**BIG DATA AND ANALYTICS IN TOURISM AND HOSPITALITY**

**GROUP-18**

1. **PAVANTEJA ANUPOJU -16322793**
2. **RAKESH NAGANDLA-16322729**
3. **PINAKAPANI REDDY-16322732**
4. **NITHISHA CHENNUPATI-16322175**

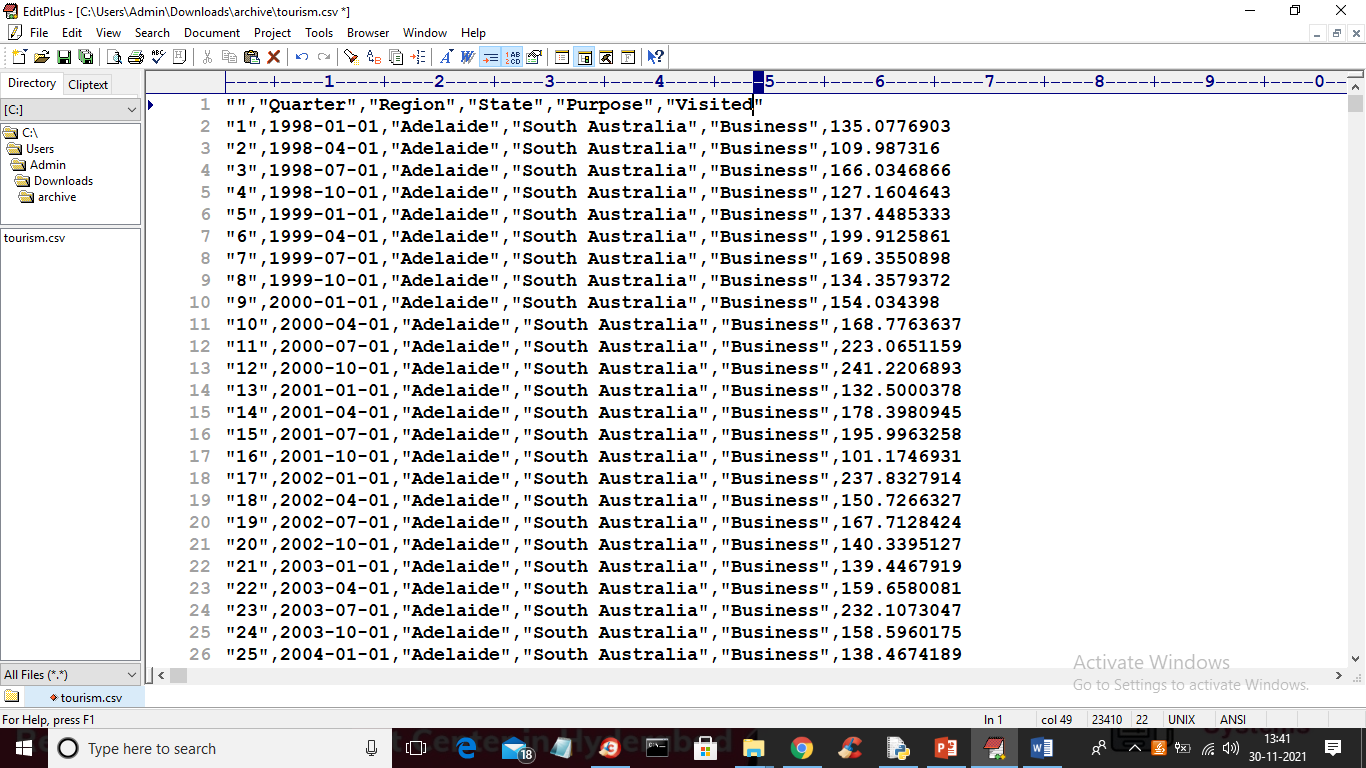
In this assignment we are working towards fast growing industry like Tourism and Hospitality which is playing major role in increasing wealth of any country. Peoples always attract towards good destination and hospitality and if we provide proper services by analysing places where peoples are attracted than it will boost tourism industry more.

To analysis tourist visits places we have used Australia tourism dataset from KAGGLE and this dataset can be downloaded from below LINK

<https://www.kaggle.com/luisblanche/quarterly-tourism-in-australia>

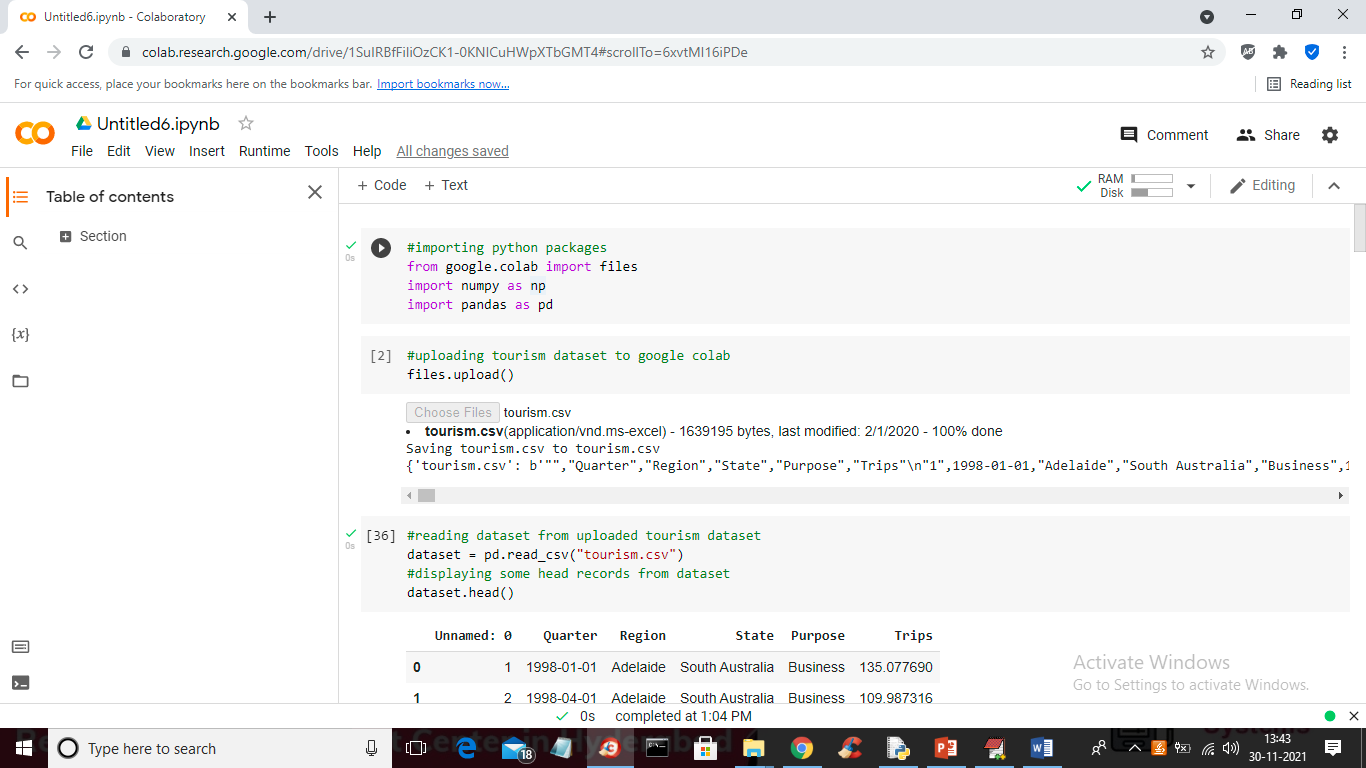
here as big data approach we are using Google Colab which provides virtual heavy resources to process any size of data and then we are analysing dataset to find places where more peoples are making trips and then finding TOP 20 most visited regions and then training this dataset with Decision Tree Regression algorithm to forecast tourist visit for next quarter for each region or place.

Below is the dataset used for this assignment analysis

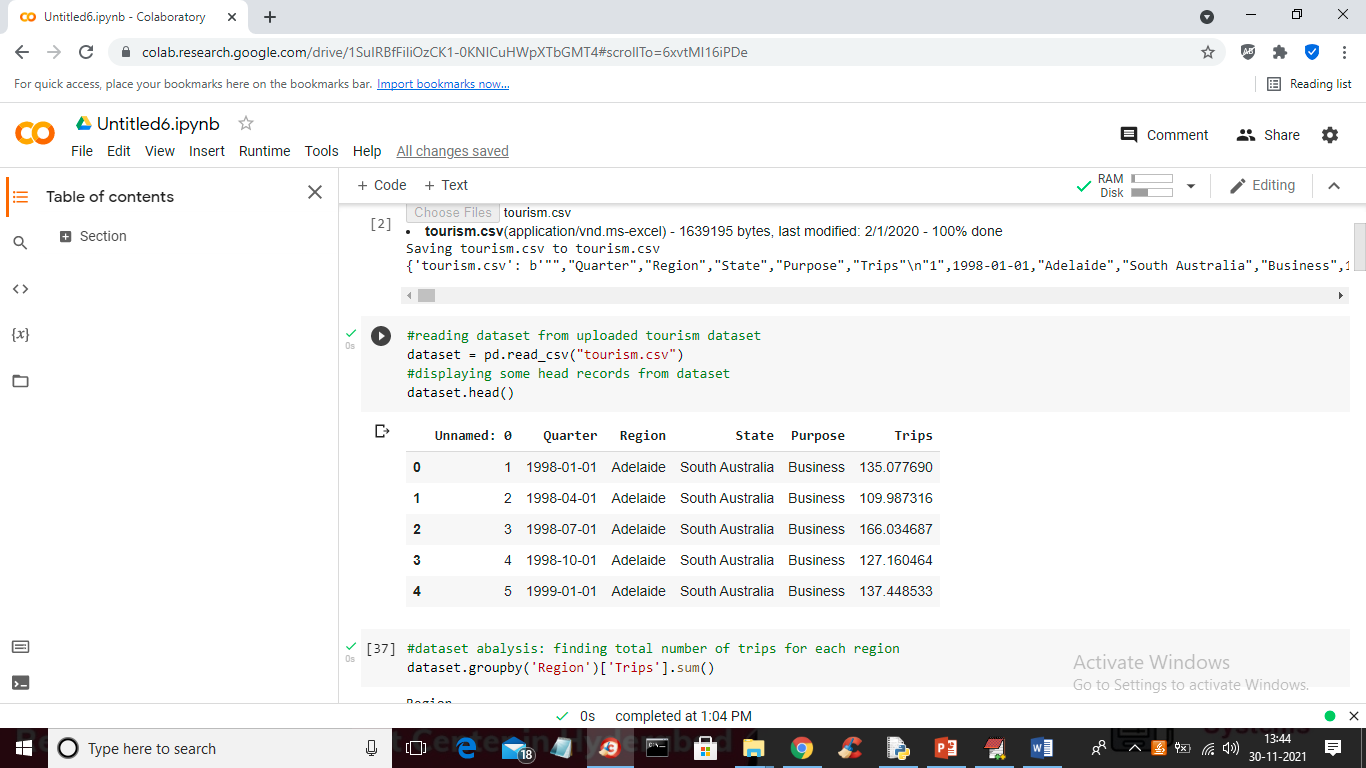


In above dataset screen first row represents dataset column names and remaining rows contains dataset values and in above screen we can see visited number of tourist in different year and region and we will used above dataset to forecast tourist visits for next quarter (in dataset each quarter contains 3 months gap).

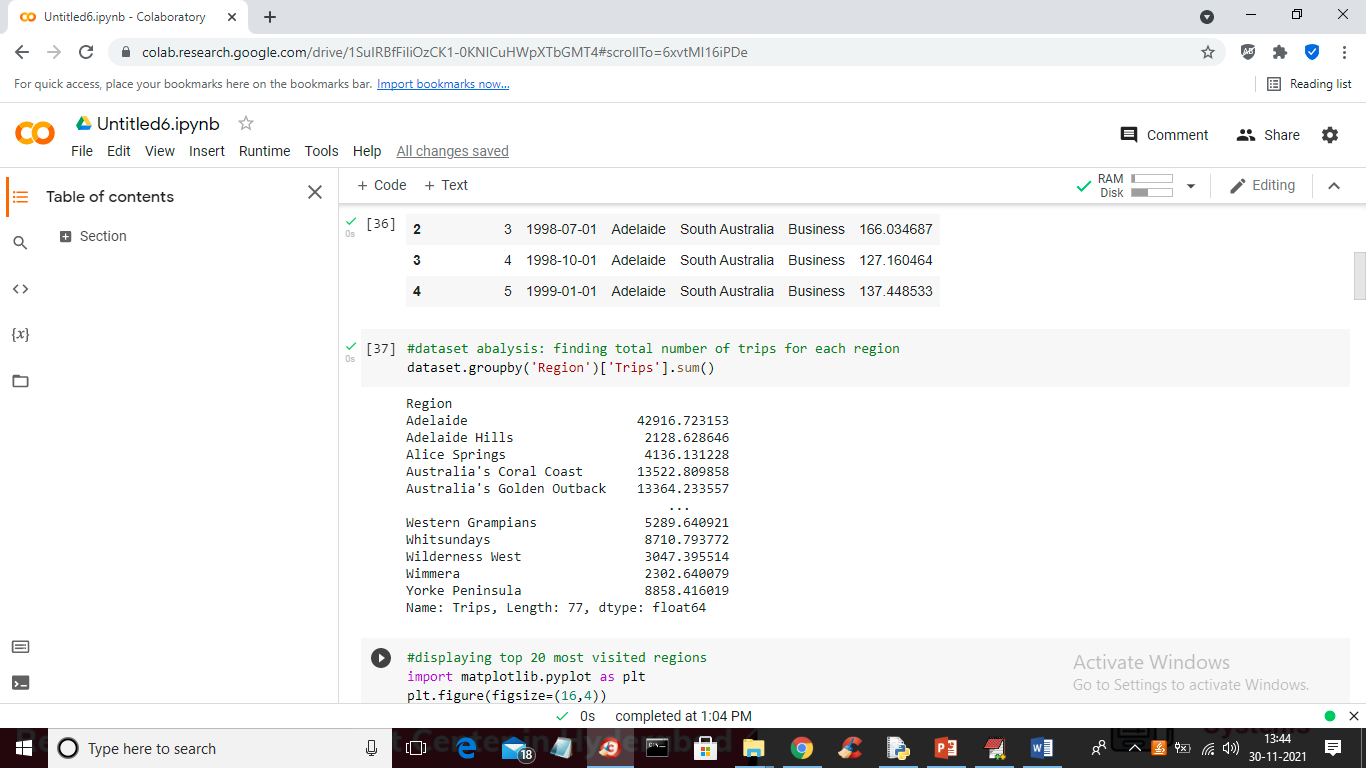
Below are the output screens of GOOGLE COLAB



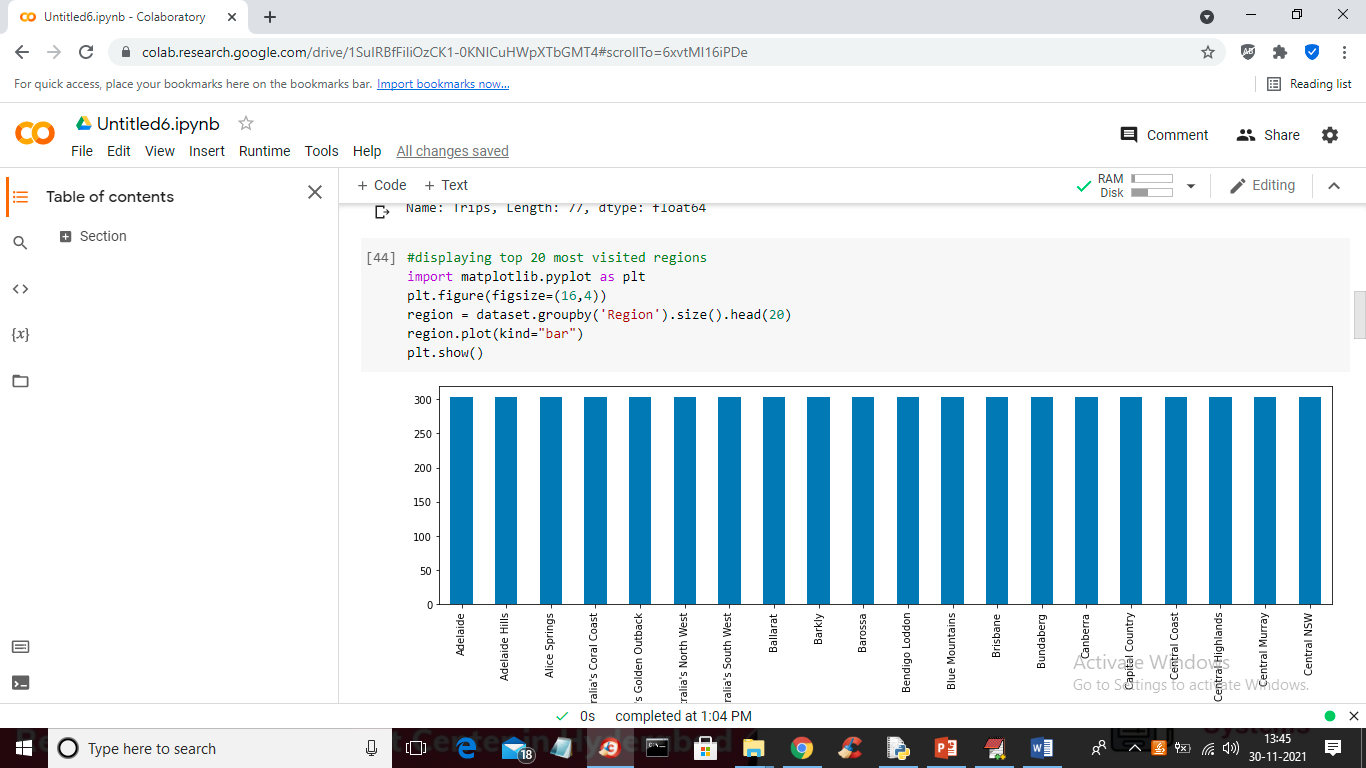
In above screen read comments associated with ‘#’ symbol to know about coding



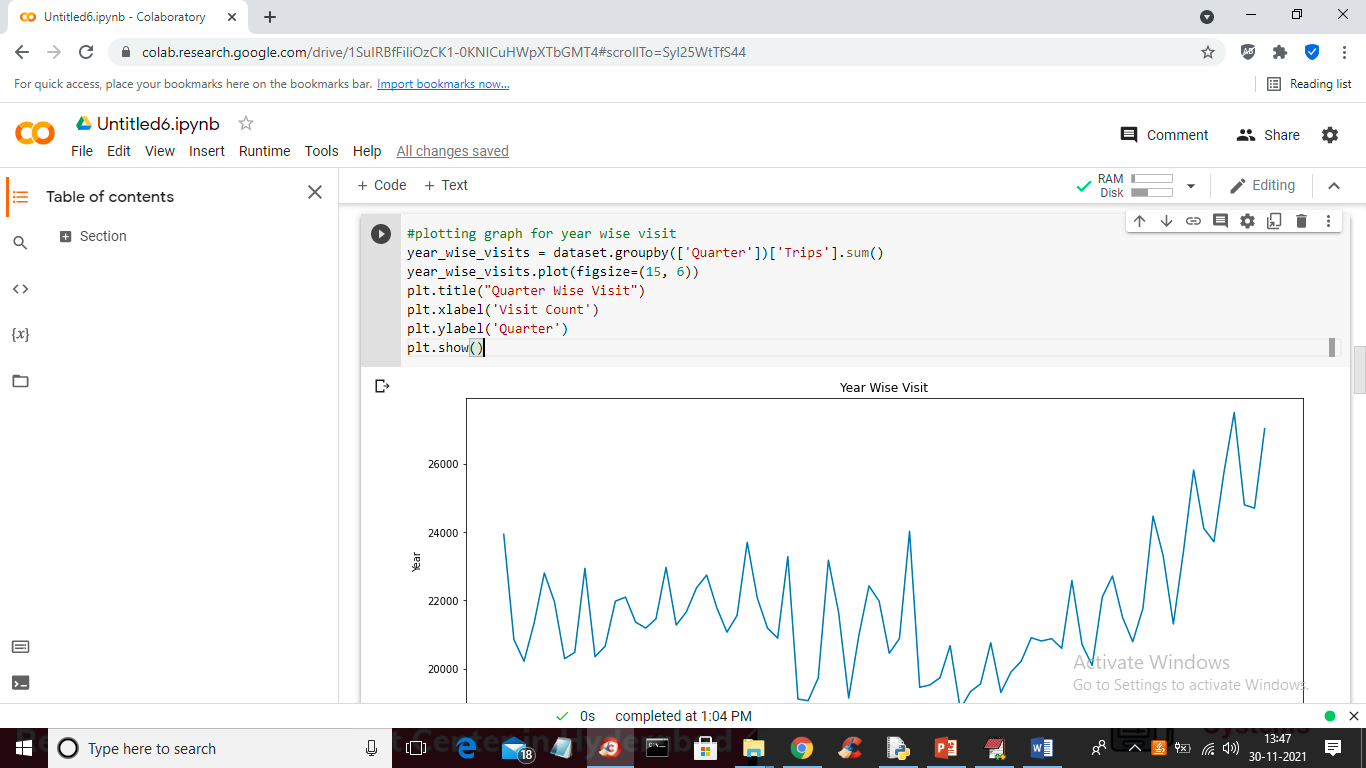
In above screen we are reading dataset and then displaying values from dataset

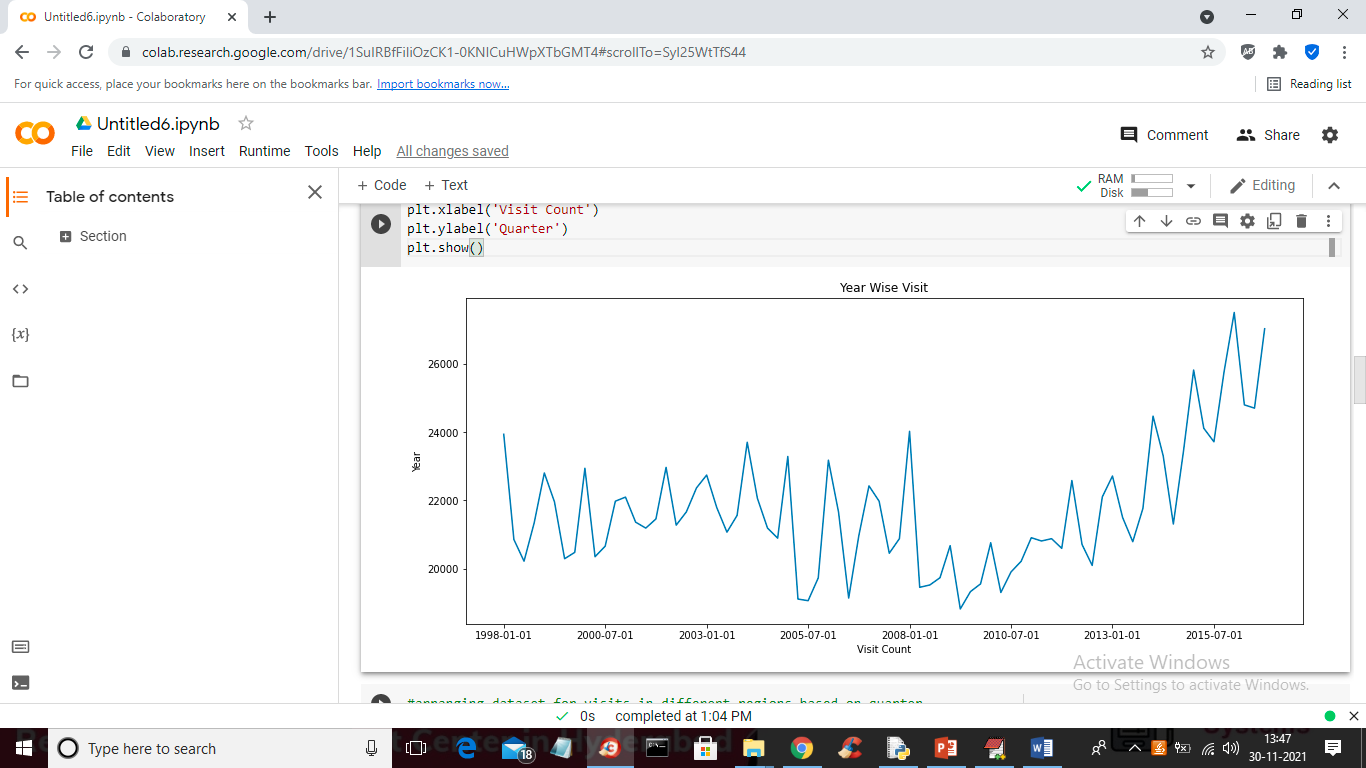


In above screen we are finding region or places where more number of peoples visited

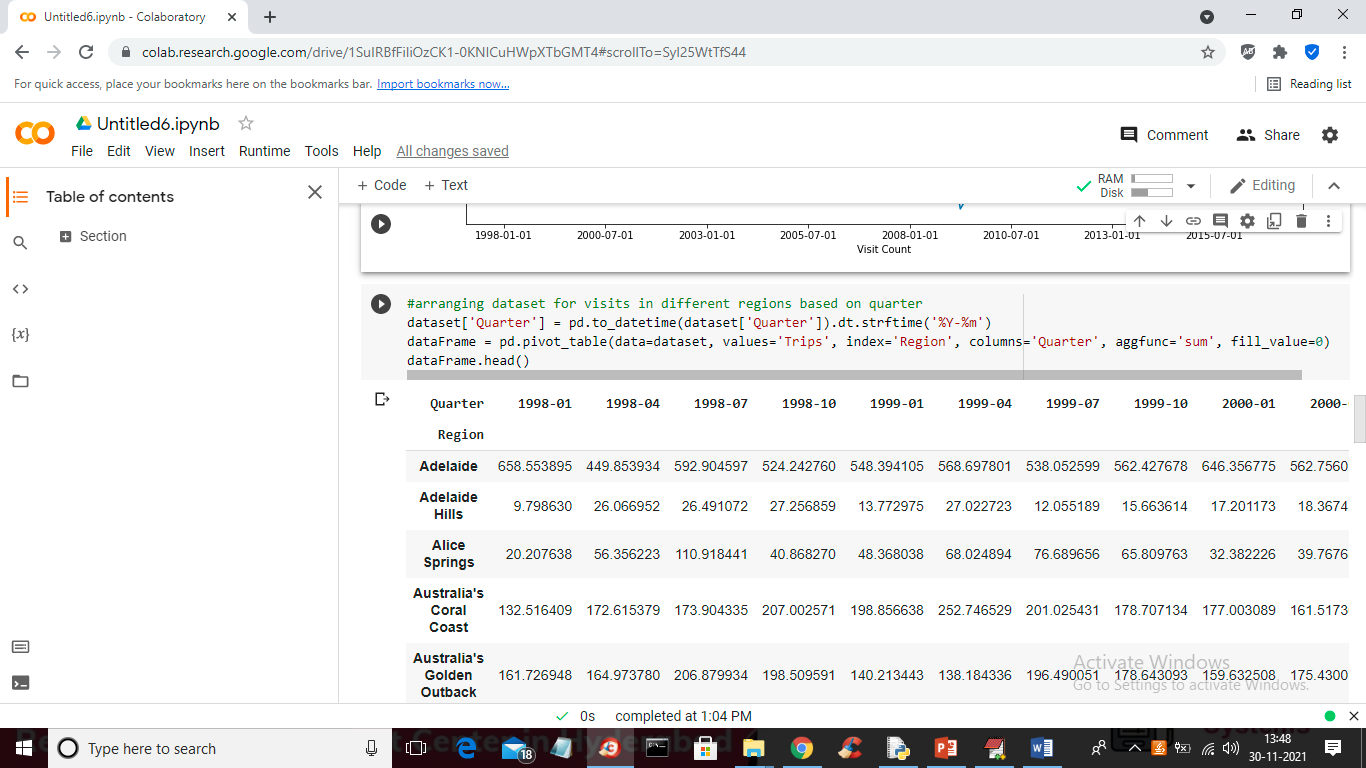


In above screen we are plotting graph with top 20 most visited regions and in above graph x-axis represent region name and y-axis represents number of peoples visited.

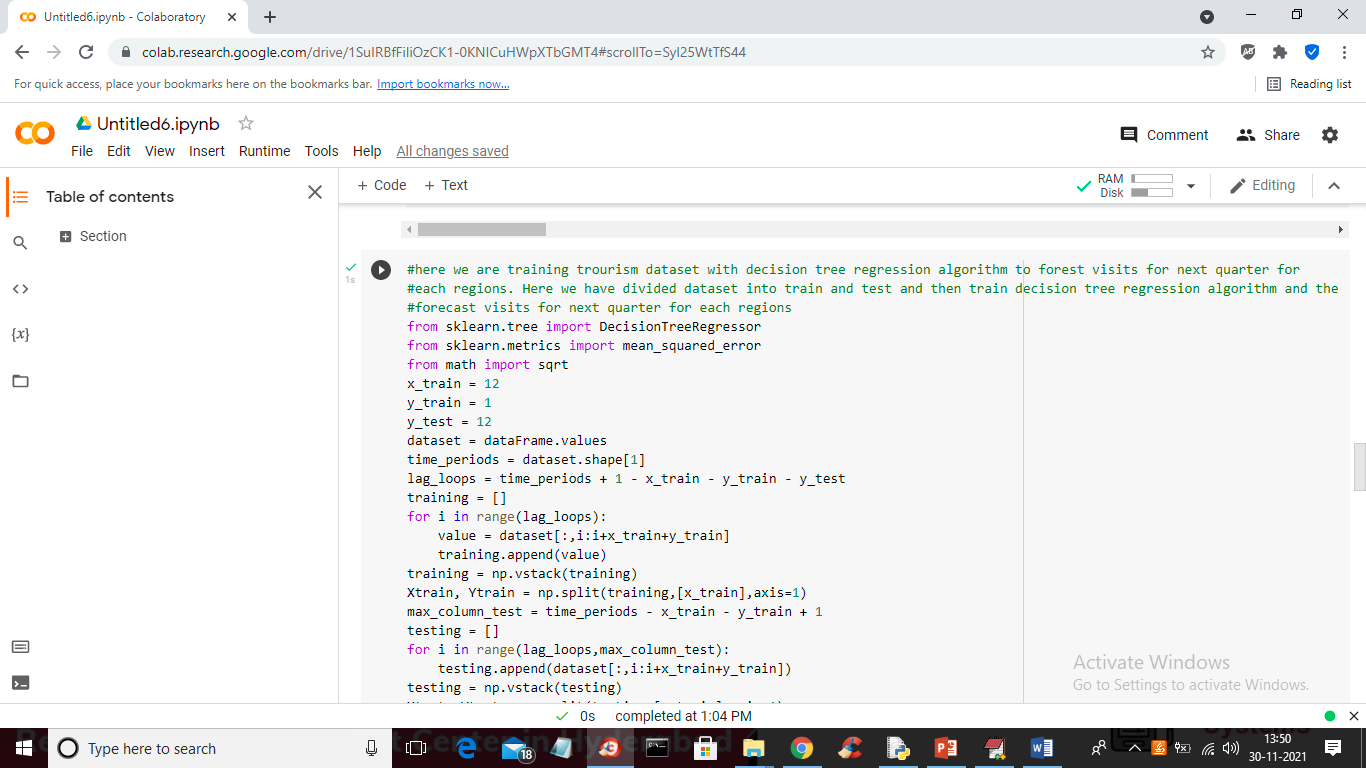




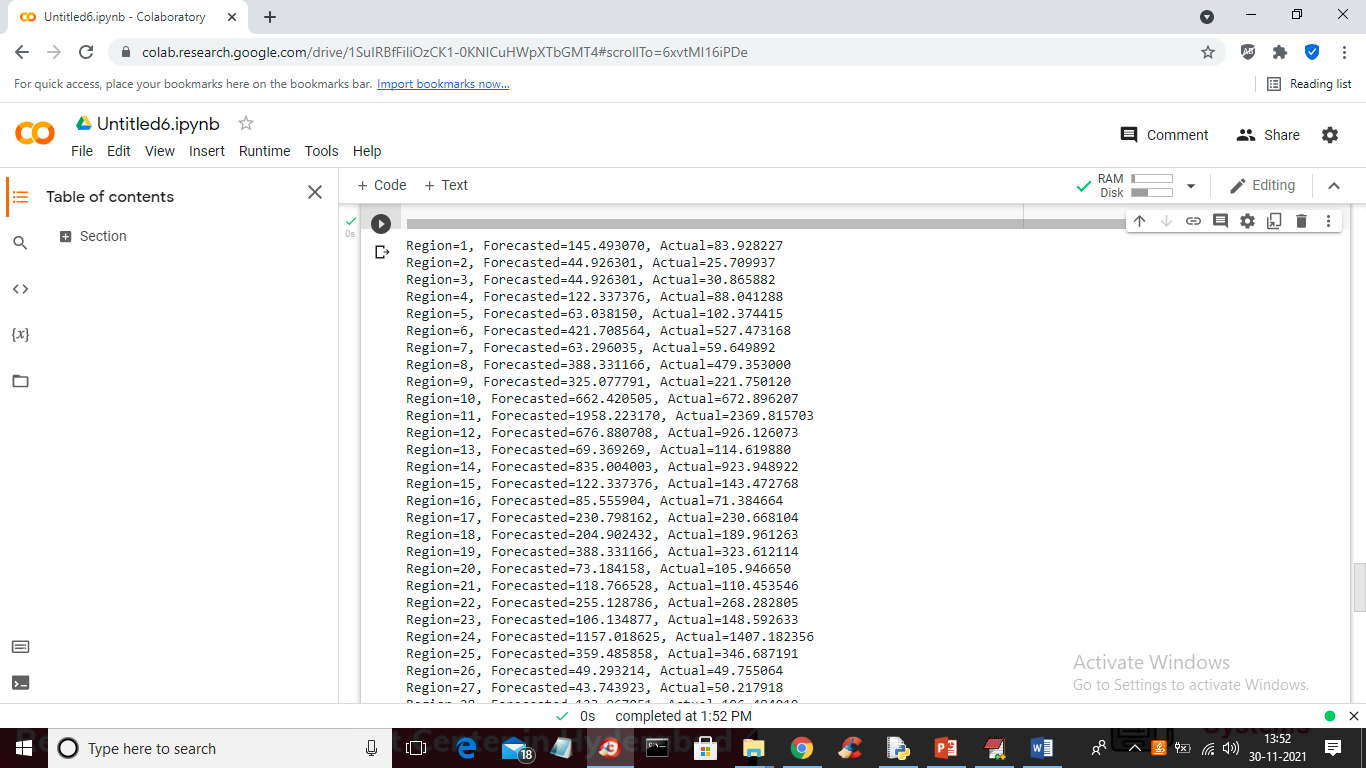
In above graph we plotting graphs for year wise tourist visits and by seeing above graph tourist manager can know which year more tourists visited



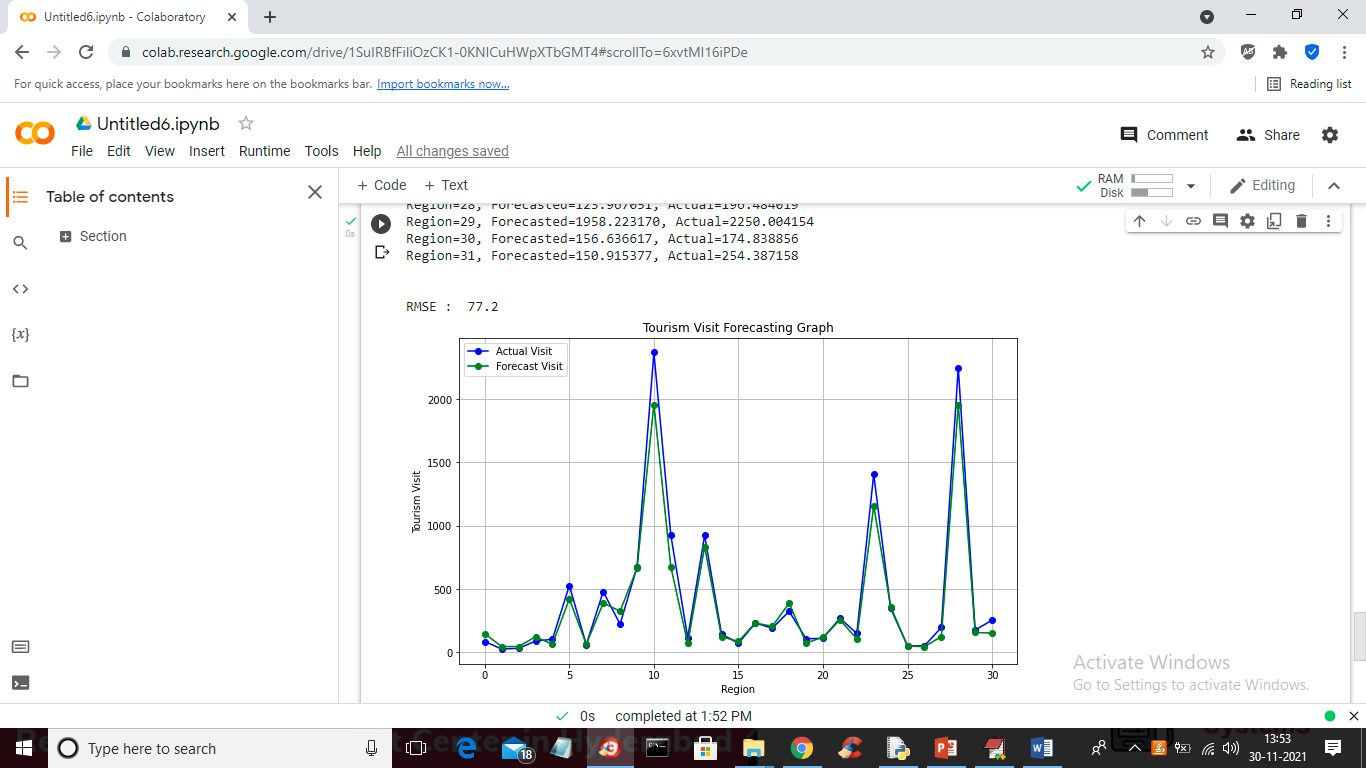
In above graph we are displaying output where user can see in which region and in which year how many tourists visited. In above output first column contains region names and other columns contains year name with number of tourist visited. Above data analysis can provide proper view of tourists visits



In above screen we wrote decision tree algorithm to train dataset and then forecast tourist visits for next quarter for each region. Here we have divided dataset into train and test and we used training data to train decision tree and used test data to forecast tourist’s visits values



In above screen for test data for each regions we forecast tourist visits and then we are displaying actual test values also and we can see there is not much difference between actual and forecasted values and below is the forecast graph comparison



In above screen we got RMSE (root mean square error) value as 77% and in above graph x-axis represents region ID and y-axis represents number of tourist visits and in above graph blue line represents actual tourist visits and green line represents forecast visits. By seeing above graph tourism manager can understand how many tourist can be expected and based on that they will get prepared for better hospitality.

Risk: if such analysis not available then tourist event peoples preparation will not be well which leads to customer dissatisfaction and may lose business for country and themselves