**Assignment 3**

Course Information

Course Title: Applications of Artificial Intelligence

Course Number EAI6010

Term and Year: Winter Part B:

Start and End Dates: Start and End Dates: March 2 – April 10, 2021

STUDENT NAME: SHIVANI ADSAR

NUID: 001399374

ASSIGNMENT: 3

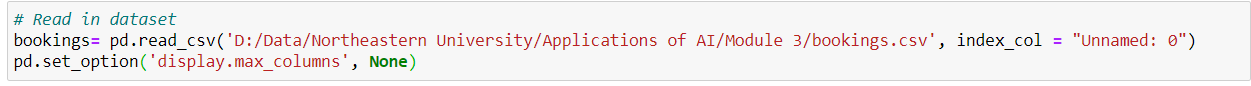
**Analysis**

Introduction:

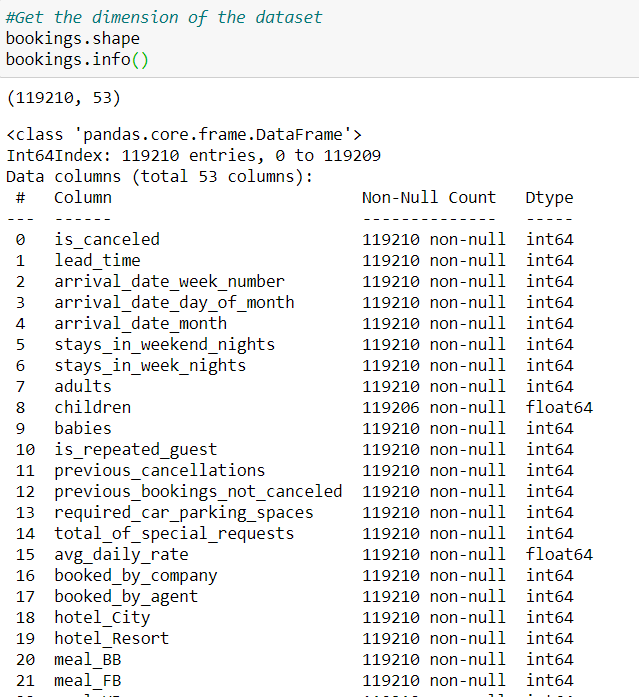
We have used the ‘bookings’ dataset to perform XG Boost on the data. This data contains information about bookings for a hotels in cities and resorts. Some information like, length of stay, number of people, date etc. have been included in the dataset.

XG Boost: This is a Machine Learning algorithm that uses decision tree base ensemble and gradient boosting for performing predictions.

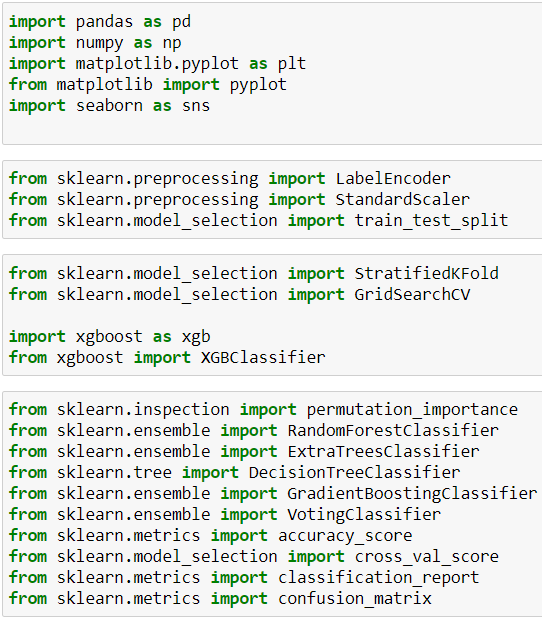
Read the dataset:



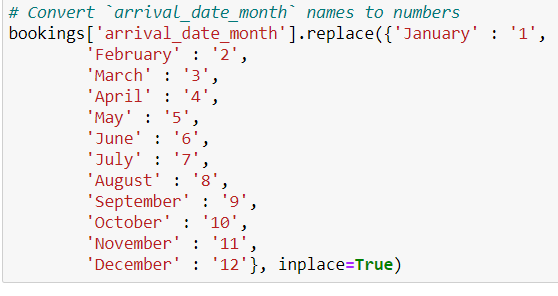
Getting to know the dimensions of the dataset:



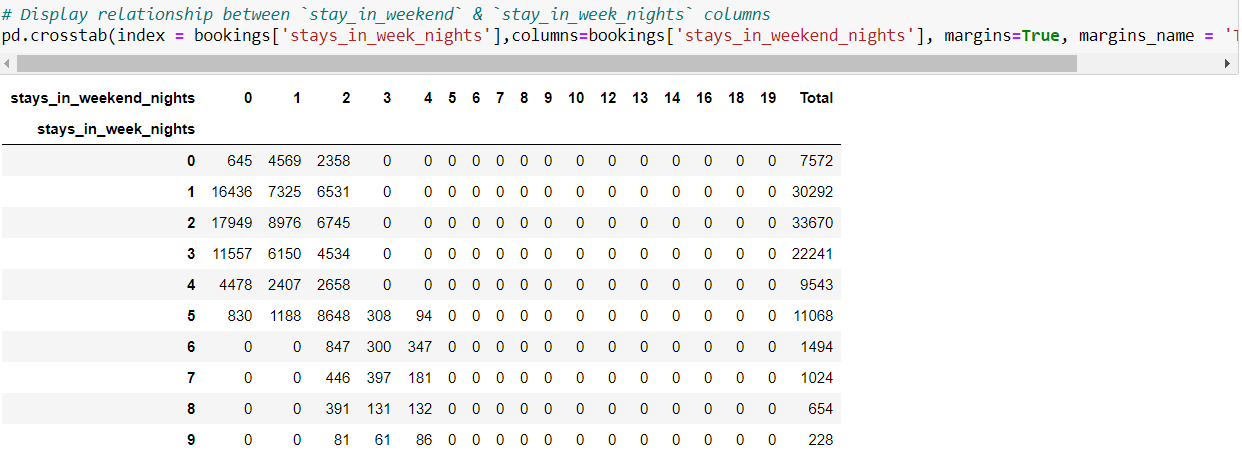
Importing the required libraries, ie. seaborn, sklearn, xgboost and matplotlib



Converting the ‘arrival date month’ parameter into numbers



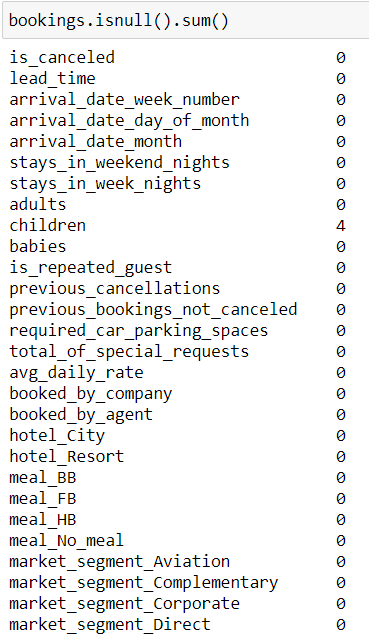
Displaying the relationship between ‘Stay in Weekend’ and ‘Stay in night’ columns.



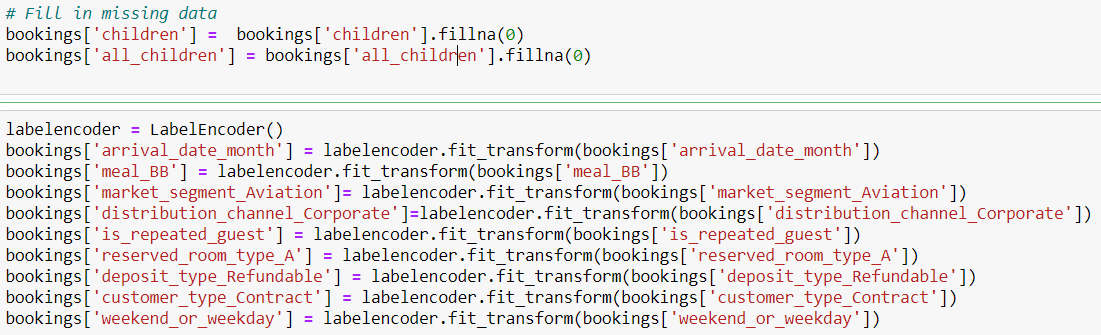
Feature Engineering:



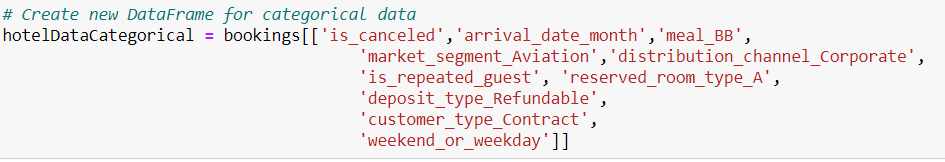
Elimination of null values:



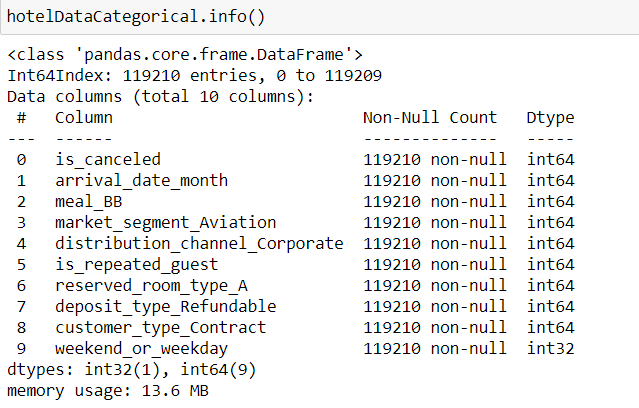
Filling the missing data:



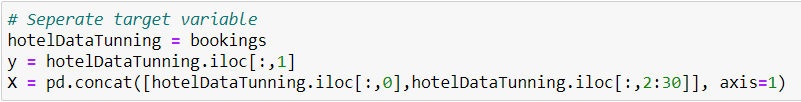
Creation of a new data frame for categorical data:



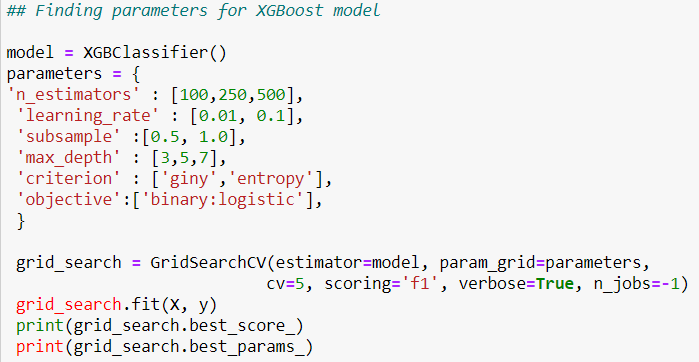
Analyzing the categorical data:



Separation of Categorical variable, ie. ‘is\_canceled’:

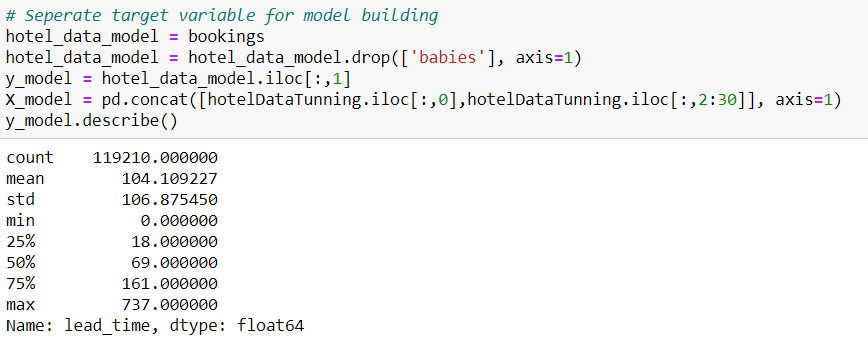


Parameters for XG Boost Model:



The ‘param\_grid’ contains a set of names in the form of keys and lists, which helps in searching over sequence of some parameter settings.

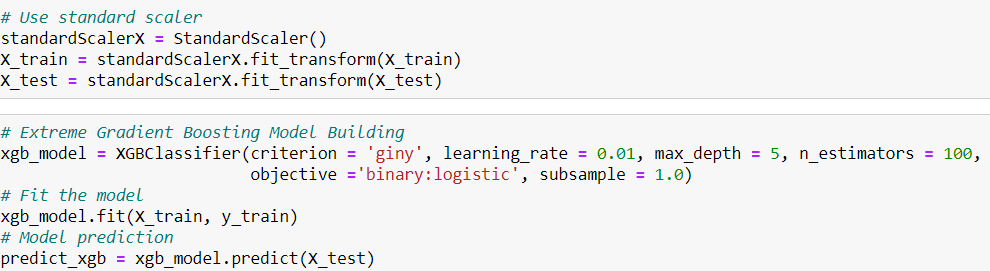
We have separated the target variable for model building:



Splitting the Train and Test data :



Now, we have used Hyperparameter Tuning in order to increase the accuracy of the XG Boost algorithm using standard scaler:



The Standard Scaler helps in data transformation in such a way that the data will have a mean of 0 and standard deviation of 1.

The accuracy of the XG Boost algorithm was predicted to be 81%. We have observed an increase in the accuracy from 76% to 81% after performing parameter tuning on the data.

