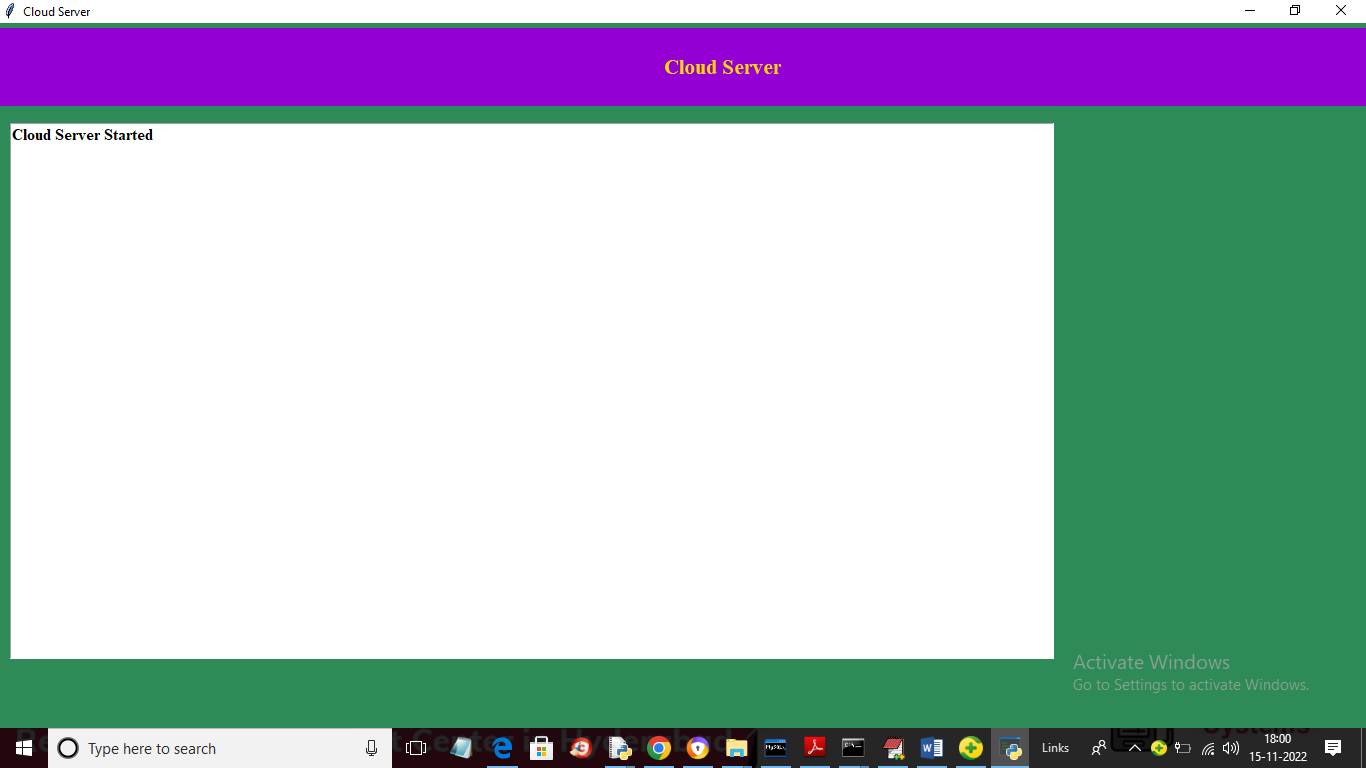
After analysis we can generate reports and graphs on sales data and by seeing this results business owner may know the sales on different dates.

To implement this project we are using ‘Restaurant Sales’ dataset and this dataset will be stored at Cloud Server. So we designed two modules called

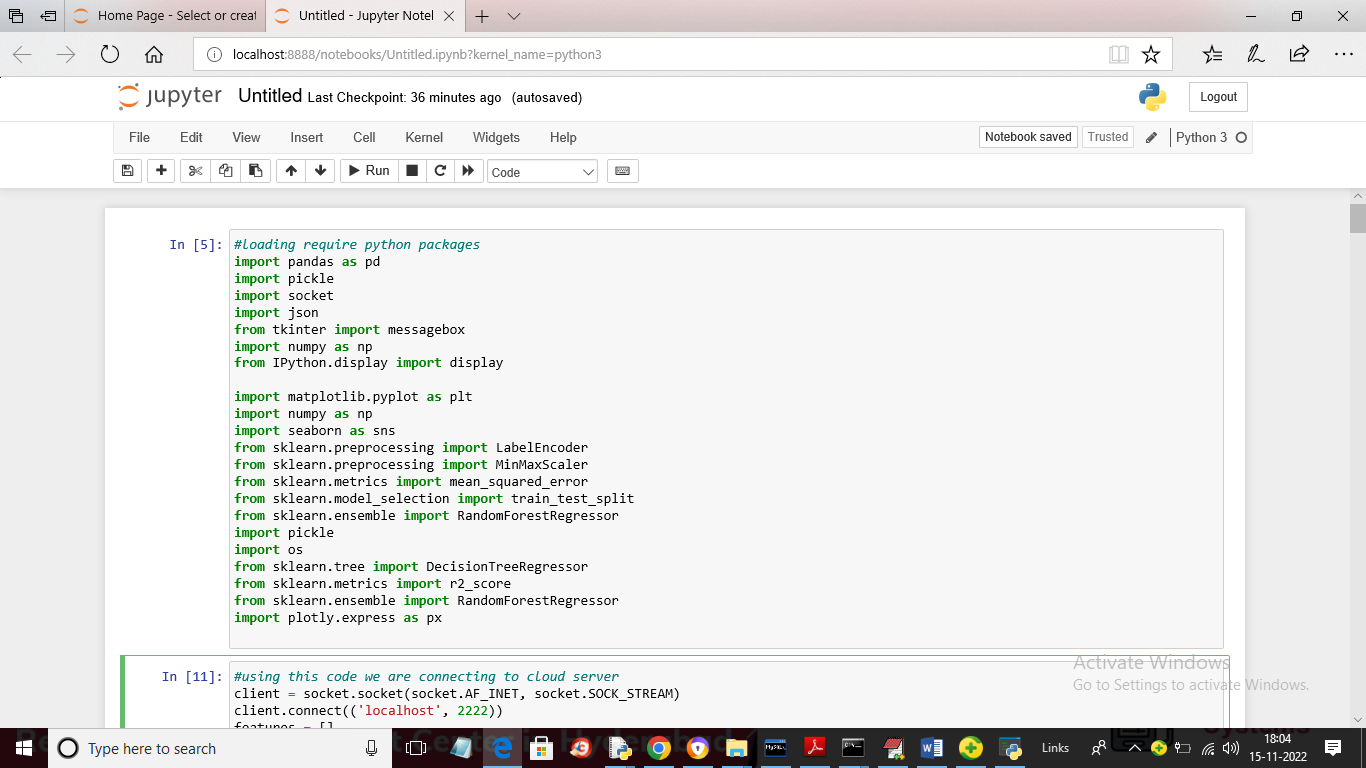
1. Cloud Server: this server will store business past data and then process this dataset and give clean data to Business System
2. Business Sales Analysis: this module will collect data from Cloud Server and then display dataset details and then plot various sales graph and then apply data mining algorithms such as Decision Tree and Random Forest to predict future sales by training itself on Past Sales History data. Decision Tree and Random Forest will get trained in past data and then predict sales for future data. Algorithm which obtained with less Mean Square Error on predicted data will be considered as accurate in sales prediction. This module will plot various sales analysis graph so we coded this module using JUPYTER notebook

SCREEN SHOTS

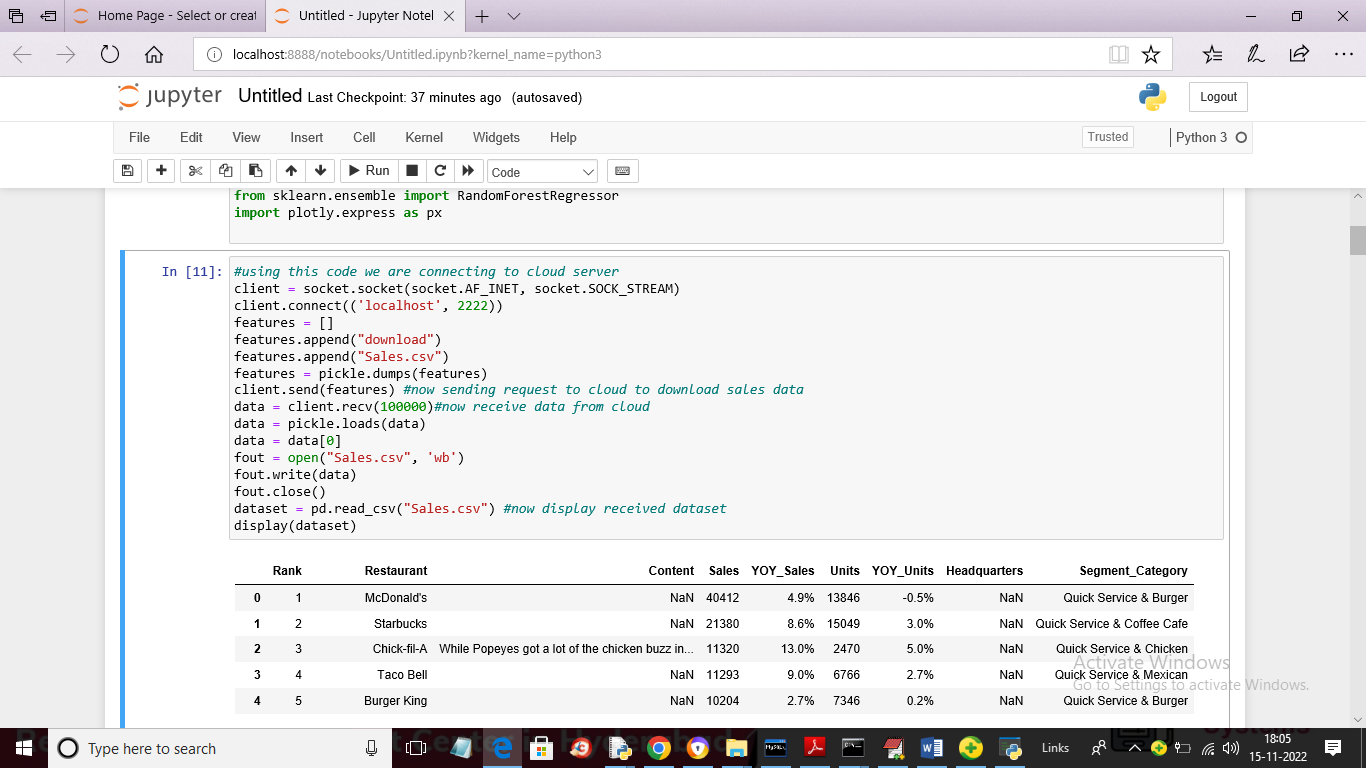
To run project double click on ‘run. Bat’ file from ‘Cloud Server’ folder to start cloud server and get below screen



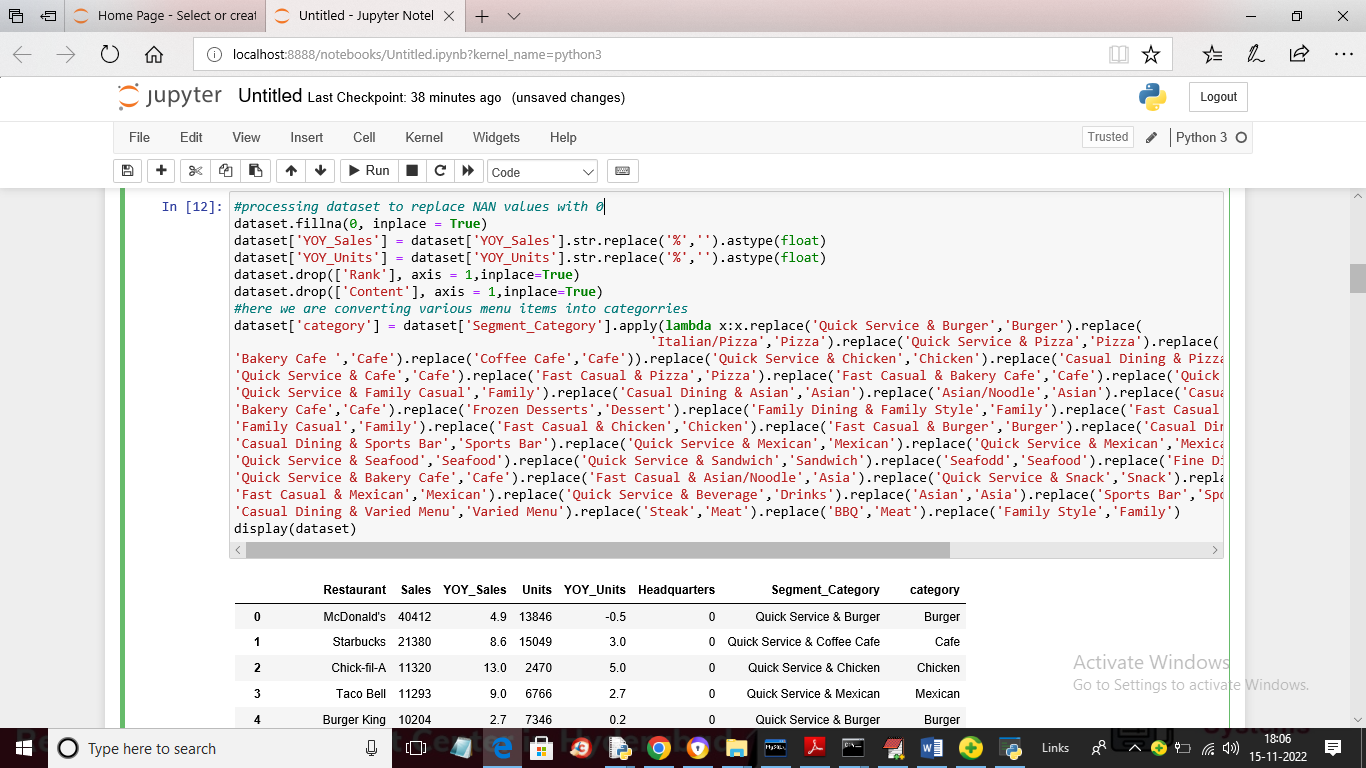
In above screen cloud server started and now double click on ‘run.bat’ file from ‘Business Sales Analysis’ folder to get output in JUPYTER output screen like below one



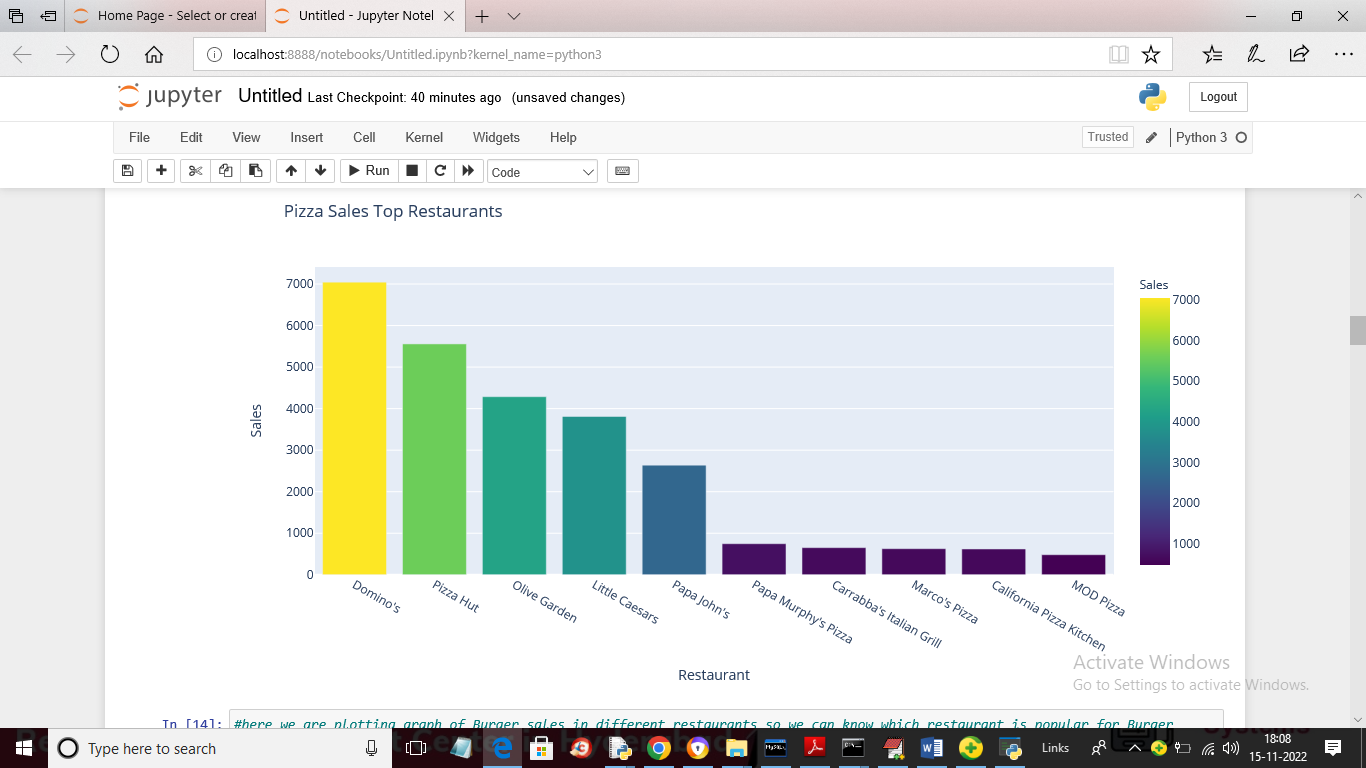
In above screen we are loading require python packages



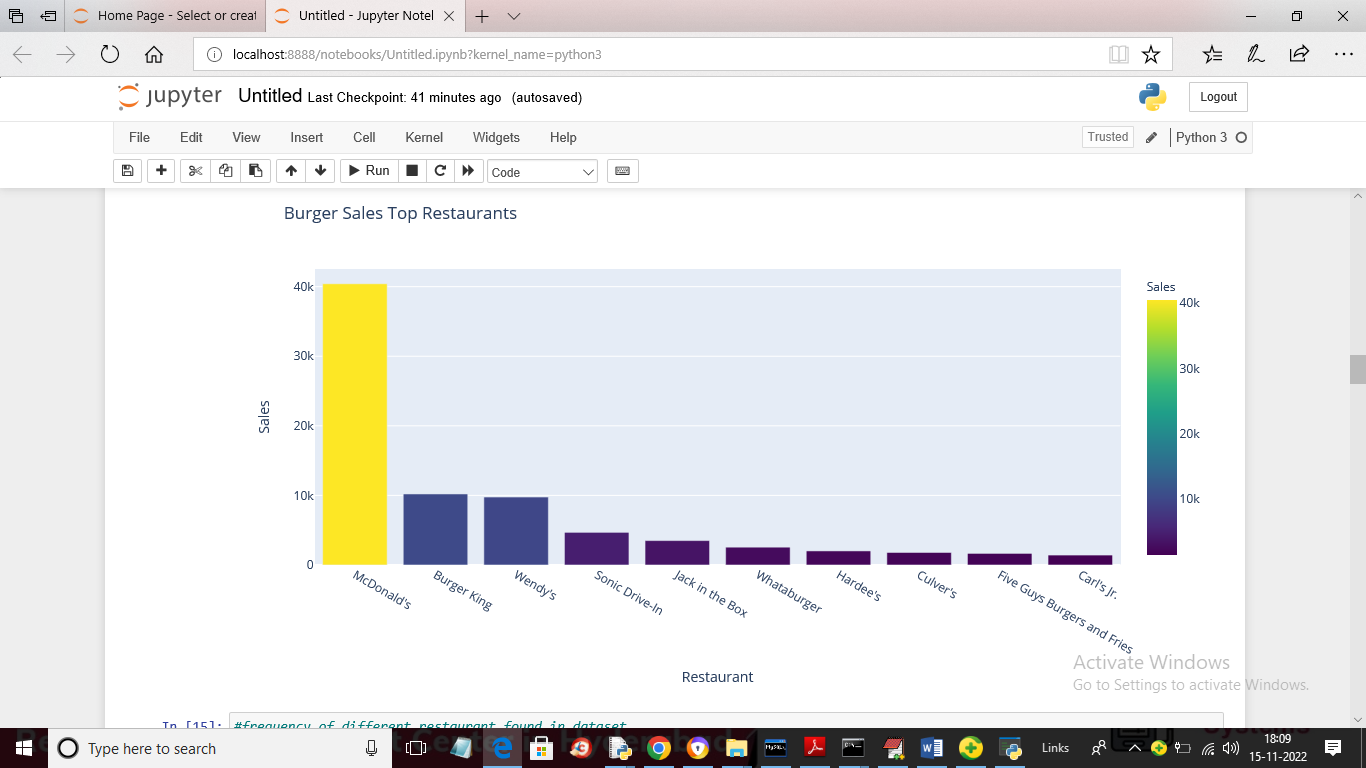
In above screen read blue colour comments where you can see we are connecting to cloud server to get past sales dataset and then displaying received dataset and in above dataset we can see NAN values are there and by processing we need to replace that data with 0



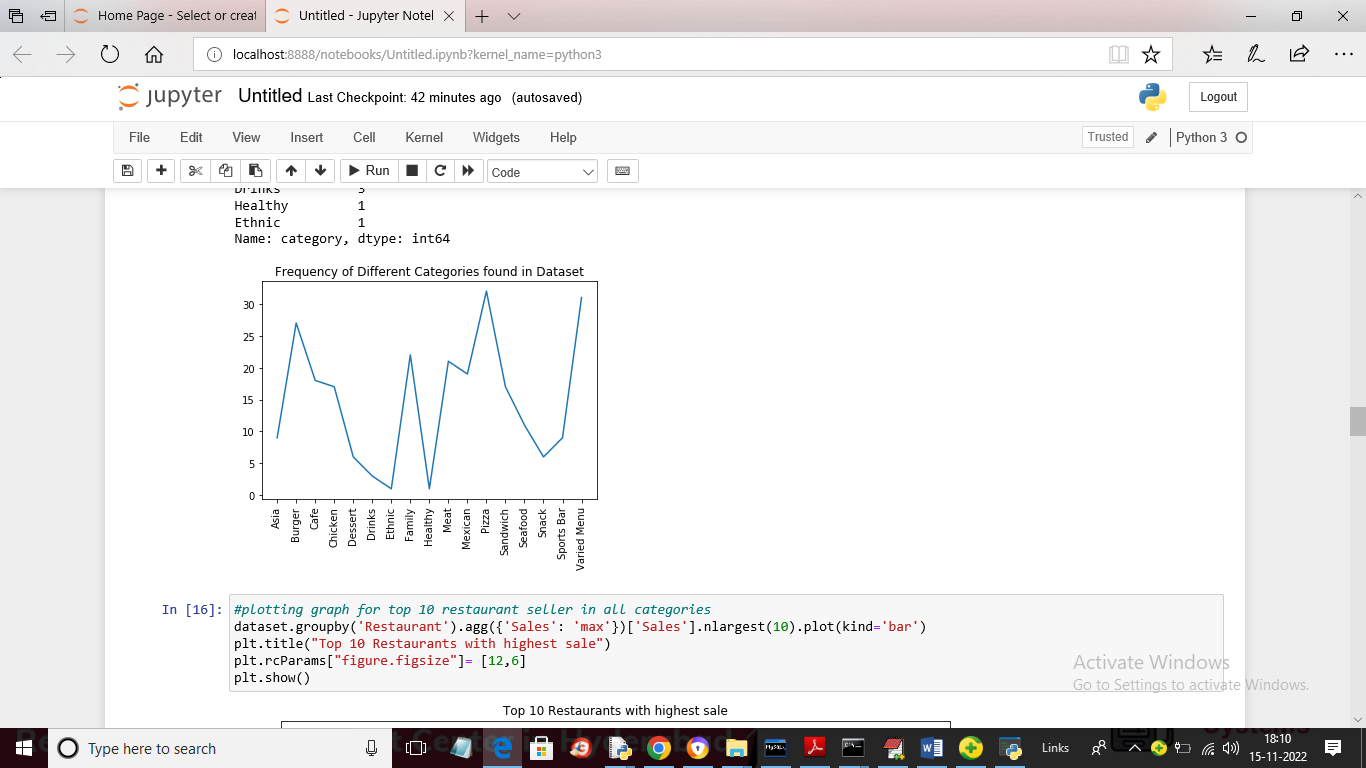
In above screen we are processing dataset to replace missing values with 0 and then converting all restaurant sales menu into categories like Burger, Pizza etc. in above screen after processing we can see all NAN values are removed



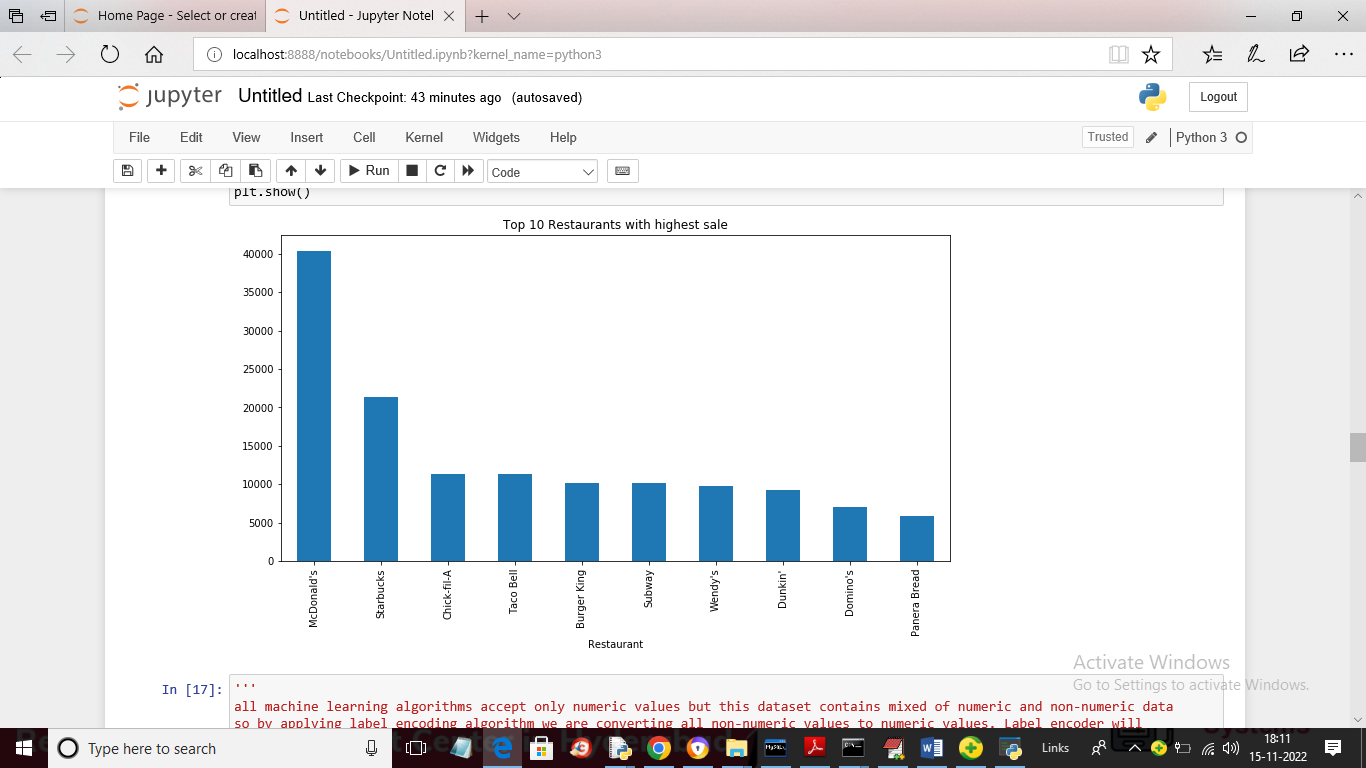
In above screen we are finding PIZZA sales done by different restaurants and by seeing above graph restaurants owners will try various technique to boost their sales and in above graph x-axis represents Restaurant Names and y-axis represents SALES



In above graph we are showing restaurants with burger sales

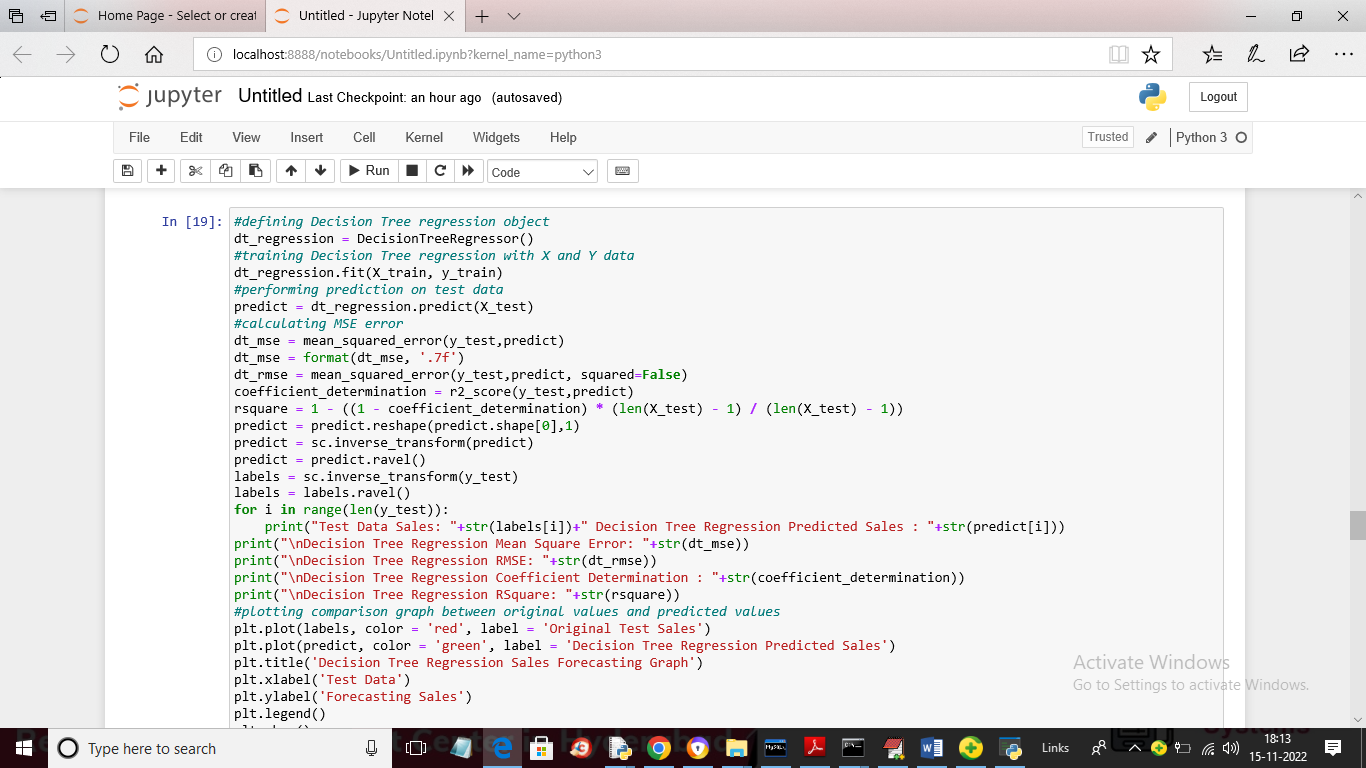


In above line graph x-axis represents menu names and y-axis represents sales and by seeing above graph owners knows which menu is in more demand and based on that they will arrange their menus

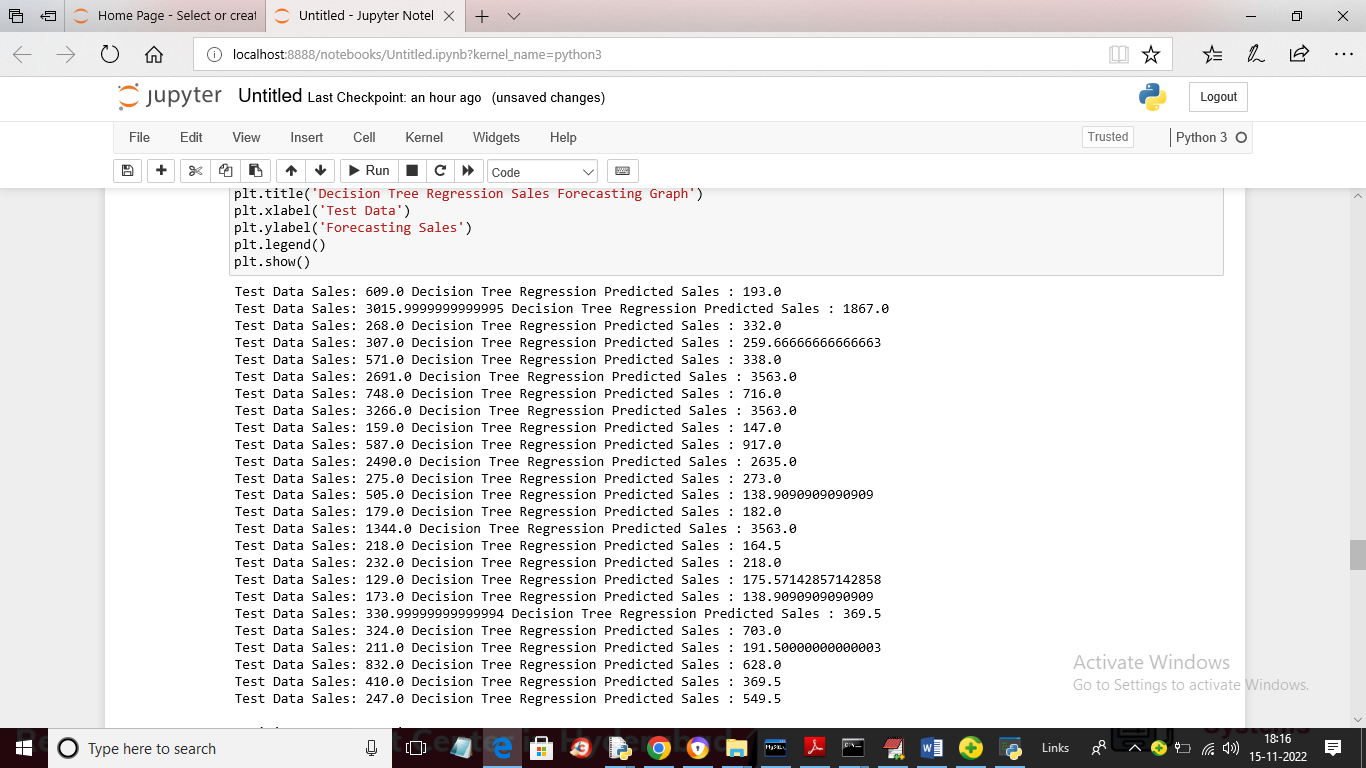


In above graph we are finding restaurants with highest sales where x-axis represents restaurant names and y-axis represents sales. So by using above graph we did sales analysis using cloud dataset.

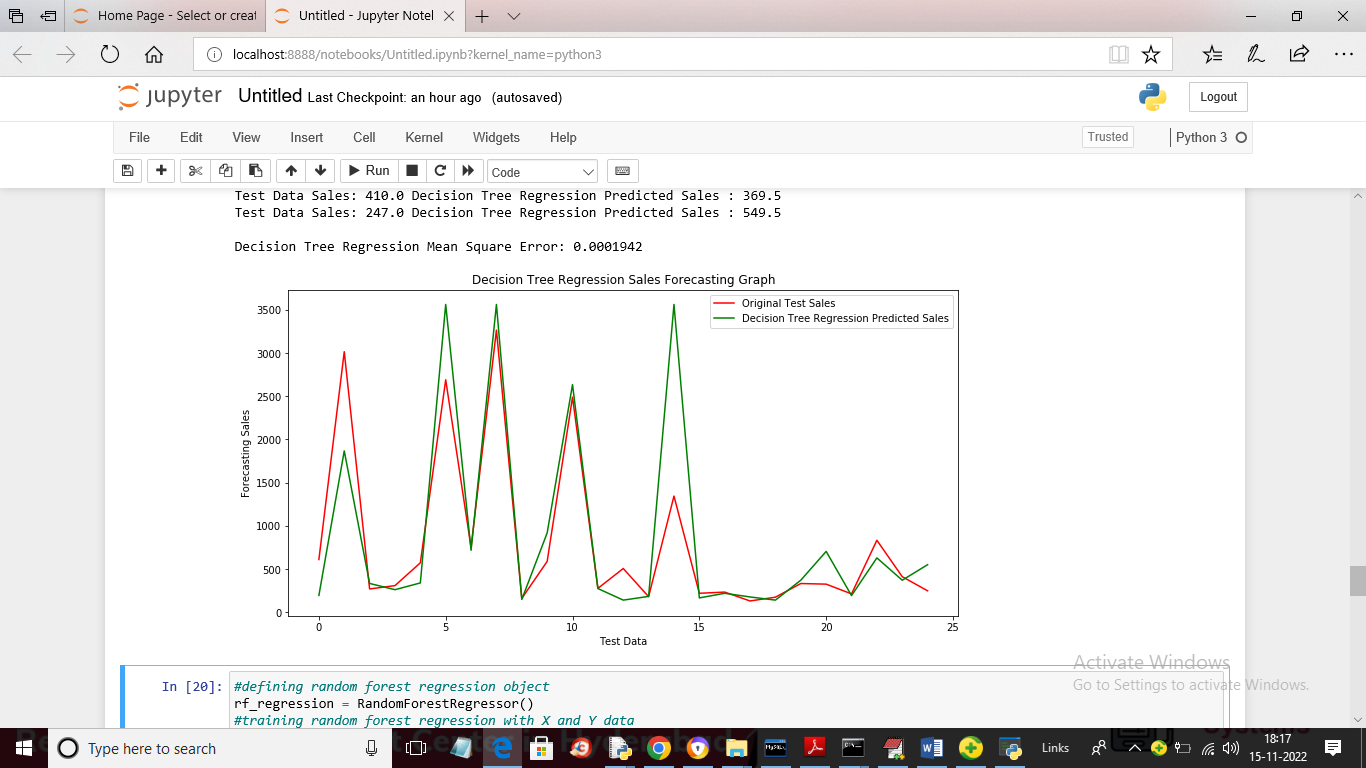
Now we will apply Data mining algorithms on above dataset to train and predict futures sales



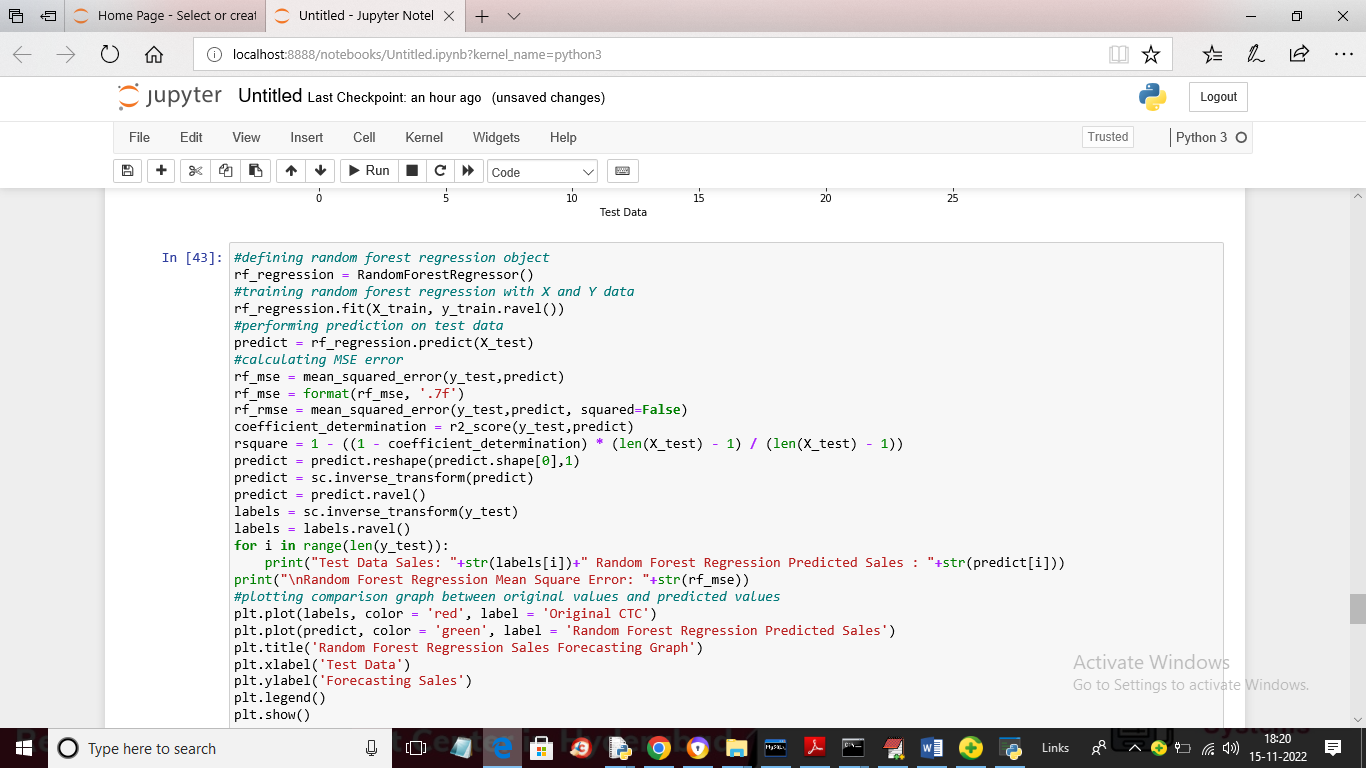
In above screen we are training decision tree data mining algorithm to trained on dataset and then perform sales prediction on TEST data to know algorithm capacity of accurate sales prediction future forecasting



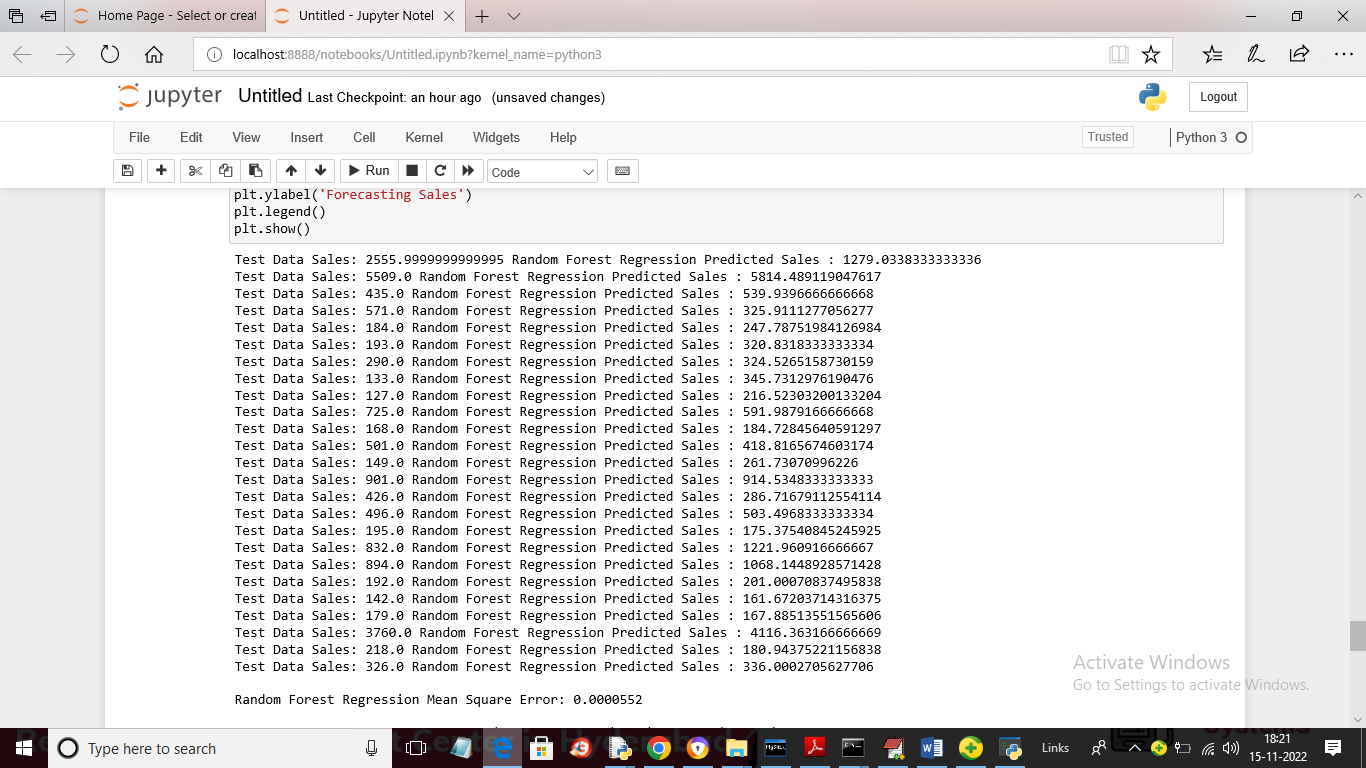
In above screen we can see Test data sales and the sales predicted by decision tree and in below graph we can see original test data sales and decision tree predicted sales



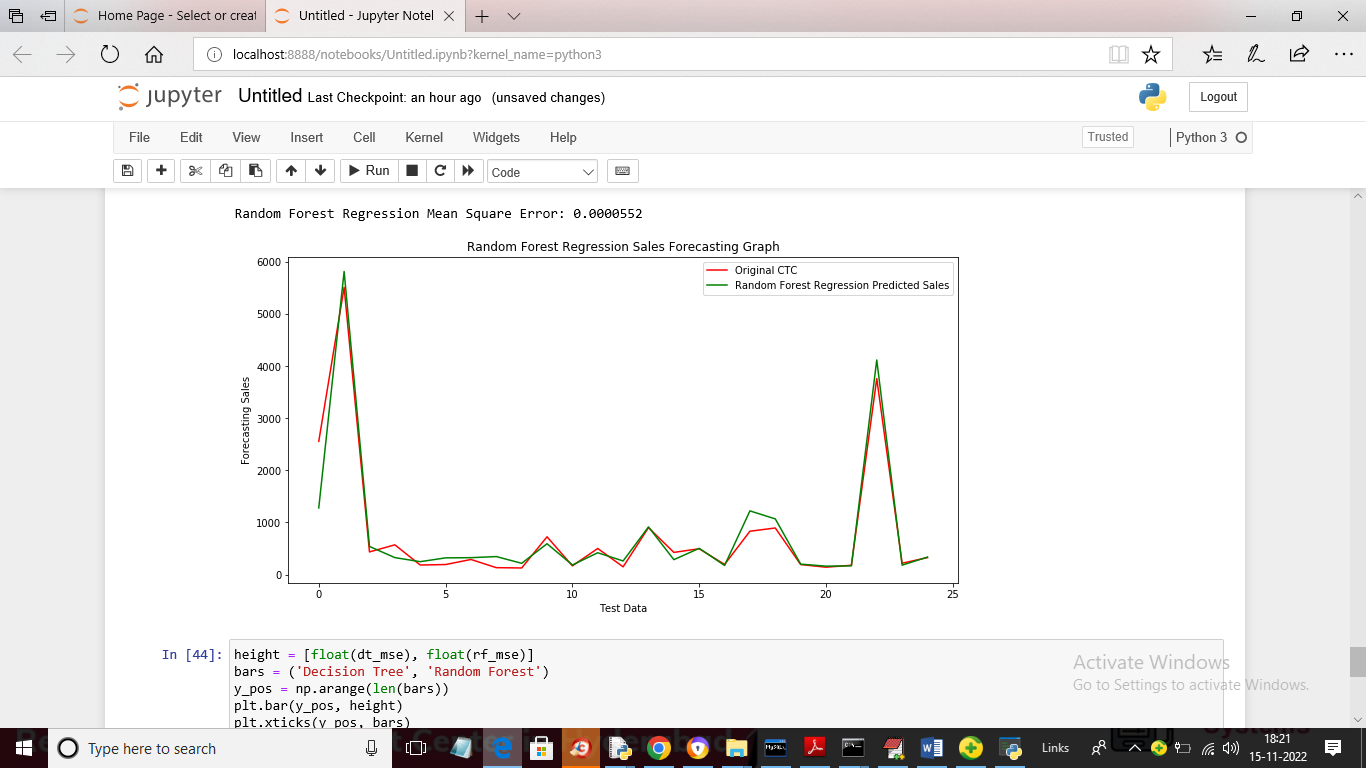
In above graph x-axis represents Number of days and y-axis represents SALES and red line represents actual TEST data sales and green line represents Decision Tree predicted sales and we can see both lines are not colliding so decision tree prediction is not accurate and if algorithm predicted accurately then it will have close difference between TEST data and predicted values. In below screen showing random forest prediction



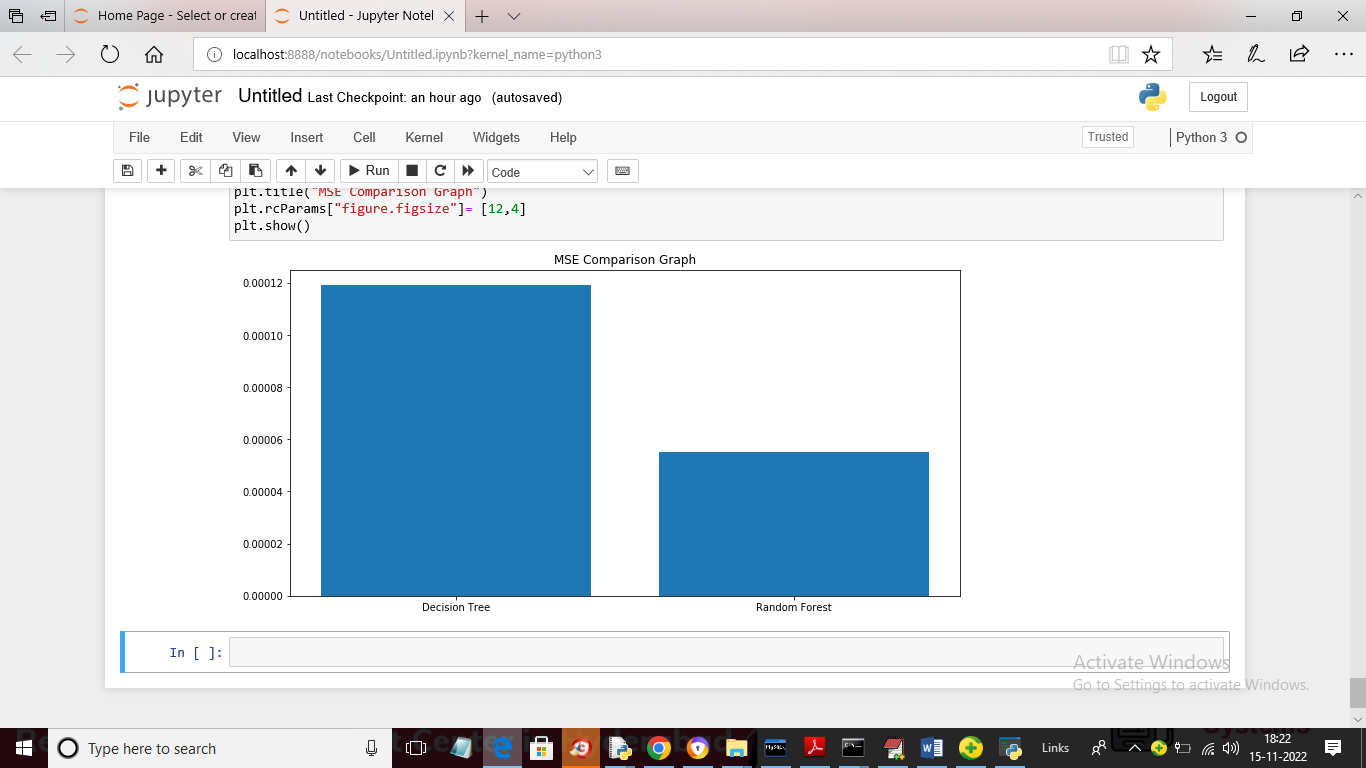
In above screen we are training Random Forest algorithm and in below screen we can see its prediction performance



In above screen we can see Original Test sales values and the sales predicted by Random Forest



In above random forest graph we can see both lines are fully overlapping with close difference so Random forest is accurate in sales prediction and below is Prediction error graph between both algorithms



In above graph x-axis represents Algorithms Names and y-axis represents error rate and in both algorithms Random Forest got less error rate so we can say random Forest is good at sales forecasting or prediction