ASSIGNMENT

The report gives brief analysis on the given dataset of product, customer and transactions. Through our analysis we can have business insights and the prediction model we could use for regression and clustering purpose.

EXPLORATORY DATA ANALYSIS:

The given code file contains the code for data analysis which includes bar graphs, scatter plots, line graphs showing the max count and number of transactions in the graphical manner.

With the help of analysis we get that most selling categories were Books and Electronics with the count 25.

The box plot gives the detailed view explaining that there are no outliers in the data and also the median is 290.

The maximum price is generated by the Books around 295 rupees and it contains the range [225-345].

The line plot in the file indicates the variation between the Total Value and Category. The maximum total value is of the Books and the shaded region shows the pricing graph.

The count plot for transactions per month is given in figure. The plot indicates that the month January has the maximum transactions.

The head map shows that all the members of the x\_train data set that all columns are independent and not correlated with each other.

LOOKLIKE MODELS:

The uploaded csv files have the measures of the of linear regression , logistic regression , ridge regression and lasso regression and among all of them lasso regression is the best among them.

CLUSTERING

We have clustered the data by using two algorithms k means and dbscan and by the analysis and graph plotting we can conclude that k means is more appropriate to this transaction data prediction.