Contents

Exhibit **1 – Data description**

Variable	Description
AccountID	account unique identifier
Churn	account churn flag (Target)
Tenure	Tenure of account
City_Tier	Tier of primary customer's city
CC_Contacted_L12m	How many times all the customers of the account has contacted customer care in last 12months
Payment	Preferred Payment mode of the customers in the account
Gender	Gender of the primary customer of the account
Service_Score	Satisfaction score given by customers of the account on service provided by company
Account_user_count	Number of customers tagged with this account
account_segment	Account segmentation on the basis of spend
CC_Agent_Score	Satisfaction score given by customers of the account on customer care service provided by company
Marital_Status	Marital status of the primary customer of the account
rev_per_month	Monthly average revenue generated by account in last 12 months
Complain_l12m	Any complaints has been raised by account in last 12 months
rev_growth_yoy	revenue growth percentage of the account (last 12 months vs last 24 to 13 month)
coupon_used_l12m	How many times customers have used coupons to do the payment in last 12 months
Day_Since_CC_connect	Number of days since no customers in the account has contacted the customer care
cashback	Monthly average cashback generated by account in last 12 months
Login_device	Preferred login device of the customers in the account

Exhibit **2 – Data preparation**

Table 1 – Data imputation

Variables	Data	Count	Action
Tenure	#	117	Removed
Gender	F	270	Replaced with “Female”
Gender	M	376	Replaced with “Male”
Account	@	332	Removed
Account Segment	Regular +	262	Replaced with “Regular Plus”
Account Segment	Super +	47	Replaced with “Super Plus”
Rev PerMonth	+	589	Removed
Day_Since_CC_connect	$	1	Replaced with “0”
Cashback	$	2	Replaced with “0”

Table 2 – Null imputation

Features	Before Cleanup	After Cleanup
AccountID	0	0
Churn	0	0
Tenure	216	0
City_Tier	112	0
CC_Contacted_LY	101	0
Payment	109	0
Gender	108	0
Service_Score	98	0
Account_user_count	442	0
account_segment	97	0
CC_Agent_Score	116	0
Marital_Status	212	0
rev_per_month	789	0
Complain_ly	357	0
rev_growth_yoy	3	0
coupon_used_for_payment	3	0
Day_Since_CC_connect	357	0
cashback	472	0
Login_device	220	0

Table 3 - Unique values

City_Tier: [3. 1. 2.]
Payment: ['Debit Card' 'UPI' 'Credit Card' 'Cash on Delivery' 'E wallet']
Gender: ['Female' 'Male']
account_segment: ['Super' 'Regular Plus' 'Regular' 'HNI' 'Super Plus']
Marital_Status: ['Single' 'Divorced' 'Married']
Login_device: ['Mobile' 'Computer']

Exhibit 3 – Univariate and multivariate analysis

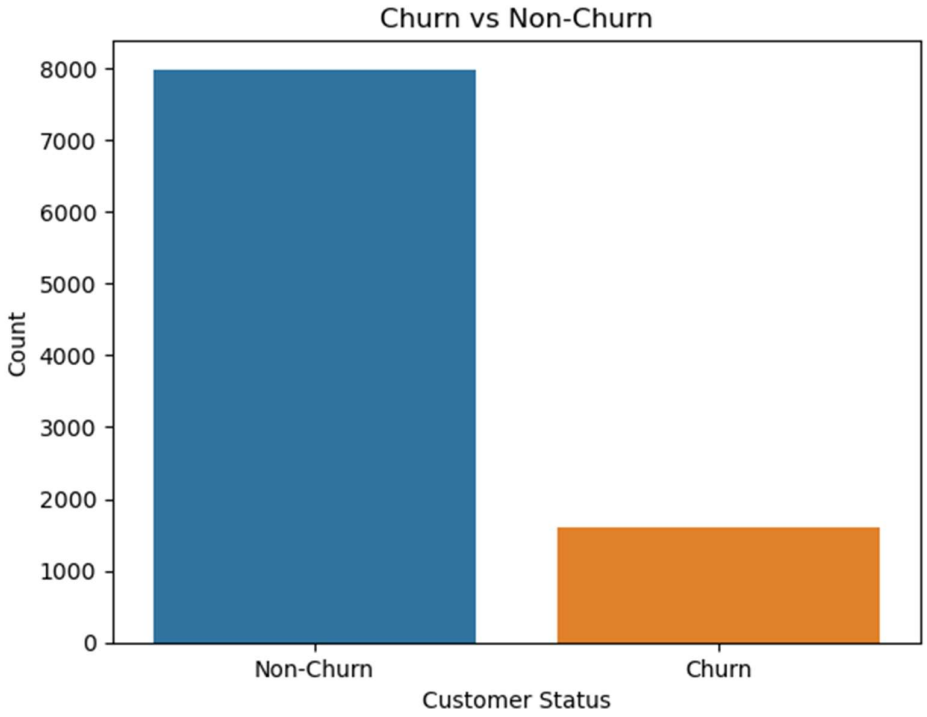

Figure 1 – Customer churn vs non-churn

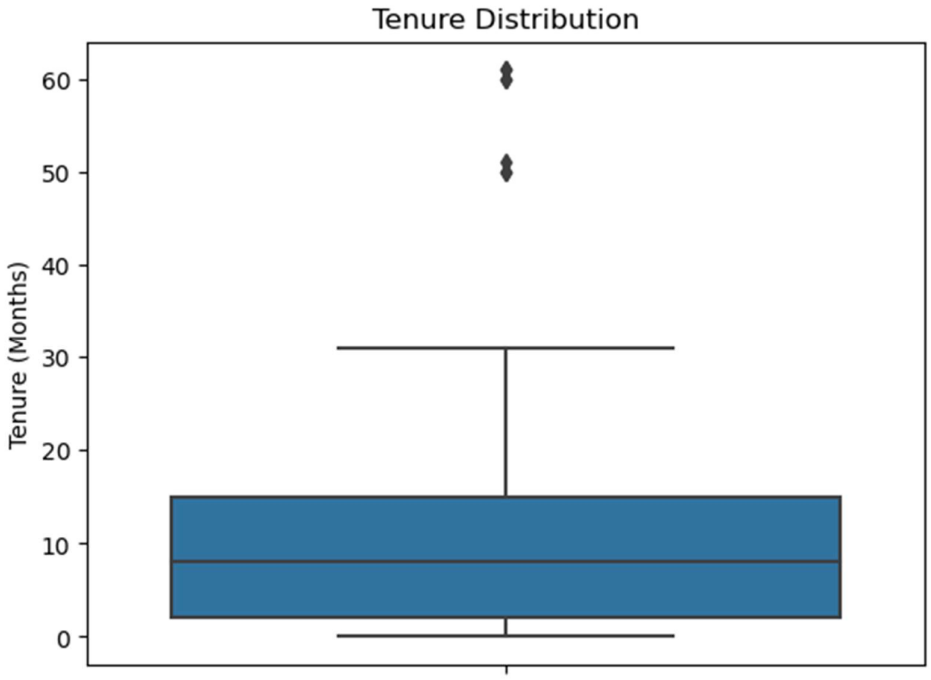

Figure 2 – Tenure distribution across customers

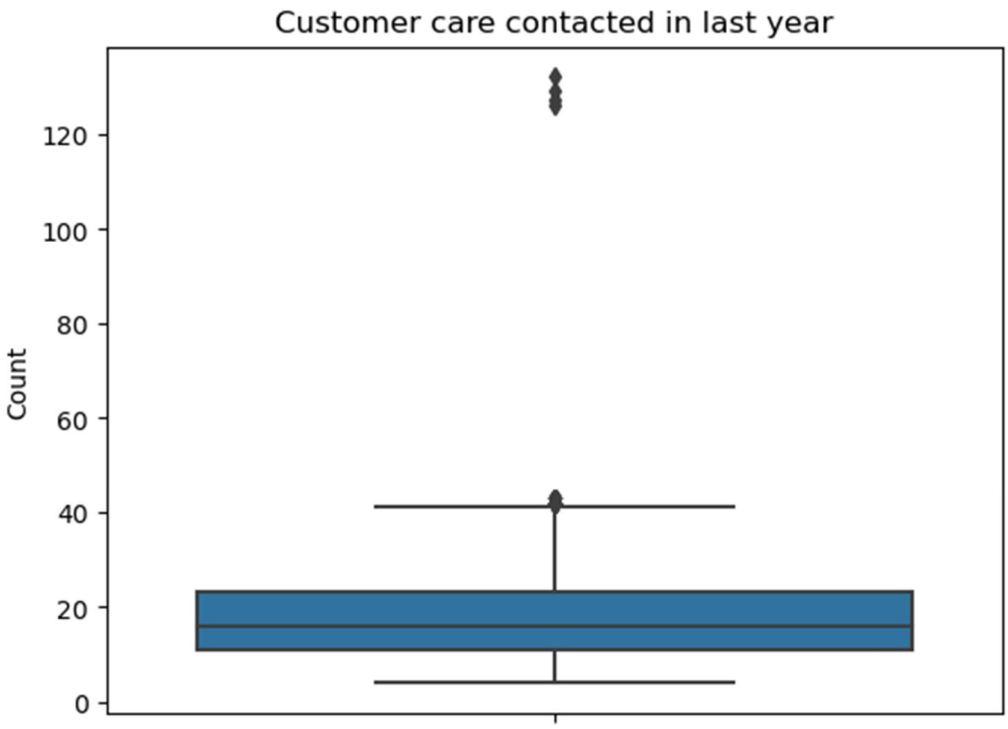

Figure 3 – Distribution of Customer who contacted customer care

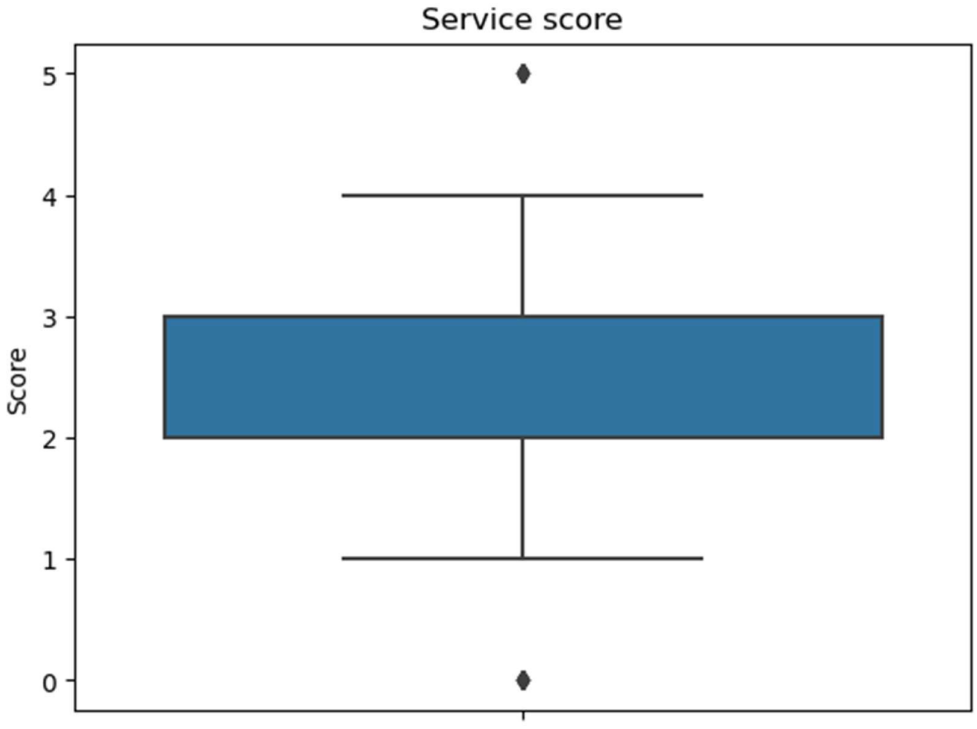

Figure 4 - Distribution of service score

Figure 5 – Distribution of account segment.

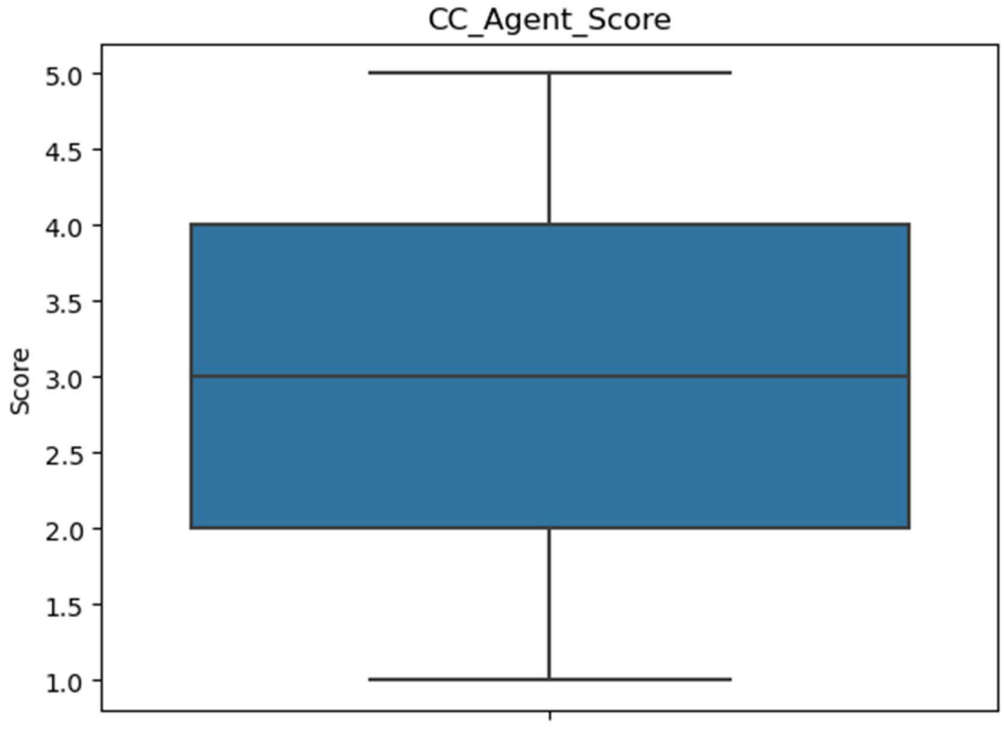

Figure 6 - Distribution of Agents score give by customer.

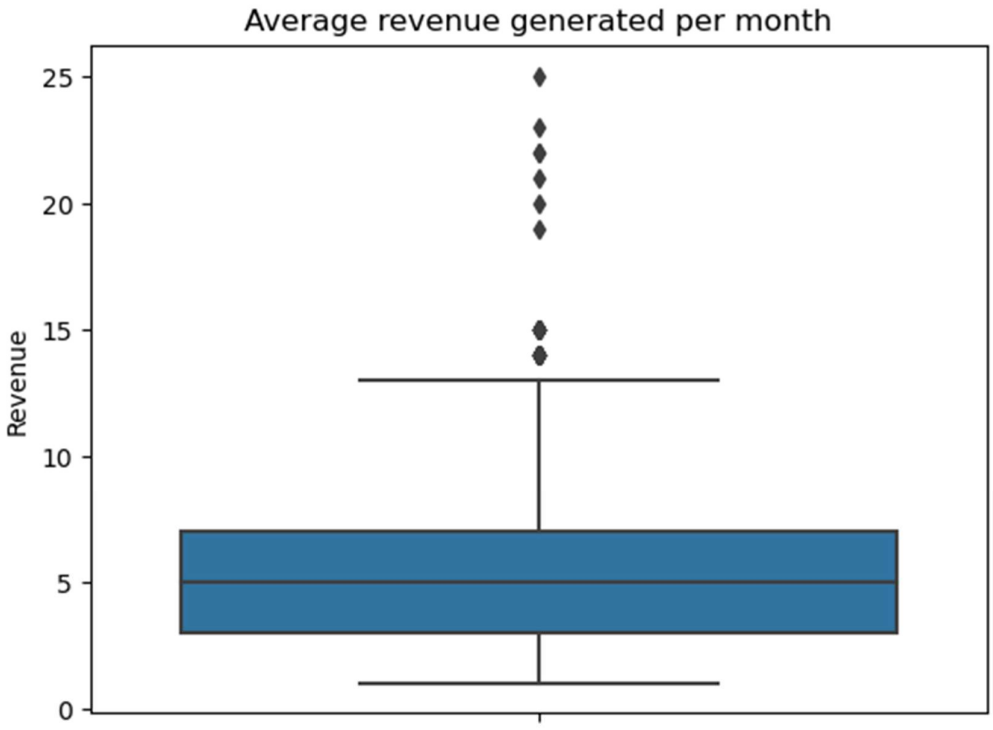

Figure 7 - Distribution of revenue generated per month

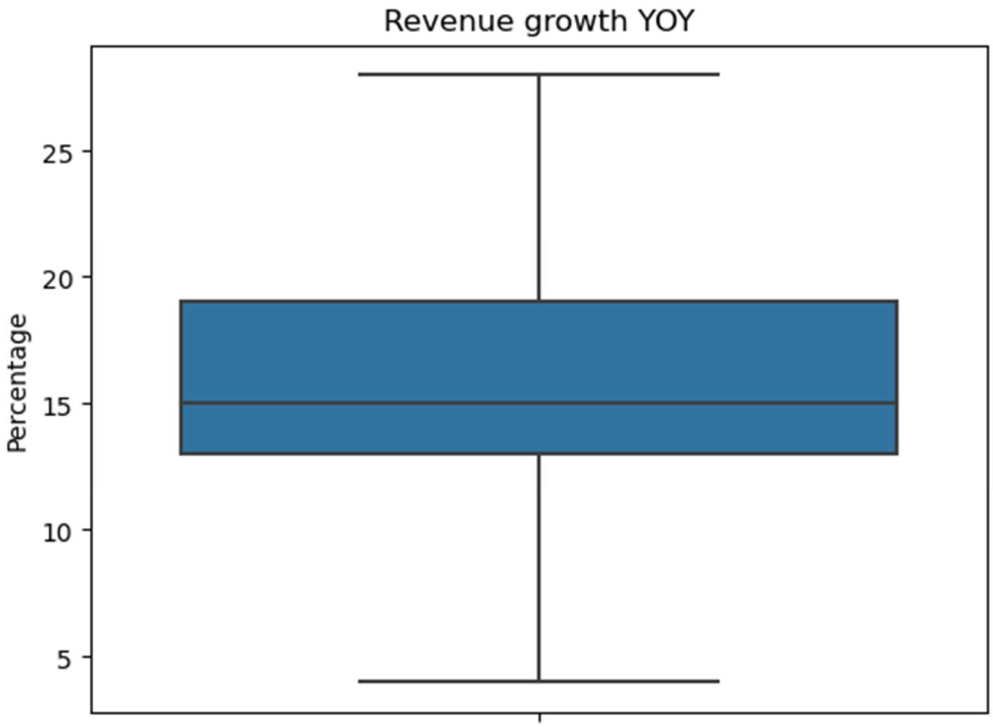

Figure 8 - Distribution of revenue growth YOY

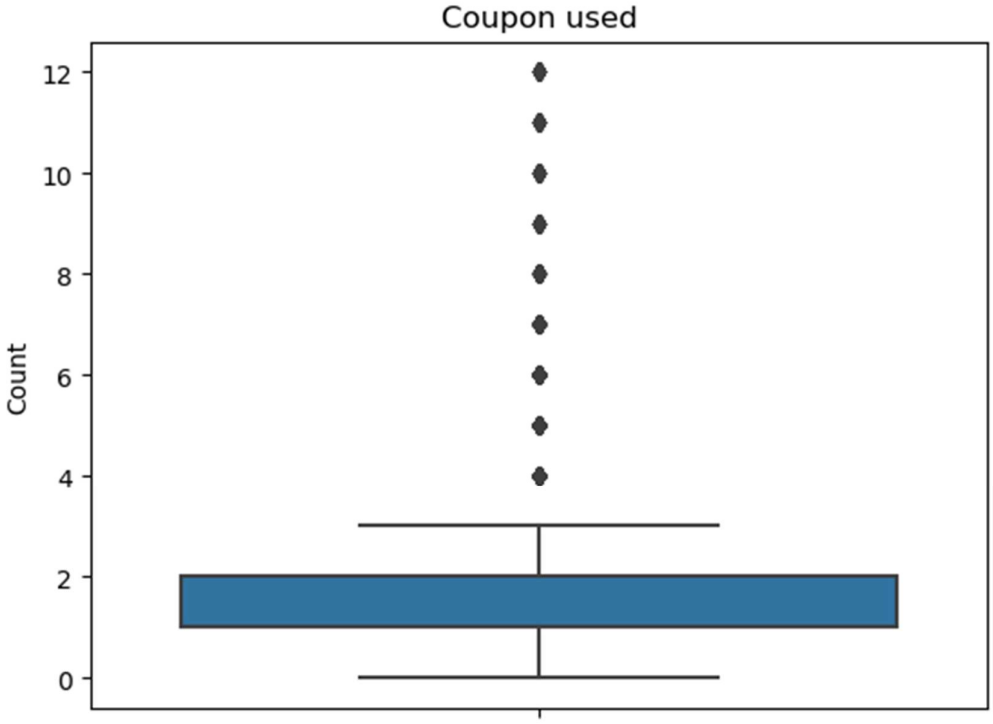

Figure 9 - Distribution of number of coupons used

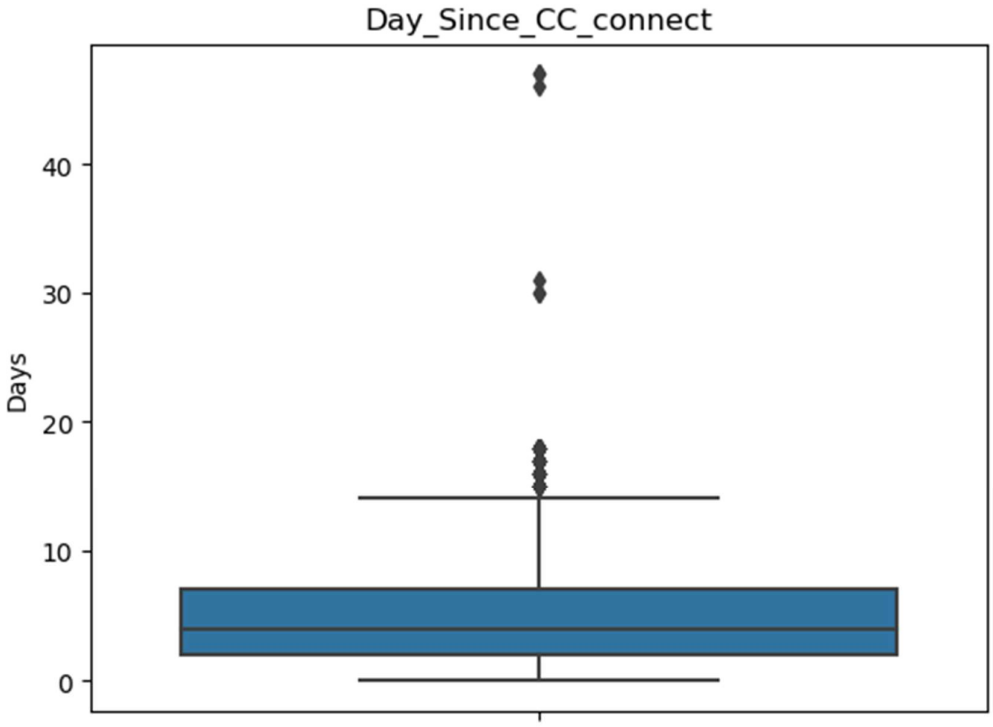

Figure 10 - Distribution of days since the customer contact customer care

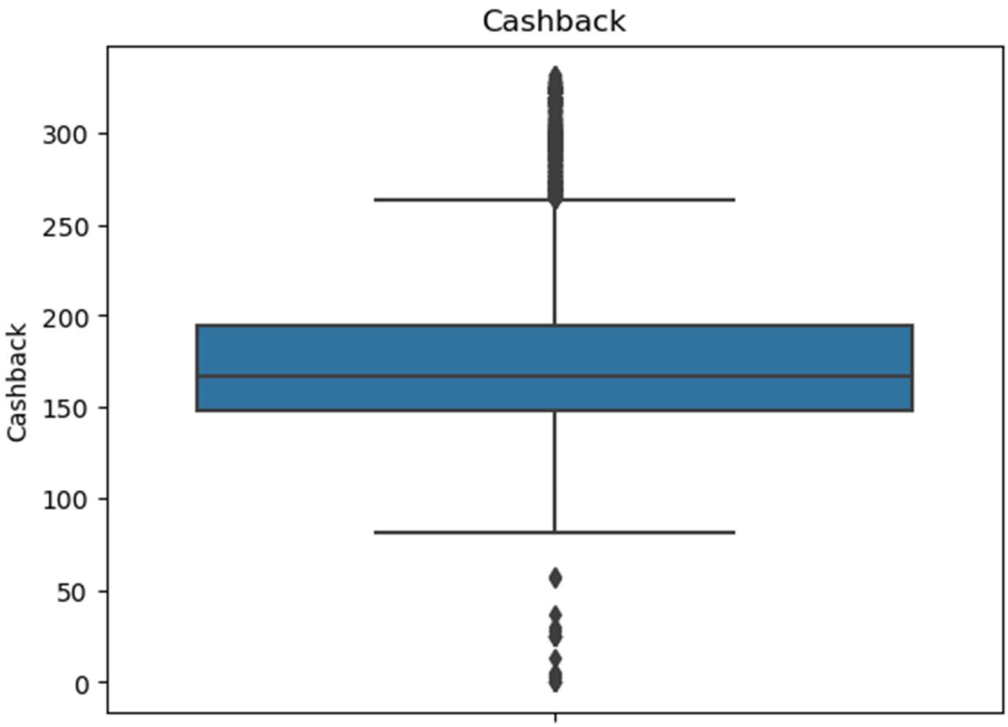

Figure 11 - Distribution of cashback received by customer.

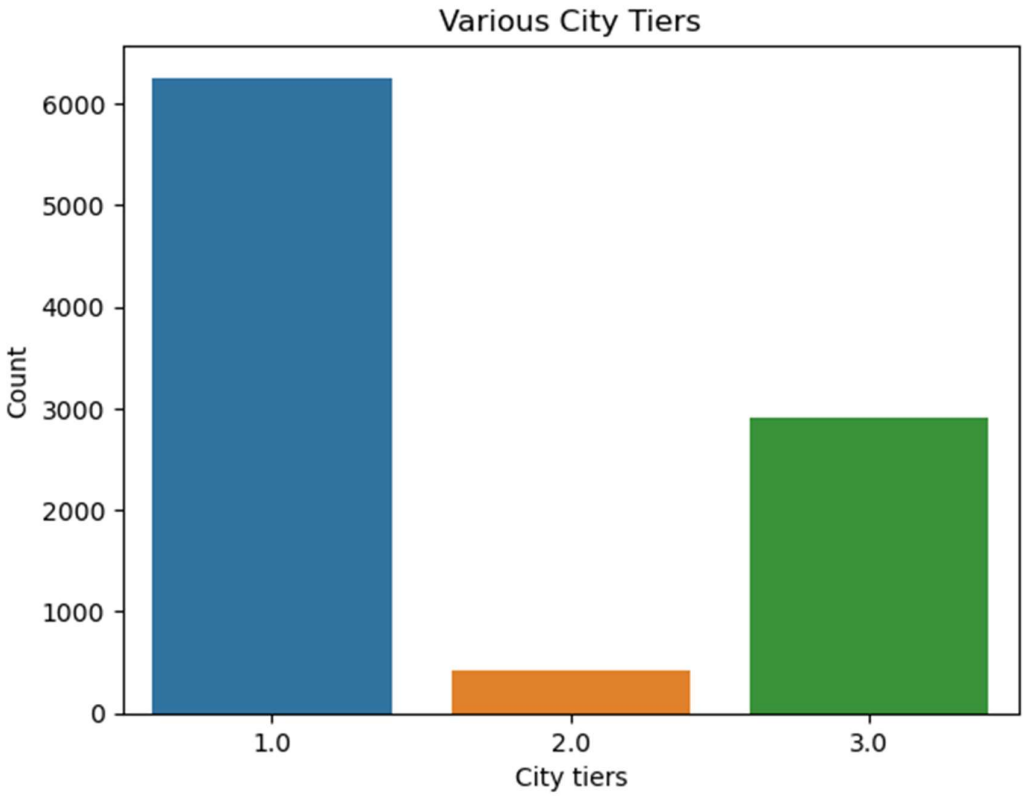

Figure 12 - Distribution of Customers across cities

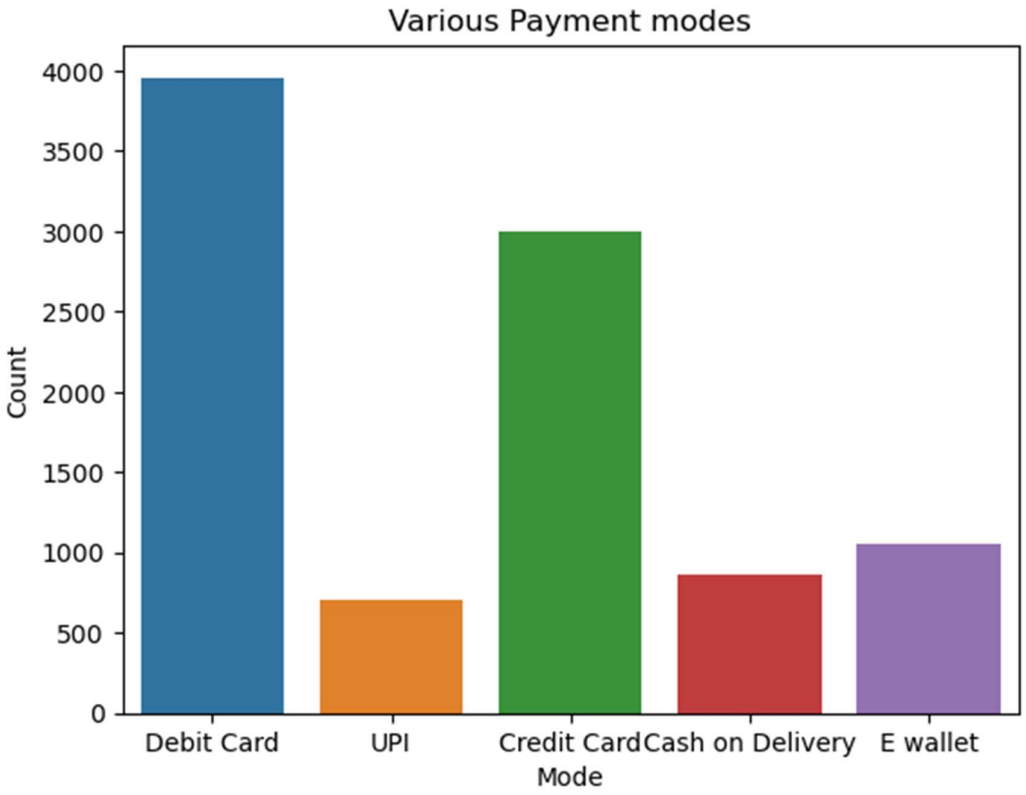

Figure 13 - distribution of payment modes

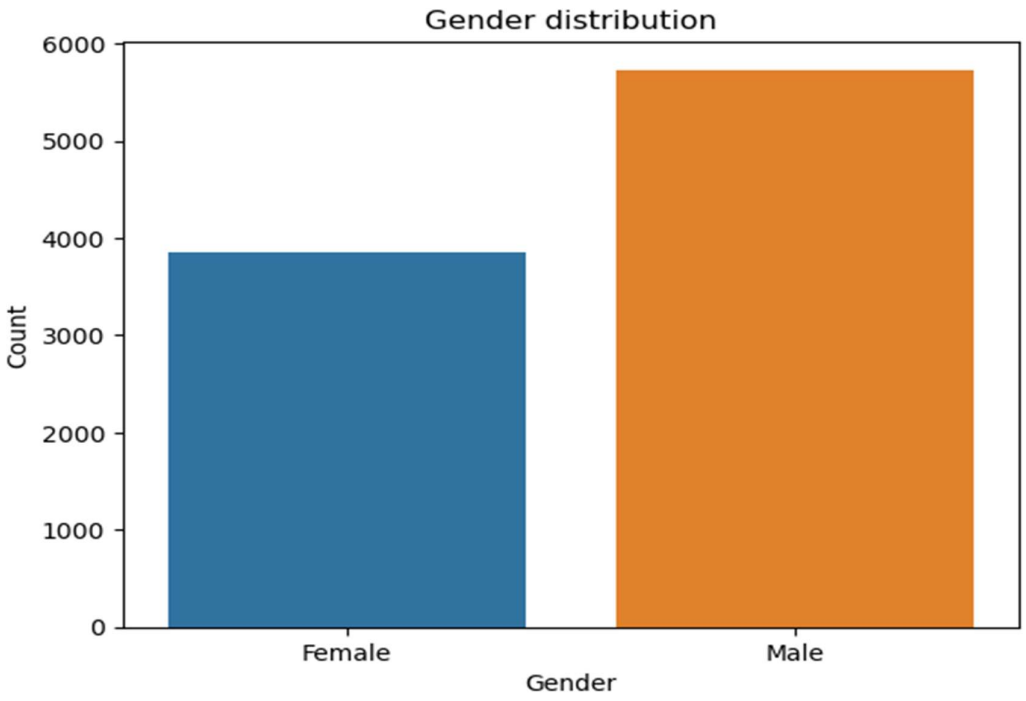

Figure 14 - distribution of gender of customers

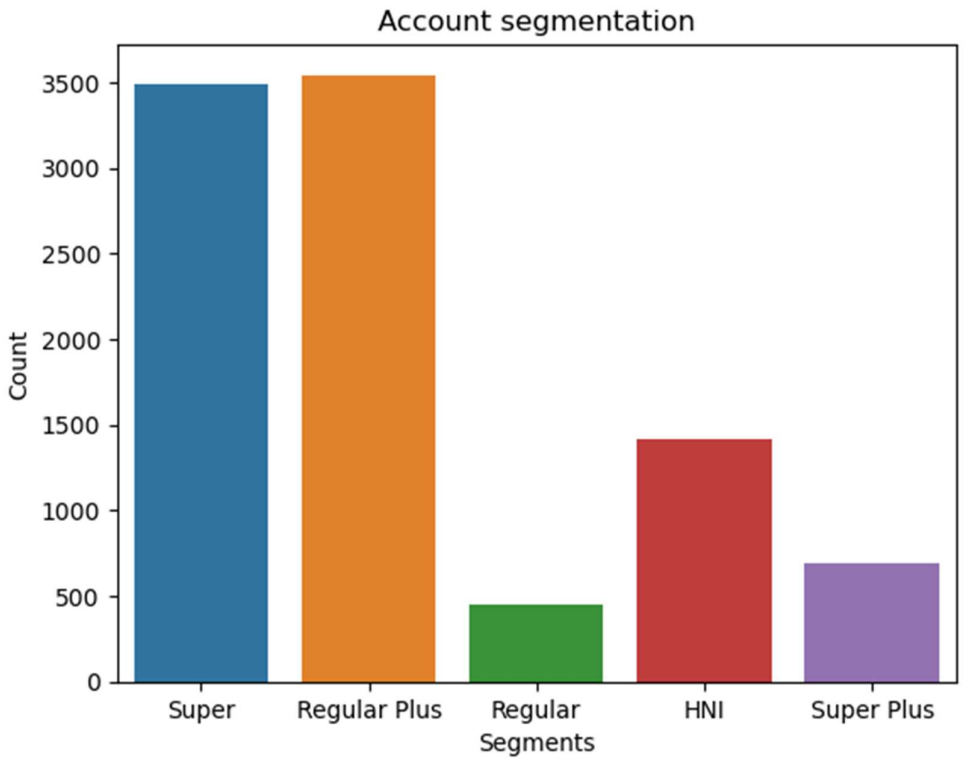

Figure 15 - distribution of customers across different segments.

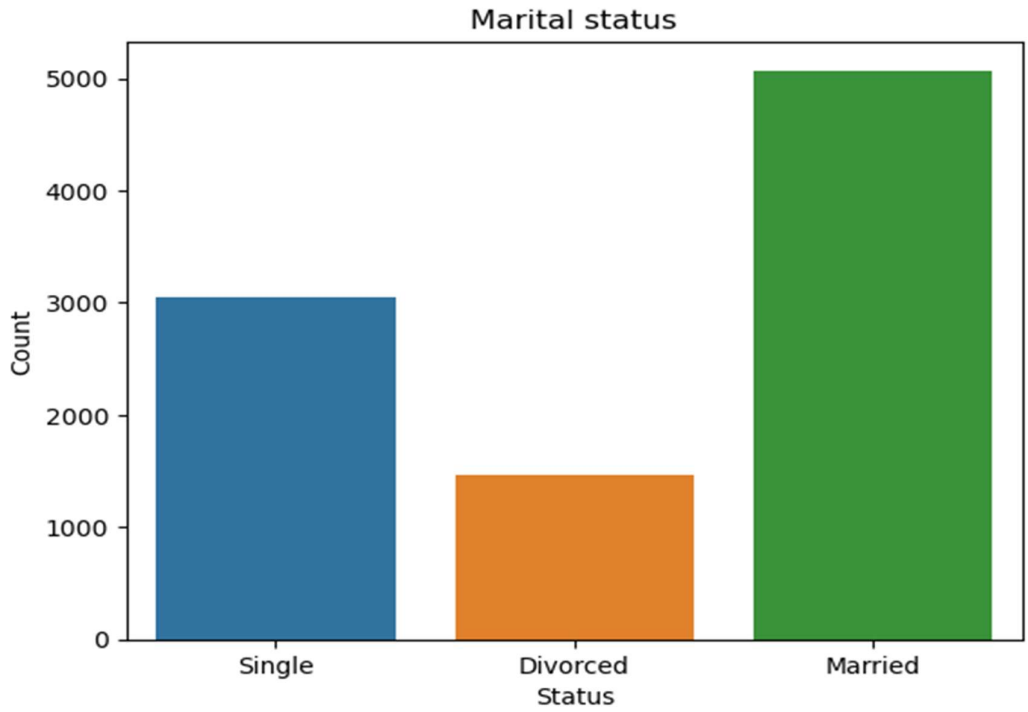

Figure 16 - distribution of customers across different Marital status

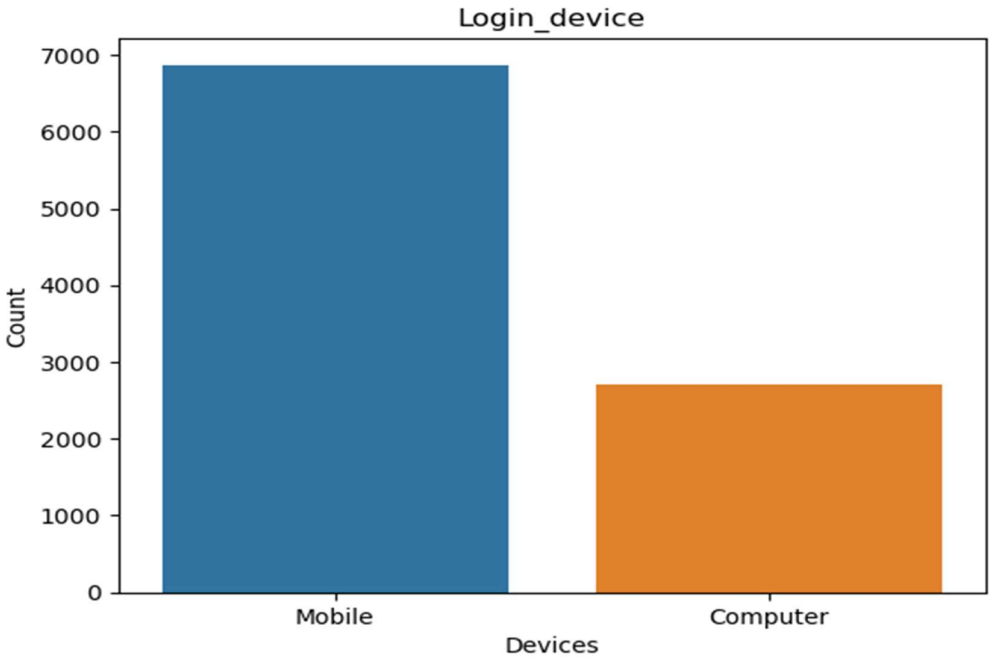

Figure 17 - Distribution of devices used by Customers

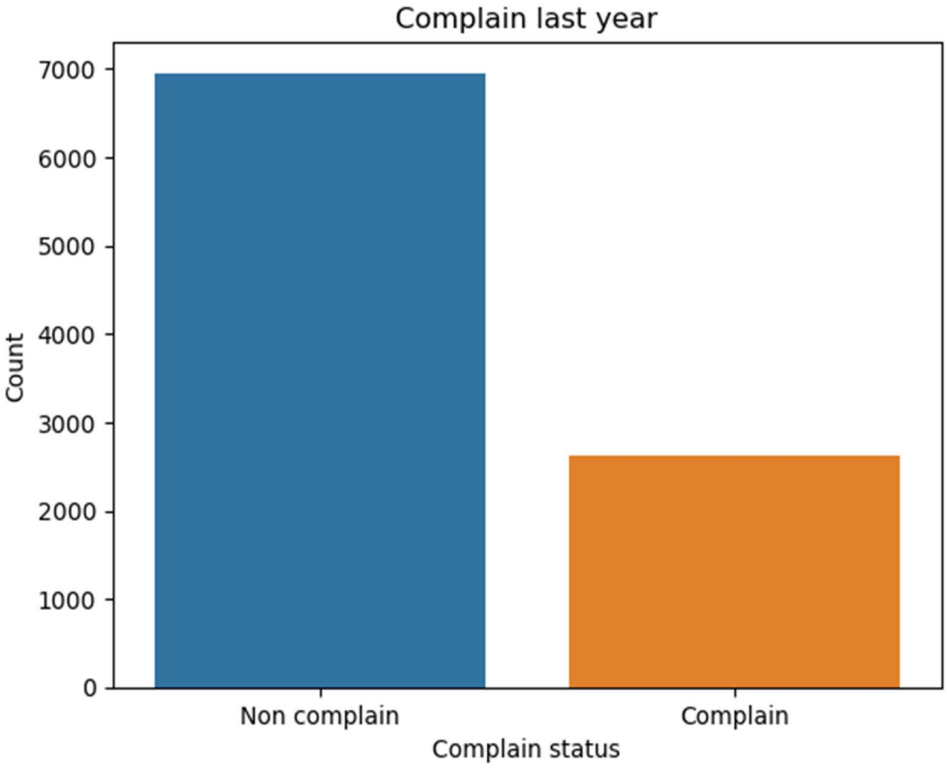

Figure 18 - distribution of Customers who complained last year

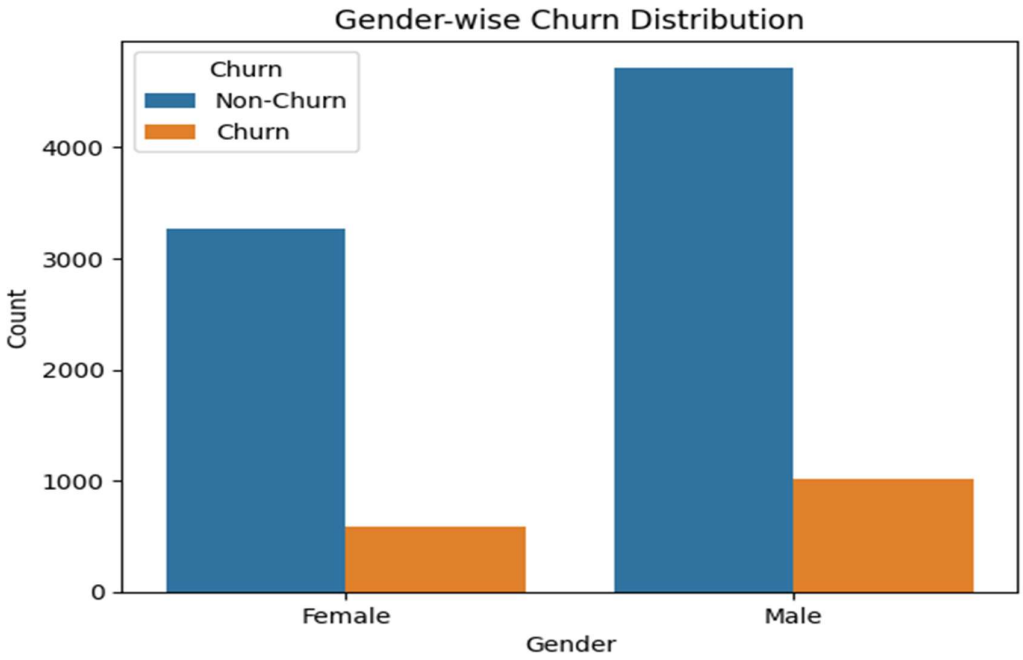

Figure 19 - distribution of churn vs Gender of the customer

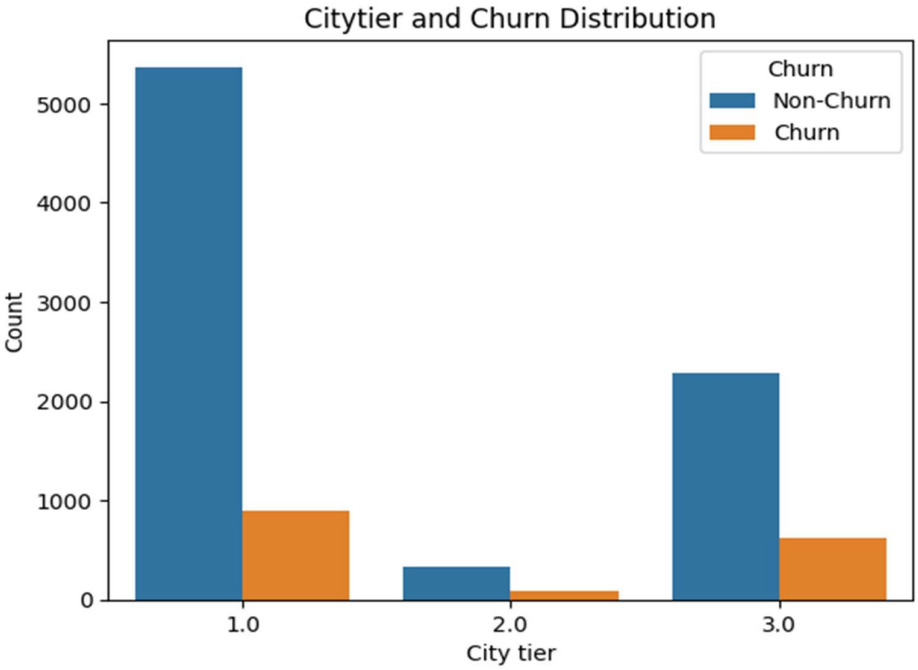

Figure 20 - distribution of Churn Vs city tier the accounts belongs.

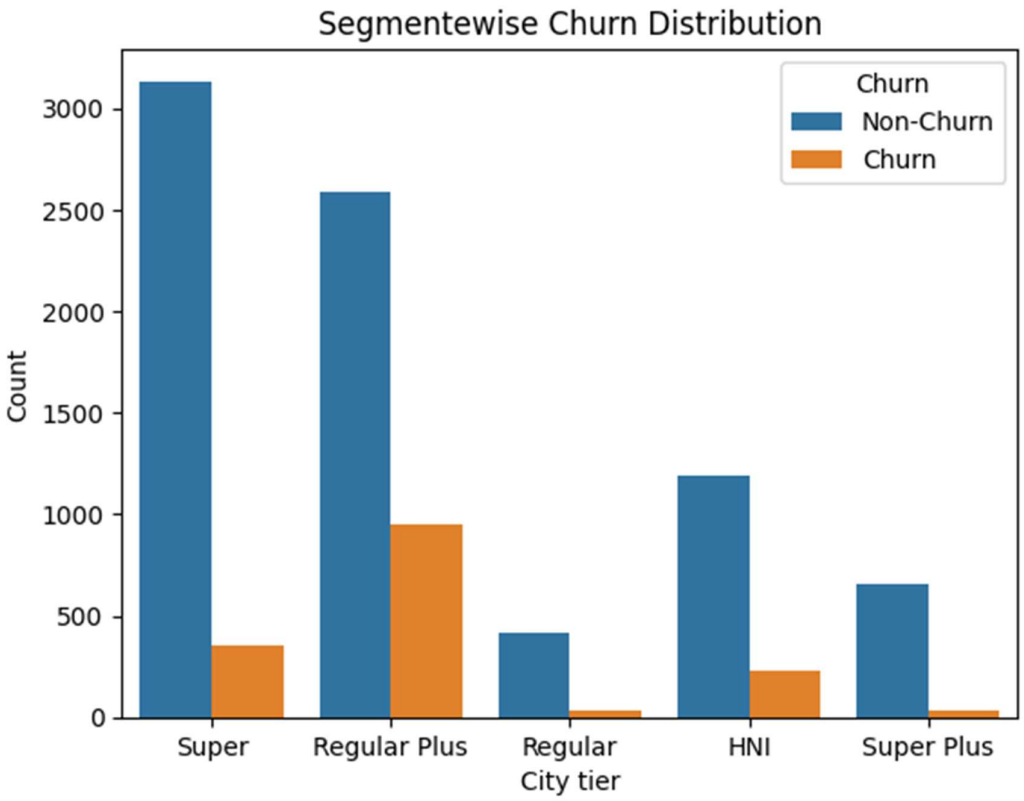

Figure 21 – Distribution of churn Vs customers across different account segments

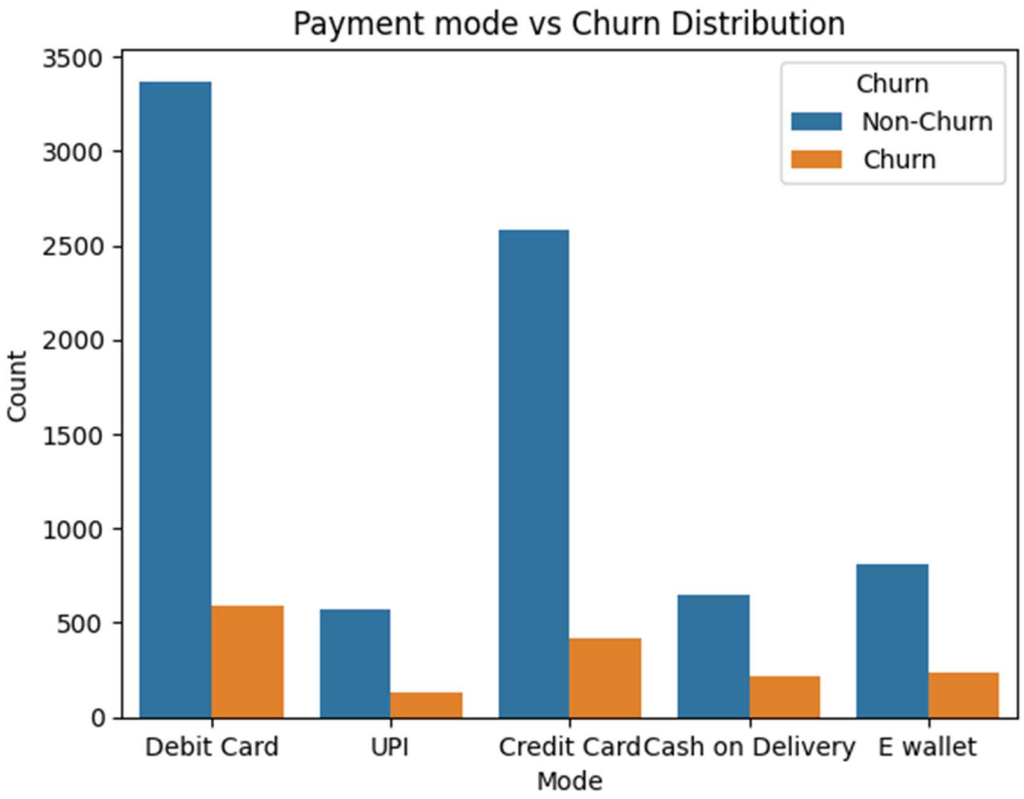

Figure 21 - Distribution churn Vs customers across different Payment

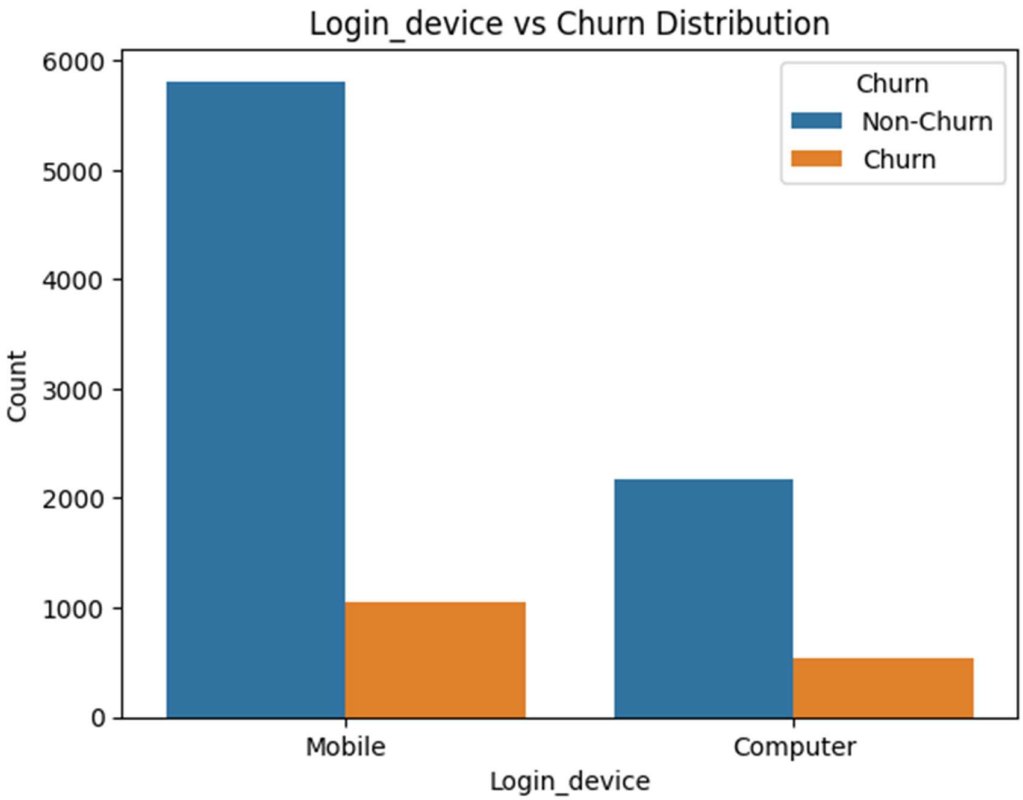

Figure 22 – distribution of churn vs device customers using.

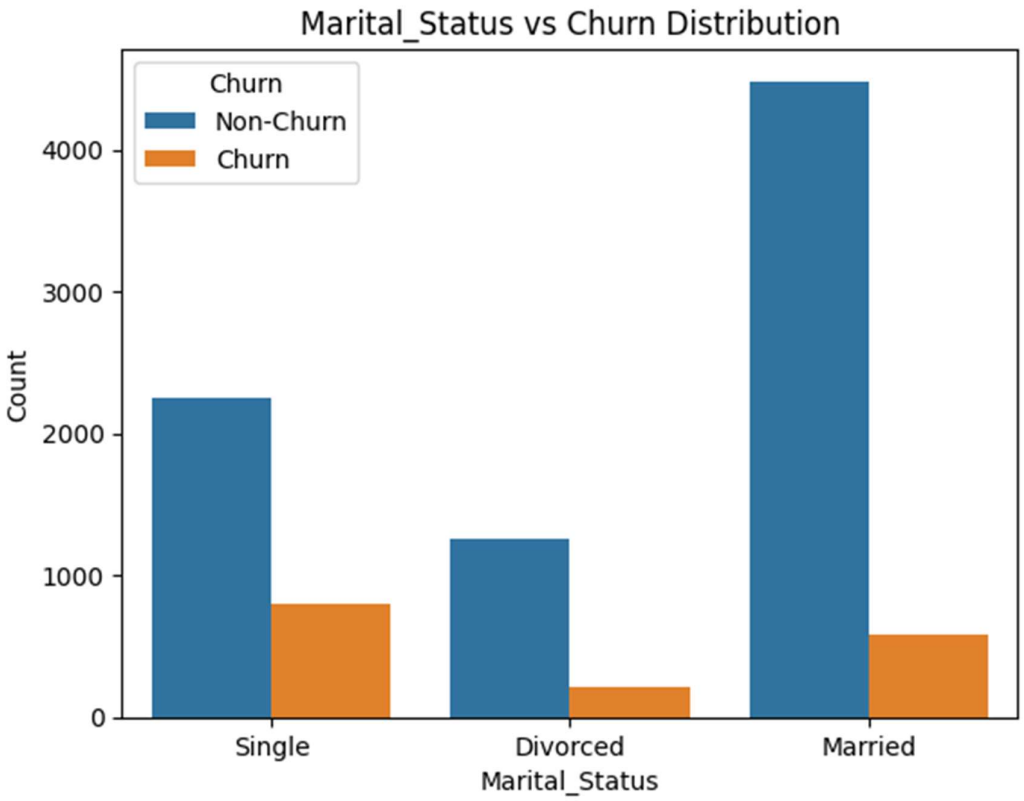

Figure 23 - distribution of churn Vs customer marital status

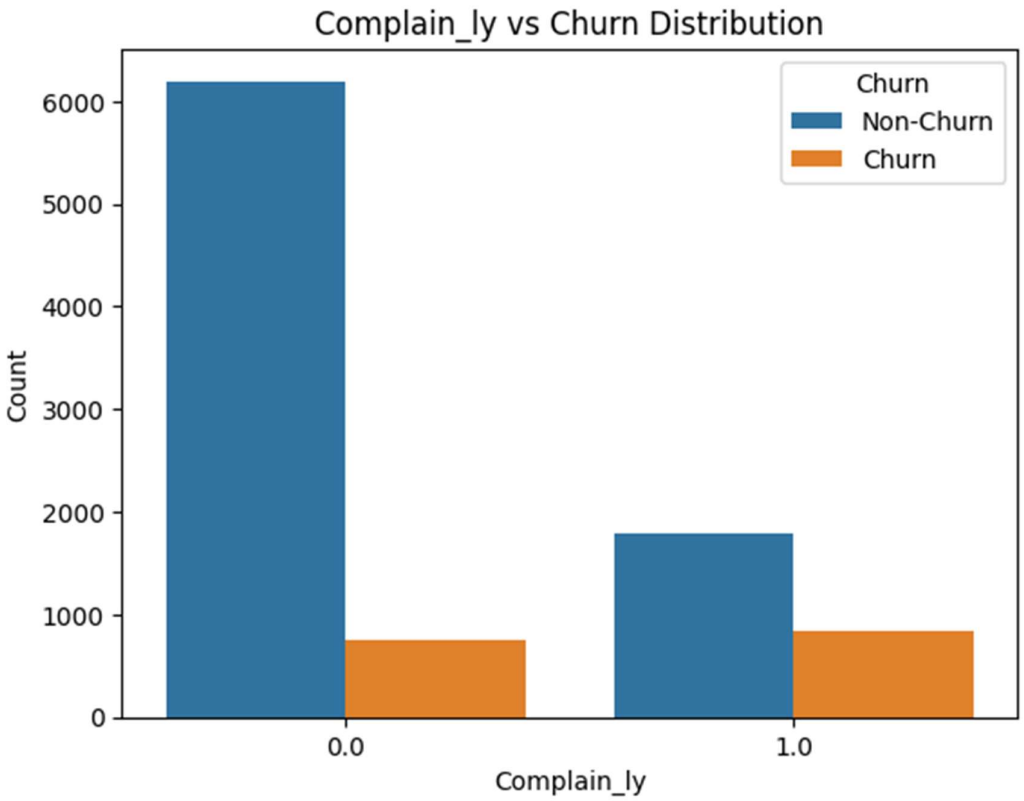

Figure 24 - Distribution of churn vs customer who complained last year.

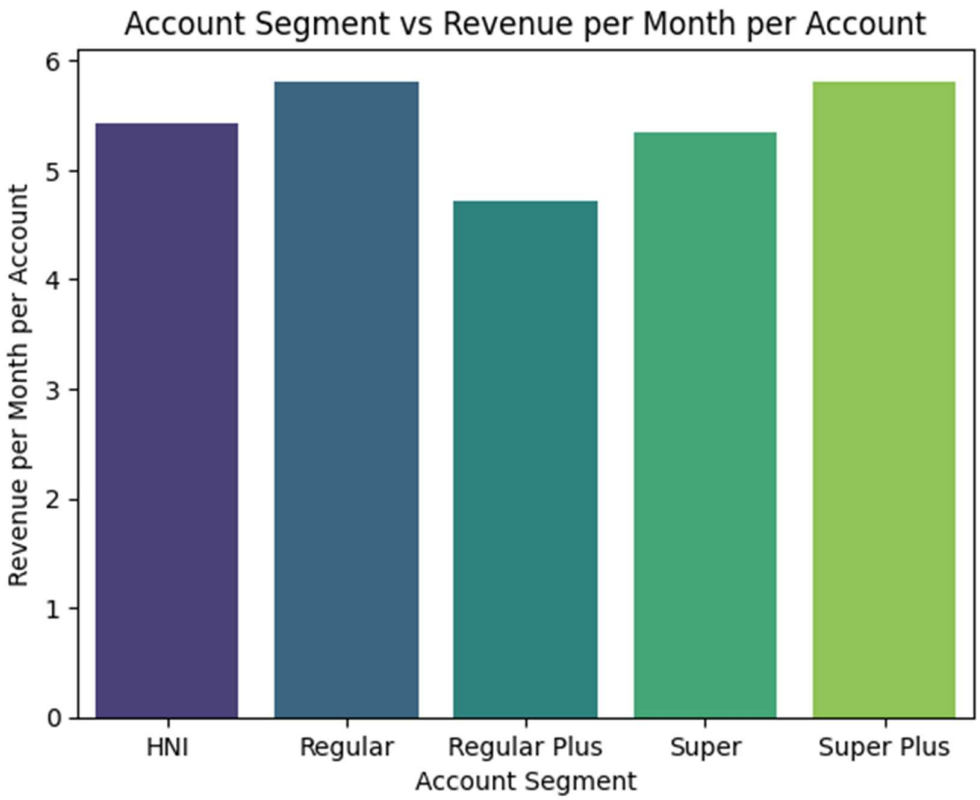

Figure 25 - distribution of average revenue per account across account segment

Exhibit 4 – Correlation matrix

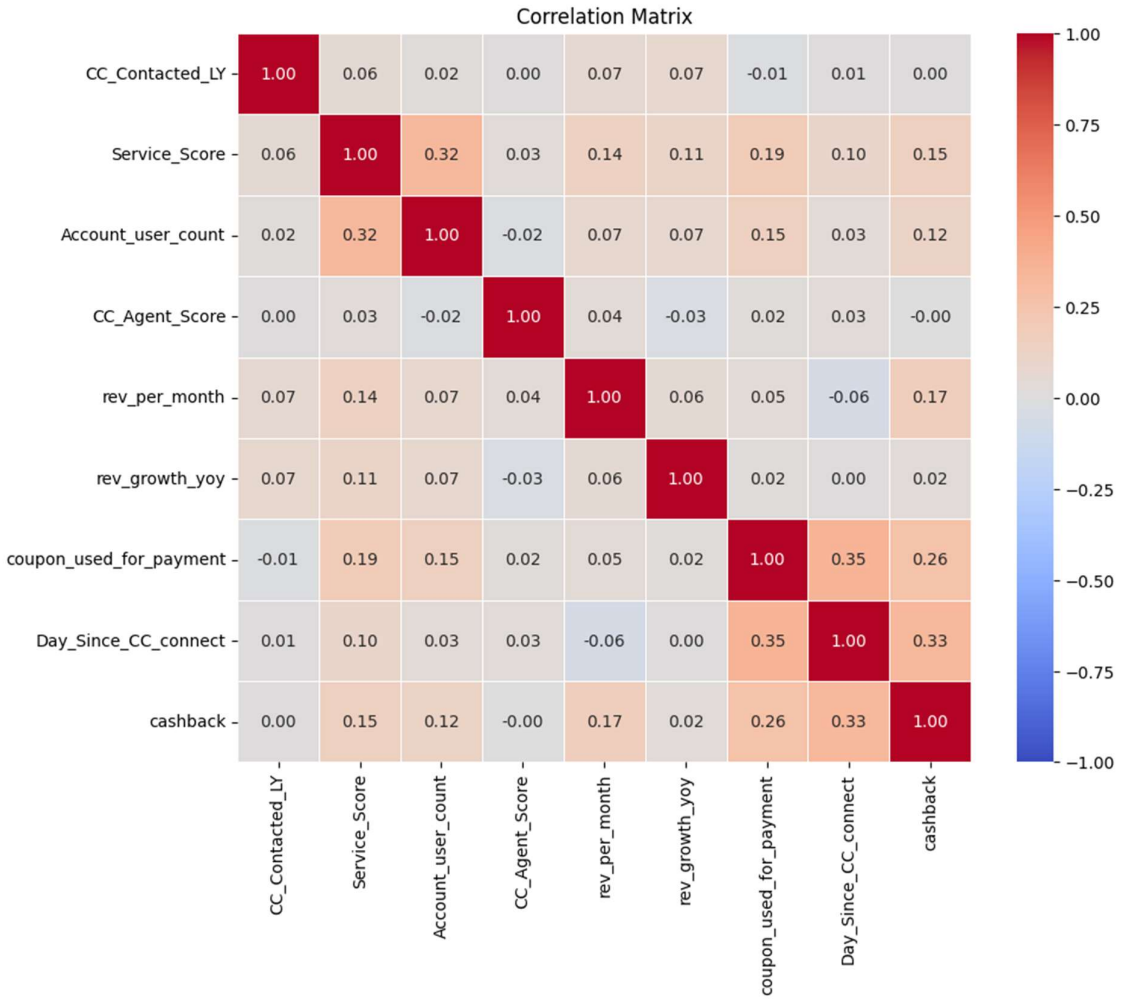

Figure 26 – Correlation matrix across all the potential numerical variables

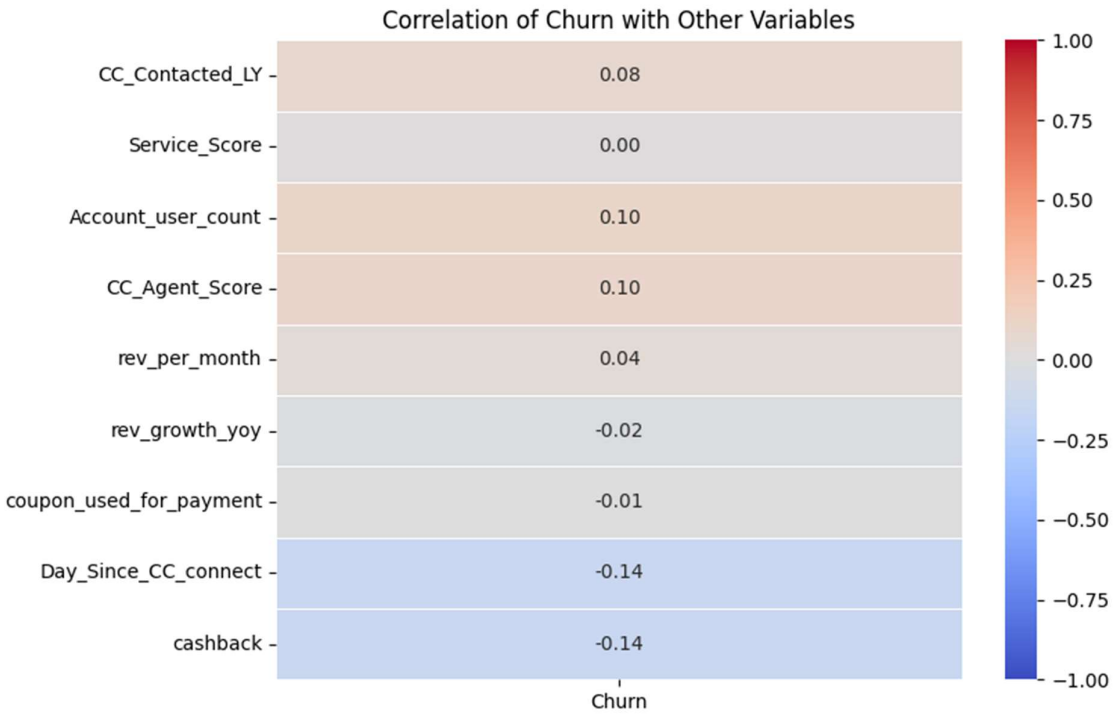

Figure 27 - Correlation of churn with other potential features.

Exhibit 5 – Model preprocessing

Converting categorical variables to factors

```
In [24]: # One-Hot Encode Categorical Variables

         data = pd.get_dummies(data, columns=['City_Tier', 'Payment', 'Gender','account_segment','Marital_Status', 'Complain_ly', 'Login_c

In [25]: print(data.columns.unique())

         Index(['AccountID', 'Churn', 'Tenure', 'CC_Contacted_LY', 'Service_Score',
                'Account_user_count', 'CC_Agent_Score', 'rev_per_month',
                'rev_growth_yoy', 'coupon_used_for_payment', 'Day_Since_CC_connect',
                'cashback', 'City_Tier_2.0', 'City_Tier_3.0', 'Payment_Credit Card',
                'Payment_Debit Card', 'Payment_E wallet', 'Payment_UPI', 'Gender_Male',
                'account_segment_Regular', 'account_segment_Regular Plus',
                'account_segment_Super', 'account_segment_Super Plus',
                'Marital_Status_Married', 'Marital_Status_Single', 'Complain_ly_1.0',
                'Login_device_Mobile'],
               dtype='object')
```

Code Snippet 1

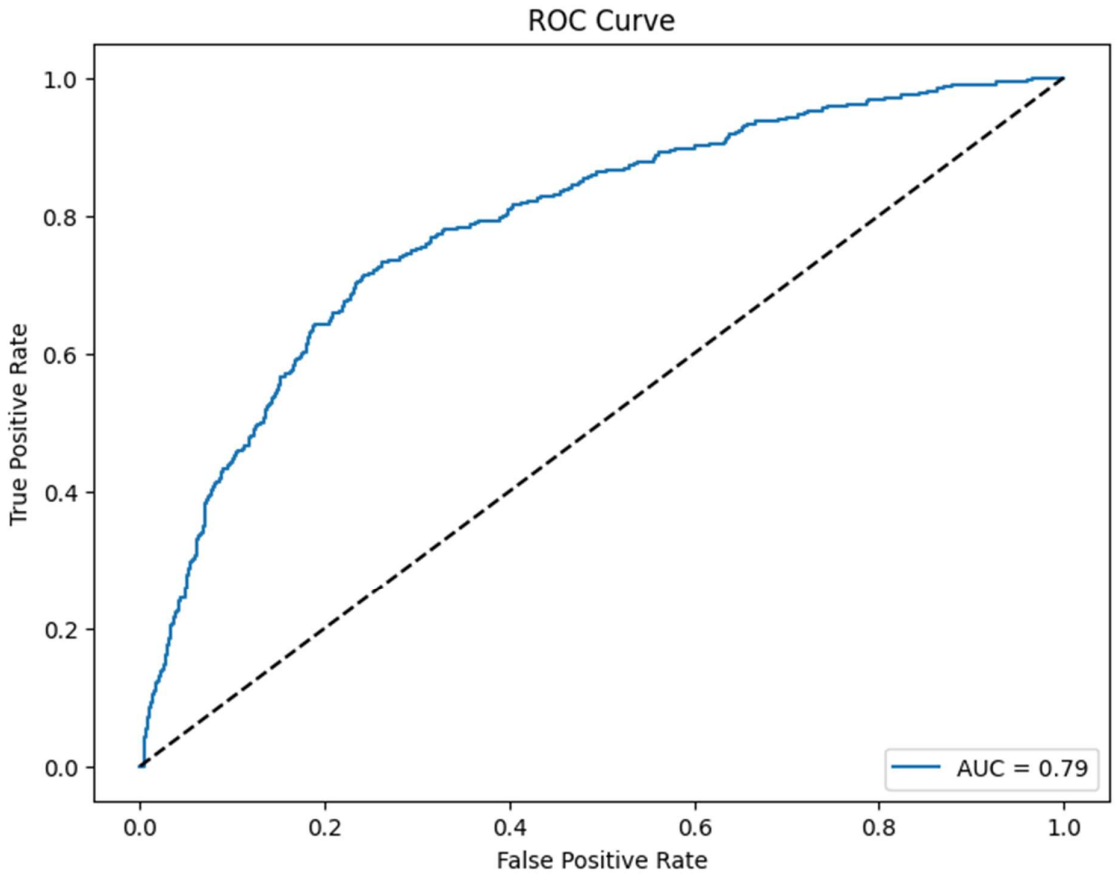

Figure 28 – ROC Curve for Logistic regression

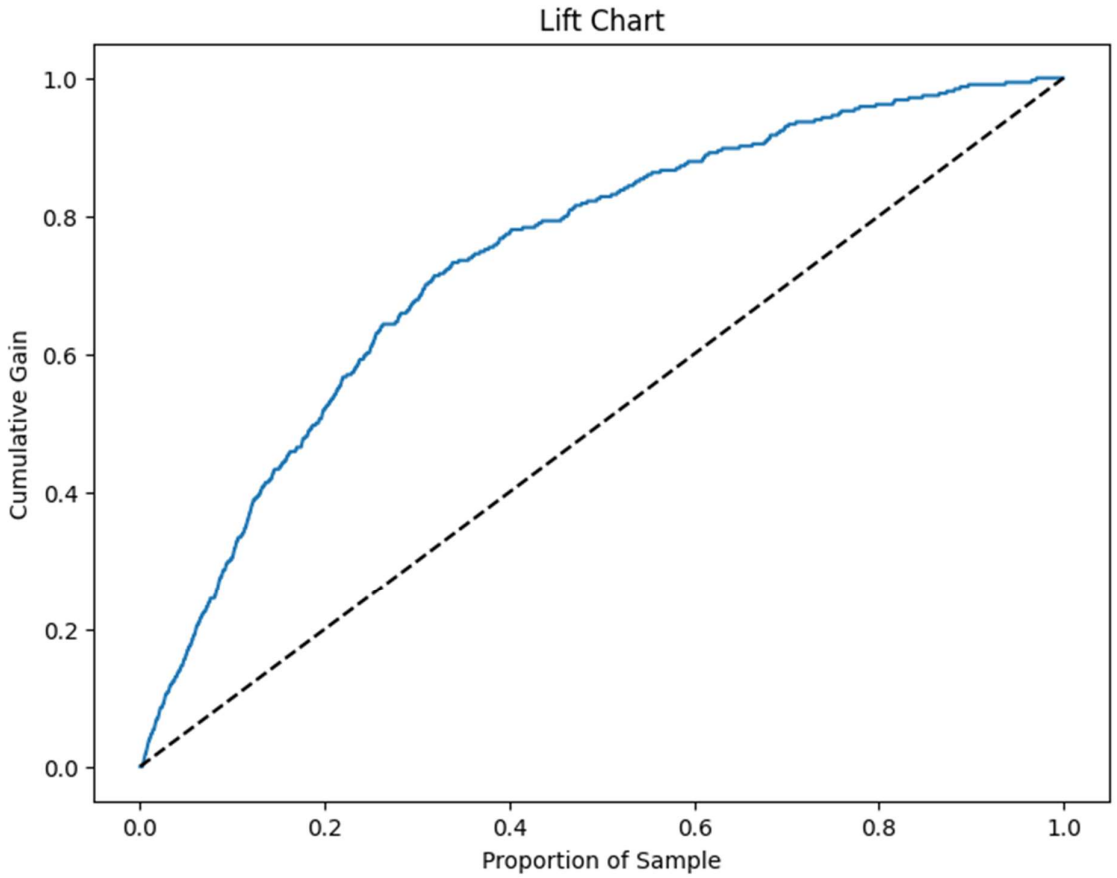

Figure 29 – LIFT curve for Logistic regression

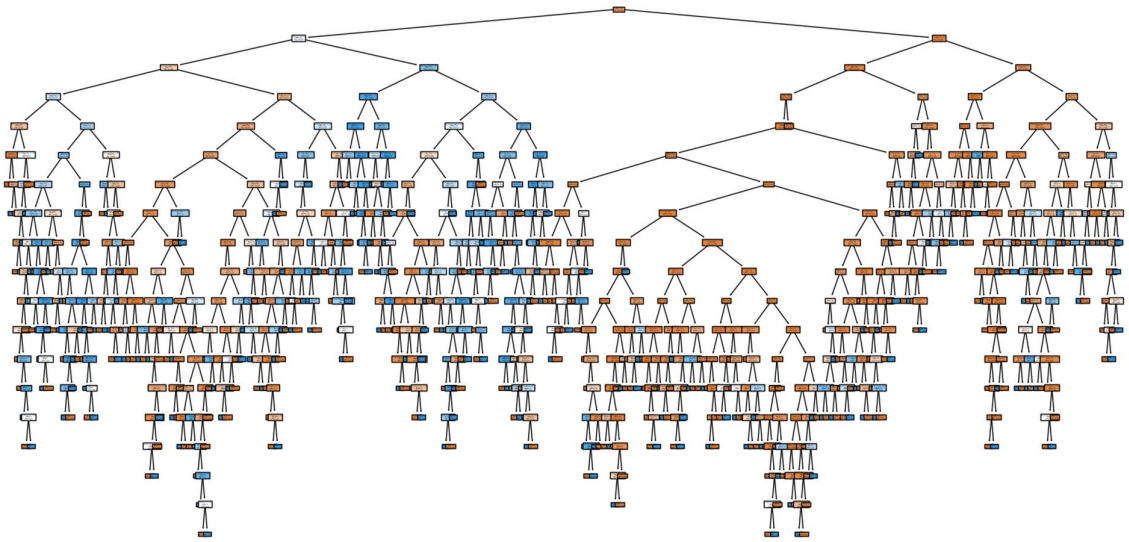

Figure 30 – Decision Tree

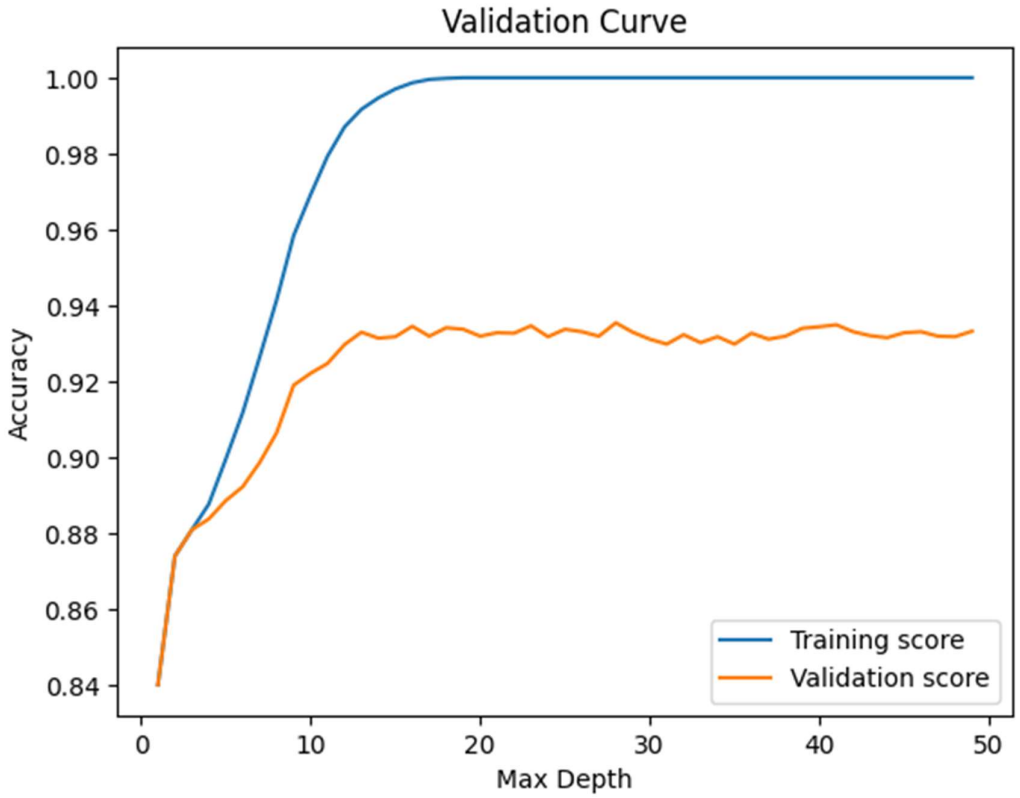

Figure 31 – Cross Validation for pruning (Maximum depth)

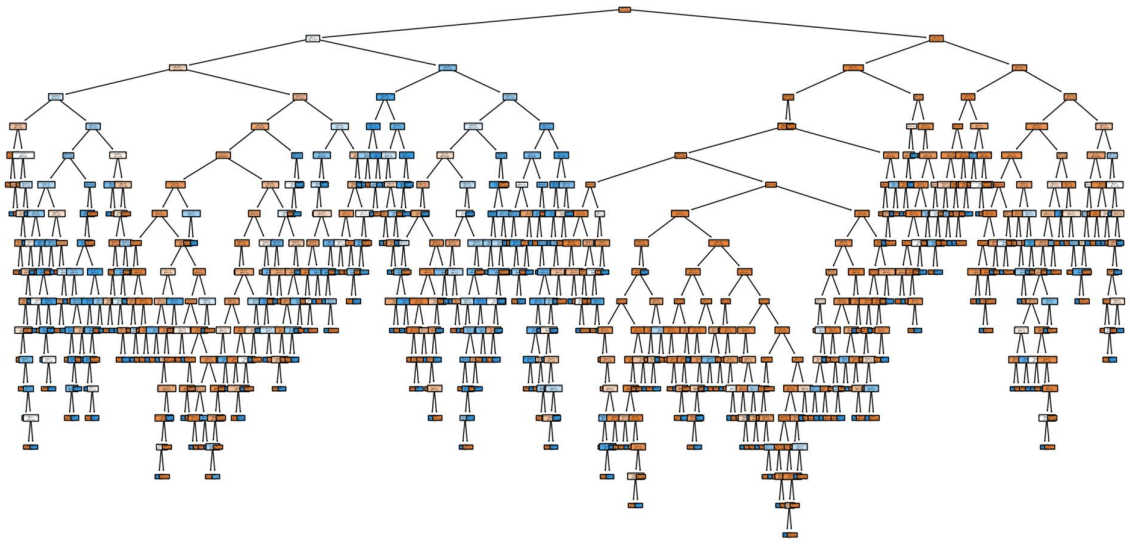

Figure 32 – Pruned decision Tree

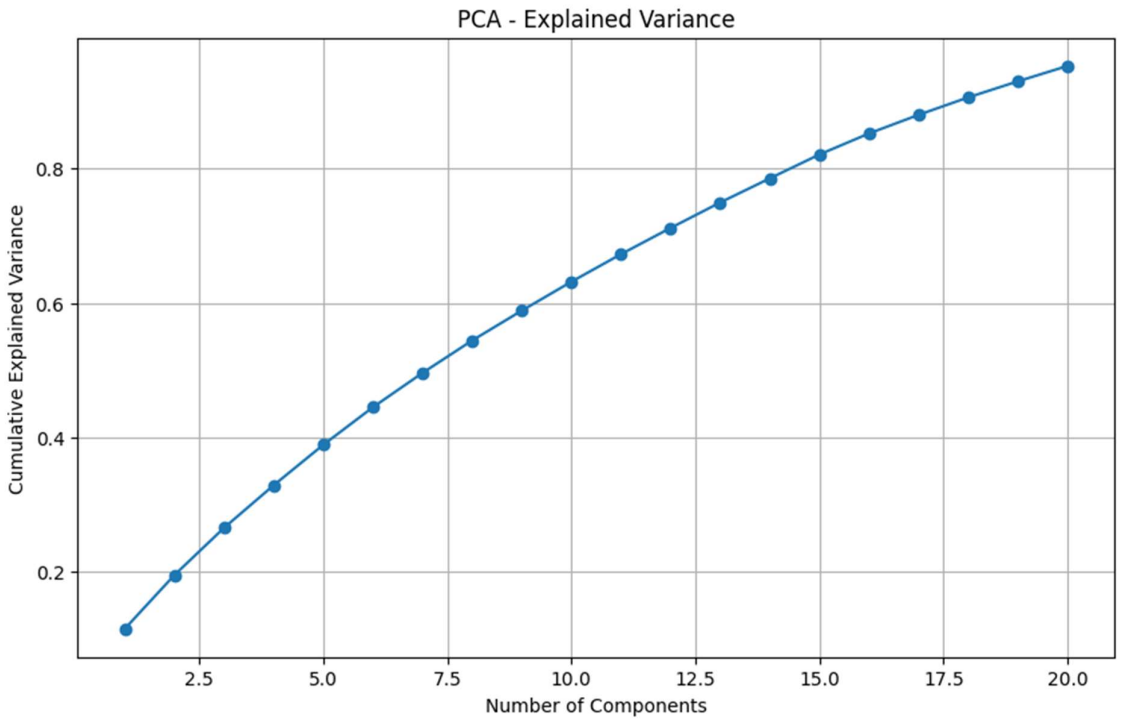

Figure 33 – PCA Curve

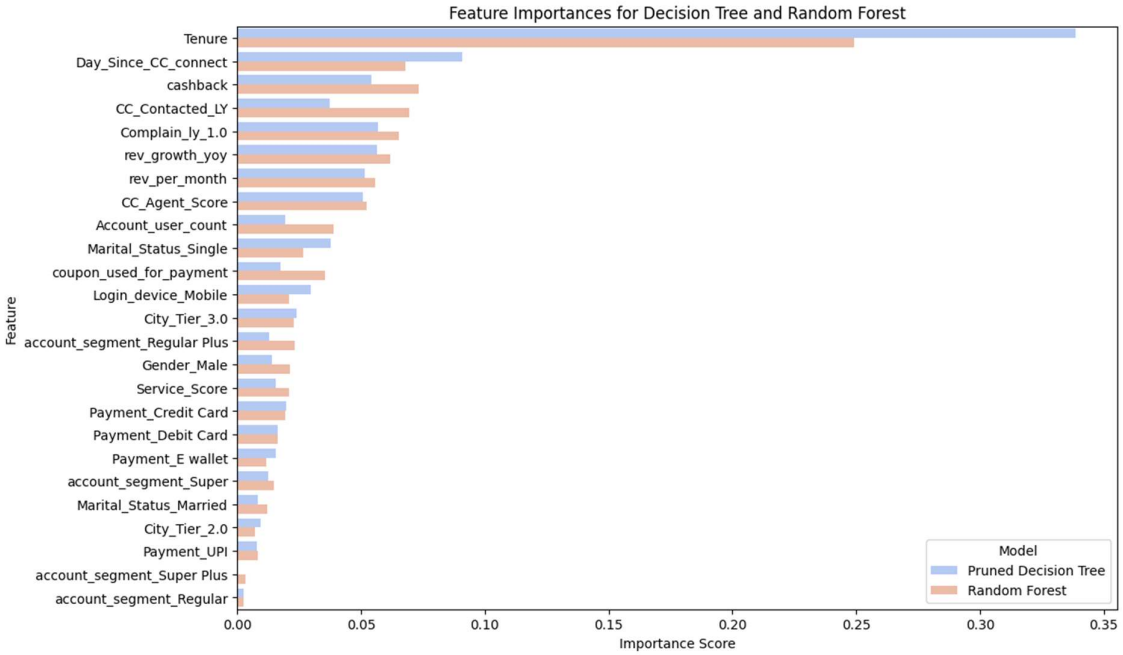

Figure 34 – Feature importance plot