Steps: EDA analysis of csv files:

Step1: Initially import matplotlib, numpy as well as pandas and seaborn library for EDA Analysis.

Step2: read the csv file using the Pandas functions here, and check the first and last elements into datasets. Check shape as well as duplicate elements present inside the datasets.

Step3: Check out null values or any null values are present there or not. Describe the datasets.

Draw pair plot and find out if there is any correlation between the variables presents or not.

Step4: Draw the histogram and boxplot between numerical columns to check out is there any outliers are presents or not. Use function describe function to view the perticuler outliers are there are not.

Step4: Drop the column Customer id as its not used in our analysis here in this case. Next we have to draw countplot based on Purchased products and count them.Now we have to make a crosstab and concatenate Products based on Ratings as well as products based on category.

Step5: Draw the bar chart between the two crosstab items .Find out the correlation between the each datasets objects and then plot a heatmap on them.

Step6: Draw the histogram between the crosstab and count the rating based on the products as well as category and group.

Steps: EDA analysis of csv files:

Step1: Import logistic regression model, classification report as well as metrics from Sklearn library.

Step2: Find out shape of the datasets. Convert the all categorical column into numerical way using dummy variable matrix concept here.

Step3: Choose dependent as well as independent variable X,Y from the dataset we have to fit the model using function classifier. Fit(X,Y).Find out classifier coefficient values here.

Step4: We have to find out the probability of the classifier model to see it will predict correctly or not using the function classifier.predict\_proba (X) here. Now we have to create a new column y\_pred and concatenating it with existing datasets.

Step5: Now we have to build our confusion matrix between actual as well as predicted value to check it will predict correctly or not. Also create a crosstab between the actual and as well as predicted values of y.

Step6: Finally check Accuracy of our model it will 74 percent here. Now Finally we have to find out the classification report based on selectivity,sensitivity,F1 Score value etc.

Step7:Finally draw a Roc Curve based on True and False Positive rate if its less than 50 percent then model is not predicting correctly here. And we also find out probability between the predicted values as well as create the Final Report of our Model.

Step8: Finally draw a crosstab between the group of people and the purchasing products and make a bar chart to see which group of people having higher purchasing cost G1 Will be the answer here. Also we see Varible2 is much more affected by the purchasing products here.

**Inference** between the models:

Now based on this datasets we have some inference that we have to draw between them those inferences are based on following-

* Here the accuracy of the model is 74 percent which means whatever the predicted value that we obtain it will correctly predicted the inference value. If it is less than fifty percent then it will not predicting it correctly there will be a possibility of falsely predicted the Purchased products based on groups as well as category here.
* Also G1 Group is effected by more no of purchased products here in this case.
* Var2 is also affected by the purchased products here in this case.