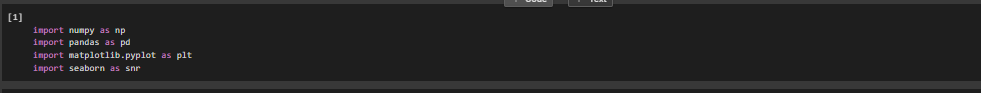
Project on Online shopping intention prediction using Machine Learning

* **Aim:-**To create a Data science Project, where we will be predicting the Online shopping intention. our objective here is to build a Machine Learning model that can help in predicting whether a customer will purchase or not, Prediction Online shopping intention with help of :-

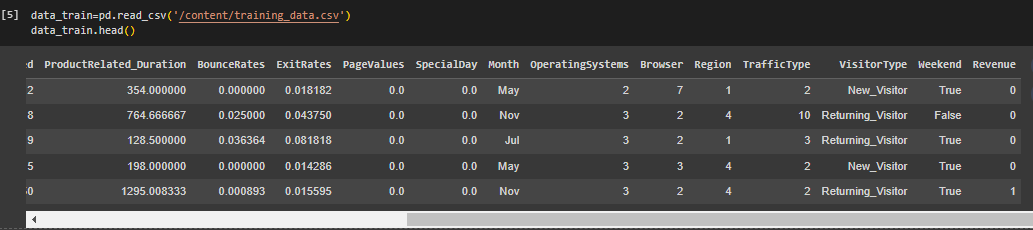
Special Day, Bounce Rate, Administrative.

Steps to be taken in the project is sub-divided into the following sections. These are:

* Importing the libraries such as ‘numpy’, ‘pandas’,‘sklearn. model’ etc.
* Loading Dataset as a CSV file for training & testing the models.
* Splitting the data set into independent & dependent sets.
* Checking if still any null values or any other data types other than float and integers are present into the dataset or not.
* Importing the train\_test\_split model from sklearn.model for splitting data into train & test sets.
* Applying the different kinds of ML Algorithms .which gives Best accuracy of model.
* Also checking with new data set for predicting the values.
* Steps of creating ML model:-
* Importing numpy as np & pandas as pd for loading and reading the data-set & using matplotlib.pyplot and Seaborn for visualization of data.

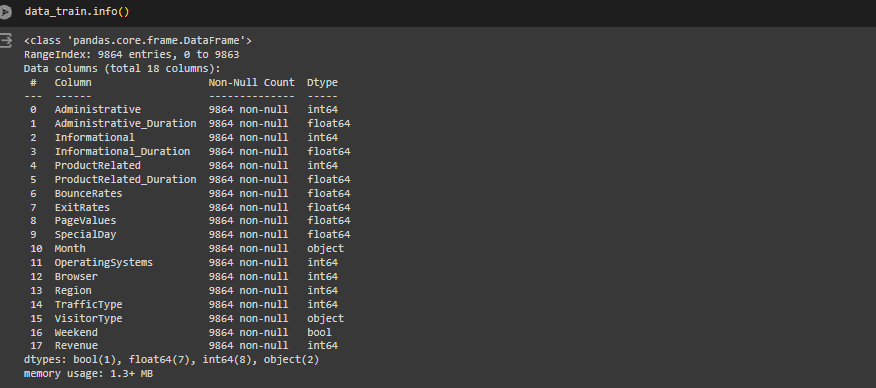


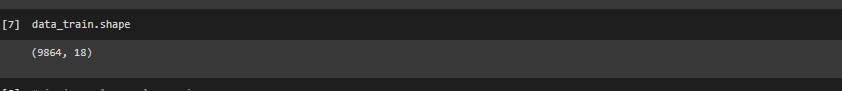
* Loading the csv-dataset in the variable name ‘data\_train’ Then viewing the data with data\_train.head()



* Checking the data such as number of columns, rows and type of data(float,integer) with help of data\_train.info()

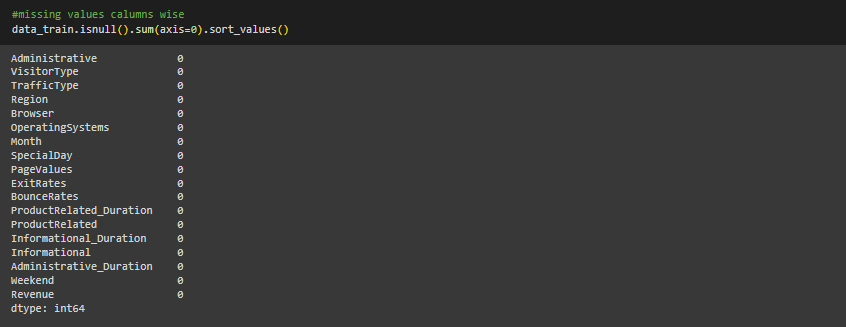
We observe that the above data have integer, object,bool and float.





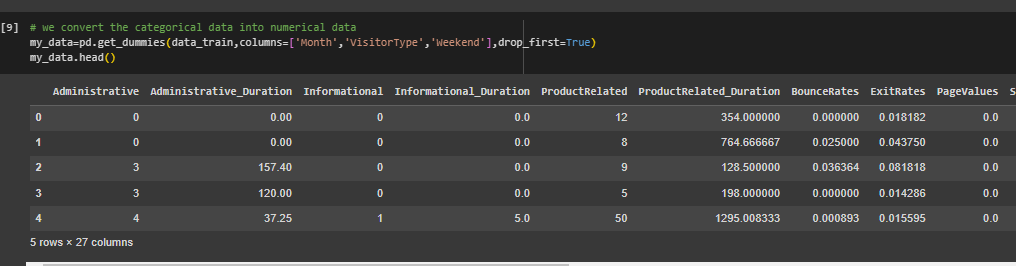
Train data have 9864 Rows and 18 columns

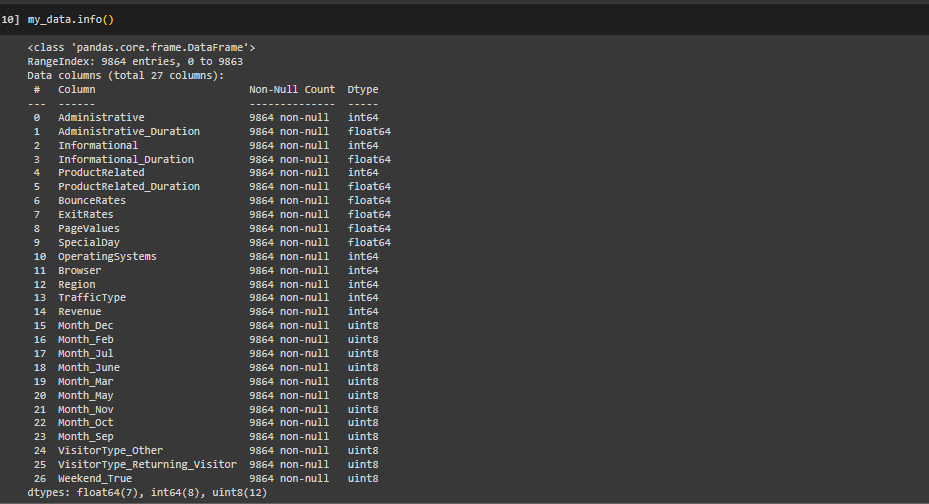
* Now checking data have Nan value or not.



We observe that the above data have not Nan value.

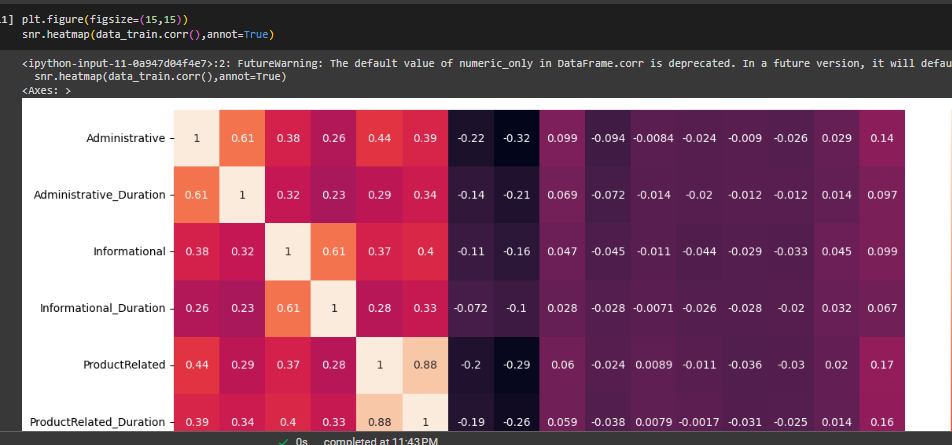
* Now,Main focus convert the categorical data into Numerical data with help of one hot encoding method.

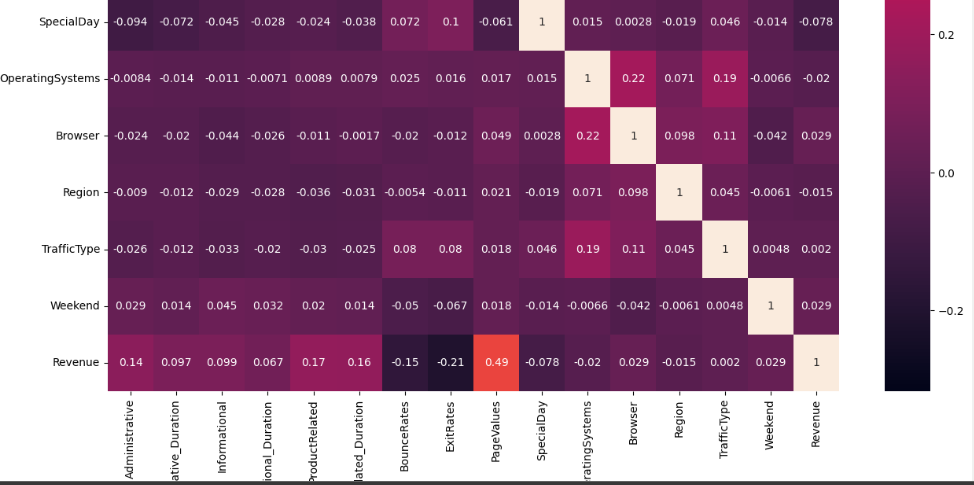




**Finally we observe the data are fully cleaned.**

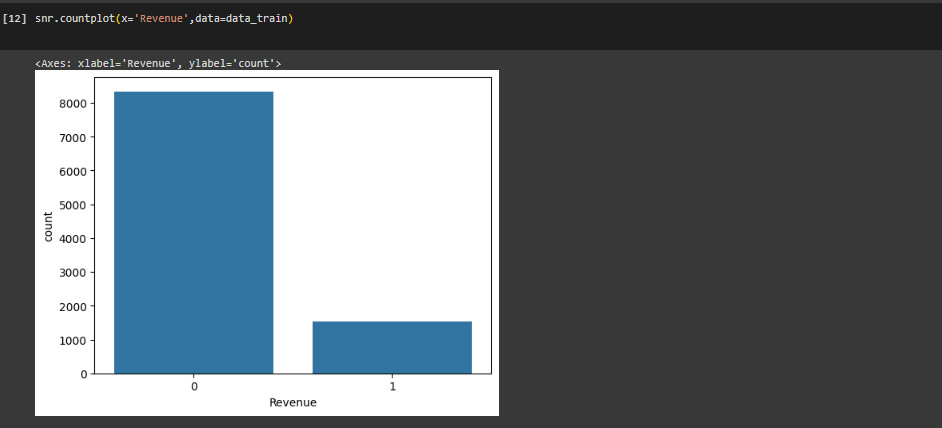
* Now we check the data dependency.





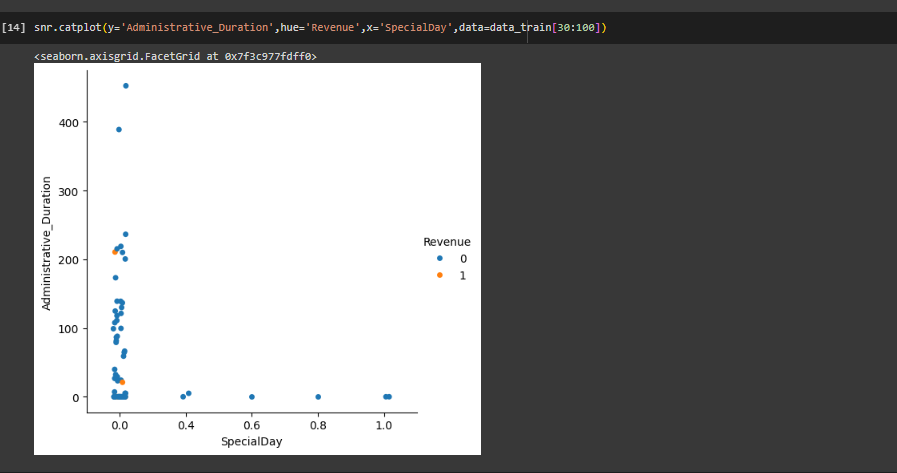
We see that data dependent each other.

* Visualizing the Online shopping intention like Special Day, Bounce Rate, Administrative.

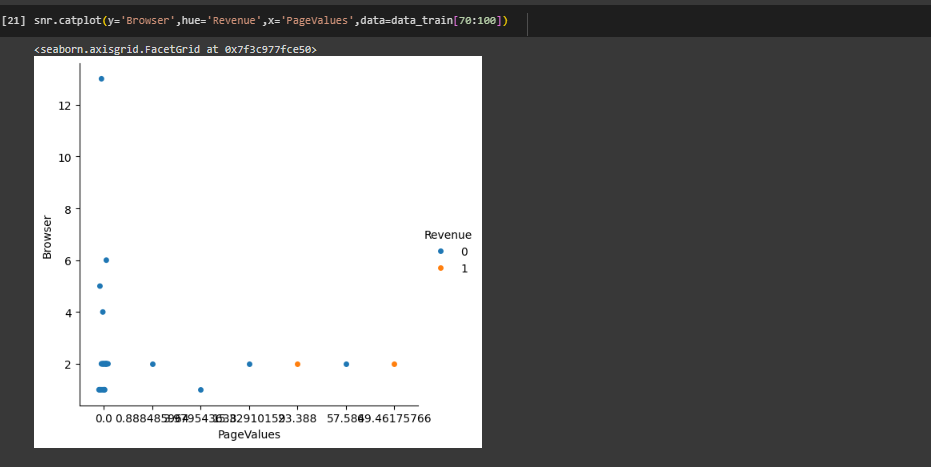


As per Visualizing the above graph, customer intention is less in Online shopping

# As per Visualizing the above graph, customer intension in weekend ..



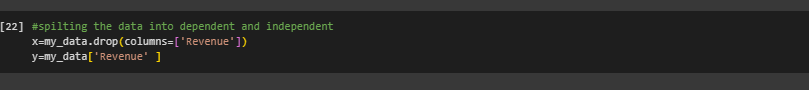
# As per Visualizing the above graph, people intension in special day..



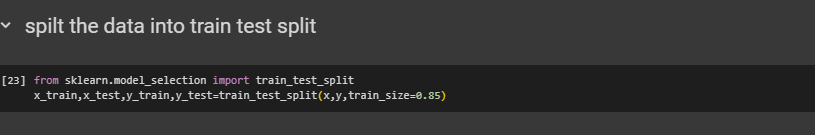
We observed that customer more search in values.

After visualization of data, we predict Online shopping intention using Machine Learning .

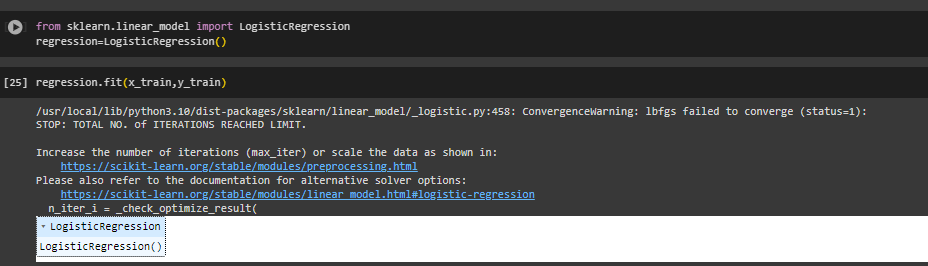
* Splitting the dataset into dependent(y) & independent(x) sets



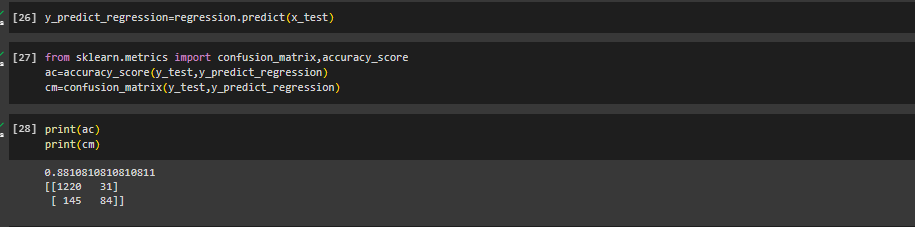
* Importing train\_test\_split from sklearn.model library for splitting the data into train and test sets. (we consider train dataset).



* Importing logistic regression from sklearn Libaray & then activating the Machine learning Model .Then used regression.fit() to training the model by providing train & test sets as x & y. And then predicted the trained model with help of MLM & the checked score as regression.score(x,y)



* Checking the accuracy with help of confusion Matrix.



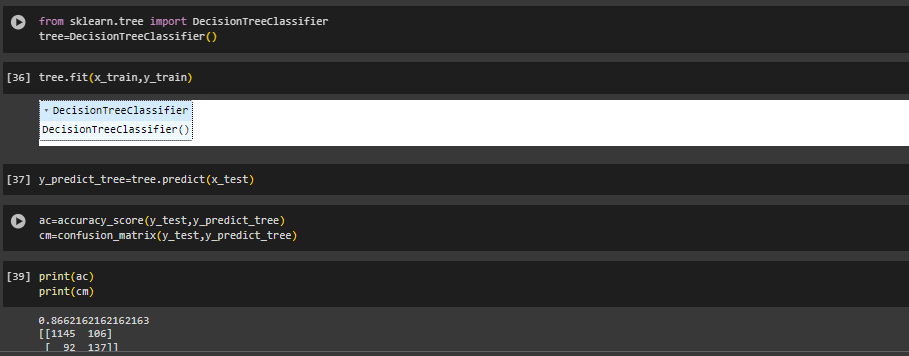
In the above model we can see that the accuracy obtained is 88%

* Now applying new algorithm Knn, then checked score.



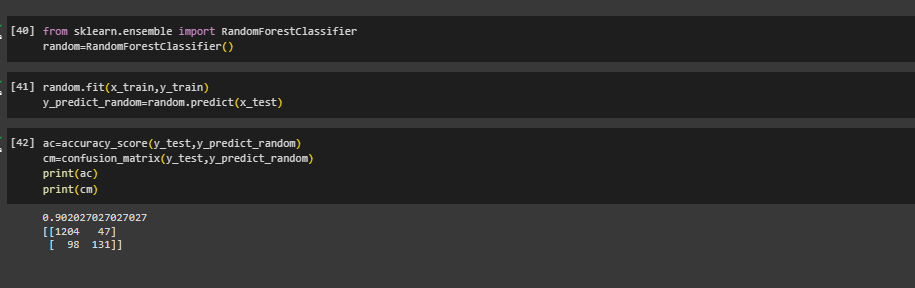
we can see that the accuracy obtained is 85%

* Now applying new algorithm DecisionTree , then checked score.



we can see that the accuracy obtained is 86%

* Now applying new algorithm RandomForest , then checked score.



we can see that the accuracy obtained with Random forest 90%

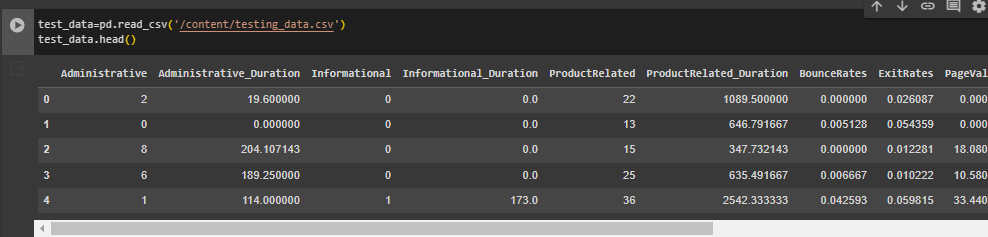
We see the accuracy is good but less than Decision Tree and Random forest algorithms.

* Now we compare all algorithms with accuracy

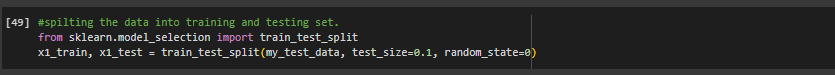
|  |  |
| --- | --- |
| Algorithms | accuracy |
| Logistic regression | 88% |
| KNN | 85% |
| Random Forest classifier | 90% |
| Decision Tree classifier | 86% |

Random Forest algorithms is better than KNN , Decision Tree and Logistic regression.

* Now recalling the test data set.
* Loading the csv-dataset in the variable name ‘test\_data’ Then viewing the data with test\_data.head()

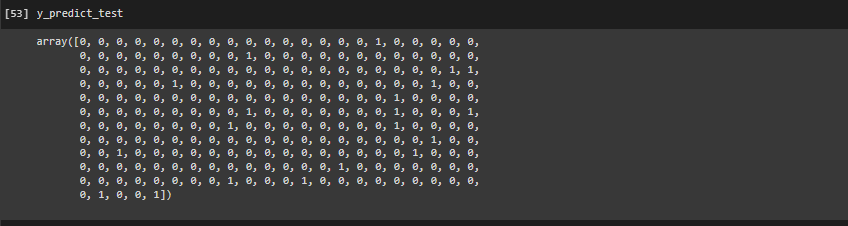


* Splitting into test & train sets as x1\_test & x1\_train. Then we find the Airline customer satisfaction using Machine Learning(Decision Tree classifier)



Applying Decision Tree classifier algorithms for predictions.





**Conclusion:- I**n this test data set we analysed the data we found the less customer intention in Online shopping intention

Thank you