FRESCO RETAIL PROJECT

Case Study By- Ali Ahmad

CONTENTS

Objective

Univariate Analysis

Bivariate Analysis

Model Building

Observations

Insights

Objective

The data analyses Fresco Retail's customers' transaction data to predict return decision using various data points like customer background, payment modes, store types, product nature, etc.

The objective is to study the influence of different transaction data points and their impact on a customer's final decision to return the products bought in a transaction.

Univariate Analysis

➢ For the target variables we see that percentage of the products that have not returned are 89.5% and the returned is only 10.50%.

➢ After checking the percentage, it implies that this Fresco Retail data is imbalanced.
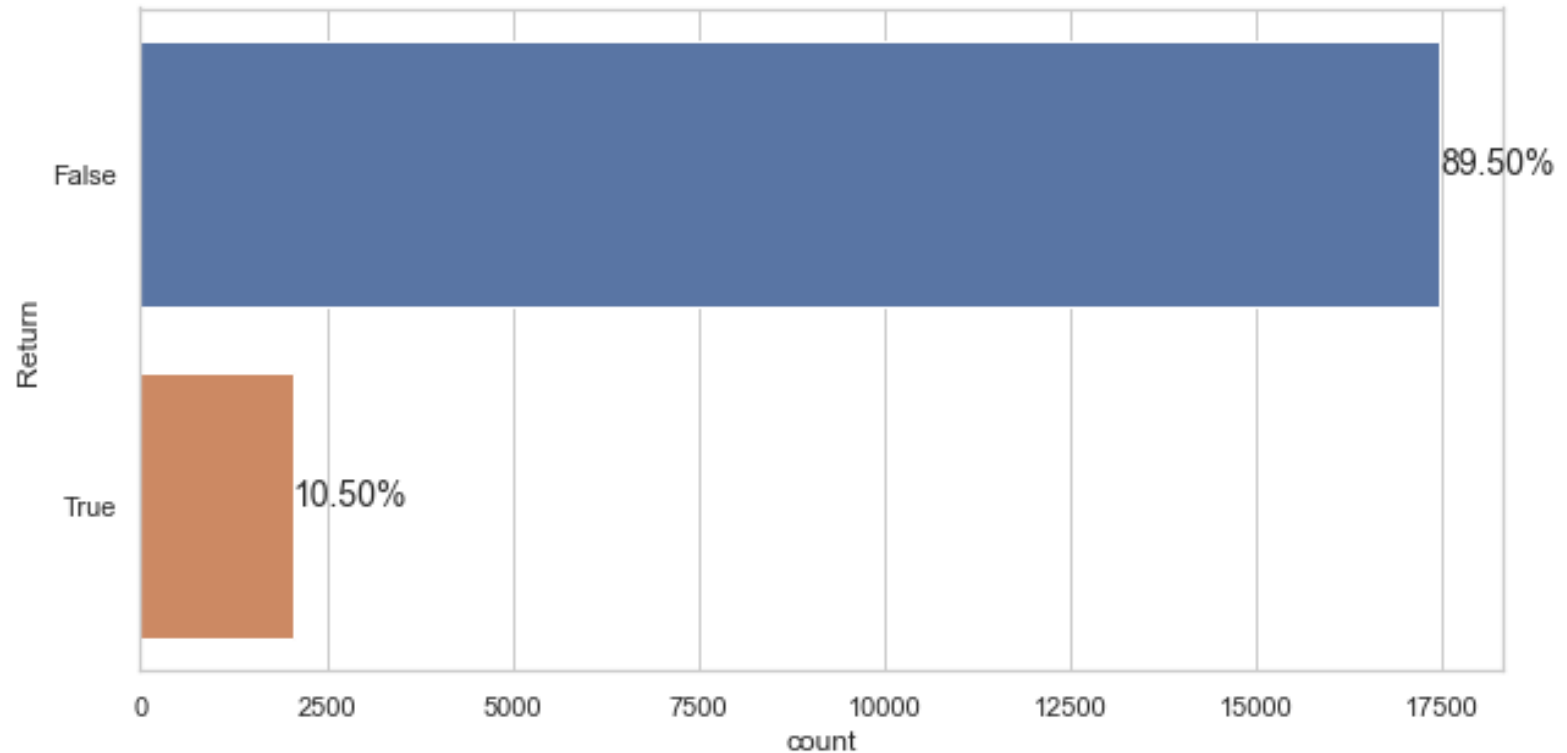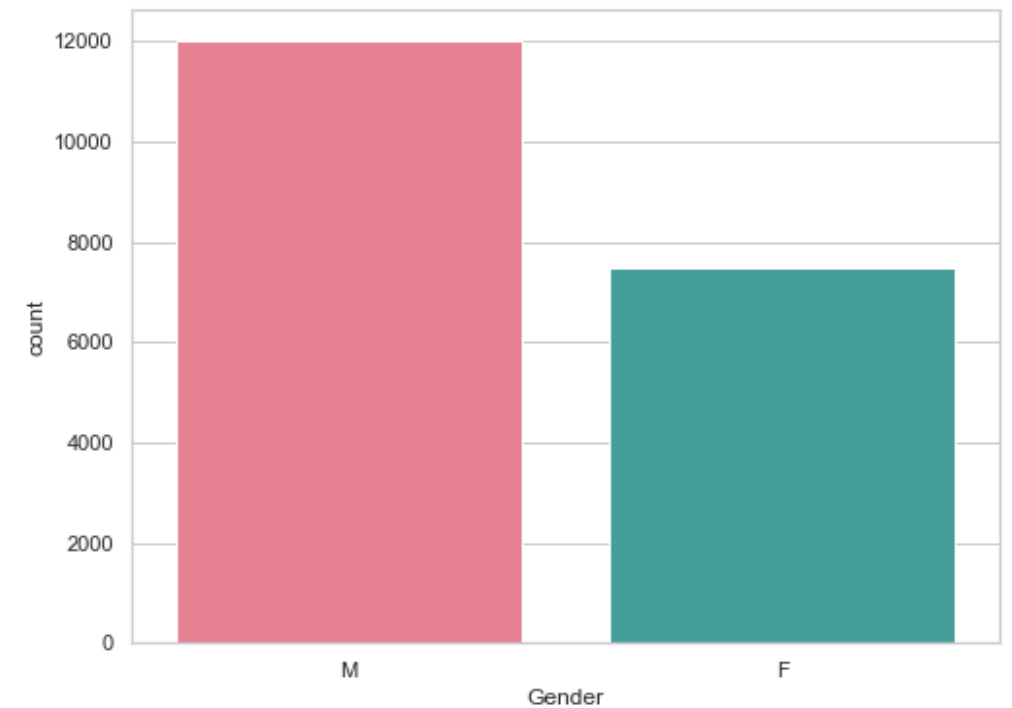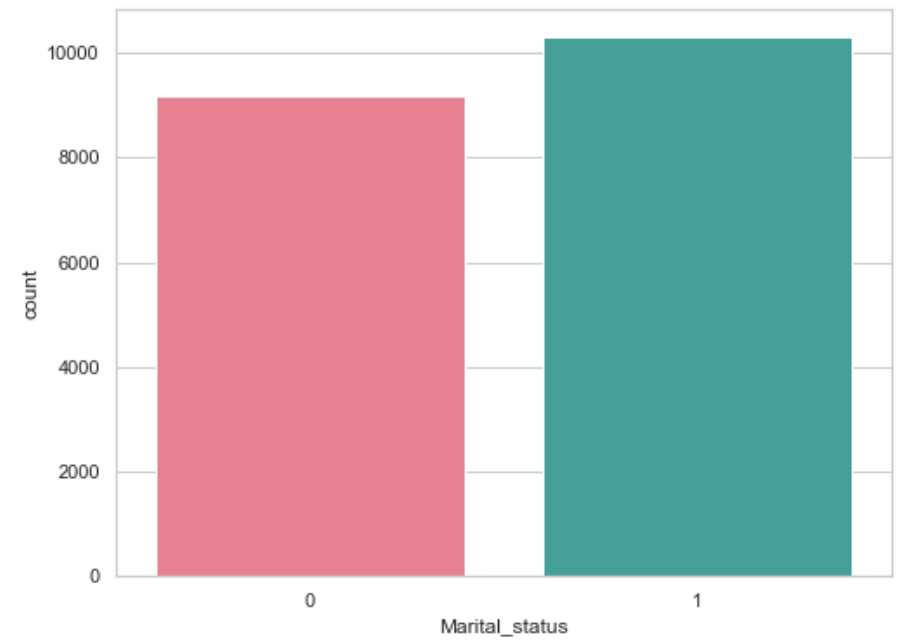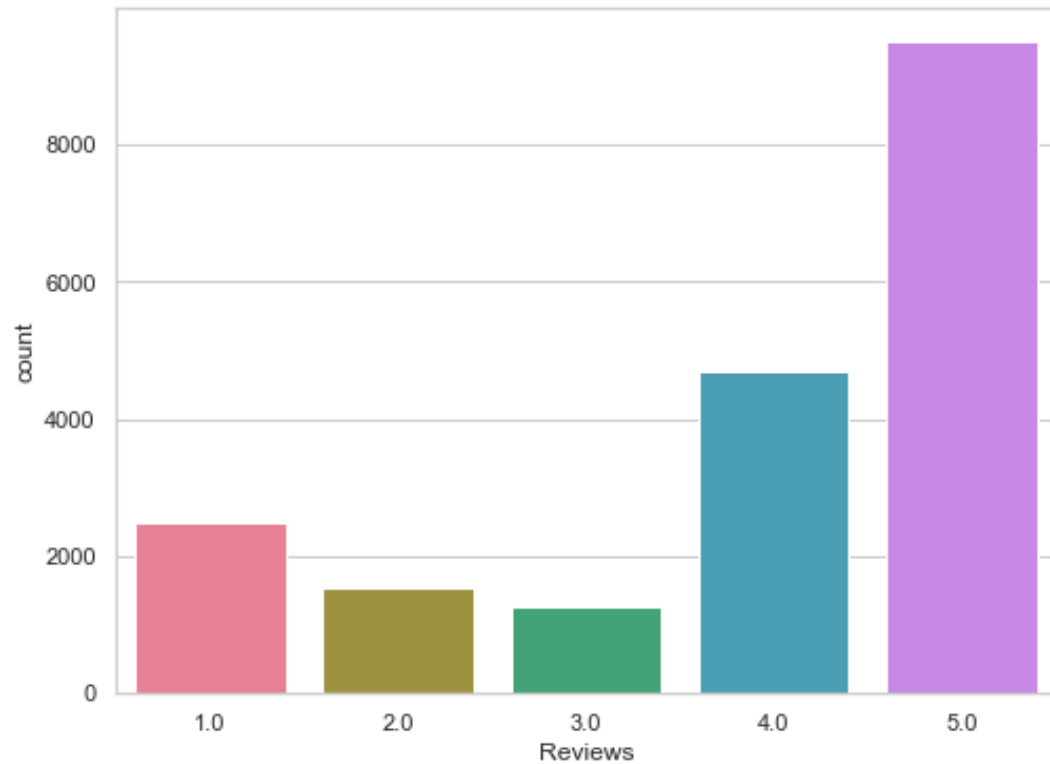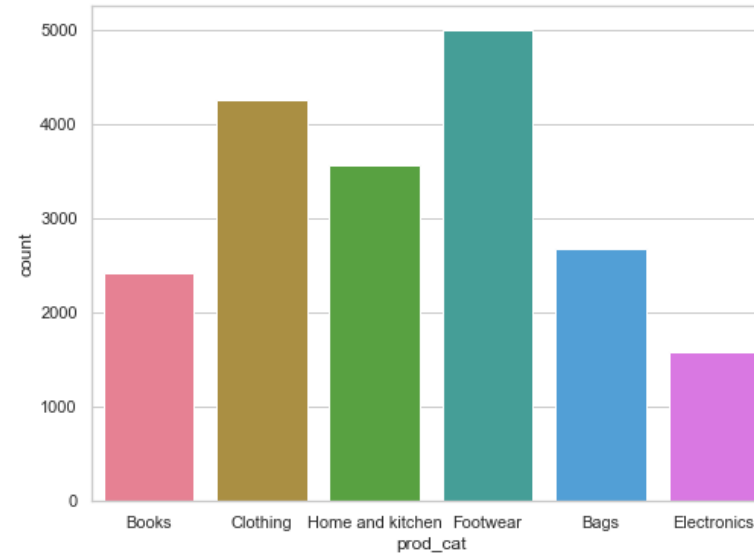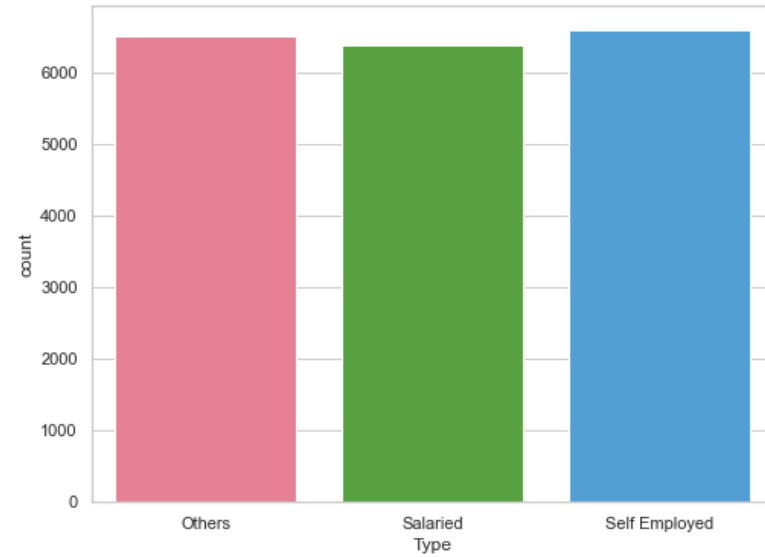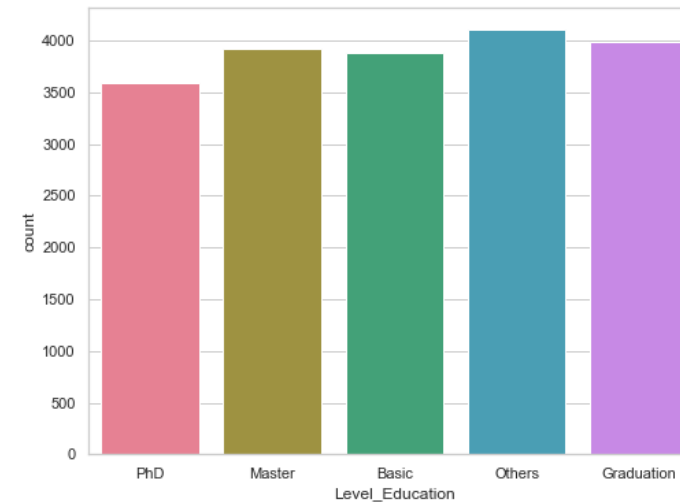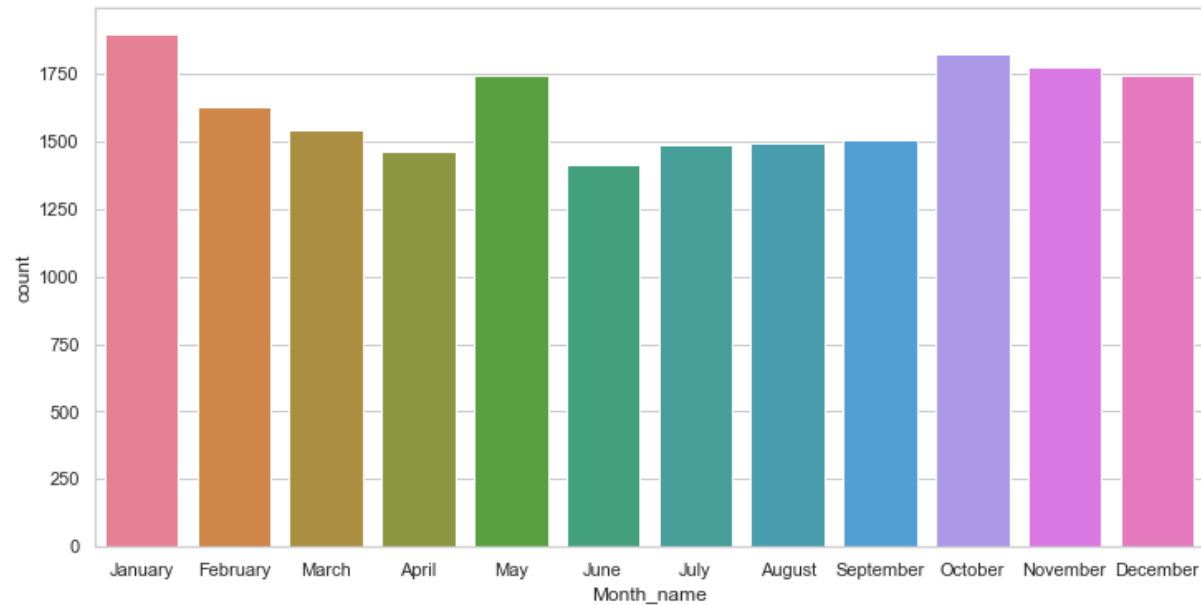
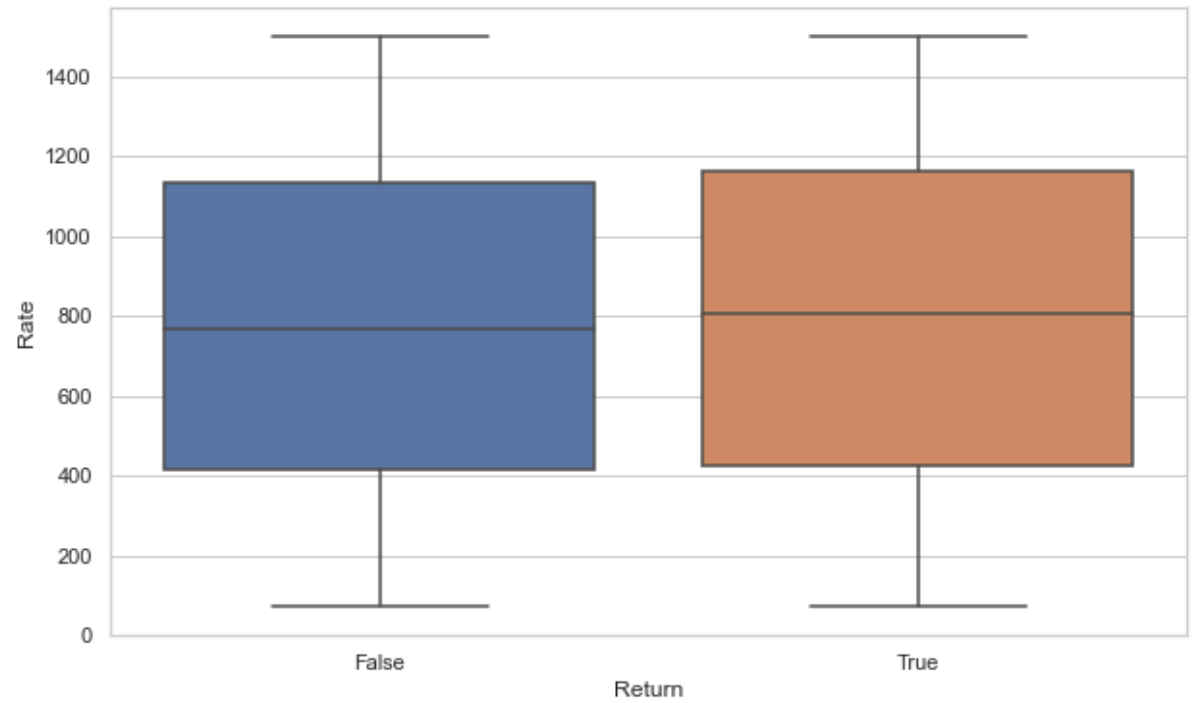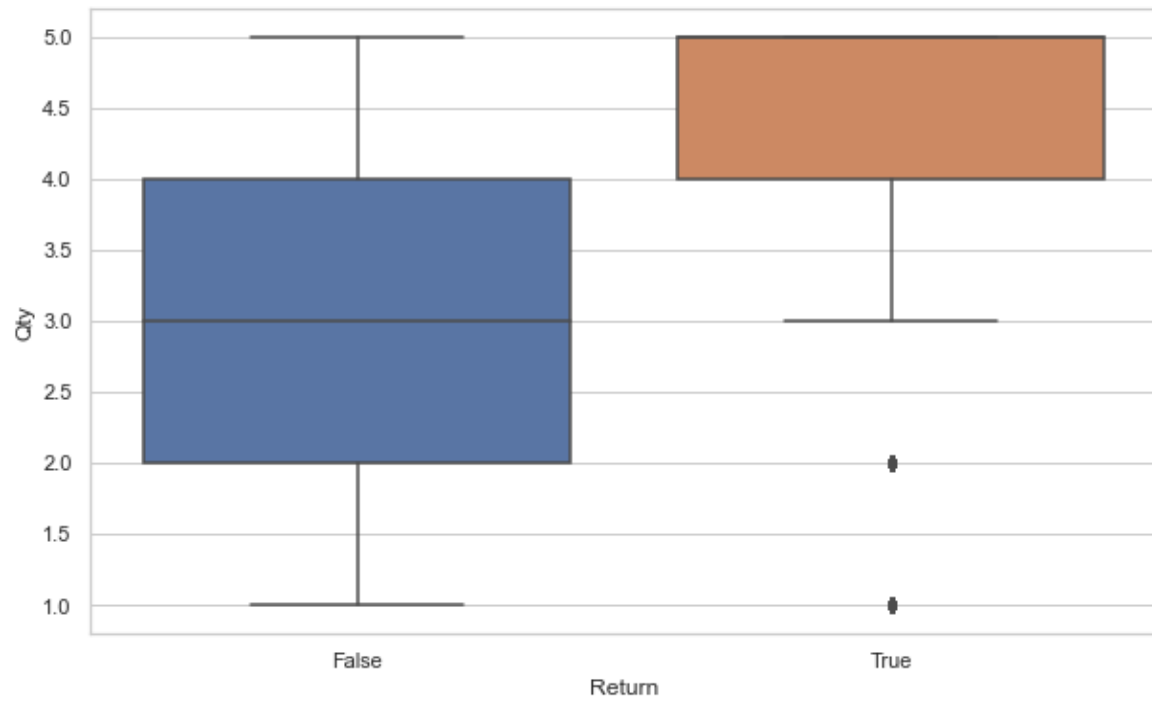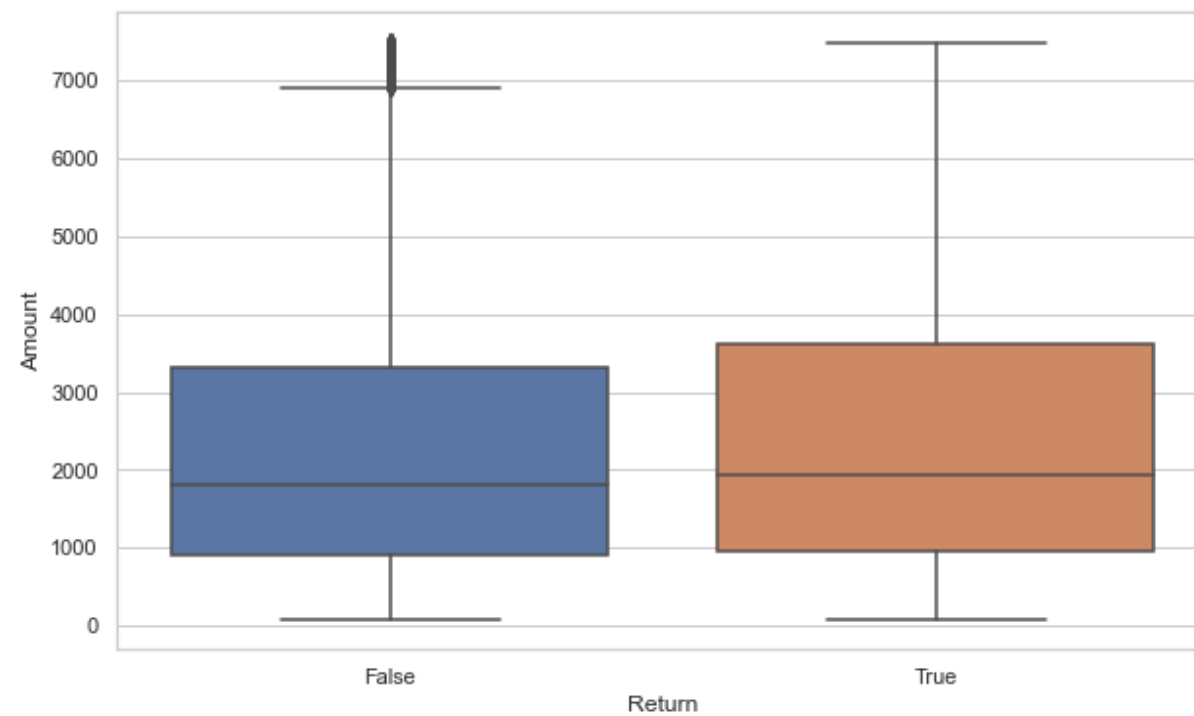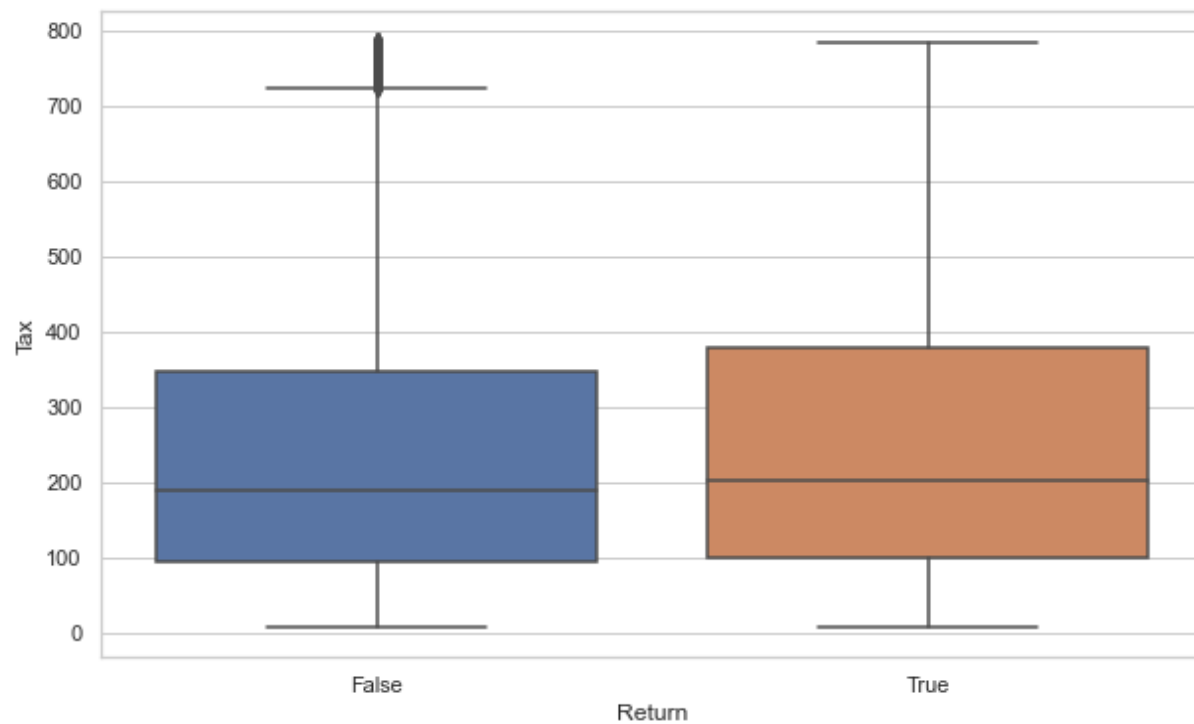# Univariate Analysis
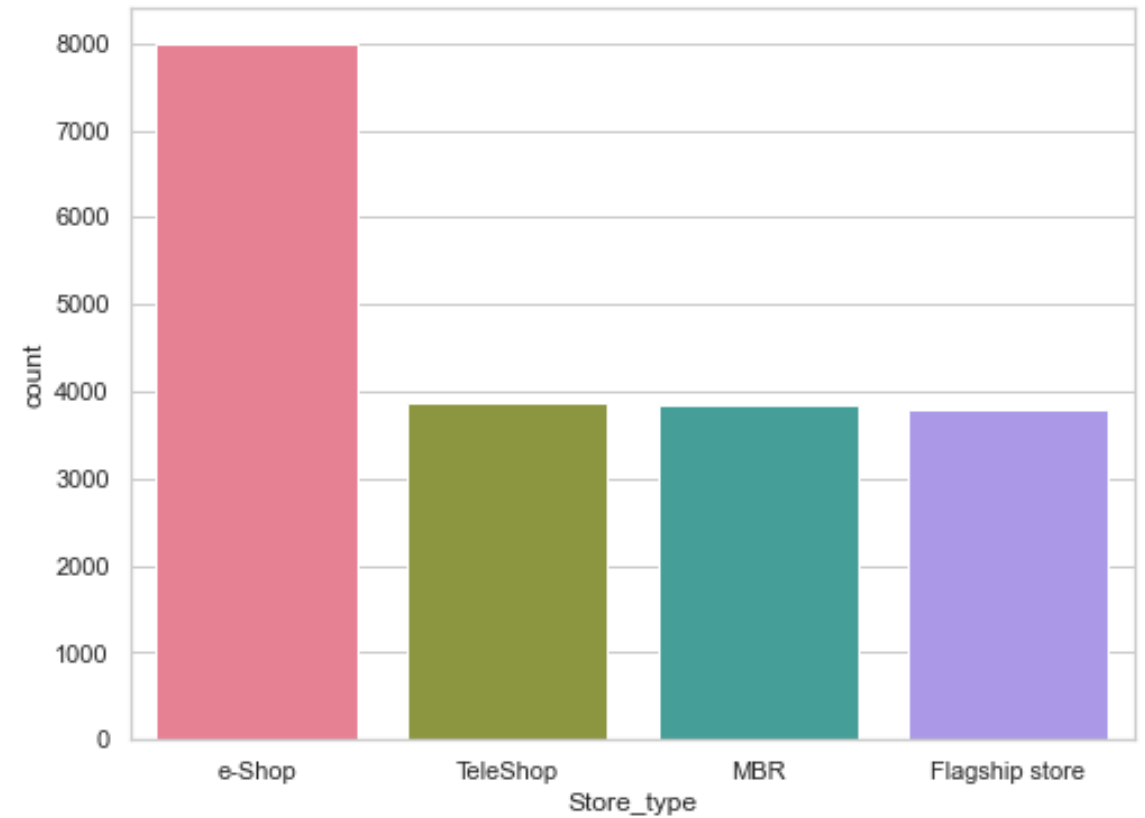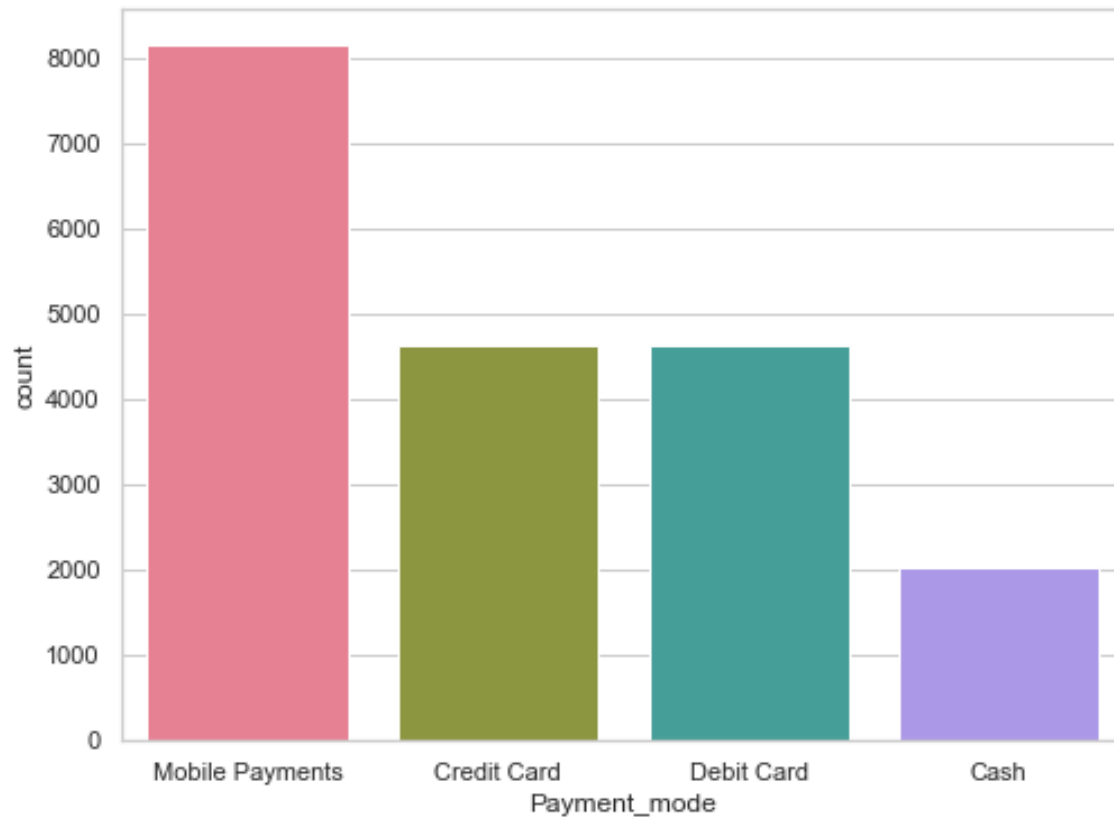
Univariate Analysis

Univariate Analysis

- The maximum return is can be seen for the Self-Employed.
- For the months, January, October and November most of the products were returned

BIVARIATE ANALYSIS

BIVARIATE ANALYSIS

BIVARIATE ANALYSIS
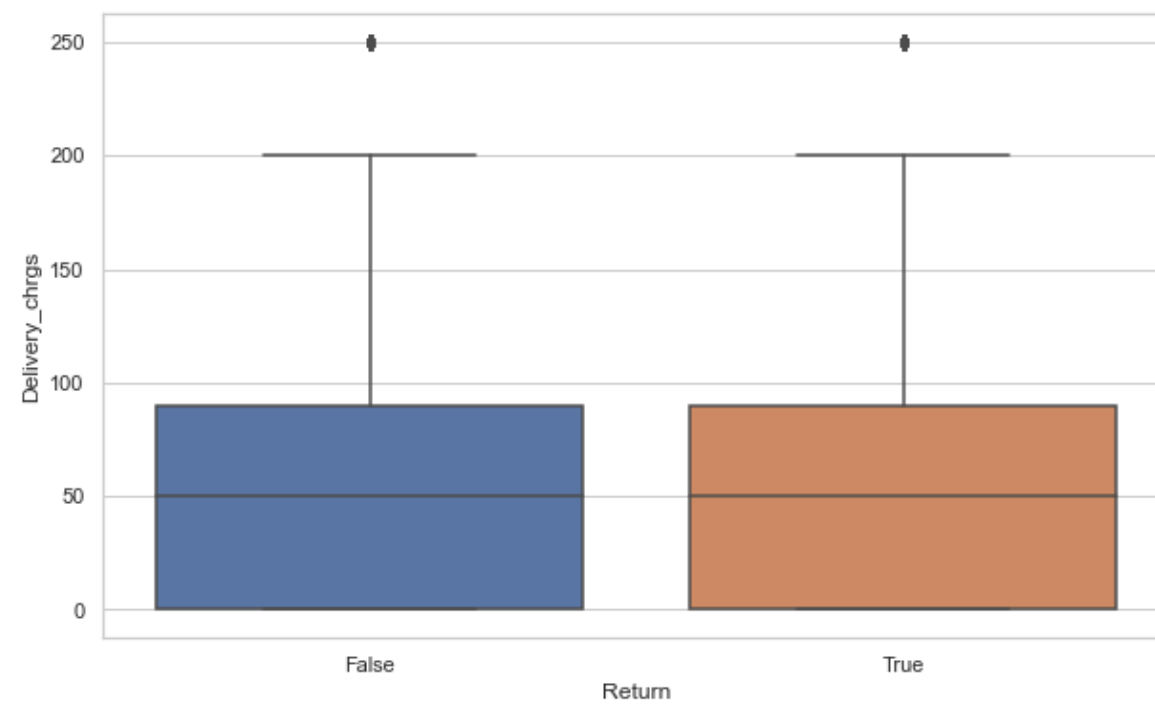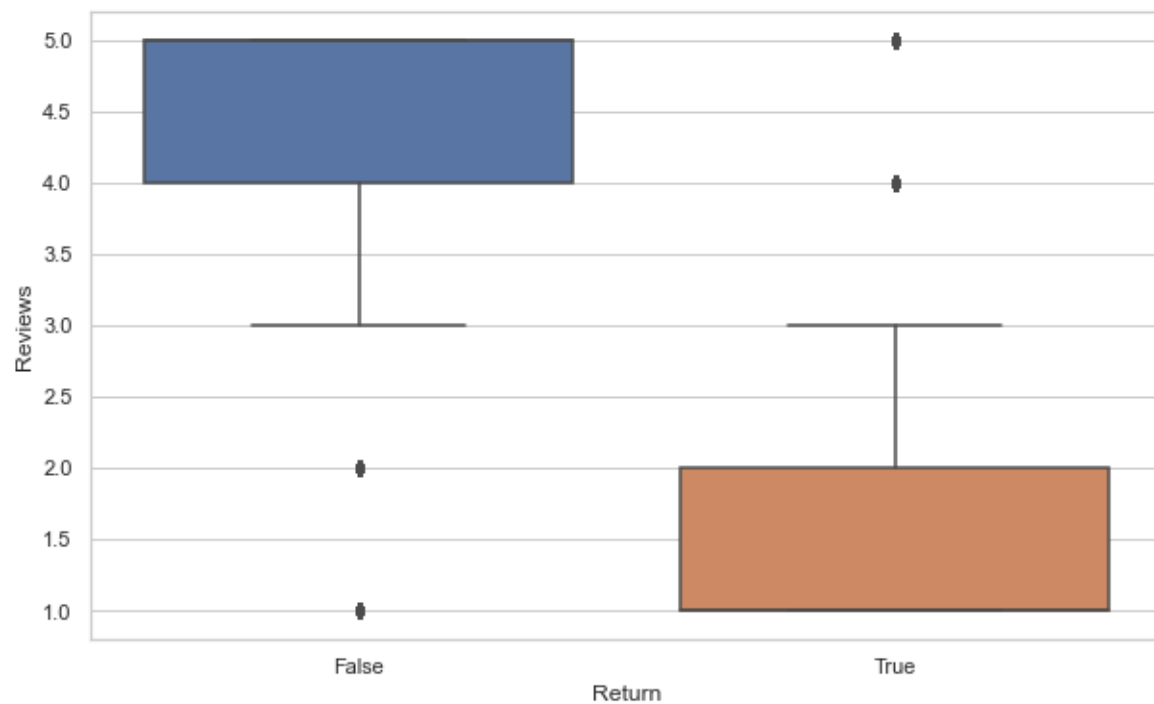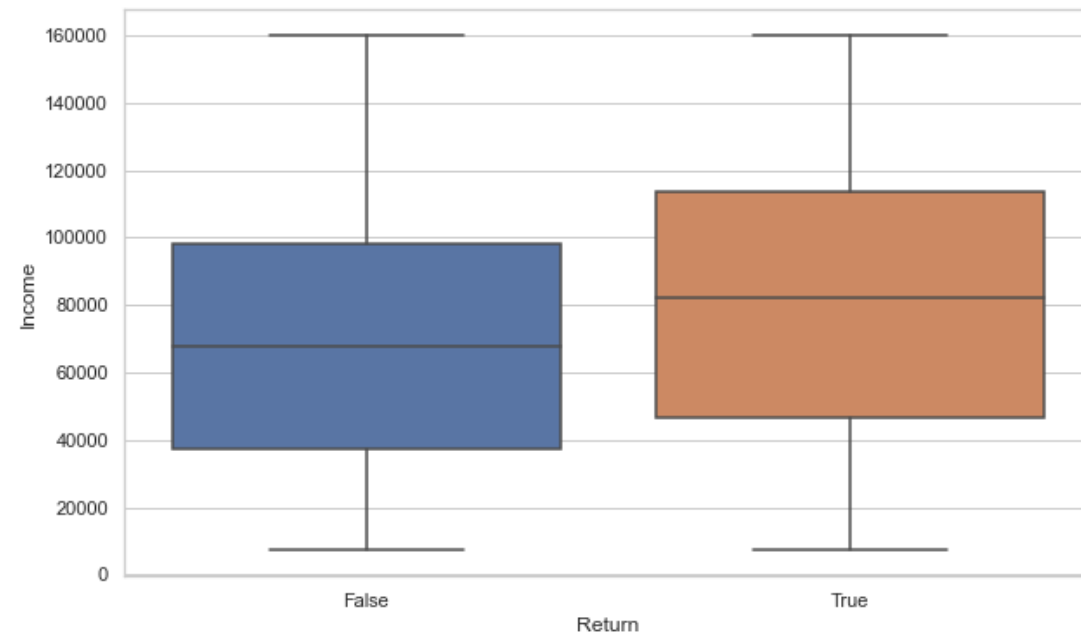
Bivariate Analysis using Pair plots

Correlation: Heatmap

Correlation matrix among variables
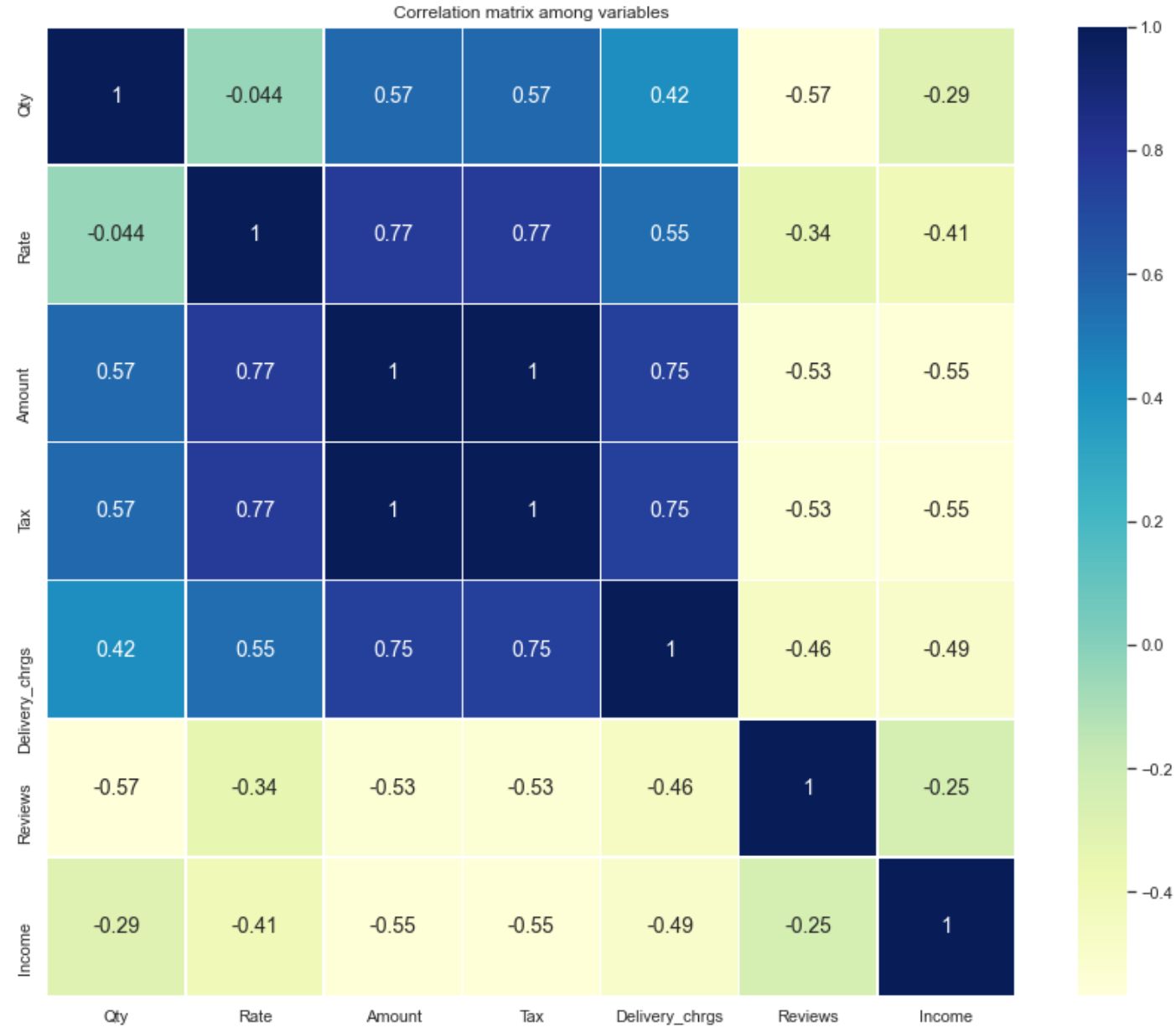
ANOVA (Analysis of Variance)

It gives the relationship between continuous and categorical columns whether they are statistically significant or not.

H0- Null hypothesis means the variables are not correlated:

1. If Small P-Value < 5% (0.05) means the variables are correlated. Null hypothesis H0 is rejected.

2. If Large P-Value > 5% (0.05) means the variables are not correlated and the Null hypothesis H0 is accepted.

Result for Anova test : In this test, variable which is not correlated with target variable.

• Delivery_chrgs is not correlated with Return i.e., P-value > 0.05, hence we can reject this variable for further analysis.

Model Building

For model building, Scikit-Learn model is used to predict the return decision of the customers.

SMOTE technique with oversampling is used to balance the dataset.

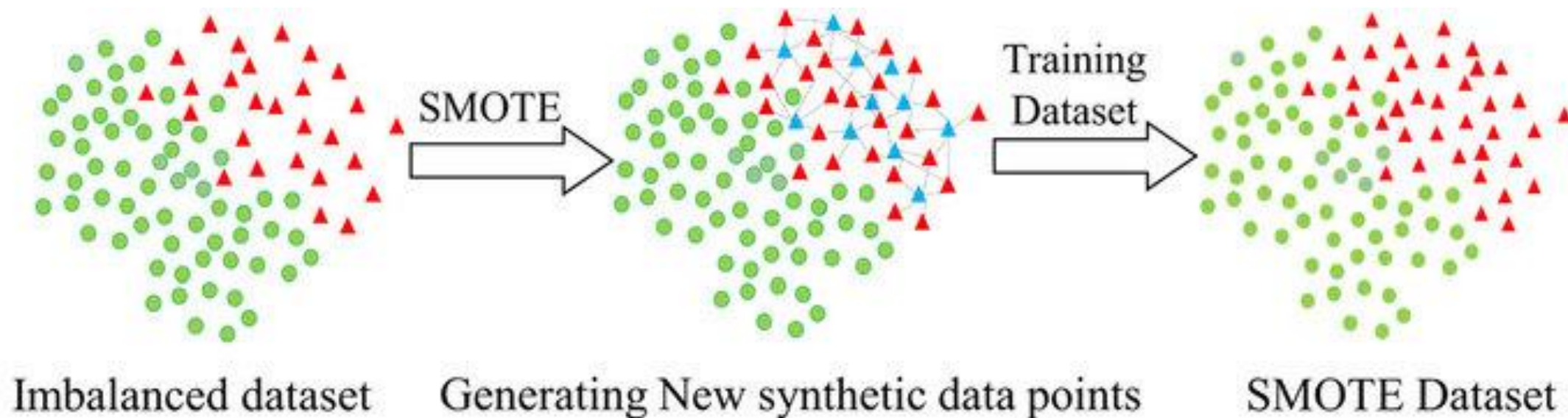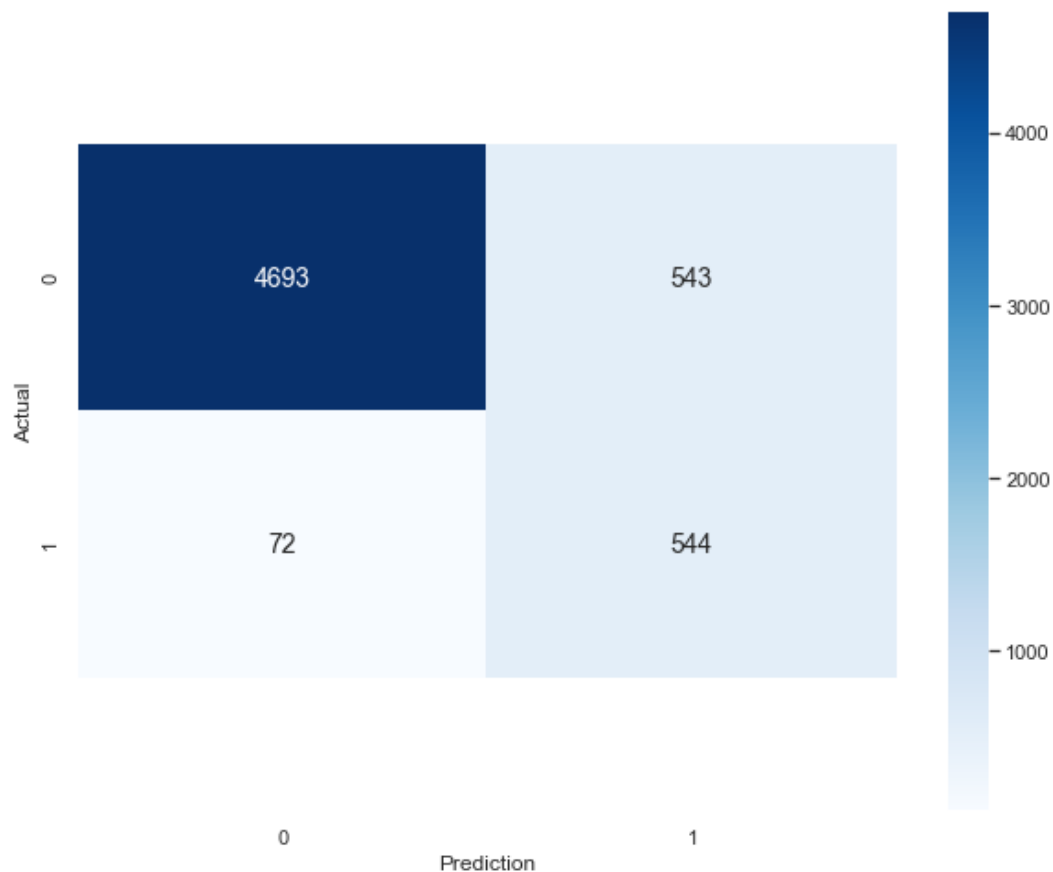Imbalanced dataset   Generating New synthetic data points   SMOTE Dataset

OBSERVATIONS

Logistic Model Accuracy : 0.89%


```
              precision   recall  f1-score   support

          0        0.98     0.90      0.94      5236
          1        0.50     0.88      0.64       616

   accuracy                          0.89      5852
  macro avg        0.74     0.89      0.79      5852
weighted avg        0.93     0.89      0.91      5852

Accuracy of logistic regression classifier on test set: 0.89
```
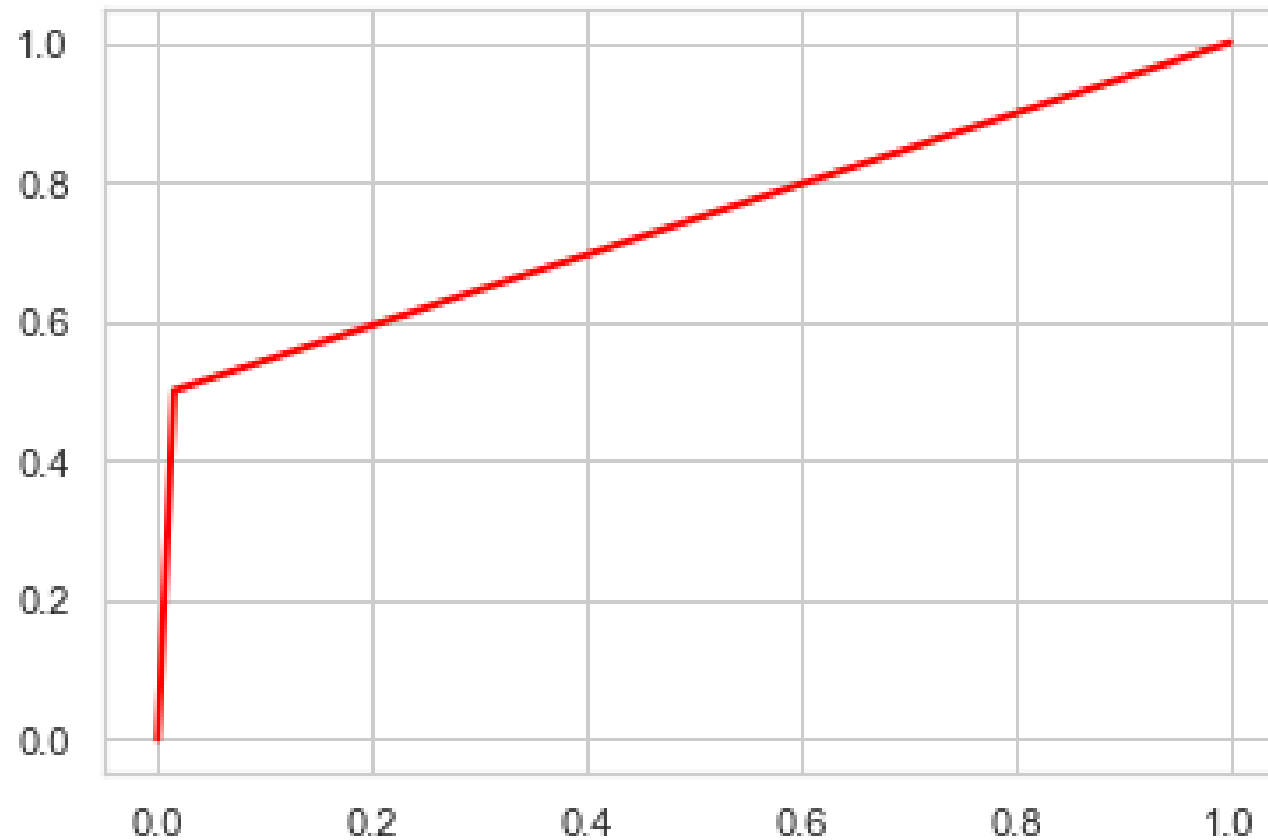
Observations for ROC Curve

Roc-auc-Score : 0.74%

💡 INSIGHTS 💡

The percentage of the products that have not returned are 89.5% and the returned is only 10.50%.

Most of the transaction has been done by Mobile payments followed by debit and credit card.

Most of the product was bought through E-shop which is the online mode.

The maximum review for the product is 5.0 and the second highest review for product is 4.0.

The Customer's education level is maximum for the miscellaneous.

The maximum return for the product can be seen for the Self-Employed.

We can see that the maximum customers are more interested in buying Footwear followed by Clothing.

Most of the products that retuned were in months of January, May, October and November.

THANKYOU