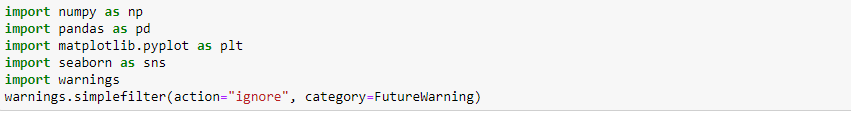
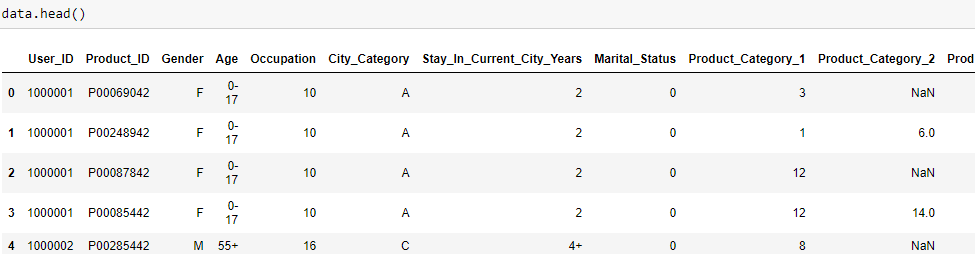
***Importing set of libraries***



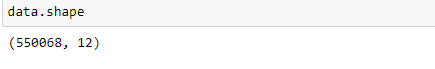
Reading the csv file which we have downloaded and placing the path below in order to display the dataset.



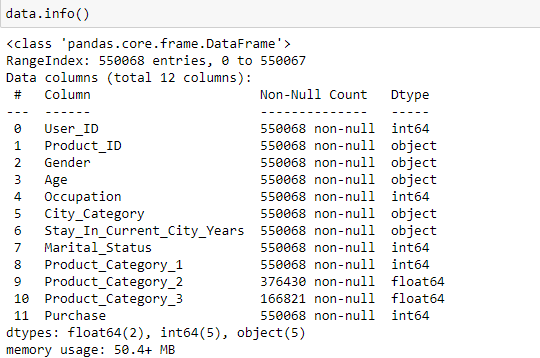
head() displays the first 5 rows of the dataset



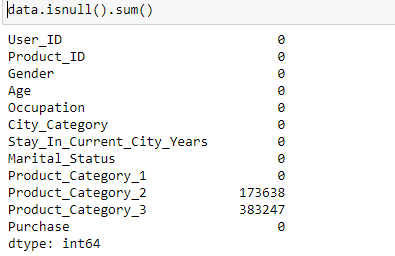
***shape attribute in pandas is used to get the shape of a DataFrame***



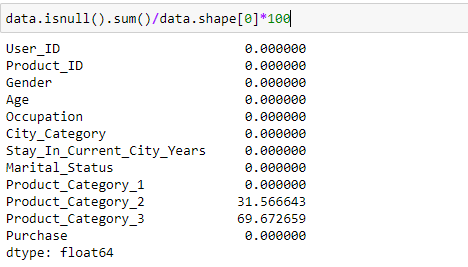
***info() is used to print the information of a DataFrame***



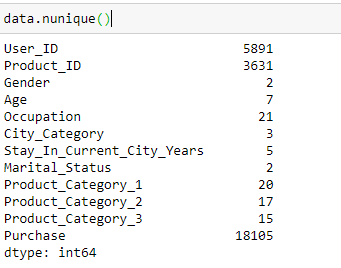
Checking whether if we have any null values in the dataset



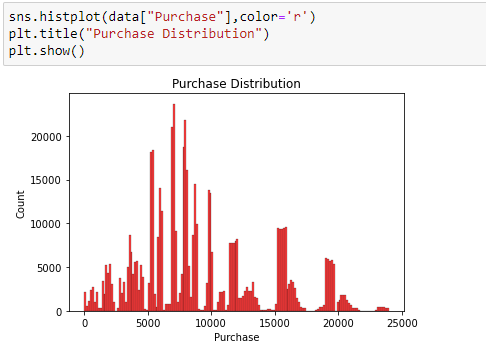
Checking null values in percentage for the dataset



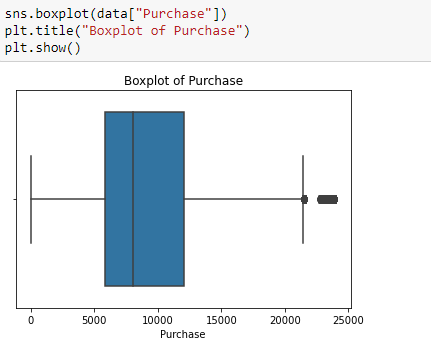
***Checking unique values in the elements***



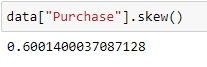
***Histogram (histplot()) is used to display the distribution of the data and shows the values of your data in series***



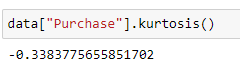
***(boxplot()) is used to display the groups of the numerical in the form quartiles***



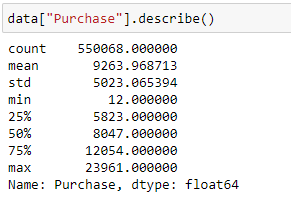
***Skewness of the data which is present in the given axis of the dataframe, it denotes an asymmetric distribution***



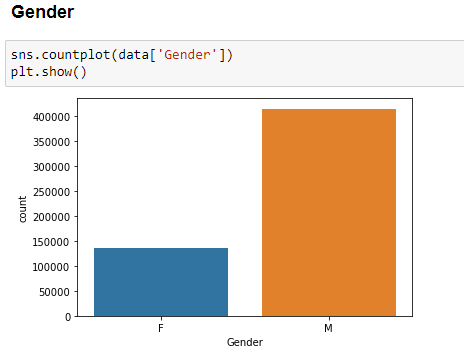
***Kurtosis is another method which is used when the data are heavy outliers or light – tailed to normal distribution***



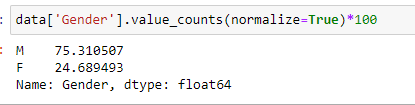
Describe() gives the entire details of the dataset



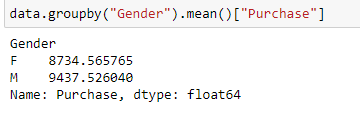
***Countplot is used to display the counts of the observations in each category i.e for female and male***



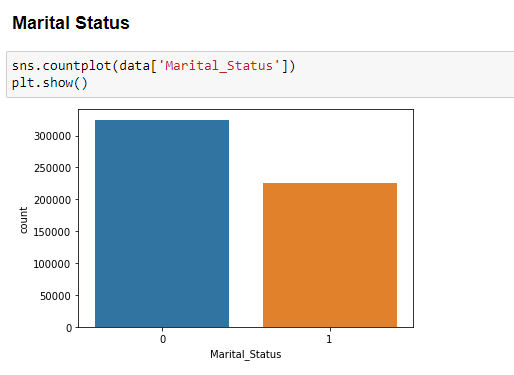
***Displaying the percentage of purchase which were made by both male and female***



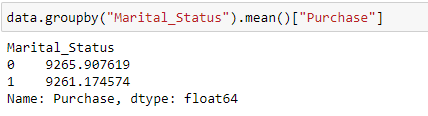
***Groupby is used for gender category so that it can separate the data into groups and mean() is used to get the average of the purchases.***



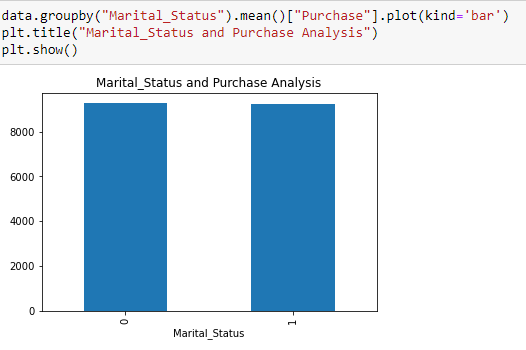
***Countplot is used to display the counts of the observations in each category i.e for married and unmarried***



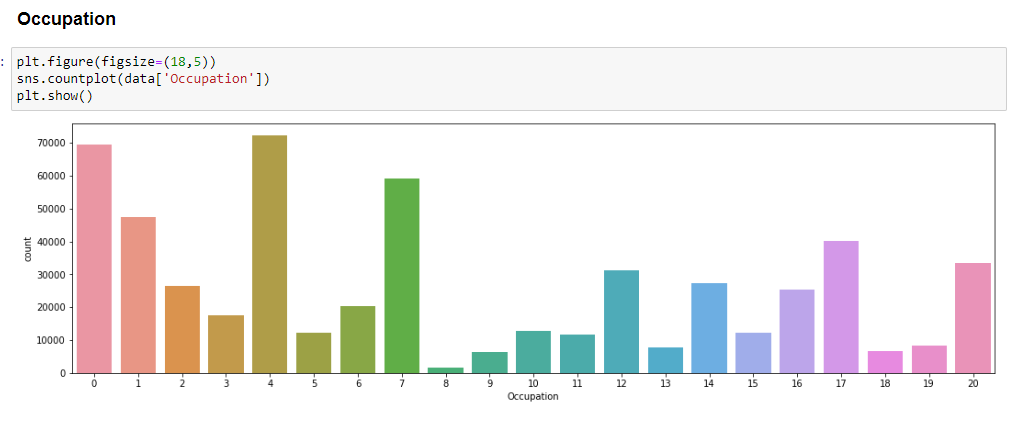
***Groupby is used for marital\_status so that it can separate the data into groups and mean() is used to get the average of the purchases made by unmarried and married people***



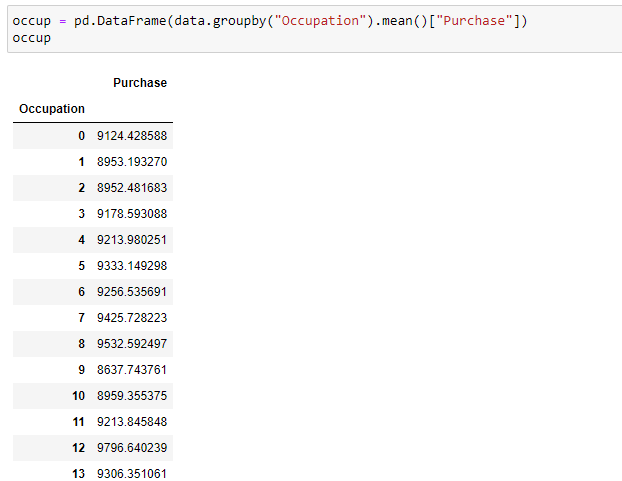
***Plotting a graph for the grouped data of marital\_status and mean of purchases***



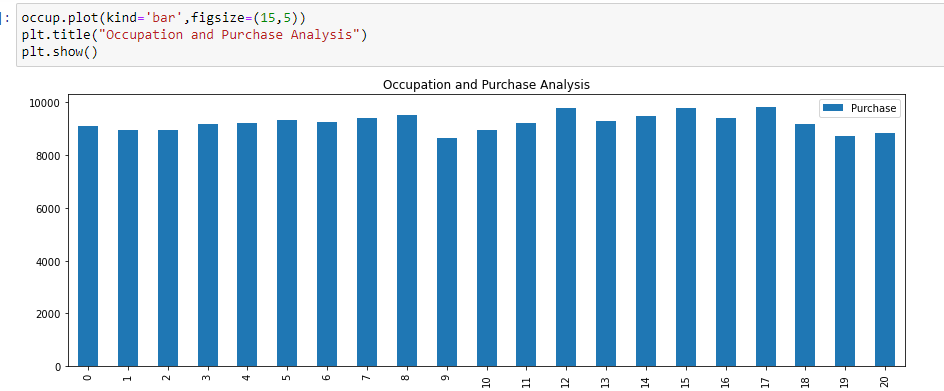
***Countplot is used to display the counts of the observations in each category i.e for occupation data***



***Groupby is used for occupation data so that it can separate the data into groups and mean() is used to get the average of the purchase of each occupation***



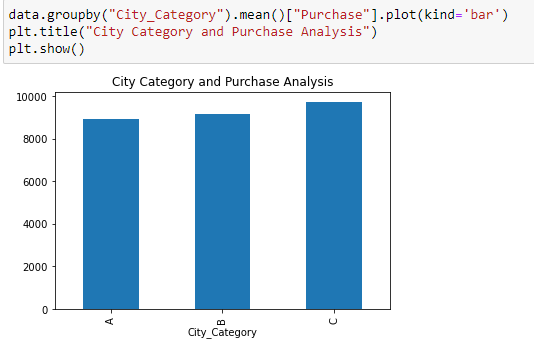
***Plotting a graph for the grouped data of occupation and mean of purchases***



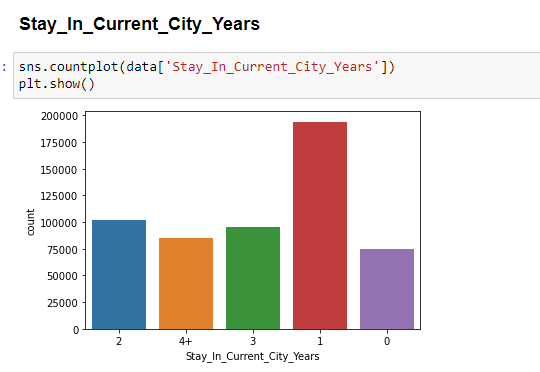
***Countplot is used to display the counts of the observations in each category i.e for City\_Category data***



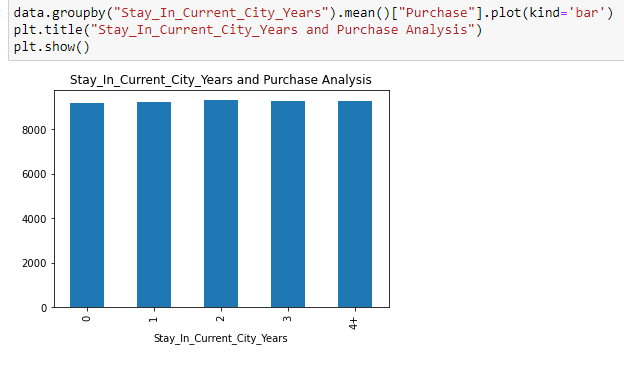
***Groupby is used for city\_category data so that it can separate the data into groups and mean() is used to get the average of the purchase***

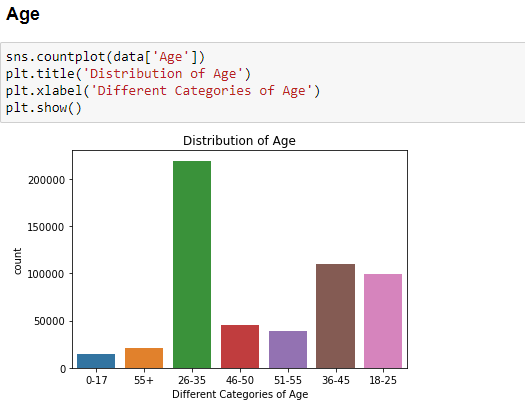


***Countplot is used to display the counts of the observations in each category i.e for stay\_in\_current\_city\_years.***

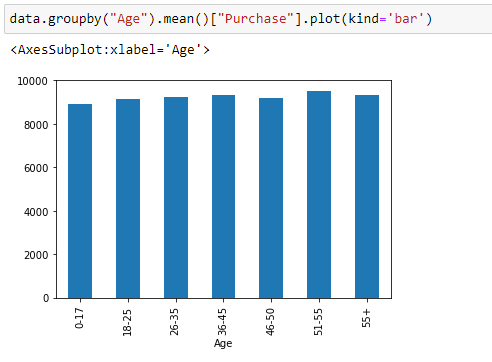


***Groupby is used for Stay\_In\_Current\_City\_Years data so that it can separate the data into groups and mean() is used to get the average of the purchase***

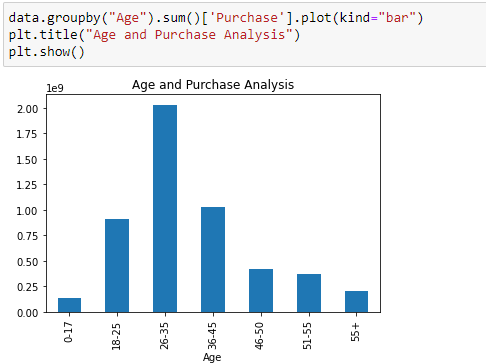


***Countplot is used to display the counts of the observations in each category i.e for diff age groups***

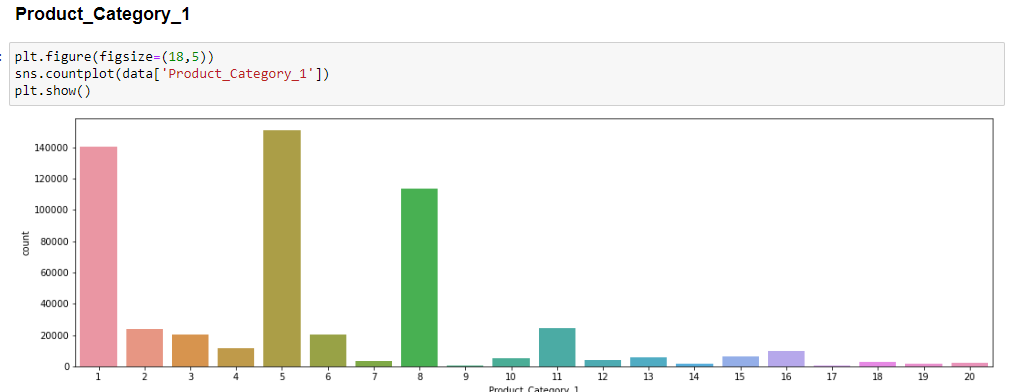
***Groupby is used for Age data so that it can separate the data into groups and mean() is used to get the average of the purchase***

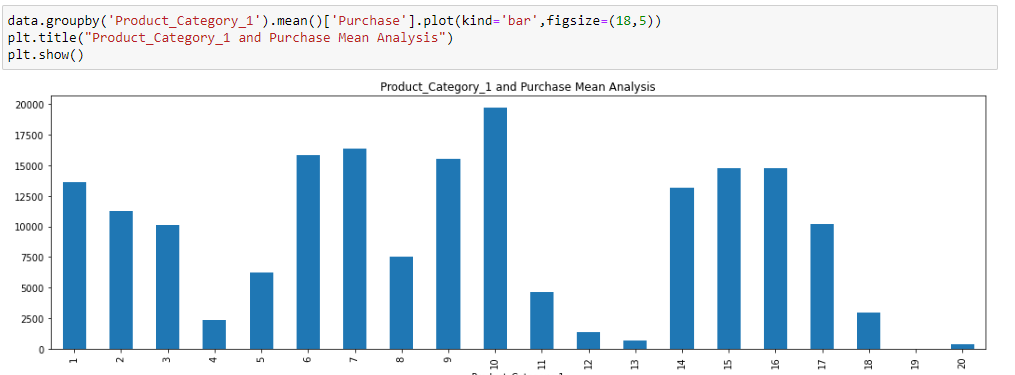
******

***Groupby is used for Age data so that it can separate the data into groups and sum() is used to get the sum of the purchase of the items***

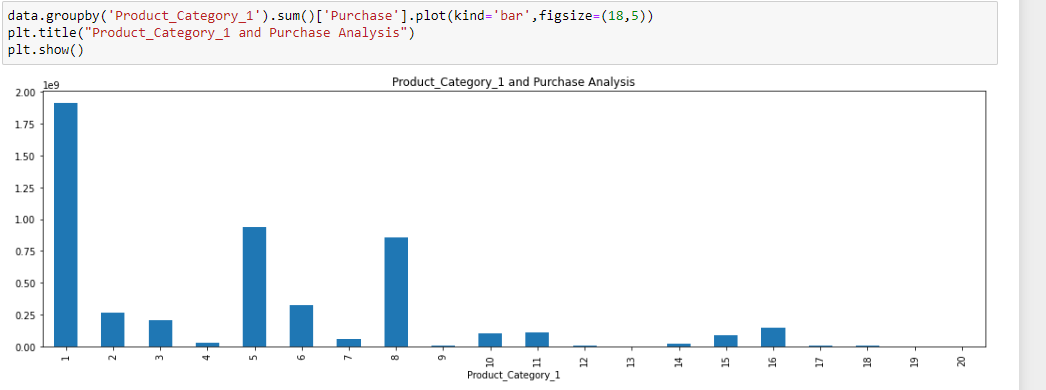
******

***Countplot is used to display the counts of the observations in each category i.e for product\_category\_1***

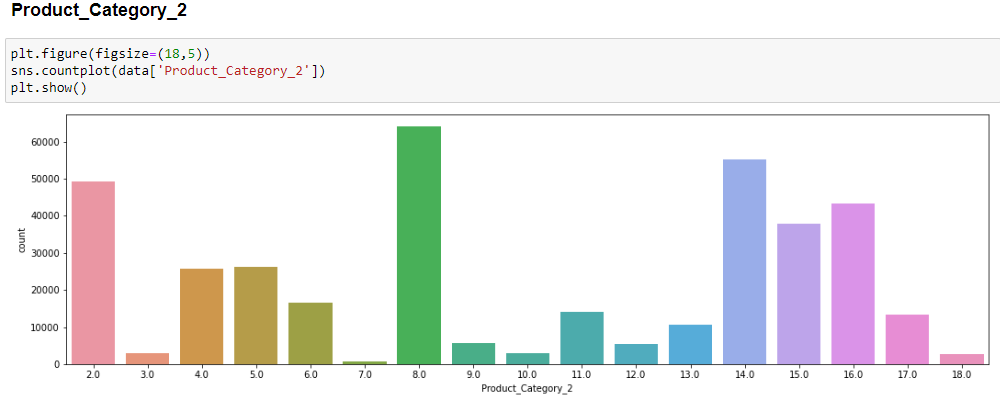
******

***Groupby is used for product\_category\_1 data so that it can separate the data into groups and mean() is used to get the average of the purchase***

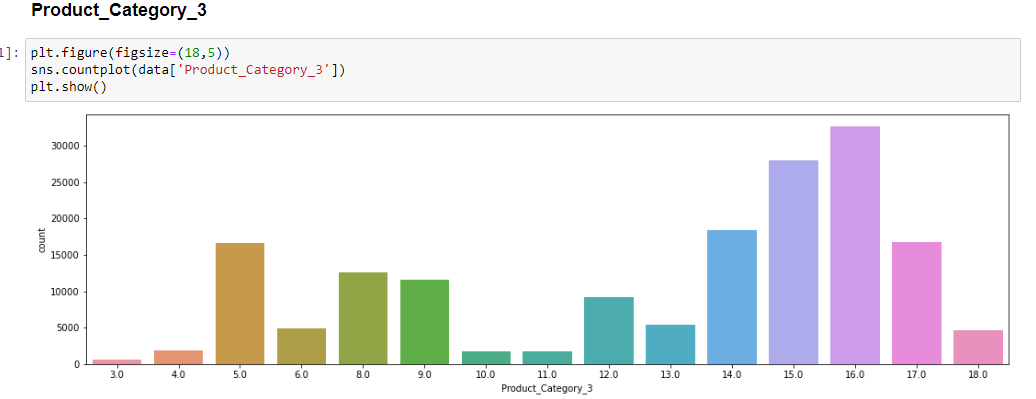
***Groupby is used for product\_category\_1 data so that it can separate the data into groups and mean() is used to get the sum of the purchase of the item***

******

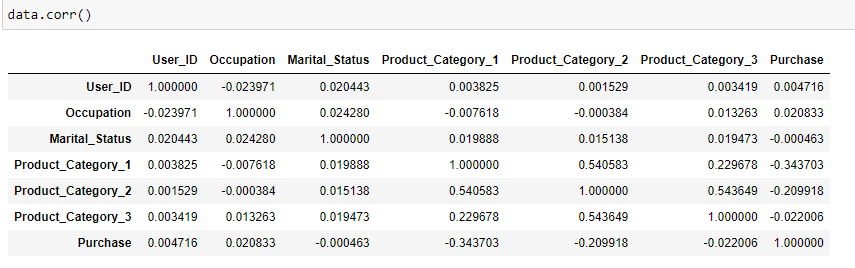
***Countplot is used to display the counts of the observations in each category i.e for product\_category\_2***

******

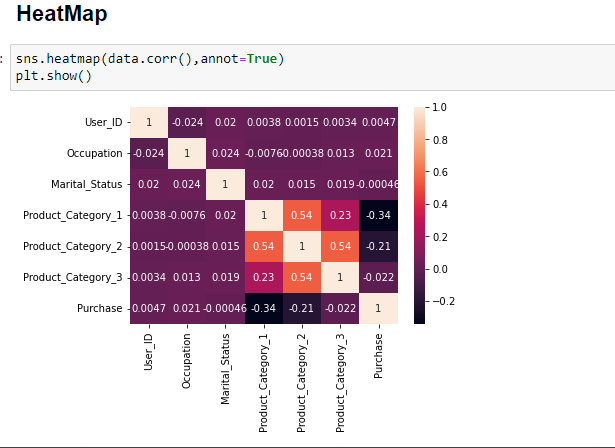
***Countplot is used to display the counts of the observations in each category i.e for product\_category\_3***

******

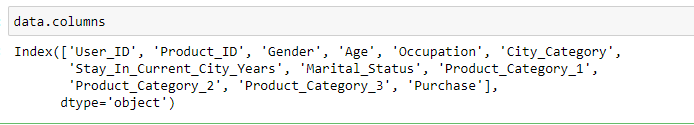
***corr() is used to find the correlation of each columns in the DataFrame. Nan values were automatically ignored.***

******

***HeatMap is used to display the correlation between the different variables in a matrix from ranging from -1 to +1. It displays the various shades of colour of each value.***

******

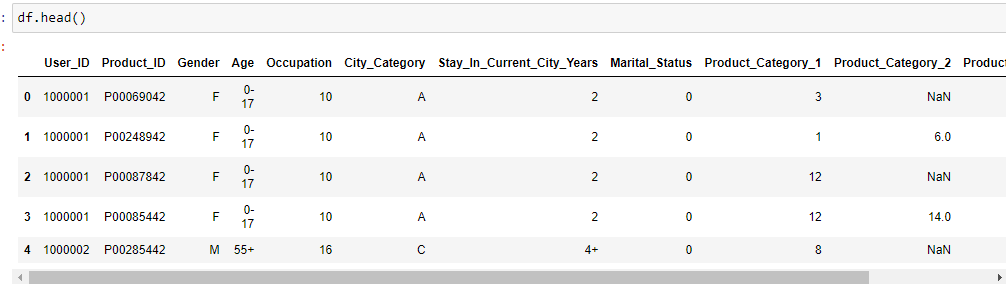
***Displaying the labels of each column below***

******

***copy() is used to create a copies of the list***

******

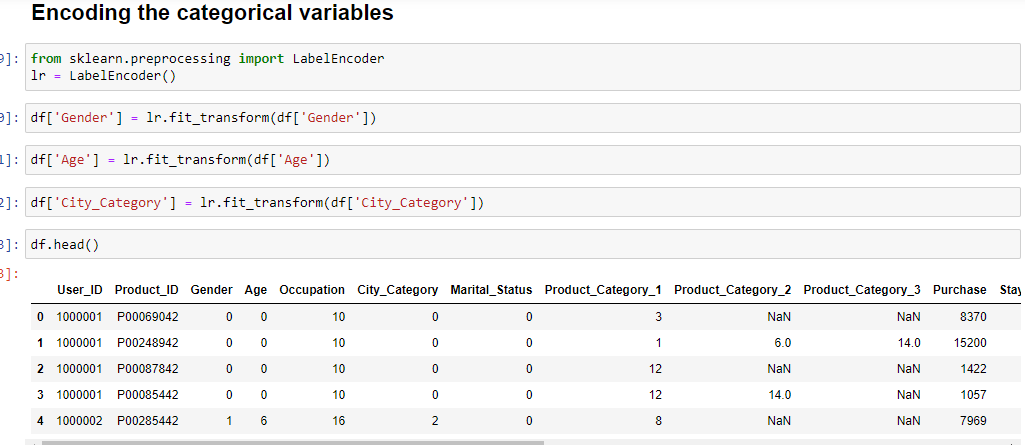
Head() displays the first 5 rows of the dataset

******

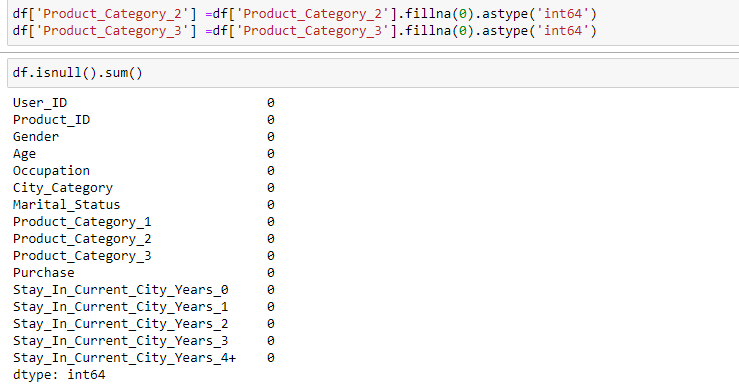
***Get\_dummies is used for manipulating the data***

******

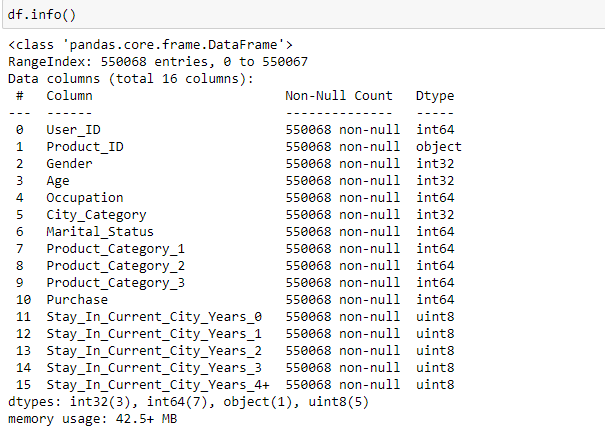
***Encoding the variables i.e converting set of labels into numeric form and displaying them using head() so that top five data is displayed***

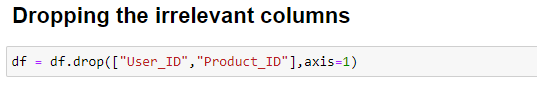
******

***fillna() is used to replace the values which are NULL with a specified value and isnull() is used to check whether any null values are present or not***

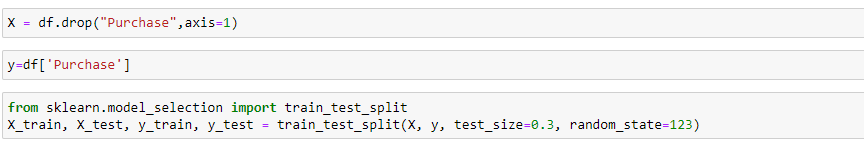
******

***info() is used to print the information of a DataFrame***

******

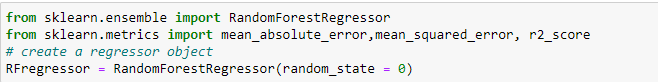
******

***Splitting data into dependent and independent variables i.e train and test***

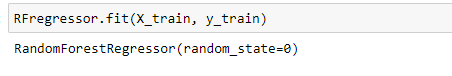
******

***A random forest regressor that fits a set of classification decision trees to different samples of data set and used to improve the accuracy of the predictions.***

***Creating and regressor object called RFregressor.***

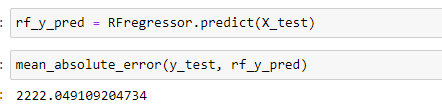
******

***Fit() is used to take the training data as set of parameters and perform set of calculations on the input data***

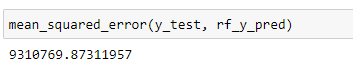
******

***Declaring a variable as rf\_y\_pred and using predict() is to predict the values of the data.***

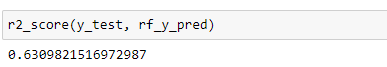
***MAE(mean\_absolute\_error) is used to calculate the summation of the absolute difference between the predicted value and the true value***

******

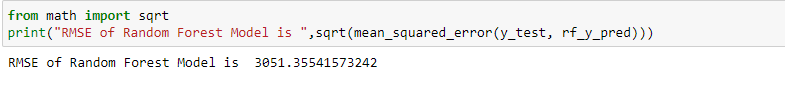
***MSE(mean\_sqaured\_error) is used to calculate the average of the square difference between the predicted value and the true value***

******

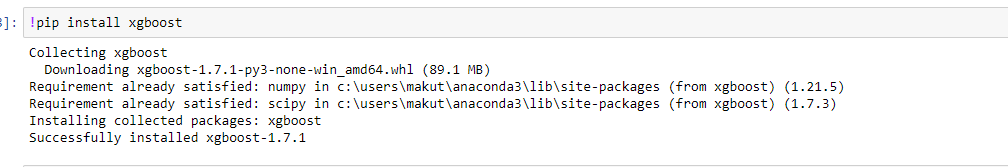
***R2\_score in regression means the proportion of variance of true variable which is explained by estimated variable***

******

***RMSE means square rooting the value of MSE***

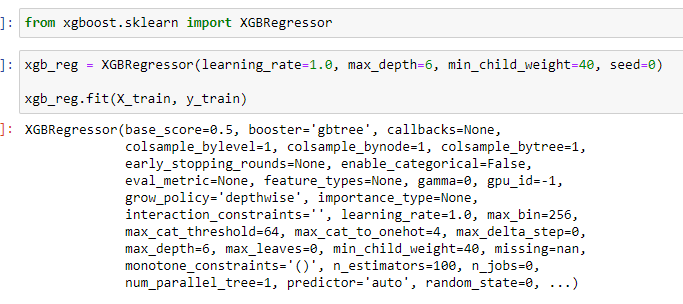
******

***Installing xgboost in order to perform xgboost regressor***

******

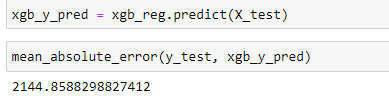
***XGBOOST means eXtreme Gradient Boosting which layout parallel tree boosting.***

***Declaring a variable and initialising the variables with some values and Fit() is used to take the training data as set of parameters and perform set of calculations on the input data***

******

***Declaring a variable as xgb\_y\_pred and using predict() is to predict the values of the data.***

***MAE(mean\_absolute\_error) is used to calculate the summation of the absolute difference between the predicted value and the true value***

******

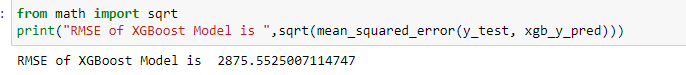
***MSE(mean\_sqaured\_error) is used to calculate the average of the square difference between the predicted value and the true value***

******

***R2\_score in regression means the proportion of variance of true variable which is explained by estimated variable***

******

***RMSE means square rooting the value of MSE***

******