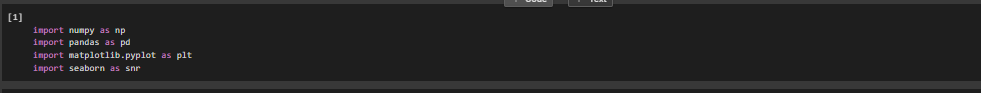
Project on Airline customer satisfaction prediction using Machine Learning

* **Aim:-**To create a Data science Project, where we will be predicting the Airline customer satisfaction. The change in airline passengers' behaviour following the pandemic crisis, travel restrictions, the ensuing economic crisis, market liberalization, high technology, and reorganization has resulted in airline services.

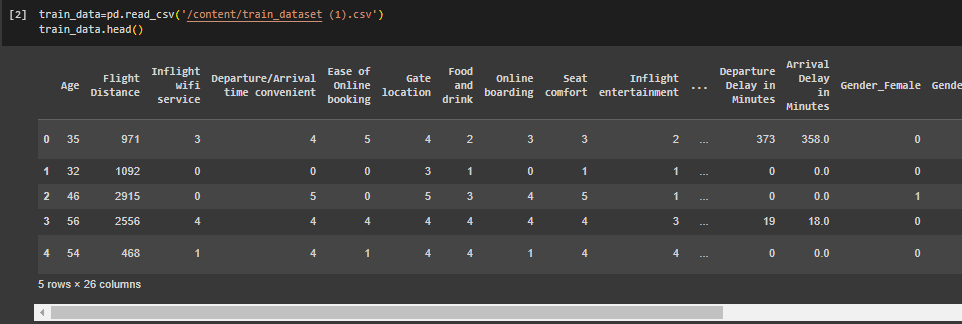
Prediction Airline customer satisfaction with help of :-

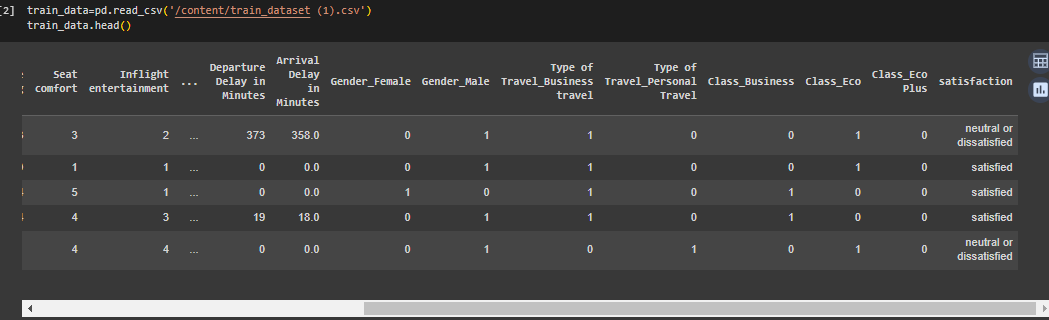
Age, Ease of Online booking, Food and drink, Departure/Arrival time convenient, Check-in service, On-board service etc. are some facility provided by Airline.

* 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.

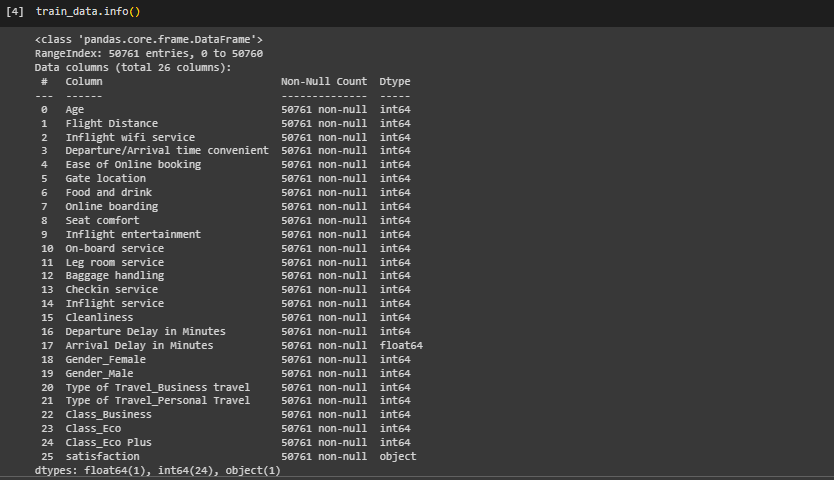


* Loading the csv-dataset in the variable name ‘train\_data’ Then viewing the data with train\_data.head()

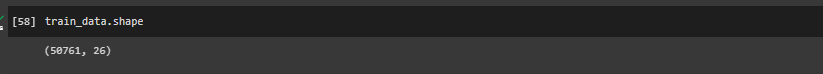




* Checking the data such as number of columns, rows and type of data(float,integer) with help of train ­\_data.info()

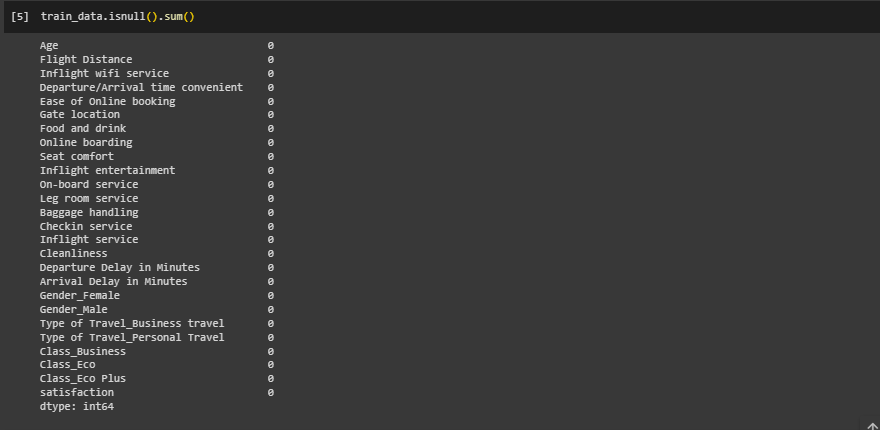


We observe that the above data have integer, object and float.



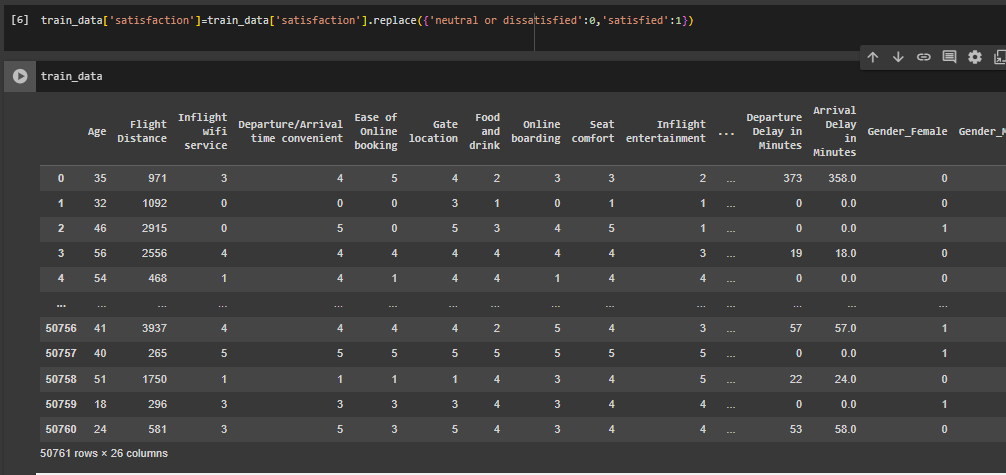
Train data have 50761 Rows and 26 columns

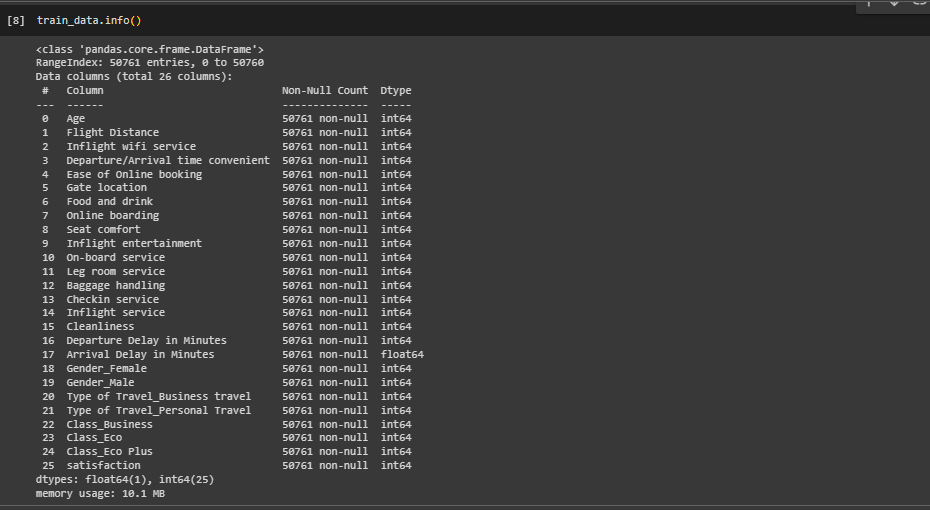
* Now checking data have Nan value or not.



We observe that the above data have not Nan value.

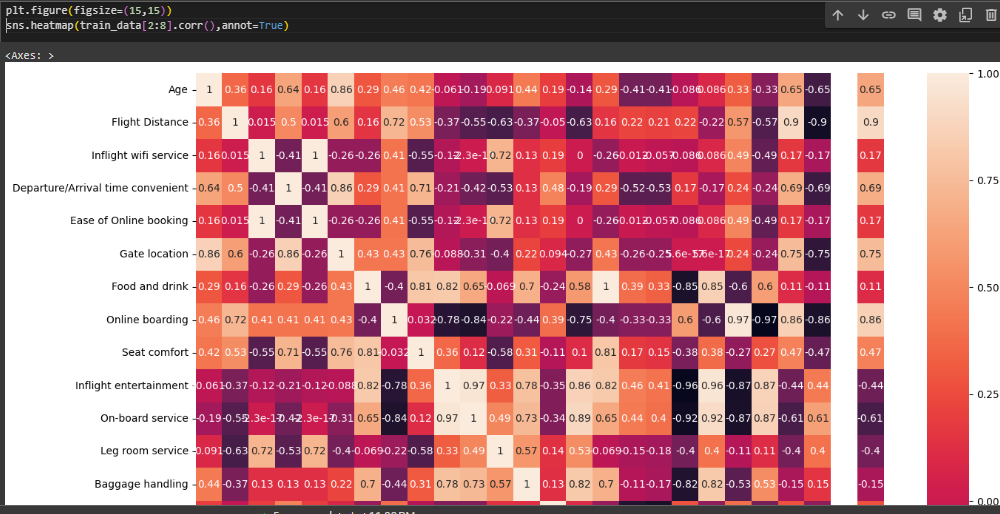
* Now,Main focus convert the categorical data into Numerical data with help of one hot encoding method.

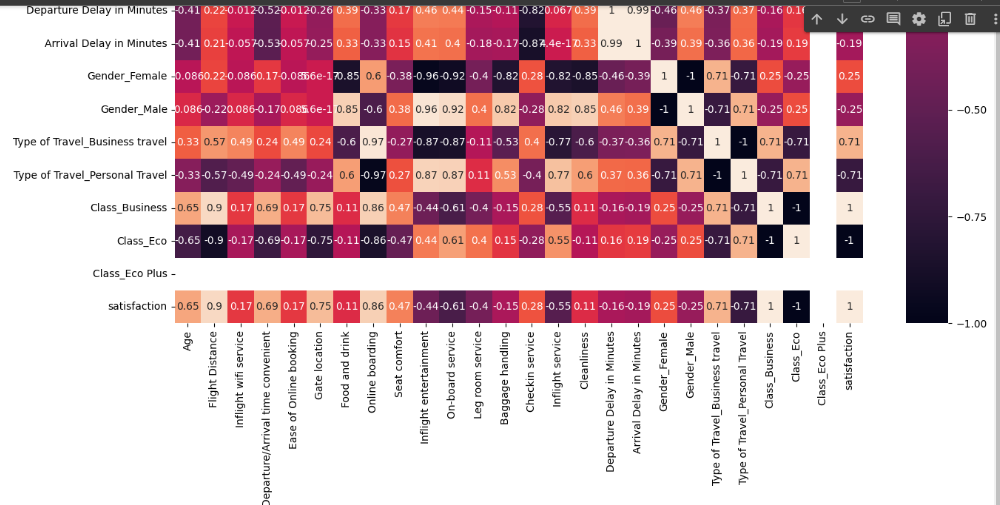




**Finally we observe the data are fully cleaned.**

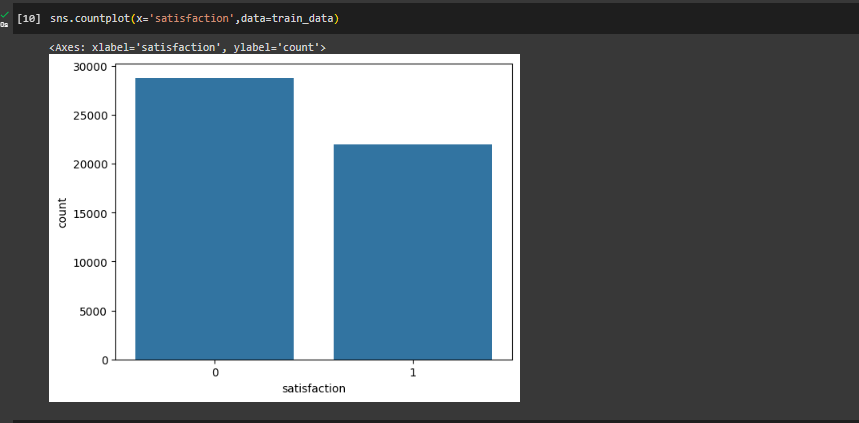
* Now we check the data dependency.



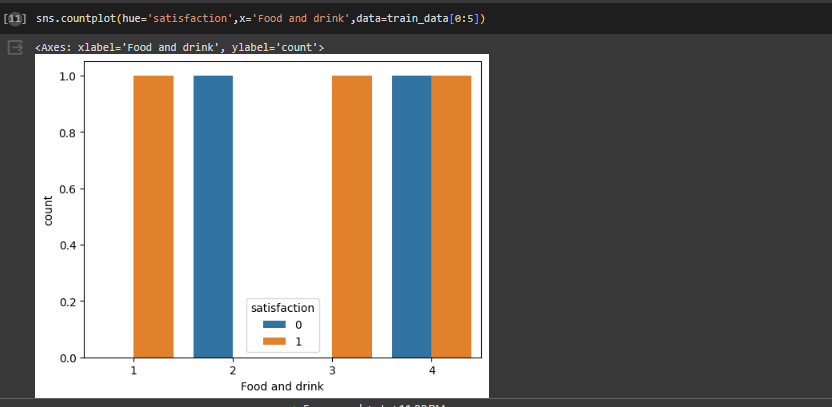


We see that data dependent each other.

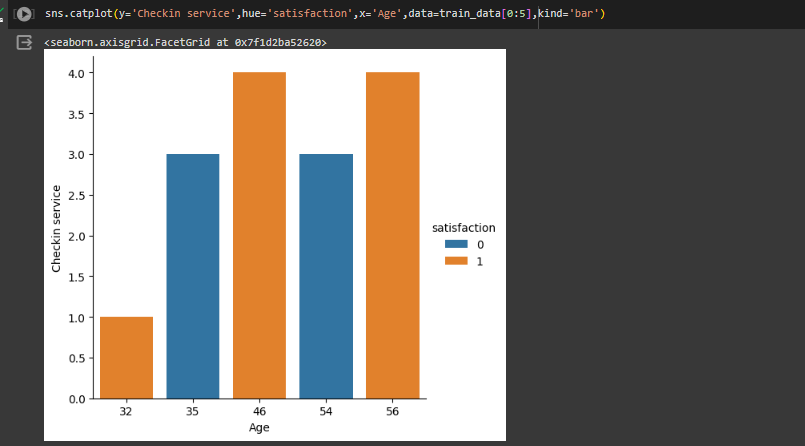
* Visualizing the Airline customer satisfaction like Age, Ease of Online booking, Food and drink, Departure/Arrival time convenient, Check-in service, On-board service etc



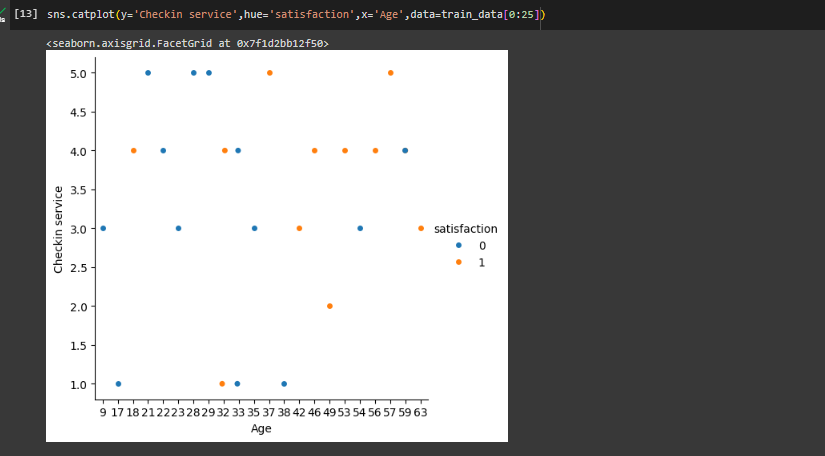
As per Visualizing the above graph, customer satisfaction is less than the neutral or dissatisfaction.



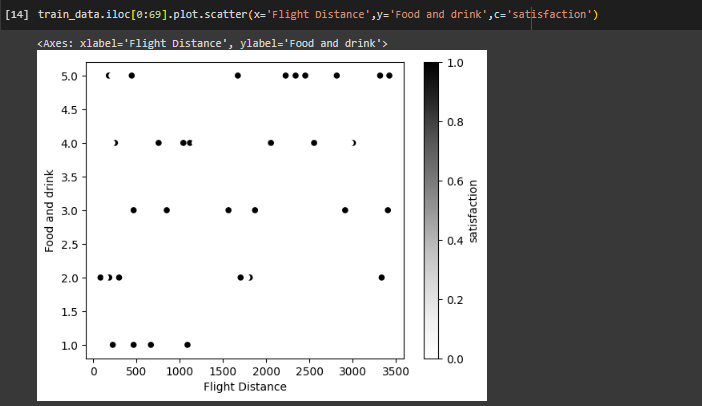
# As per Visualizing the above graph, customer satisfied with food and drink..



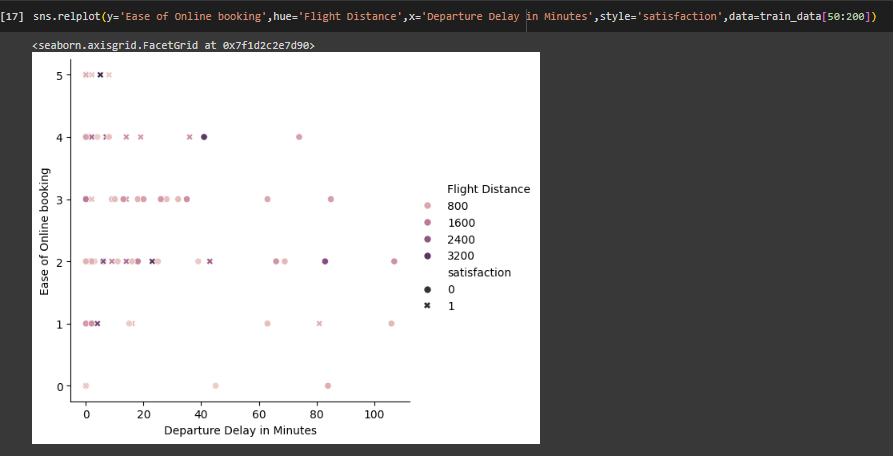
# As per Visualizing the above graph, people satisfied is in age range 32, 46 and 56 in checking service.



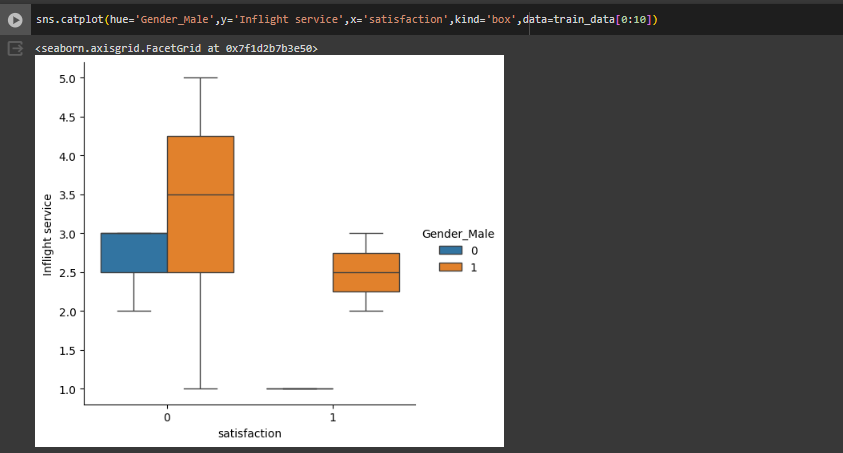
We observed that customer have age above 30 is more satisfied in with best checking service rate



We observe the above graph, customer satisfied in more distance traveling with good food and drink..



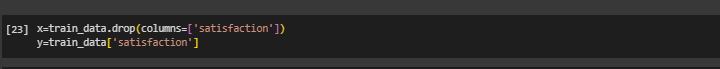
We observe the above graph, customer is satisfy with less departure delay with short distance travel and average ease of online booking service.



We observe the above graph, Male customer is more interest than female for traveling in air service.

After visualization of data, we predict Airline customer satisfaction using Machine Learning .

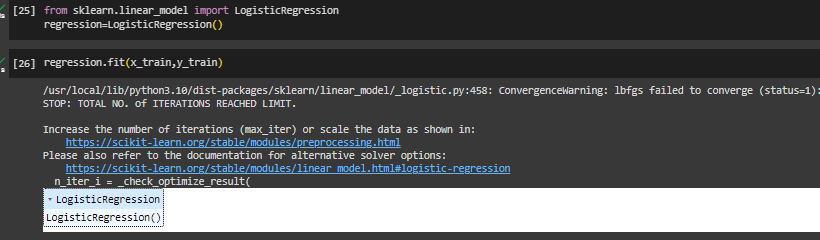
* Splitting the dataset into dependent(y) & independent(x) sets



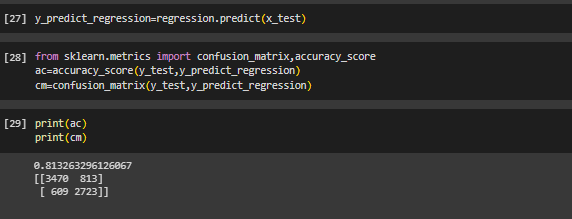
* Importing train\_test\_split from sklearn.model library for splitting the data into train and test sets. (we consider train dataset).



* 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)

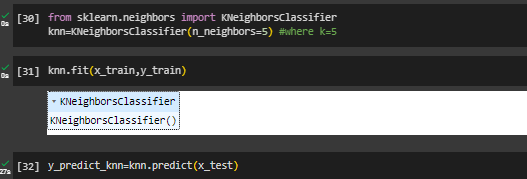


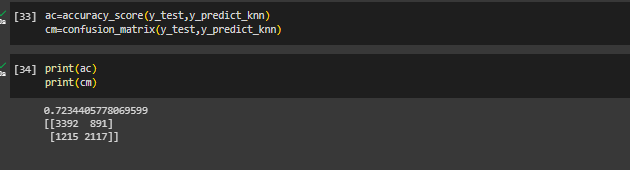
* Checking the accuracy with help of confusion Matrix.



In the above model we can see that the accuracy obtained is 81%

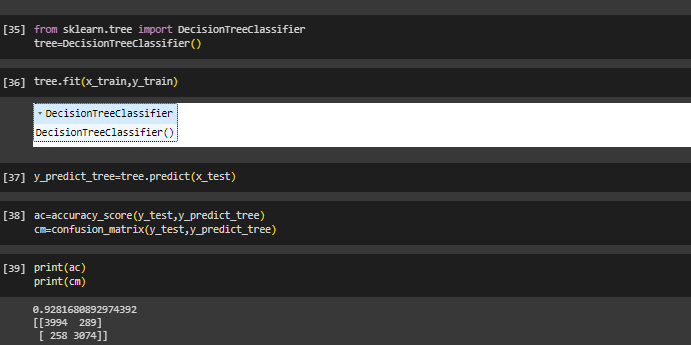
* Now applying new algorithm Knn, then checked score.





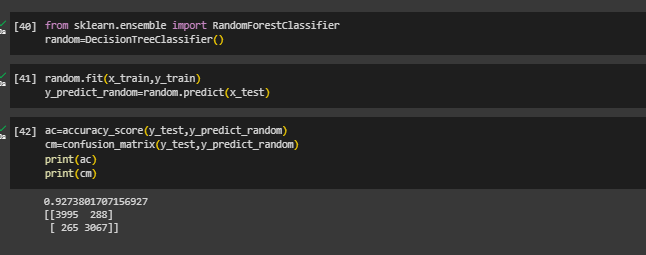
we can see that the accuracy obtained is 72%

* Now applying new algorithm DecisionTree , then checked score.



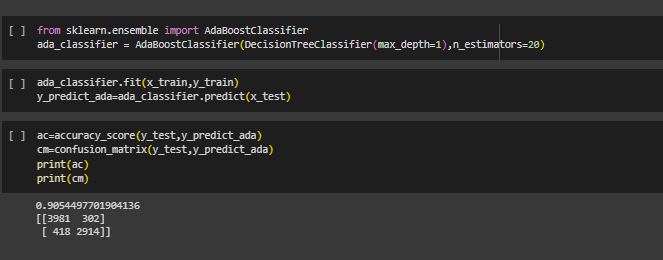
we can see that the accuracy obtained is 92%,is best for test data set.

* Now applying new algorithm RandomForest , then checked score.



we can see that the accuracy obtained with Random forest and decision tree both are same approx.92%

* We want to check with other algorithms(AdaBoost) for best accuracy



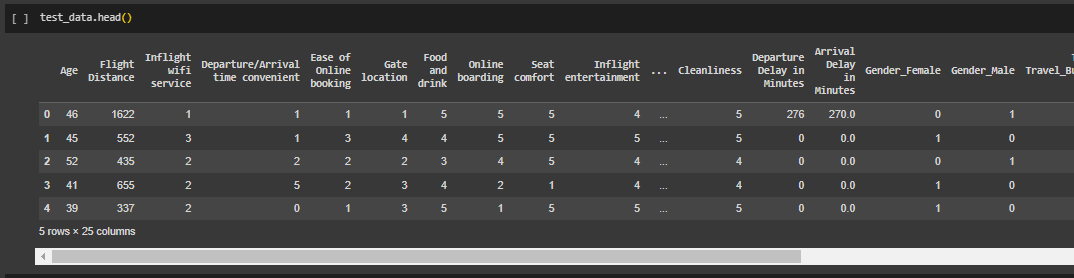
We see the accuracy is good but less than Decision Tree and Random forest algorithms.

* Now we compare all algorithms with accuracy

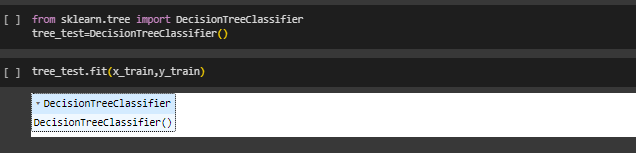
|  |  |
| --- | --- |
| Algorithms | accuracy |
| Logistic regression | 81% |
| KNN | 72% |
| Ada Boost | 90% |
| Random Forest classifier | 92.7% |
| Decision Tree classifier | 92.8% |

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

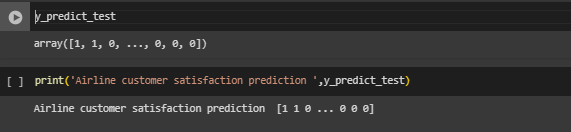
* Now recalling the test data set.
* Loading the csv-dataset in the variable name ‘test\_data’ Then viewing the data with test\_data.head()



* 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 maximum customer are neutral or dissatisfied with airline service.

Thank you