greatlearning
Power Ahead

CUSTOMER CHURN PREDICTION

CAPSTONE PROJECT

BY NANCY GUPTA

Problem Statement :

An E Commerce company or DT is facing a lot of competition in the current market and it has become a challenge to retain the existing customers in the current situation. Hence, the company wants to develop a model through which they can do churn prediction of the accounts and provide segmented offers to the potential churners. In this company, account churn is a major thing because 1 account can have multiple customers. hence by losing one account the company might be losing more than one customer. You have been assigned to develop a churn prediction model for this company and provide business recommendations on the campaign. Your campaign suggestion should be unique and be very clear on the campaign offer because your recommendation will go through the revenue assurance team. If they find that you are giving a lot of free (or subsidized) stuff thereby making a loss to the company; they are not going to approve your recommendation. Hence be very careful while providing campaign recommendation.

Executive Summary of the Problem :

In the highly competitive E-Commerce/DTH market, customer retention has become a major challenge. The company is experiencing significant account churn, which impacts multiple customers per account. This report presents a **churn prediction model** and **data-driven campaign recommendations** aimed at reducing churn while maintaining profitability. The proposed strategy ensures minimal financial risk while maximizing customer retention.

- Customer churn leads to **revenue loss** and increased **customer acquisition costs**.
- Each **churned account** can impact **multiple customers**.
- The company requires a predictive **churn model** to identify at-risk accounts **before they leave**.
- A targeted **retention strategy** must be implemented, ensuring **cost-effectiveness** while maximizing retention.

Data Report

```
df.head()
```

	AccountID	Churn	Tenure	City_Tier	CC_Contacted_LY	Payment	Gender	Service_Score	Account_user_count	account_segment	CC_Agent_Score	Marital_Status	rev_per_month	Complain_ly	rev_growth_yoy	coupon
0	20000	1	4	3.0	6.0	Debit Card	Female	3.0	3	Super	2.0	Single	9	1.0	11	
1	20001	1	0	1.0	8.0	UPI	Male	3.0	4	Regular Plus	3.0	Single	7	1.0	15	
2	20002	1	0	1.0	30.0	Debit Card	Male	2.0	4	Regular Plus	3.0	Single	6	1.0	14	
3	20003	1	0	3.0	15.0	Debit Card	Male	2.0	4	Super	5.0	Single	8	0.0	23	
4	20004	1	0	1.0	12.0	Credit Card	Male	2.0	3	Regular Plus	5.0	Single	3	0.0	11	

- This is how the first 5 rows of the dataset look like.

```
no. of rows:  11260
no. of columns:  19
```

- There are 11260 rows and 19 columns in this Dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11260 entries, 0 to 11259
Data columns (total 19 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   AccountID                11260 non-null  int64
 1   Churn                    11260 non-null  int64
 2   Tenure                   11158 non-null  object
 3   City_Tier                11148 non-null  float64
 4   CC_Contacted_LY          11158 non-null  float64
 5   Payment                  11151 non-null  object
 6   Gender                   11152 non-null  object
 7   Service_Score            11162 non-null  float64
 8   Account_user_count       11148 non-null  object
 9   account_segment          11163 non-null  object
 10  CC_Agent_Score           11144 non-null  float64
 11  Marital_Status           11048 non-null  object
 12  rev_per_month            11158 non-null  object
 13  Complain_ly              10903 non-null  float64
 14  rev_growth_yoy           11260 non-null  object
 15  coupon_used_for_payment  11260 non-null  object
 16  Day_Since_CC_connect     10903 non-null  object
 17  cashback                 10789 non-null  object
 18  Login_device             11039 non-null  object
dtypes: float64(5), int64(2), object(12)
memory usage: 1.6+ MB
```

➢ Out of the 19 columns of the Dataset there are 5 columns which are float type, 2 columns integer type and 12 columns are found to be object type.

➢ This clearly explains that there are anomalies and missing values in the dataset that needs to be treated.

	0
cashback	471
Complain_ly	357
Day_Since_CC_connect	357
Login_device	221
Marital_Status	212
CC_Agent_Score	116
Account_user_count	112
City_Tier	112
Payment	109
Gender	108
Tenure	102
CC_Contacted_LY	102
rev_per_month	102
Service_Score	98
account_segment	97
Churn	0
AccountID	0
rev_growth_yoy	0
coupon_used_for_payment	0

dtype: int64

➢ The above picture gives us total number of null values or missing values each feature contains.

```
Number of duplicate rows = 259
```

➢ Further on we see the Dataset contains 259 duplicate rows that needs to be dropped.

➢ After dropping the duplicate rows, the number of rows in the dataset remains 11001 and the number of columns in the dataset remains 18.

➢ Our Target feature is the column “Churn” and out of 11001 accounts only 9149 accounts are active. 1852 accounts are already churned.

➢ The column “Tenure” has 116 ‘#’ values and I have replaced the ‘#’ values with 1 that is the maximum tenure value.

➢ Tier of Primary Customer’s City has 3 types – 1, 2 and 3.

- Customers use 5 types of payment methods. Debit Card is used the most by the customers followed by Credit card. E-Wallet comes next followed by Cash on Delivery. UPI is least preferred by the customers.
- We have 6548 Male customers and 4345 Female customers in the account.
- Satisfaction score given by customers of the account on service provided by company is majorly between 2 and 4. It is an average score which is not good for the company's reputation.
- Number of customers tagged with one account is majorly between 3 and 5.
- Account segmentation on the basis of spend has been done as Regular, Regular Plus, Super, Sper Plus and High-net-worth Individual (HNI). We see 4014 Regular plus accounts, 3961 Super accounts, 1615 HNI accounts, 803 Super Plus accounts and 511 Regular accounts.
- Satisfaction score distribution figure given by customers of the account on customer care service provided by company is below.

CC_Agent_Score	
3.0	3270
1.0	2261
5.0	2126
4.0	2064
2.0	1164

- The next picture is the count chart of Marital Status of the Customers.

Marital_Status	
Married	5710
Single	3412
Divorced	1668

Complain_ly	
0.0	7602
1.0	3042

- 3042 complains has been raised in the last 12 months.

Login_device	
Mobile	7308
Computer	2933
Others	539

- Customers mainly prefer Mobile to purchase the products, raise complaints and many more followed by computer.

➢ After dropping all the rows from the dataset that contains missing values now, we have

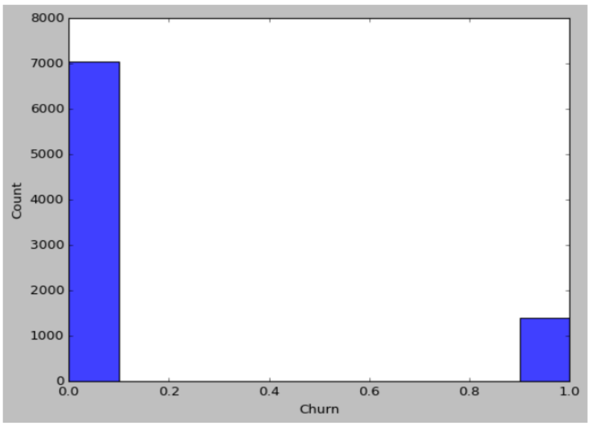
```
no. of rows:  8447
no. of columns:  18
```

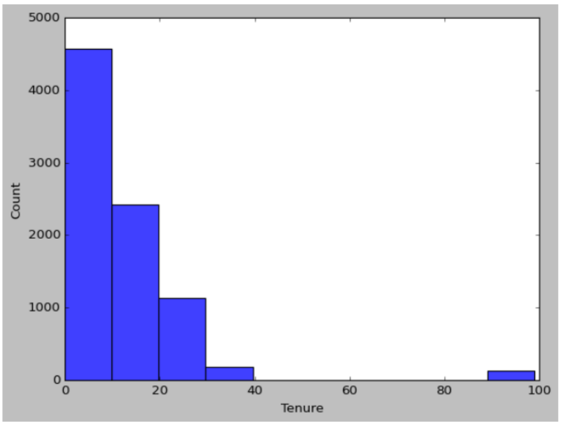
➢ Detailed Description of the cleaned Dataset

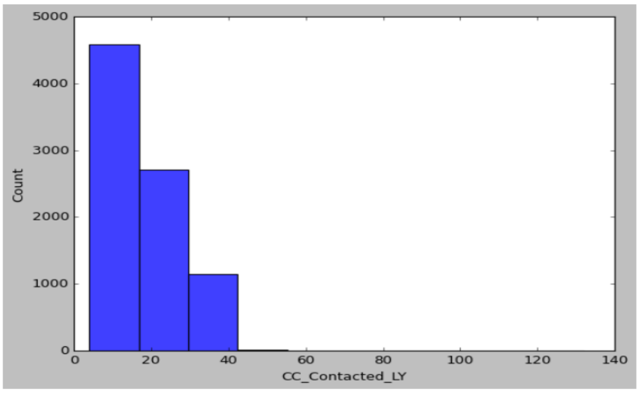
	count	mean	std	min	25%	50%	75%	max
Churn	8447.0	0.166213	0.372294	0.0	0.00	0.00	0.000	1.00
Tenure	8447.0	11.316444	13.949807	0.0	2.00	9.00	16.000	99.00
City_Tier	8447.0	1.656564	0.917293	1.0	1.00	1.00	3.000	3.00
CC_Contacted_LY	8447.0	17.919025	8.929066	4.0	11.00	16.00	23.000	132.00
Service_Score	8447.0	2.902451	0.727135	0.0	2.00	3.00	3.000	5.00
Account_user_count	8447.0	3.823369	1.200406	1.0	3.00	4.00	4.000	7.00
CC_Agent_Score	8447.0	3.047472	1.382974	1.0	2.00	3.00	4.000	5.00
rev_per_month	8447.0	6.575589	13.167177	1.0	3.00	5.00	7.000	140.00
Complain_ly	8447.0	0.283533	0.450739	0.0	0.00	0.00	1.000	1.00
rev_growth_yoy	8447.0	16.205162	3.764174	4.0	13.00	15.00	19.000	28.00
coupon_used_for_payment	8447.0	1.812715	2.005490	0.0	1.00	1.00	2.000	16.00
Day_Since_CC_connect	8447.0	4.659642	3.706566	0.0	2.00	3.00	8.000	47.00
cashback	8447.0	179.384497	49.631708	0.0	147.15	165.11	199.255	331.26

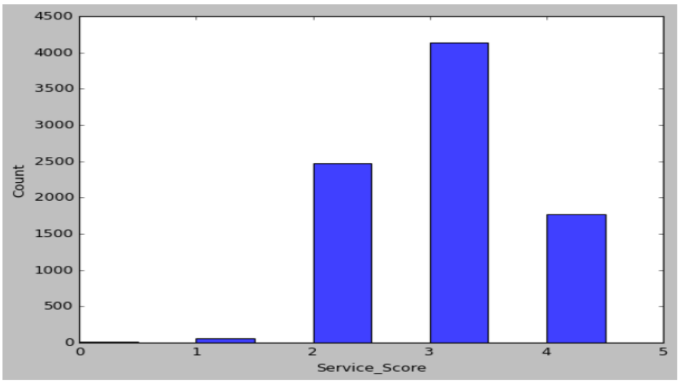
Exploratory Data Analysis

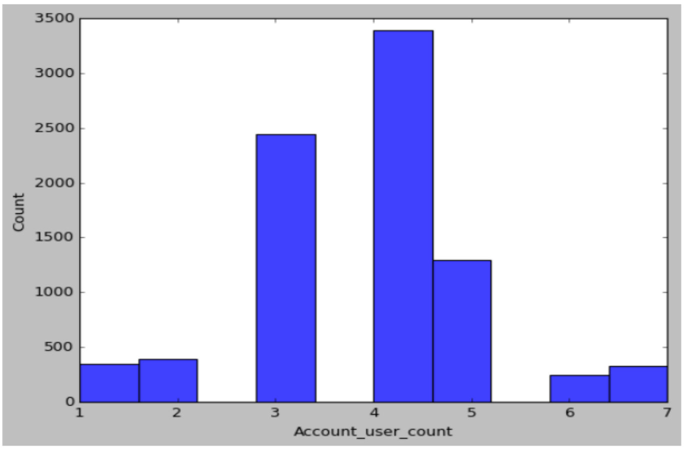
Univariate Analysis: Below are the histograms of all the numerical variables of the Dataset.

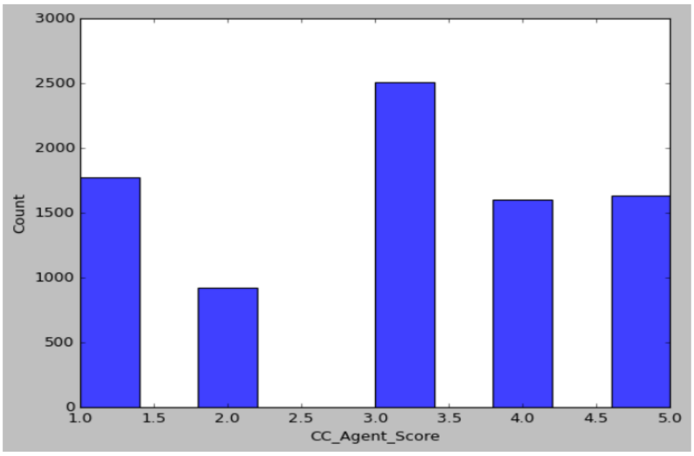


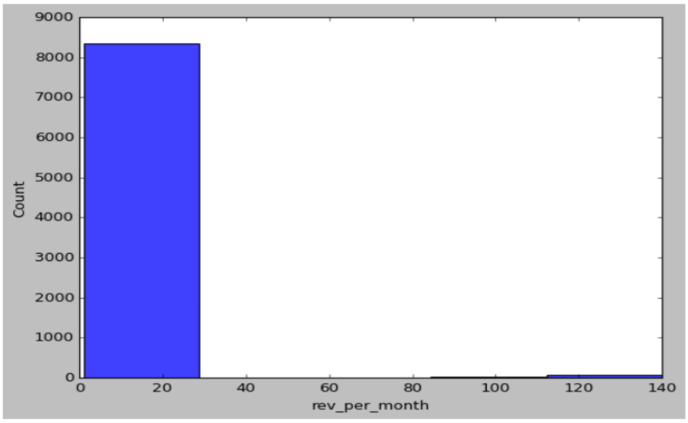


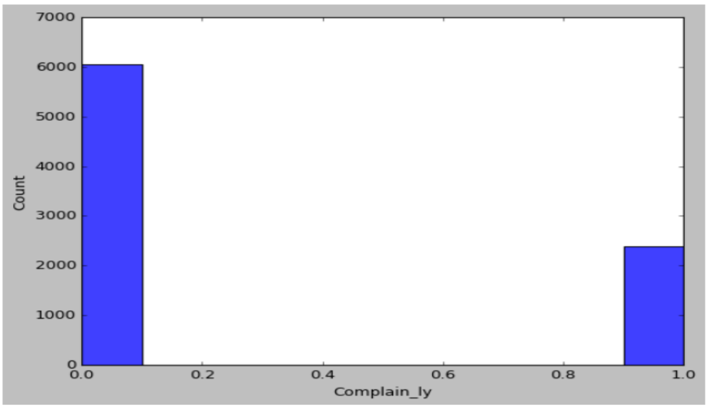

5000
4000
3000
2000
1000
0
Count
0 20 40 60 80 100 120 140
CC_Contacted_LY

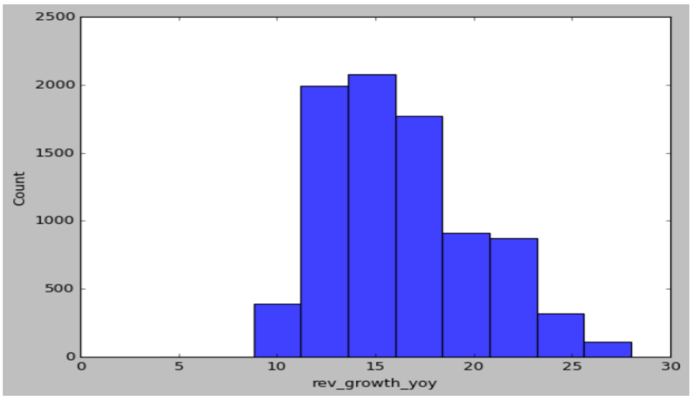

4500
4000
3500
3000
2500
2000
1500
1000
500
0
Count
0 1 2 3 4 5
Service_Score

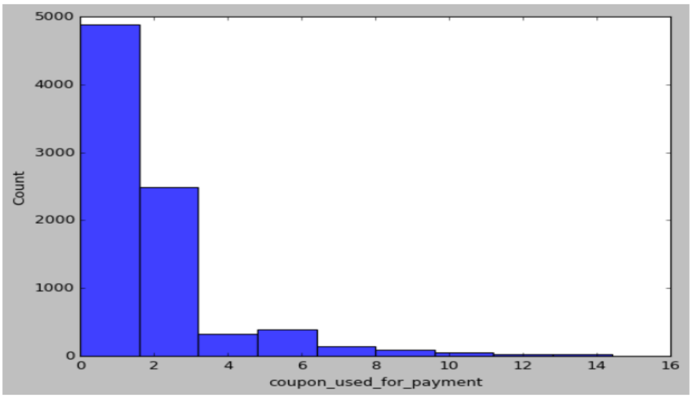

3500
3000
2500
2000
1500
1000
500
0
Count
1 2 3 4 5 6 7
Account_user_count

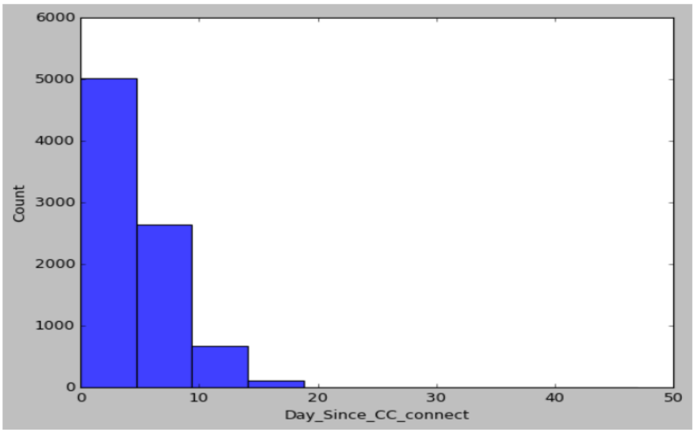

3000
2500
2000
1500
1000
500
0
Count
1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
CC_Agent_Score

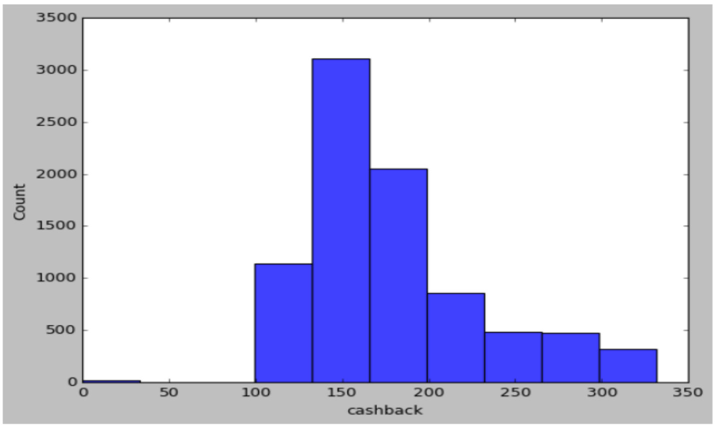

9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Count
0 20 40 60 80 100 120 140
rev_per_month


7000
6000
5000
4000
3000
2000
1000
0
Count
0.0 0.2 0.4 0.6 0.8 1.0
Complain_ly

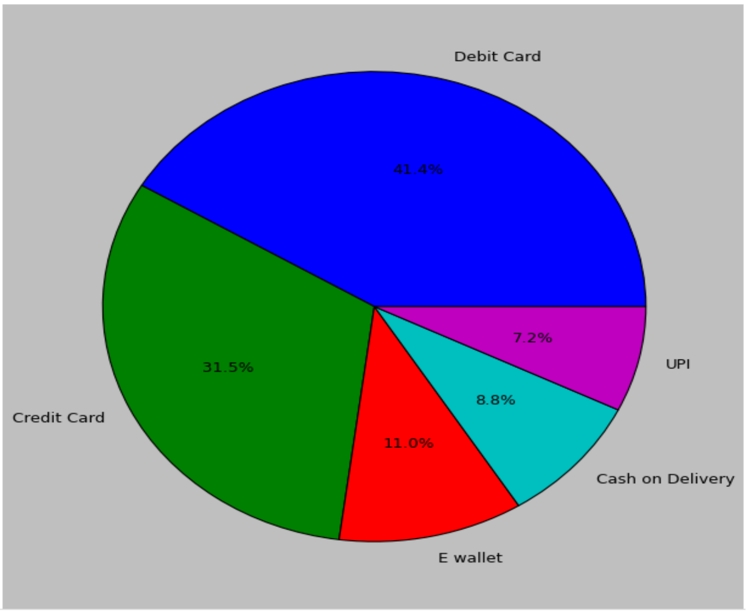


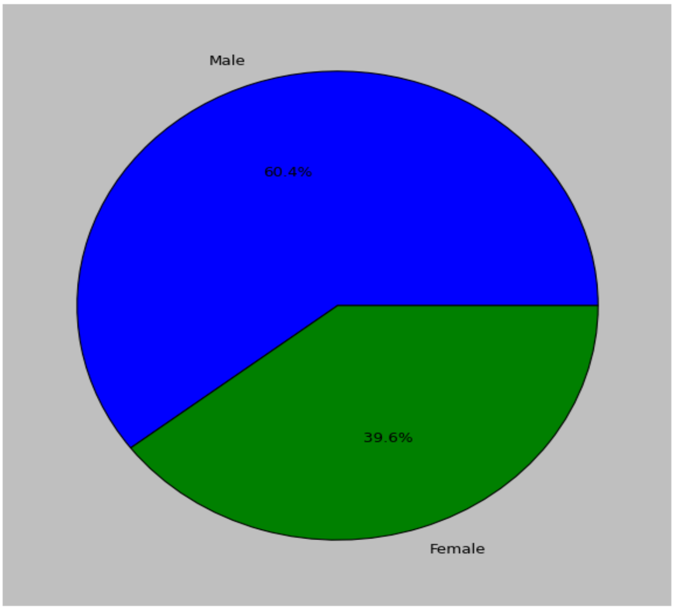


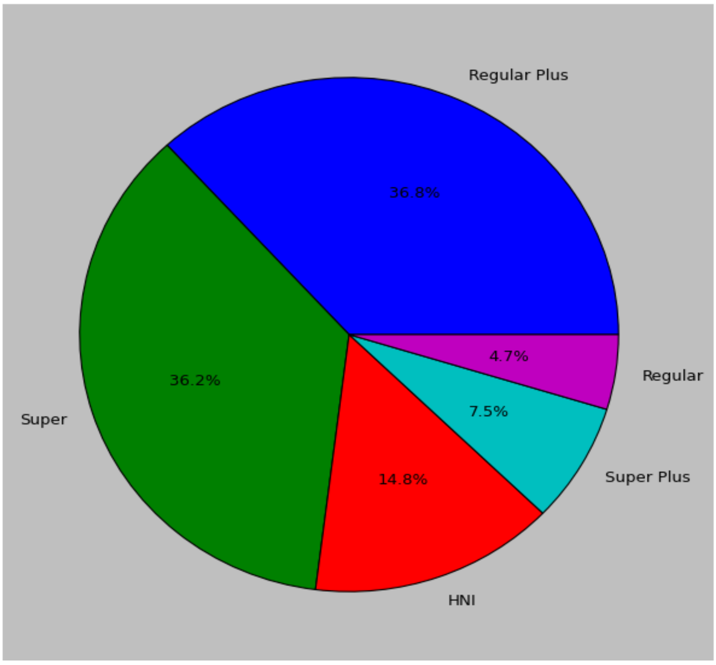


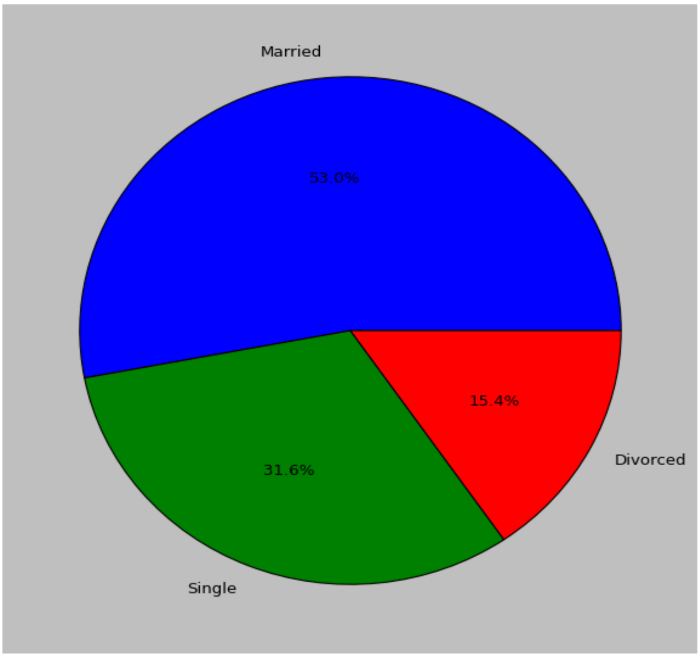


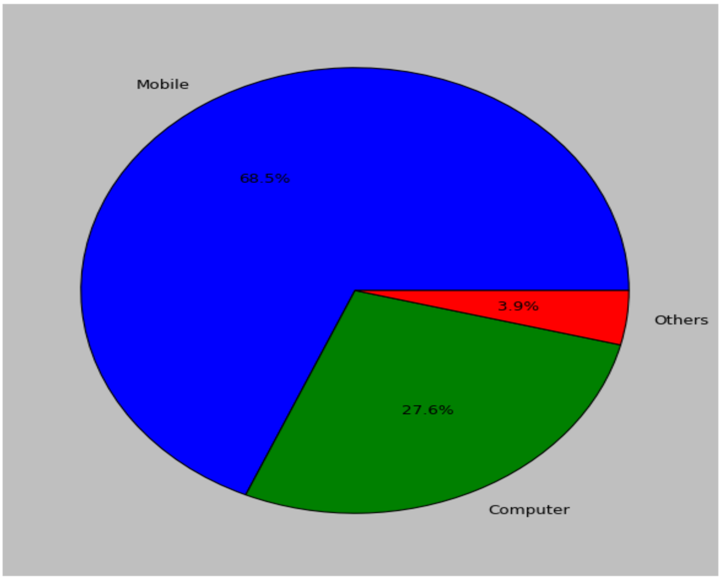
➢ Below are the pie charts of all the categorical variables in the Dataset.







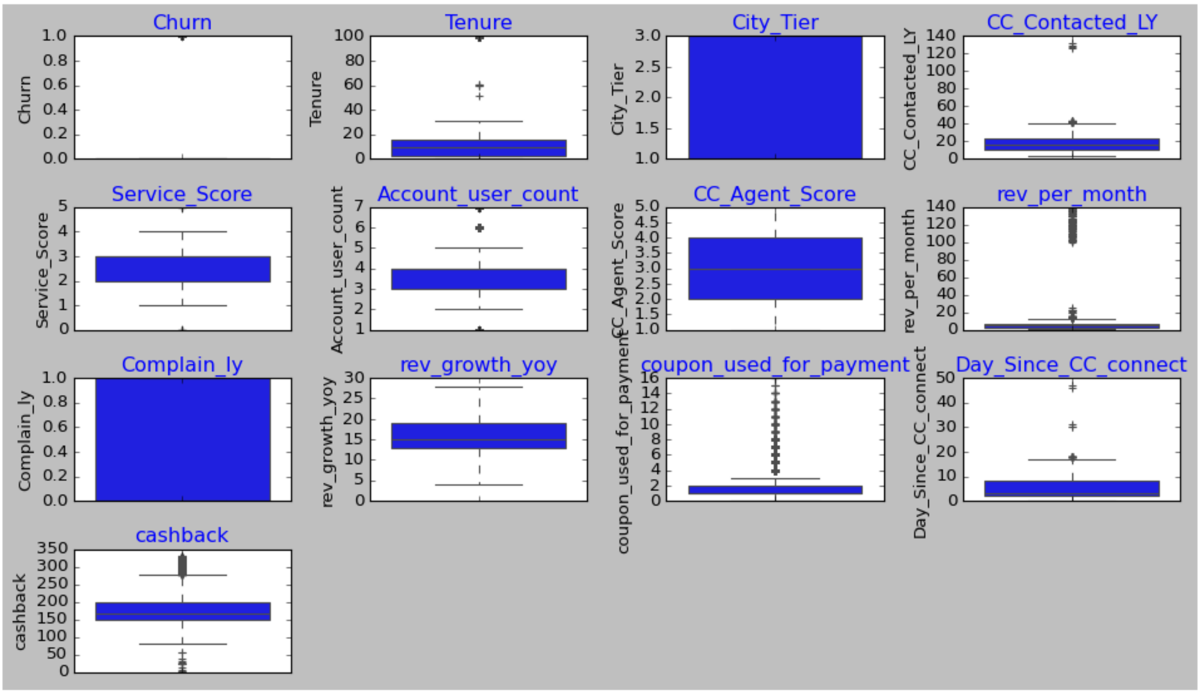




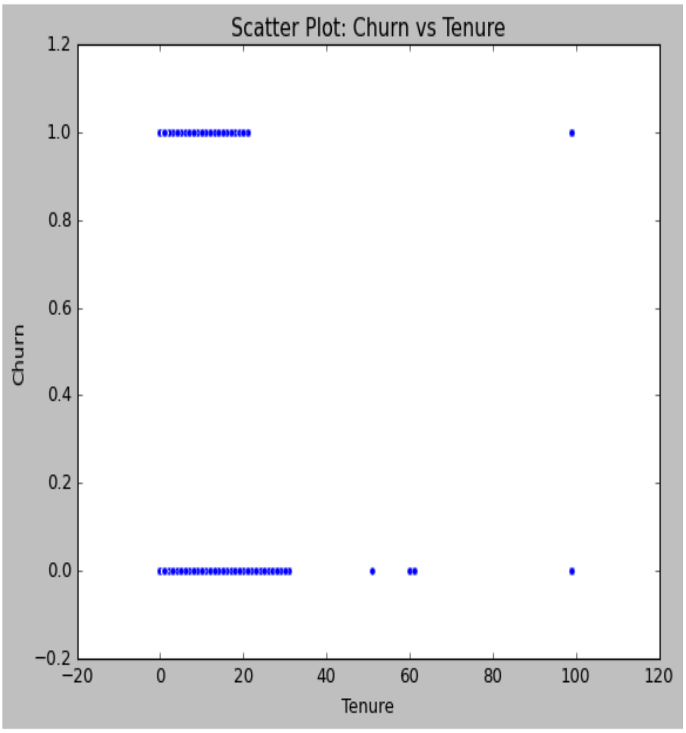
Insights

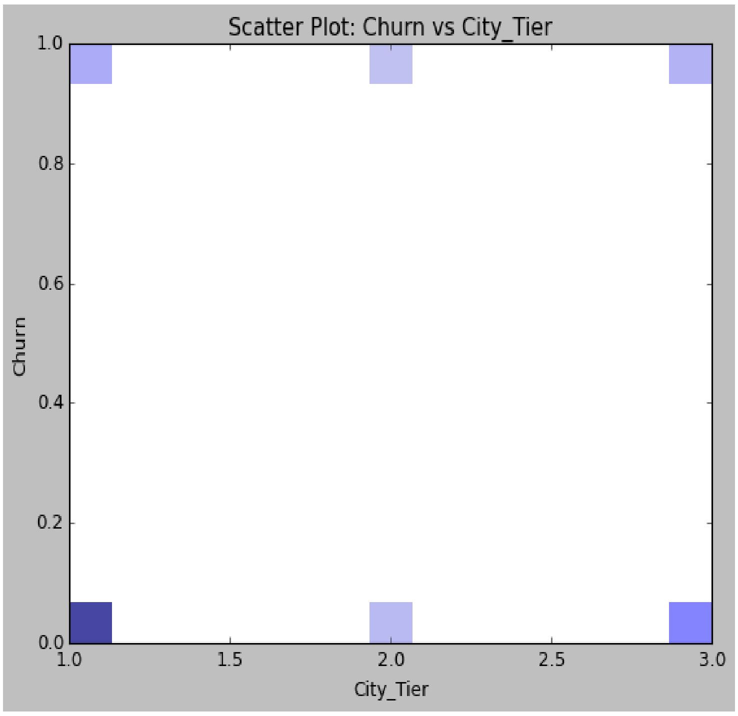
- From the above diagrams we find 68.5% of customers use mobile to login.
- 73% of the account holders are Regular Plus and Super customers combined.
- 53% of the customers are married and 31.6% are Single.
- Customers use Debit card and Credit card mainly to make the payments.
- Percentage of Male customers are higher than Female Customers.

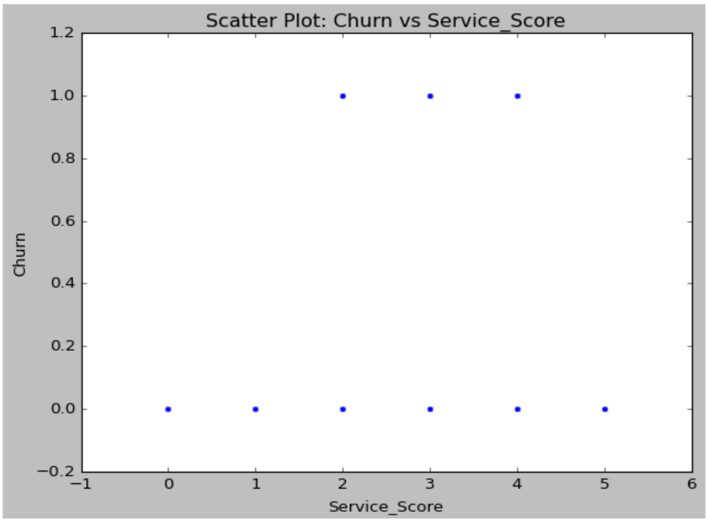
Bivariate Analysis:

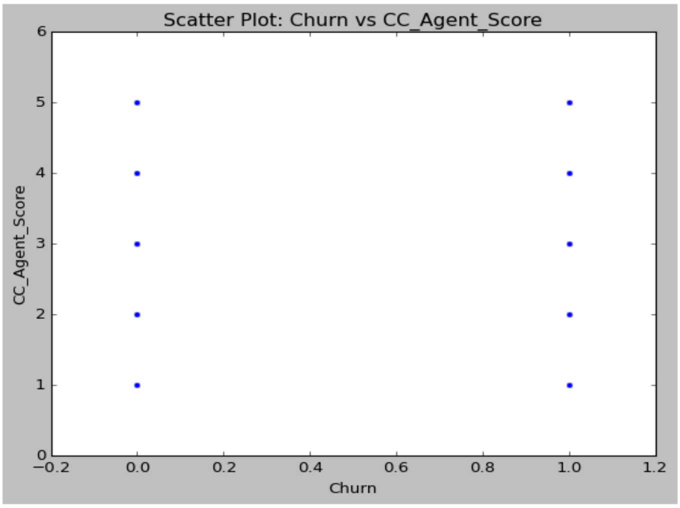


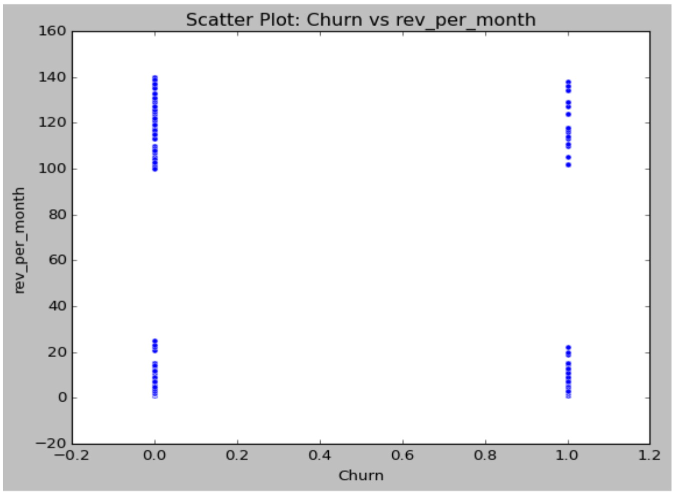
Boxplot of all the numerical features of the Dataset.

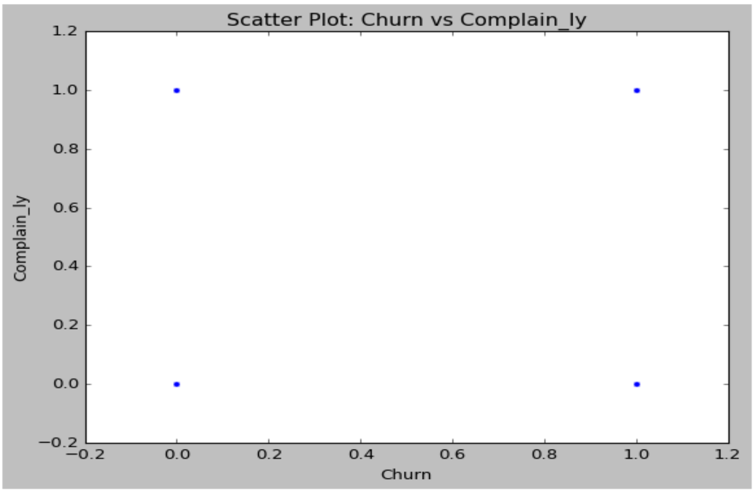


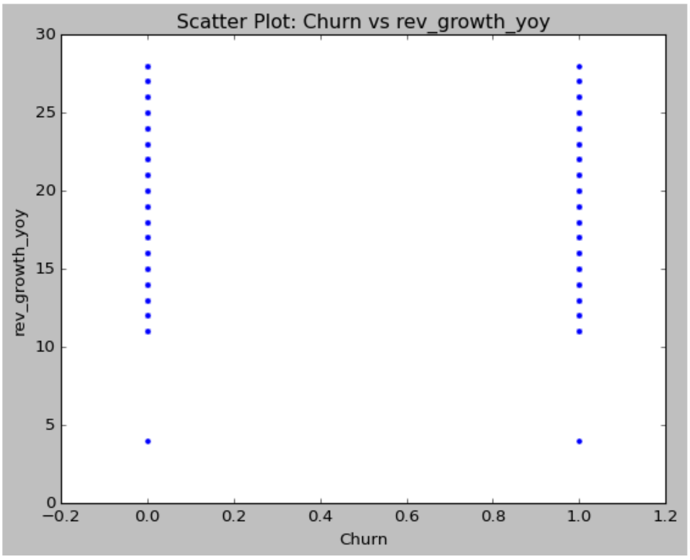


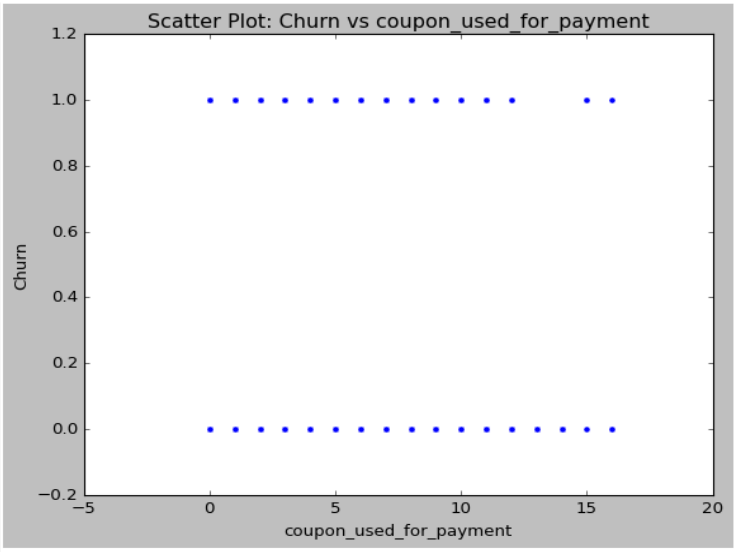

Scatter Plot: Churn vs Service_Score
Churn
Service_Score

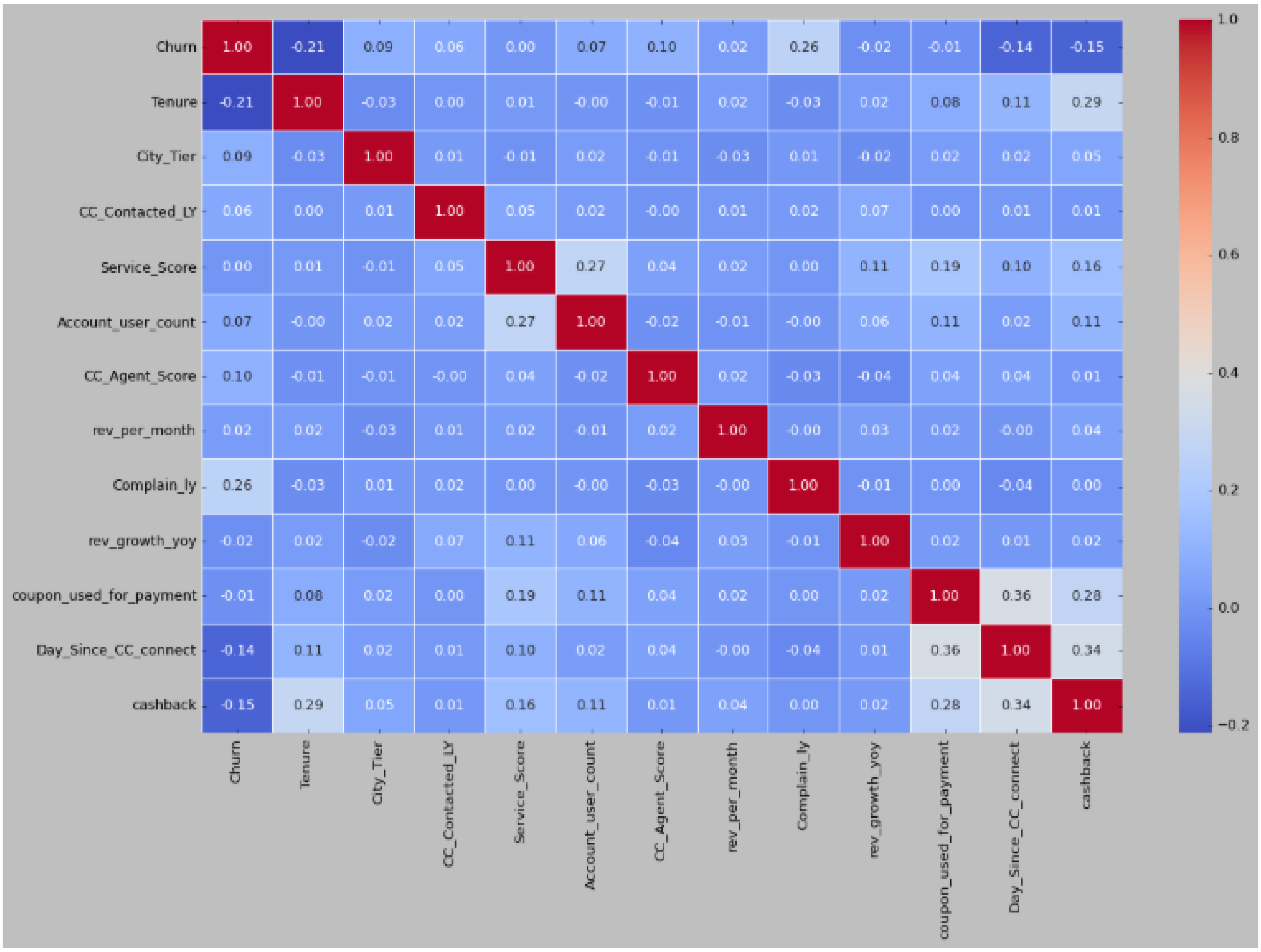

Scatter Plot: Churn vs CC_Agent_Score
CC_Agent_Score
Churn


Scatter Plot: Churn vs rev_per_month
rev_per_month
Churn


Scatter Plot: Churn vs Complain_ly
Complain_ly
Churn

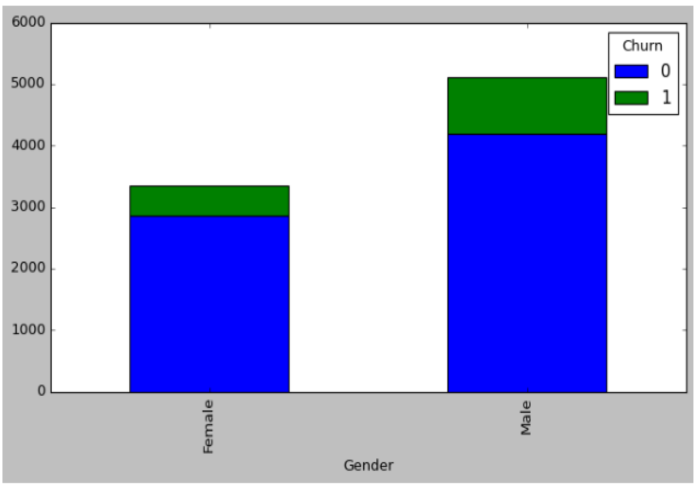

Scatter Plot: Churn vs rev_growth_yoy
rev_growth_yoy
Churn

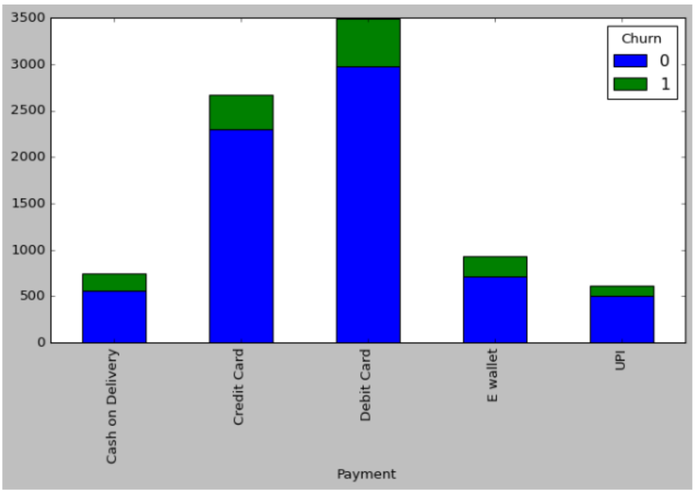

Scatter Plot: Churn vs coupon_used_for_payment
Churn
coupon_used_for_payment

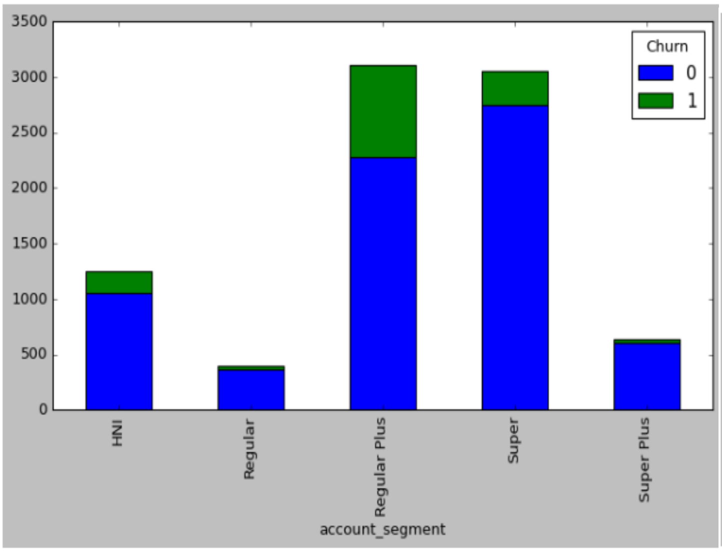


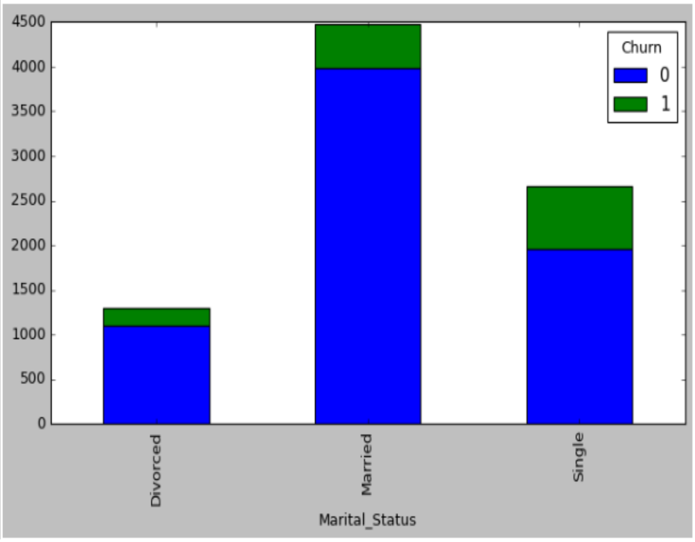
Correlation matrix

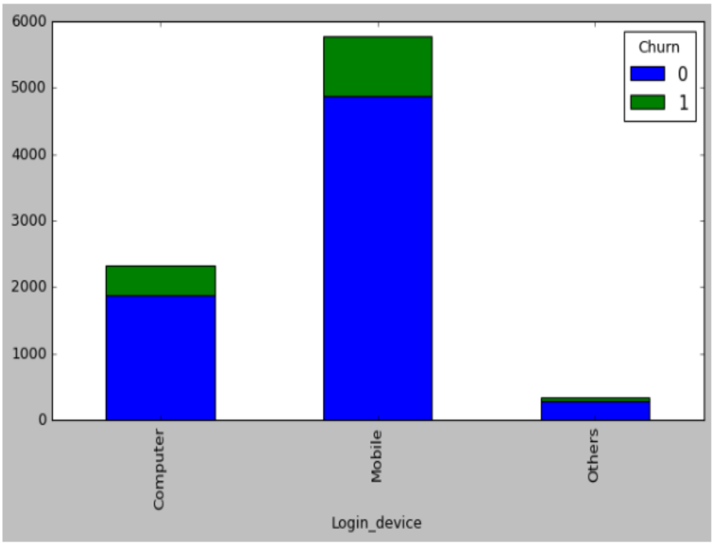
➢ Below are the crosstabs, scatter plots and Violin Plots of different features:

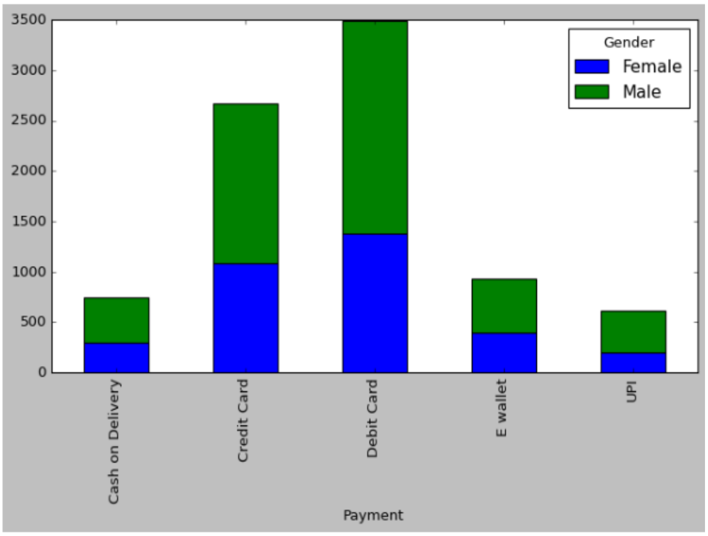


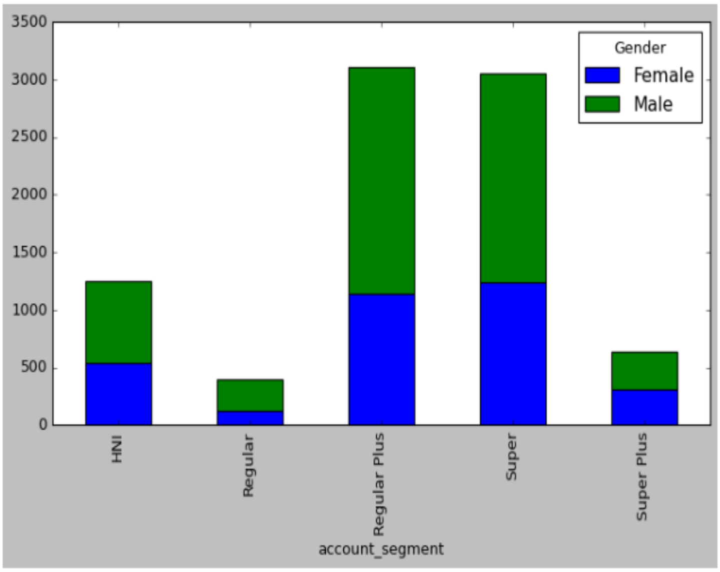


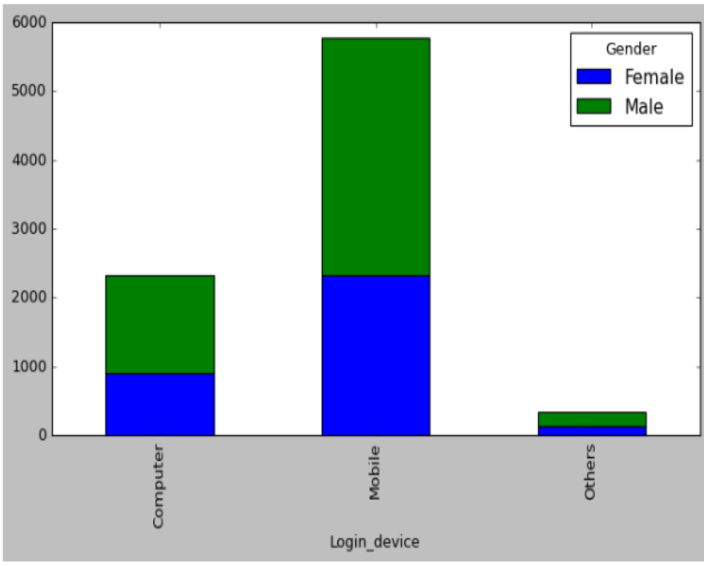

3500
3000
2500
2000
1500
1000
500
0
Churn
0
1
HNI
Regular
Regular Plus
Super
Super Plus
account_segment

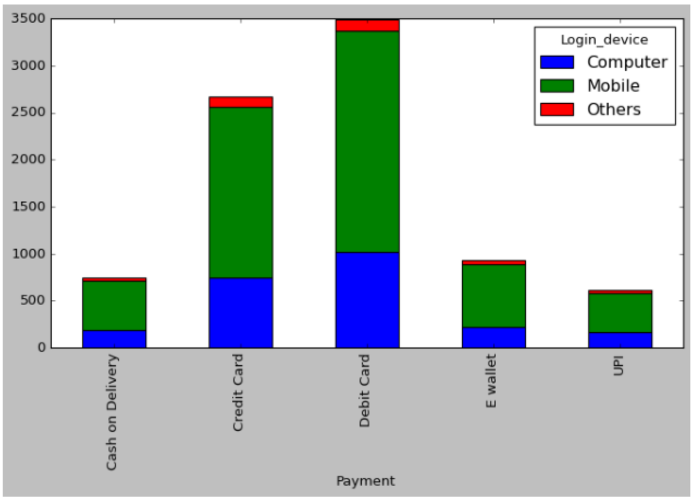

4500
4000
3500
3000
2500
2000
1500
1000
500
0
Churn
0
1
Divorced
Married
Single
Marital_Status

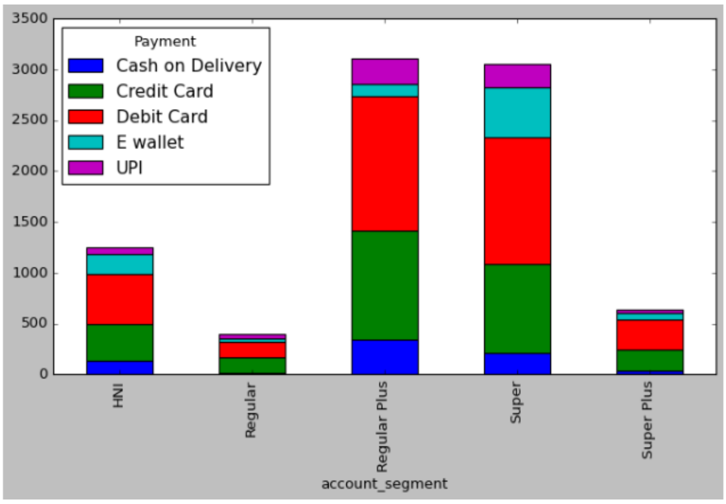

6000
5000
4000
3000
2000
1000
0
Churn
0
1
Computer
Mobile
Others
Login_device

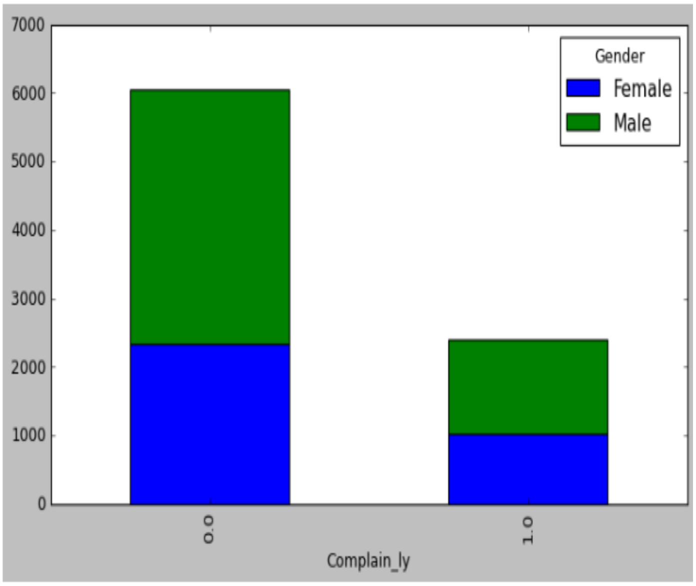

3500
3000
2500
2000
1500
1000
500
0
Gender
Female
Male
Cash on Delivery
Credit Card
Debit Card
E wallet
UPI
Payment


3500
3000
2500
2000
1500
1000
500
0
Gender
Female
Male
HNI
Regular
Regular Plus
Super
Super Plus
account_segment

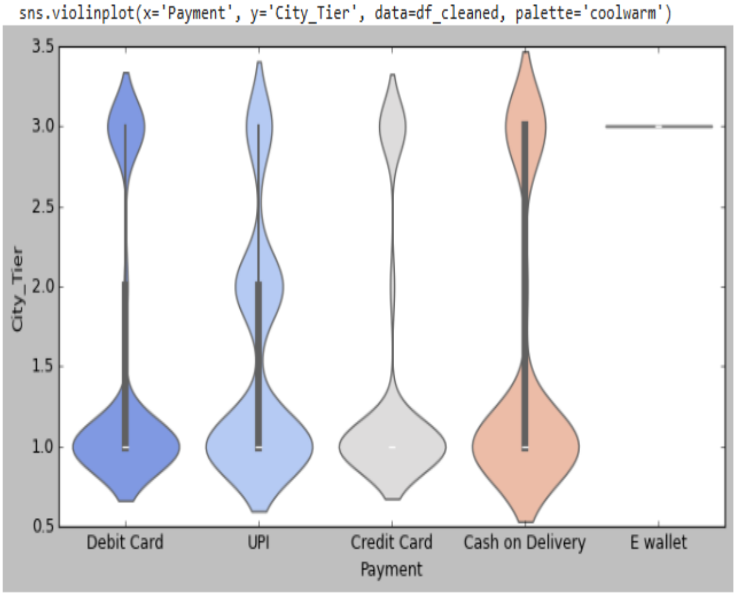

6000
5000
4000
3000
2000
1000
0
Gender
Female
Male
Computer
Mobile
Others
Login_device



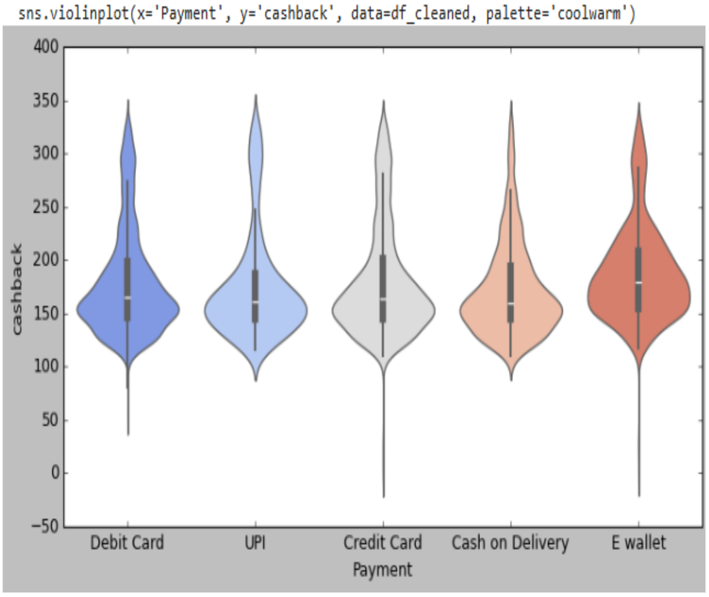




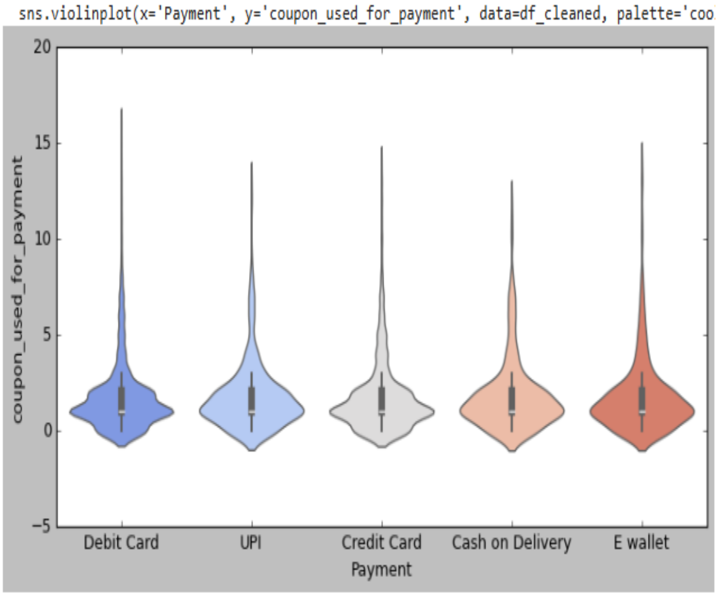

```
sns.violinplot(x='Payment', y='City_Tier', data=df_cleaned, palette='coolwarm')
```

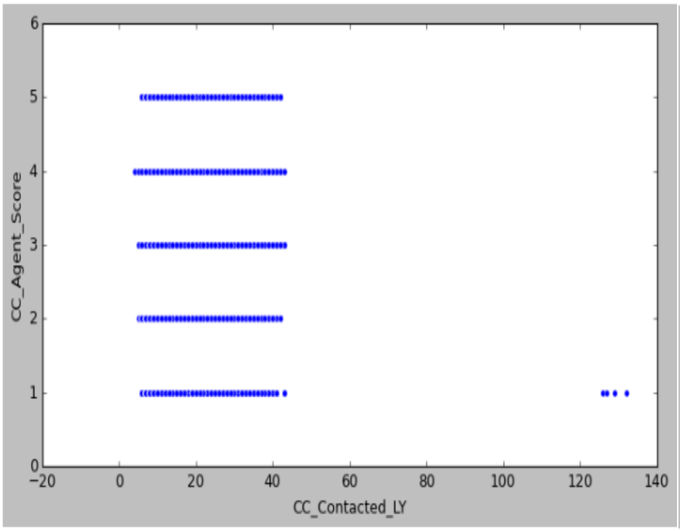


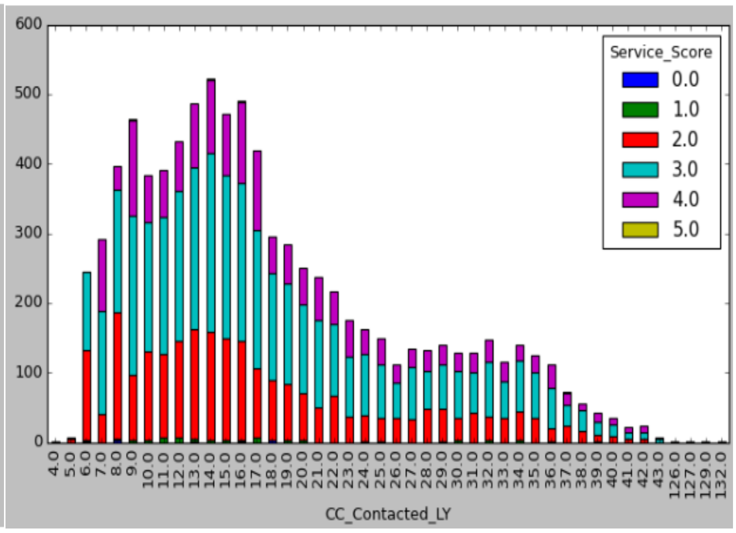

```
sns.violinplot(x='Payment', y='cashback', data=df_cleaned, palette='coolwarm')
```

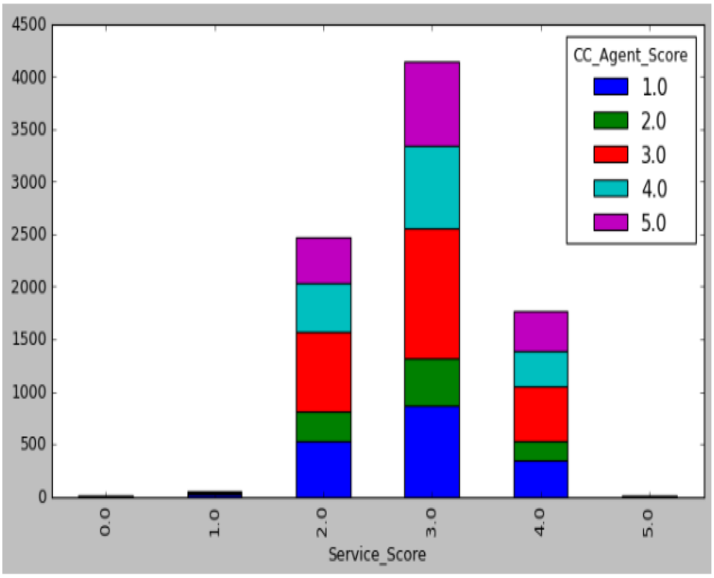



```
sns.violinplot(x='Payment', y='coupon_used_for_payment', data=df_cleaned, palette='coo
```

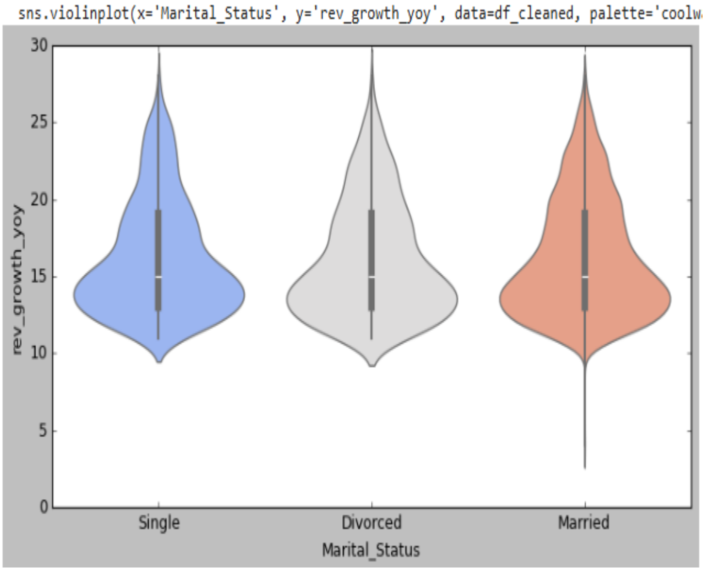


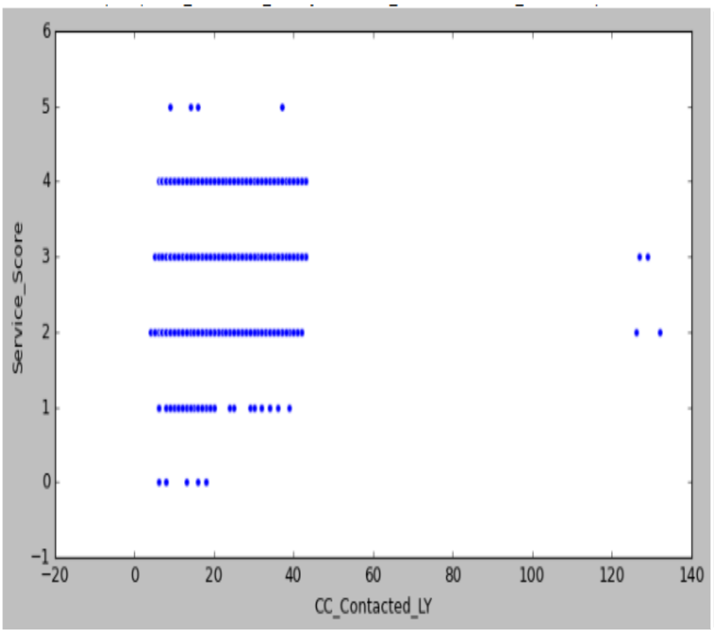


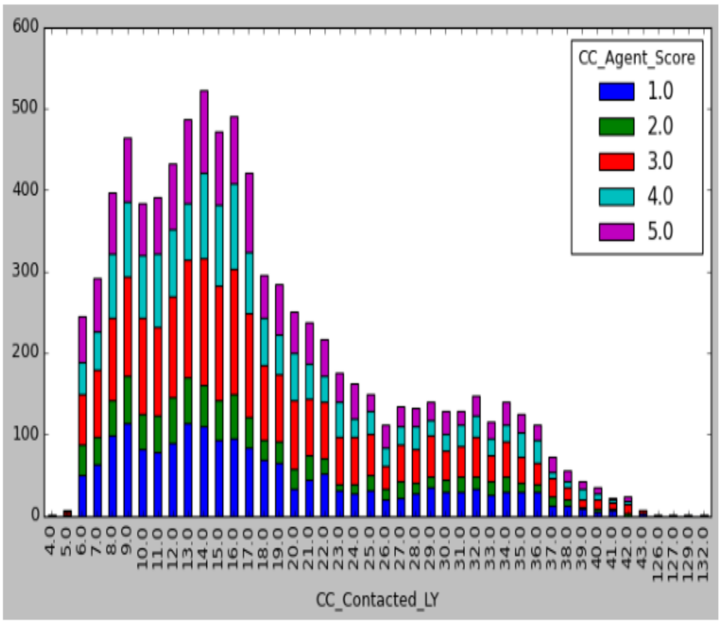





```
sns.violinplot(x='Marital_Status', y='rev_growth_yoy', data=df_cleaned, palette='coolw
```





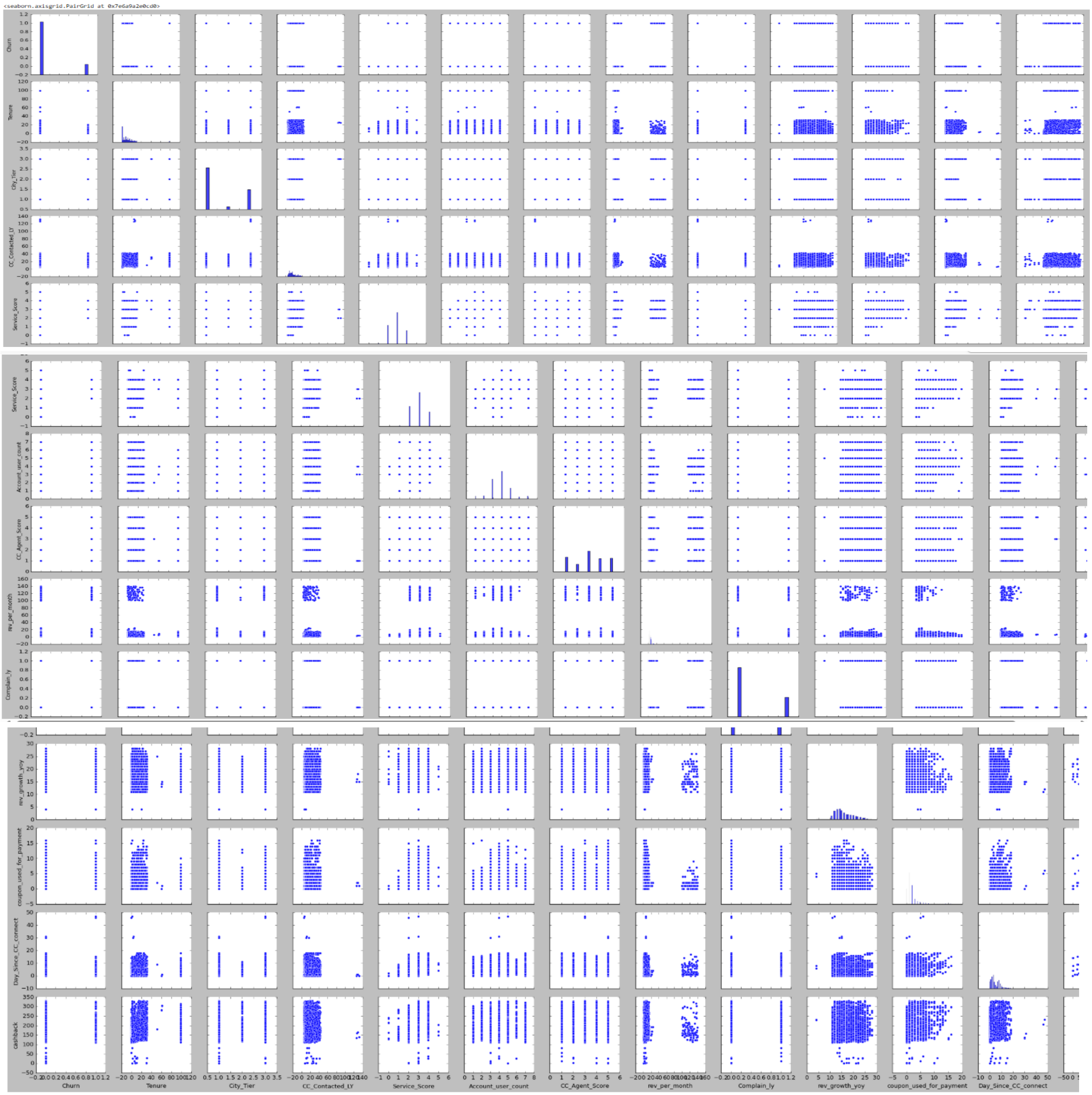


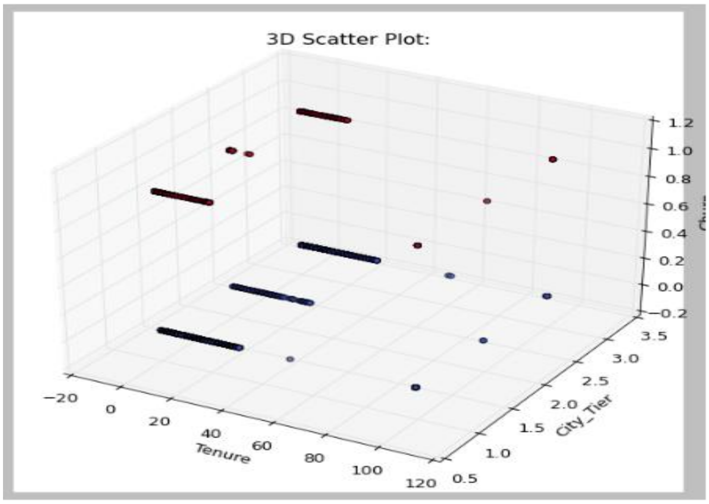
Insights

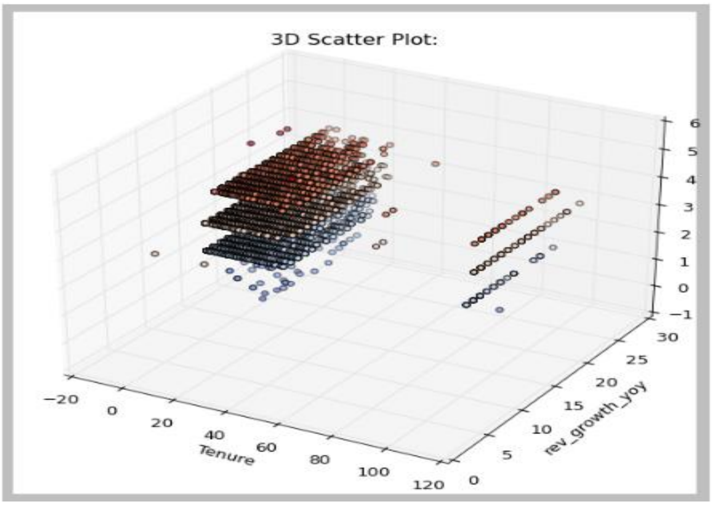
- In order to perform bivariate analysis, I have represented the features through scatterplot, boxplot, violin plot and crosstabs.
- If we check the boxplot of all numerical features, we find there are outliers present in the Dataset that needs to be treated.
- Service Score is majorly scattered between 2 and 4.
- Tenurity of the account, Day since CC Connect and Cashback are negatively correlated with Churn which is our target variable.
- Day since CC connect and coupon used for payment has positive correlation.
- Gender and payment method do not seem to be the reason for affecting churning of the customers.
- Regular Plus customers are more likely to leave the services or are in our high priority list.
- Customers who are Single, needs a special attention.
- Both Male and Female customers use Debit card and Credit card more than any other payment method for making the payments.
- We majorly have Regular Plus and Sper customers who are using the product and the services.
- Super customers are a bit inclined towards E-wallet as their payment method.
- City Tier 1 customers are more likely to proceed with the payments in time and so they are our loyal customers followed by city tier 3. City Tier 2 customers need attention.
- Cashback for the customers range between 100 and 350. This needs to be managed.
- Customers have mainly rated 3 for the services provided by the company which is a very average score.
- The service score majorly ranges between 2 and 4.
- Customer care service rating of the company is also between 2 and 4 which is not a decent score.
- The monthly average revenue generated by the company is between 1 to 10k out of which majorly the revenue flow is in the 3k box.

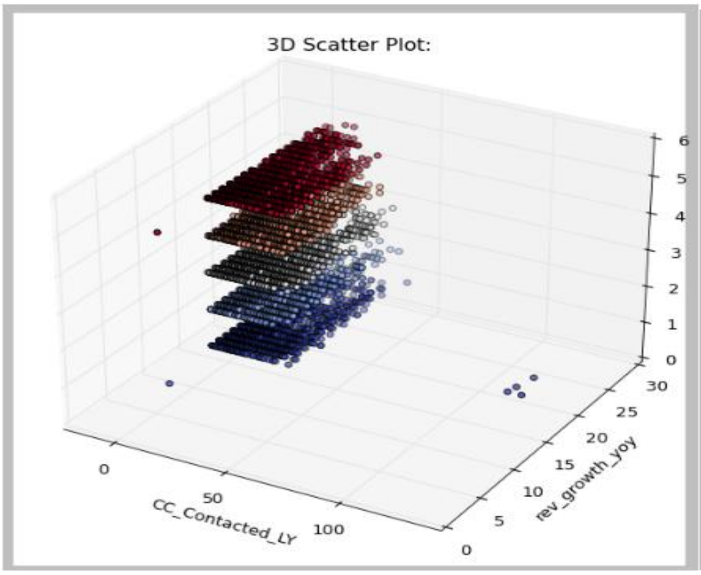
Multivariate Analysis:

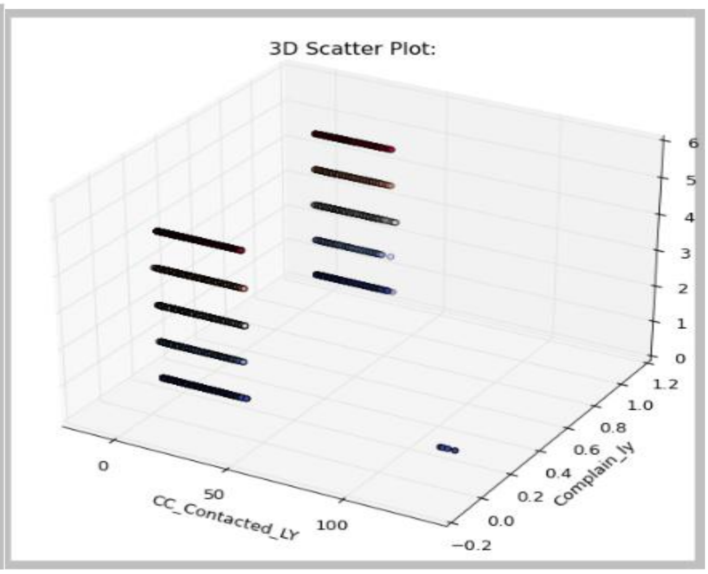
Below is the pairplot of all the features in the Dataset:

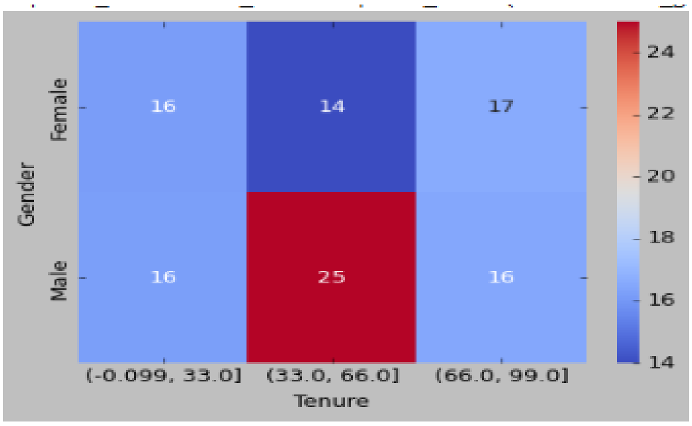


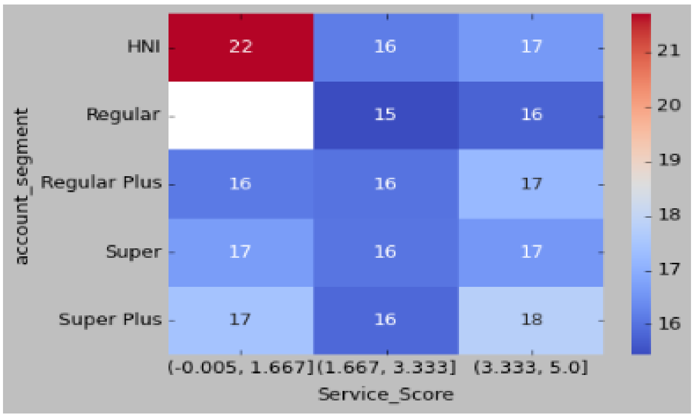








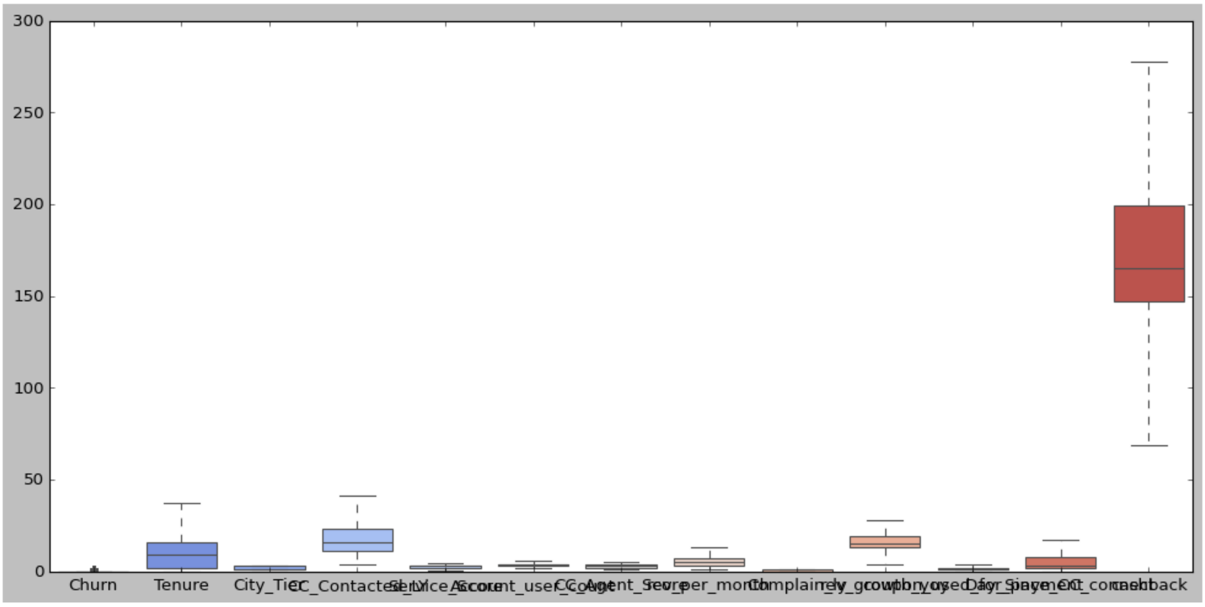




Insights

- In order to go ahead with multivariate analysis, I have used Pairplot, 3D Scatter plot and Pivot Table.

- Tenurity of the account majorly lies between 0 to 20 months and is equally distributed among all the three City Tiers.
- Churning however is majorly seen in Tier1 and Tier 3 cities.
- Tier 2 cities are not that active and needs special attention
- Revenue growth percentage lies between 10 and 30 with respect to Tenurity of the account.
- All the customers of the account have contacted Customer care around 0 to 50 times in the last 12 months.
- Customer contacted customer care and complaints raised are directly proportional to each other.
- Male customers with Tenurity 33 to 66 months have revenue growth of 25%.
- Female customers on the other hand with Tenurity 33 to 66 months have revenue growth of only 14%.
- Revenue growth is 22% for High-Net-Worth Individuals but the service score is very poor. If not taken care the company might lose its HNI.
- After the Outliers Treatment, this is what the boxplot of the features of the Data looks like



- After Data processing the dataset now has 8447 rows and 18 columns. Out of these 8447 rows 7043 are active accounts and 1404 are already churned.
- Split the dataset into Training(70%) and Testing(30%) sets.

Different Models and its Performance Matrix:

Logistic Regression Model:

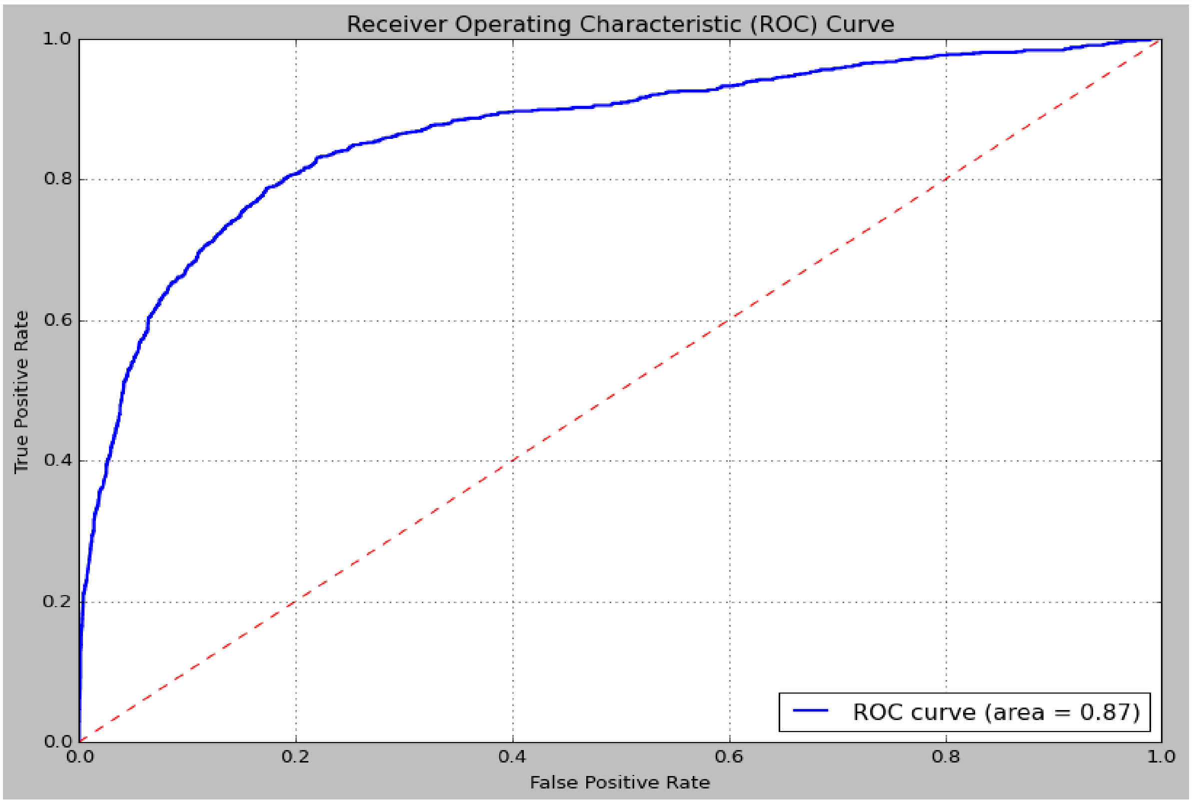
- Confusion matrix and Classification Report on Train Dataset.

```
0.8795669824086604
[[4755  174]
 [ 538  445]]
              precision    recall  f1-score   support

           0       0.90      0.96      0.93      4929
           1       0.72      0.45      0.56       983

    accuracy                           0.88      5912
   macro avg       0.81      0.71      0.74      5912
weighted avg       0.87      0.88      0.87      5912
```

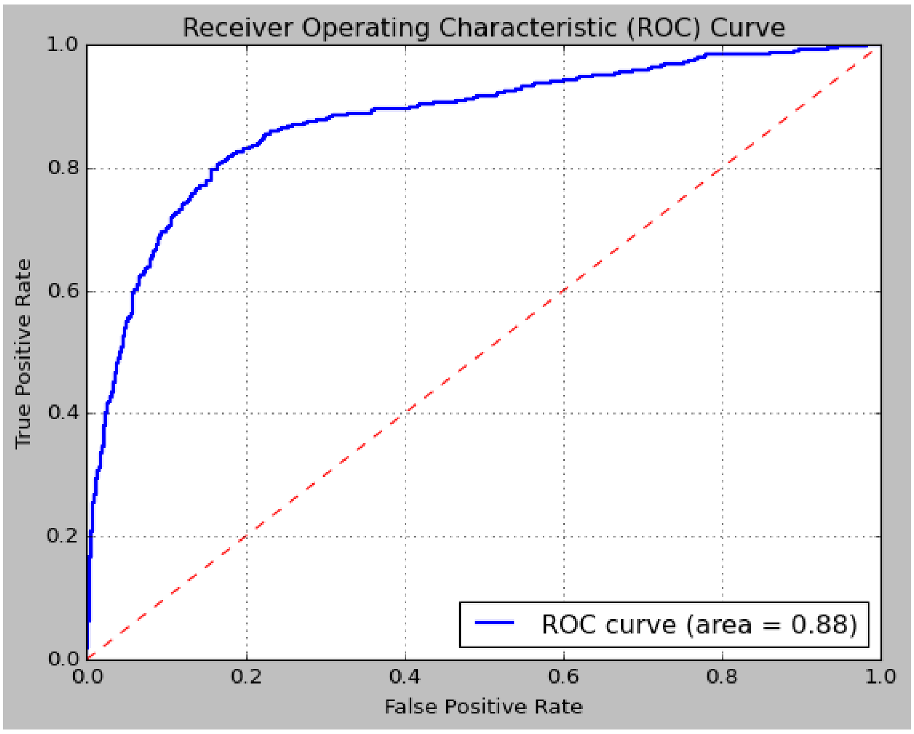
- Confusion matrix and Classification Report on Test Dataset.

```
0.8800788954635108
[[2043   71]
 [ 233  188]]
              precision    recall  f1-score   support

           0       0.90      0.97      0.93      2114
           1       0.73      0.45      0.55       421

    accuracy                           0.88      2535
   macro avg       0.81      0.71      0.74      2535
weighted avg       0.87      0.88      0.87      2535
```

- AUC-ROC Curve on Train Dataset.



➢ AUC-ROC Curve on Test Dataset.



Linear Discriminant Analysis:

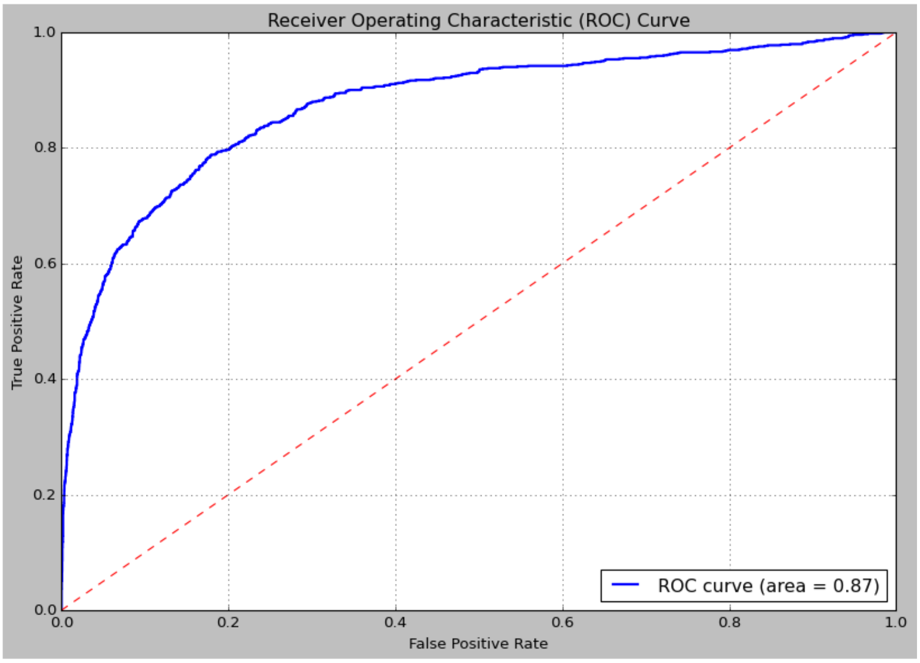
➢ Confusion matrix and Classification Report on Train Dataset.

```
0.888700947225981
[[4800  129]
 [ 529  454]]
              precision    recall  f1-score   support

           0       0.90      0.97      0.94      4929
           1       0.78      0.46      0.58       983

    accuracy                           0.89      5912
   macro avg       0.84      0.72      0.76      5912
weighted avg       0.88      0.89      0.88      5912
```

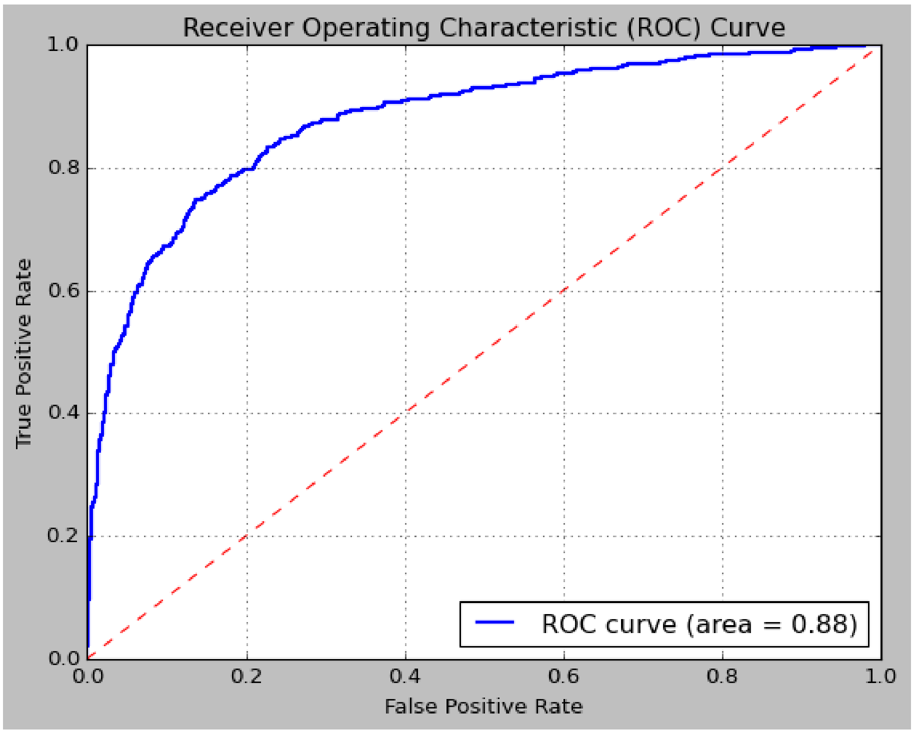
➢ Confusion matrix and Classification Report on Test Dataset.

```
0.886785009861933
[[2054   60]
 [ 227  194]]
              precision    recall  f1-score   support

           0       0.90      0.97      0.93      2114
           1       0.76      0.46      0.57       421

    accuracy                           0.89      2535
   macro avg       0.83      0.72      0.75      2535
weighted avg       0.88      0.89      0.87      2535
```

➢ AUC-ROC Curve on Train Dataset.



➢ AUC-ROC Curve on Test Dataset.



Decision Tree Classifier - CART Model:

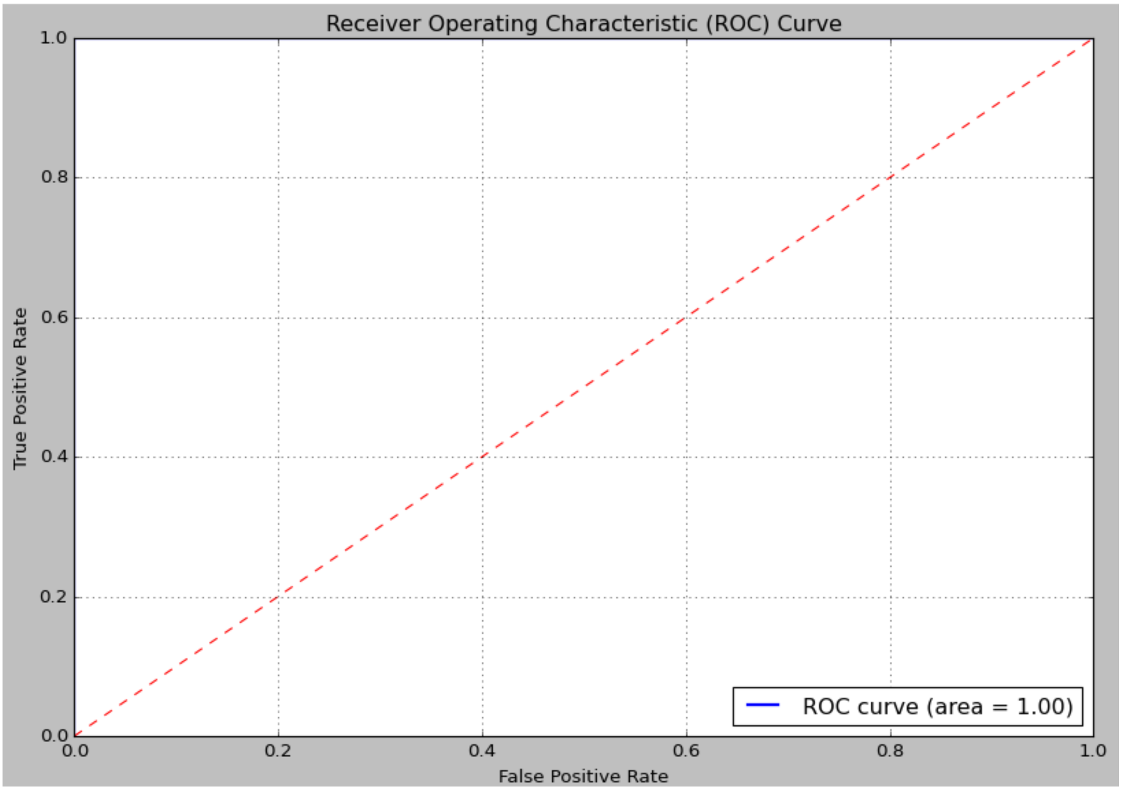
➢ Confusion matrix and Classification Report on Train Dataset.

```
1.0
[[4929    0]
 [   0  983]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      4929
           1       1.00      1.00      1.00       983

    accuracy                           1.00      5912
   macro avg       1.00      1.00      1.00      5912
weighted avg       1.00      1.00      1.00      5912
```

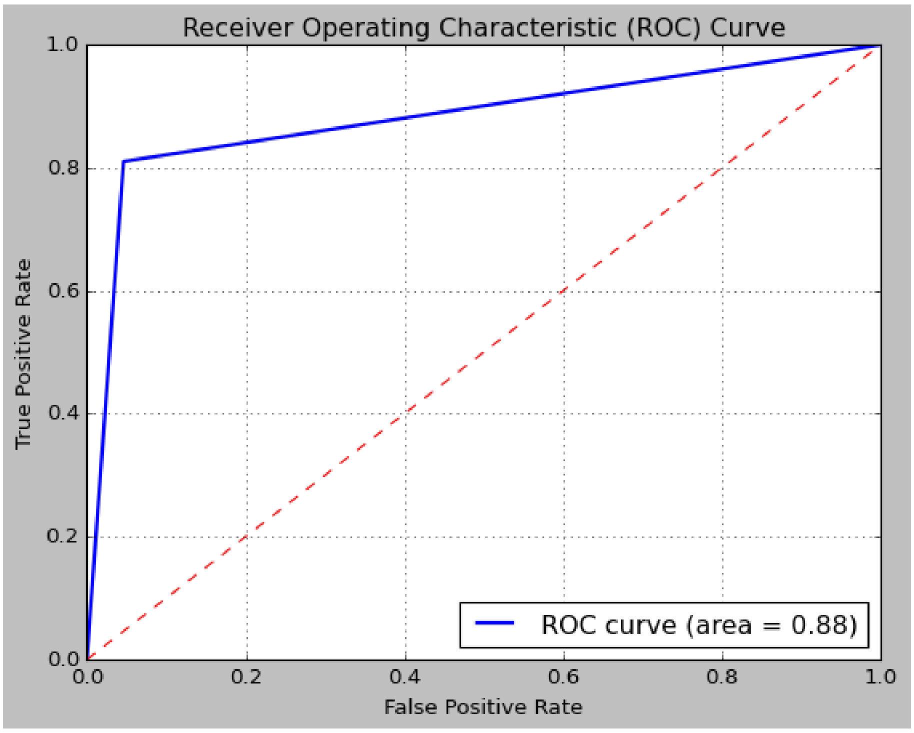
➢ Confusion matrix and Classification Report on Test Dataset.

```
0.9297830374753452
[[2016   98]
 [  80  341]]
              precision    recall  f1-score   support

           0       0.96      0.95      0.96      2114
           1       0.78      0.81      0.79       421

    accuracy                           0.93      2535
   macro avg       0.87      0.88      0.88      2535
weighted avg       0.93      0.93      0.93      2535
```

➢ AUC-ROC Curve on Train Dataset.



➢ AUC-ROC Curve on Test Dataset.



Naive Bayes Model:

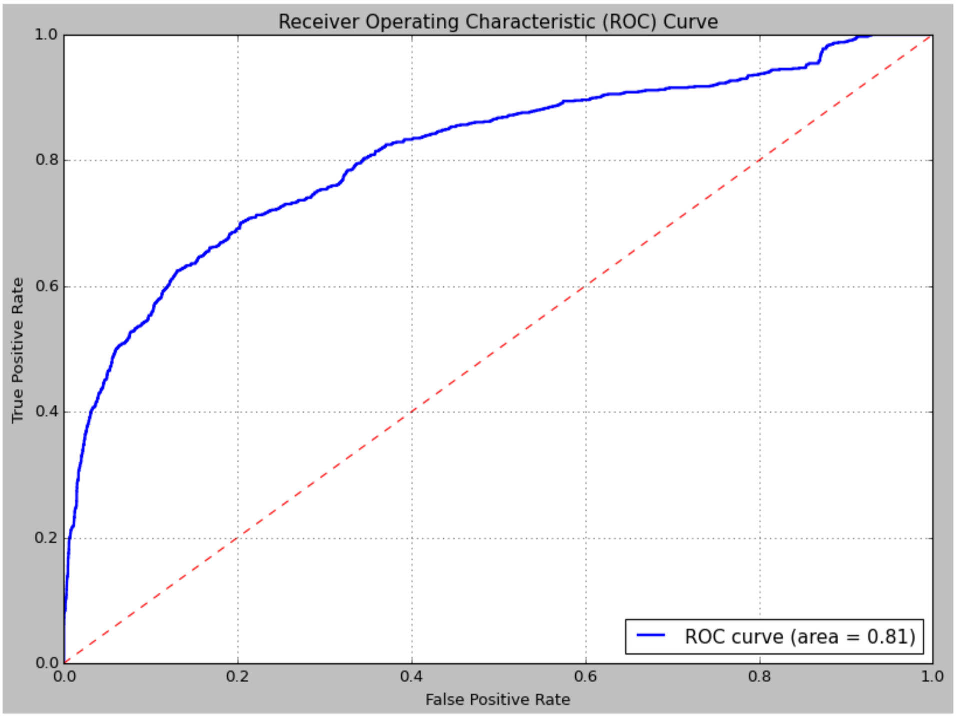
➢ Confusion matrix and Classification Report on Train Dataset.

```
0.7763870094722598
[[3897 1032]
 [ 290  693]]
              precision    recall  f1-score   support

           0       0.93      0.79      0.85      4929
           1       0.40      0.70      0.51       983

    accuracy                           0.78      5912
   macro avg       0.67      0.75      0.68      5912
weighted avg       0.84      0.78      0.80      5912
```

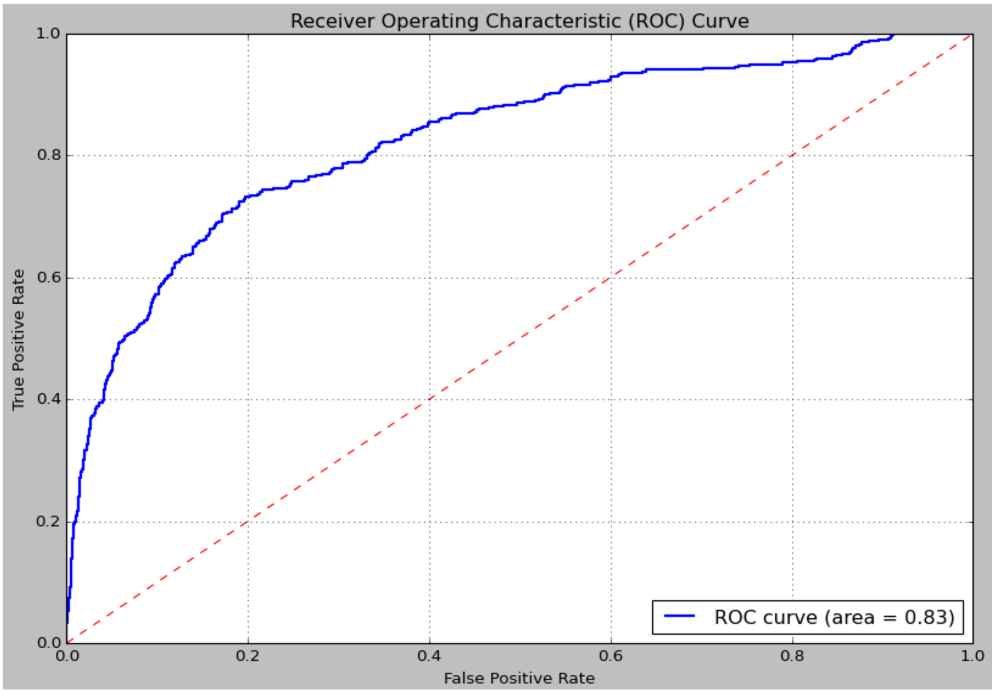
➢ Confusion matrix and Classification Report on Test Dataset.

```
0.7869822485207101
[[1687  427]
 [ 113  308]]
              precision    recall  f1-score   support

           0       0.94      0.80      0.86      2114
           1       0.42      0.73      0.53       421

    accuracy                           0.79      2535
   macro avg       0.68      0.76      0.70      2535
weighted avg       0.85      0.79      0.81      2535
```

➢ AUC-ROC Curve on Train Dataset.



➢ AUC-ROC Curve on Test Dataset.



KNN Model:

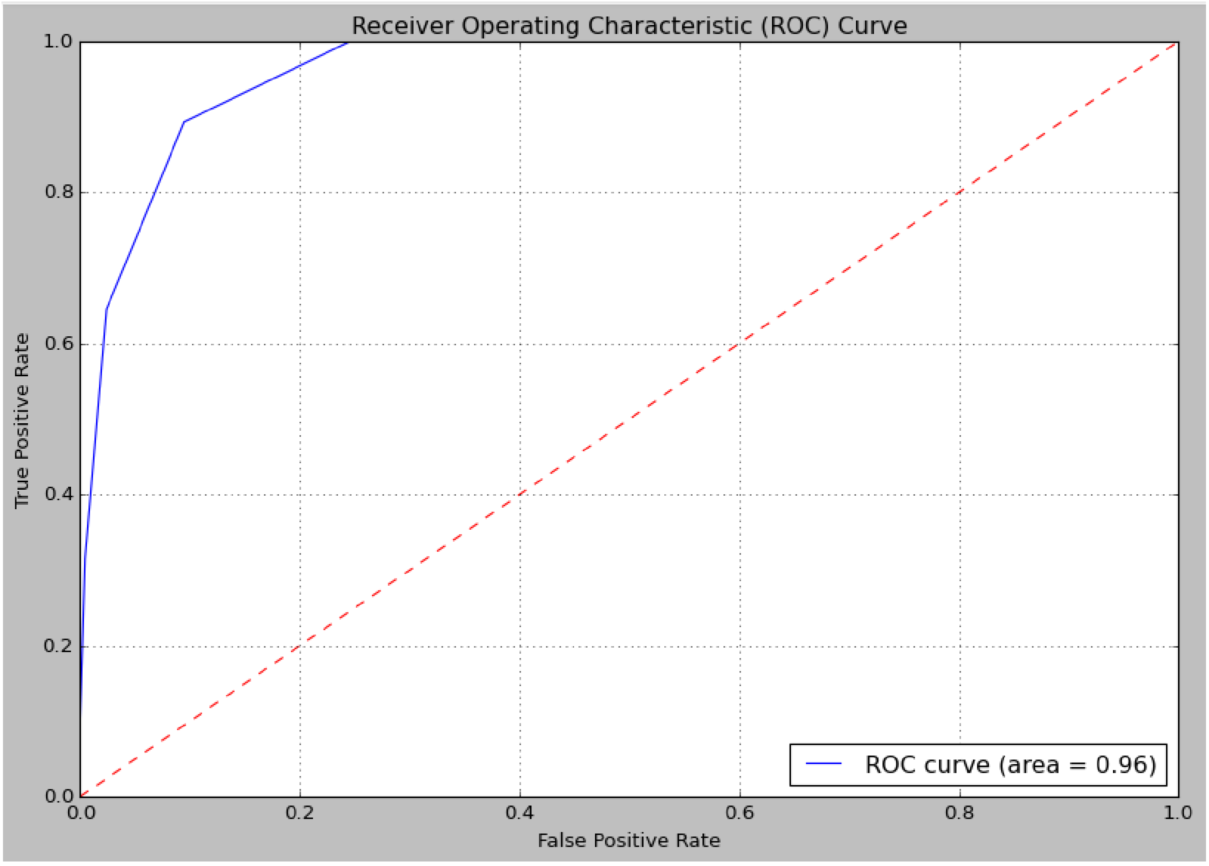
➢ Confusion matrix and Classification Report on Train Dataset.

```
0.9211772665764547
[[4812  117]
 [ 349  634]]
              precision    recall  f1-score   support

           0       0.93      0.98      0.95      4929
           1       0.84      0.64      0.73       983

    accuracy                           0.92      5912
   macro avg       0.89      0.81      0.84      5912
weighted avg       0.92      0.92      0.92      5912
```

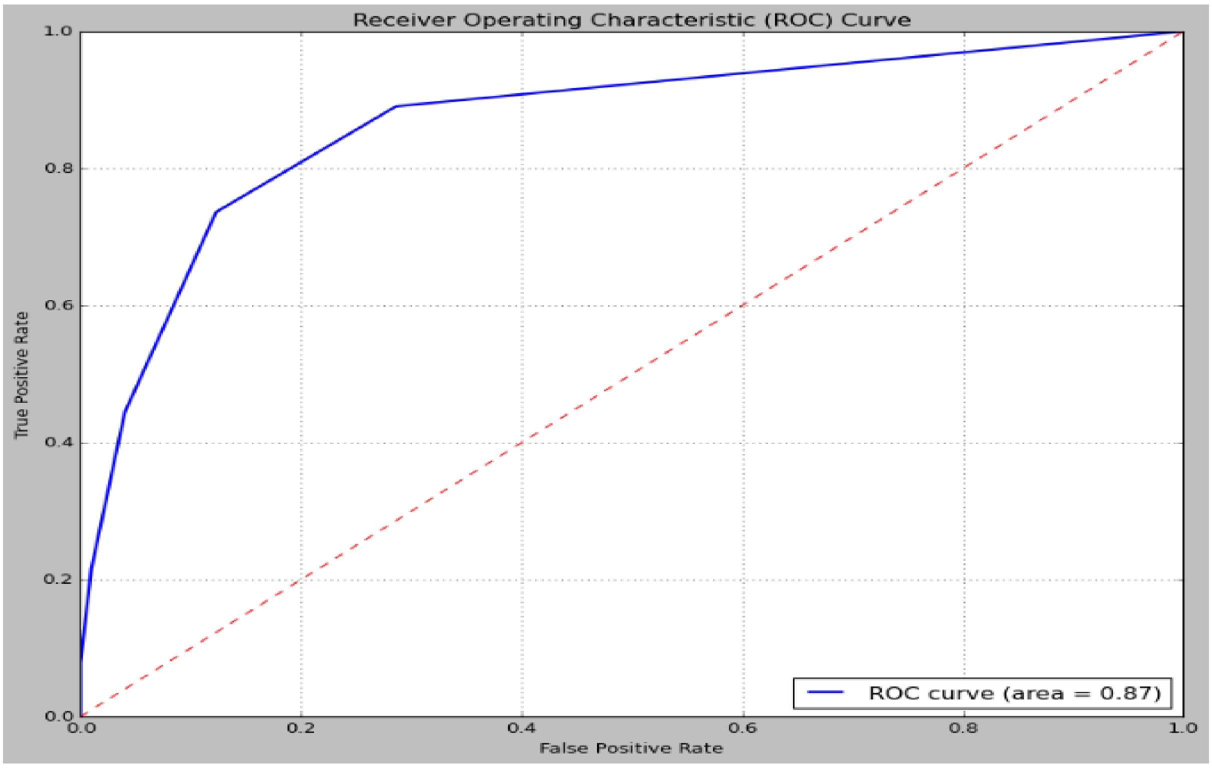
- Confusion matrix and Classification Report on Test Dataset.

```
0.8741617357001973
[[2029   85]
 [ 234  187]]
              precision    recall  f1-score   support

           0       0.90      0.96      0.93      2114
           1       0.69      0.44      0.54       421

    accuracy                           0.87      2535
   macro avg       0.79      0.70      0.73      2535
weighted avg       0.86      0.87      0.86      2535
```

- AUC-ROC Curve on Train Dataset.



➢ AUC-ROC Curve on Test Dataset.



Random Forest Model:

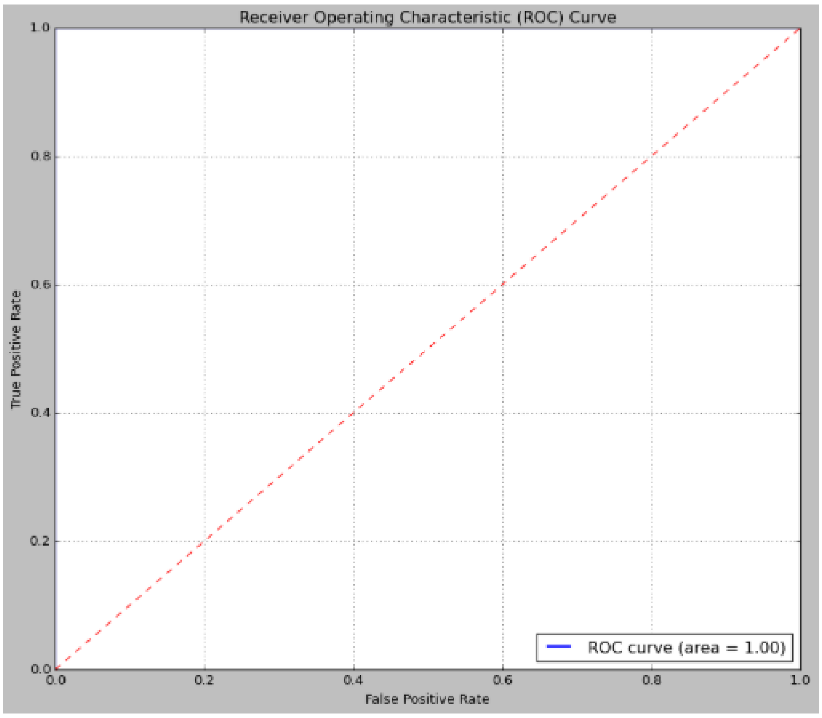
➢ Confusion matrix and Classification Report on Train Dataset.

```
1.0
[[4929    0]
 [   0  983]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      4929
           1       1.00      1.00      1.00       983

    accuracy                           1.00      5912
   macro avg       1.00      1.00      1.00      5912
weighted avg       1.00      1.00      1.00      5912
```

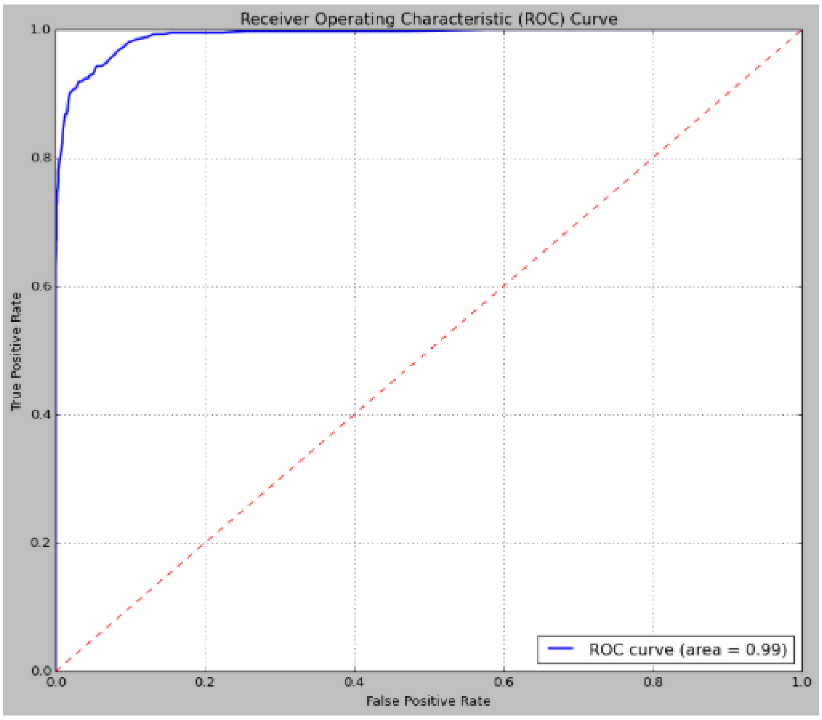
➢ Confusion matrix and Classification Report on Test Dataset.

```
0.9621301775147929
[[2100   14]
 [  82  339]]
              precision    recall  f1-score   support

           0       0.96      0.99      0.98      2114
           1       0.96      0.81      0.88       421

    accuracy                           0.96      2535
   macro avg       0.96      0.90      0.93      2535
weighted avg       0.96      0.96      0.96      2535
```

➢ AUC-ROC Curve on Train Dataset.



➢ AUC-ROC Curve on Test Dataset.



<u>Boosting Classifier Model using Gradient Boost:</u>

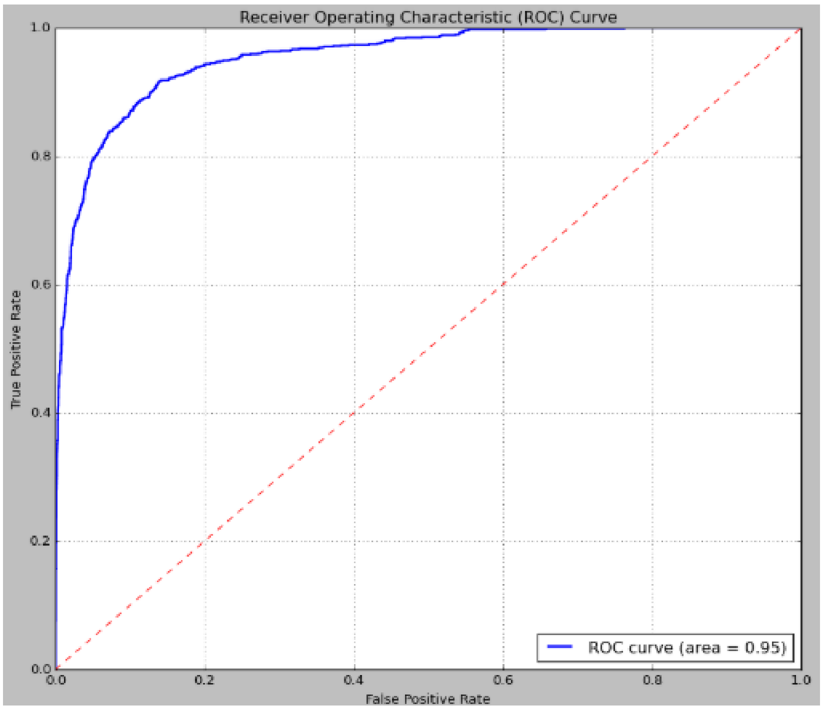
➢ Confusion matrix and Classification Report on Train Dataset.

```
0.9235453315290933
[[4828  101]
 [ 351  632]]
              precision    recall  f1-score   support

           0       0.93      0.98      0.96      4929
           1       0.86      0.64      0.74       983

    accuracy                           0.92      5912
   macro avg       0.90      0.81      0.85      5912
weighted avg       0.92      0.92      0.92      5912
```

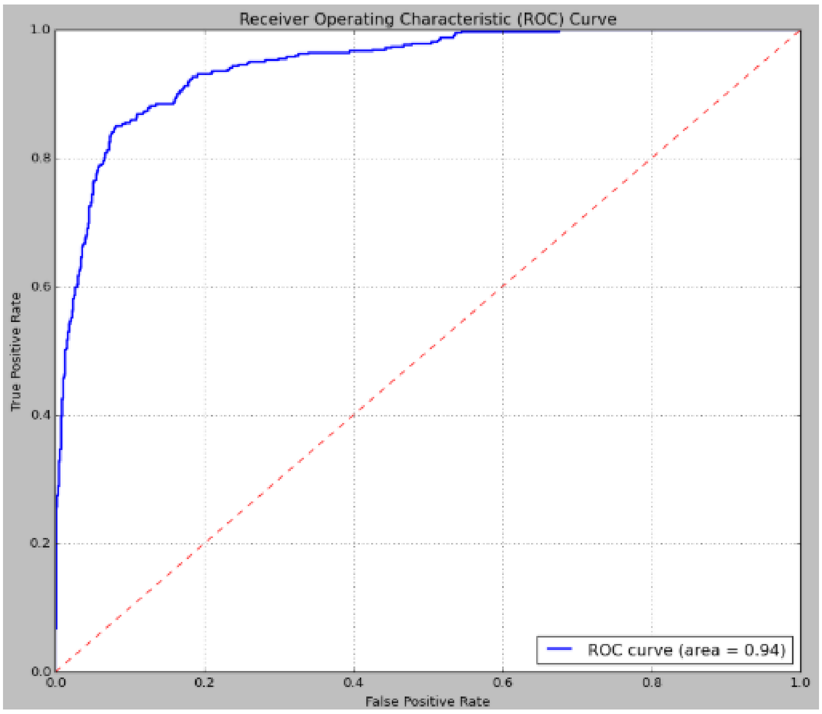
➢ Confusion matrix and Classification Report on Test Dataset.

```
0.9108481262327416
[[2049   65]
 [ 161  260]]
              precision    recall  f1-score   support

           0       0.93      0.97      0.95      2114
           1       0.80      0.62      0.70       421

    accuracy                           0.91      2535
   macro avg       0.86      0.79      0.82      2535
weighted avg       0.91      0.91      0.91      2535
```

➢ AUC-ROC Curve on Train Dataset.



- AUC-ROC Curve on Test Dataset.



Key Findings:

- As we compare all the seven models build for this dataset, we find that the best performing model is Random Forest Model with 100% and 96% accuracy for Training and Testing Data respectively which is fairly the best followed by Decision Tree model with the model score 100% and 93% for Training and Testing Data respectively.
- The recall Values for both the models are same that is 100% and 81% for training and testing data, However the precision and the AUC score for Random Forest Model is slightly better than the Decision Tree Model. Hence, I choose to go with Random Forest model for our outcome in this Dataset.
- Gradient Boost is also a good model with Accuracy 92% overall. The recall value and the precision value for the model is average. AUC score however is fairly good that is 95%.
- The other three models KNN model, LDA and Logistic Regression are average performing models to which we can apply SMOTE to check if their efficiency increases.
- Naive Bayes Model is the worst performing model for this Dataset.
- Performed the Chi-Square test on all the features against the target variable. p-value (p): A small p-value (≤ 0.05) suggests a significant association between the feature and the target variable. Hence Gender (p-value 0.0) and Tenure (p-value 0.0) has significant association.
- Revenue growth percentage of the account (p-value 0.0) and Monthly average revenue generated by account (p-value 0.2) has significant association with churning of the customers.
- Service score(p-value 0.02) and coupon used for payment(p-value 0.08) also impacts our target variable significantly.

Important Features:

- Tenurity of the account, Monthly average cashback generated by account, Any Complaints raised by the account in last 12 months and the number of times all the customer of the account contacted customer care in the last 12 months are some of the most important features to be closely and deeply looking at.
- The Customers are definitely not very happy with the services provided by the company. It is the major concern and needs to be worked on in order to stop the customers from leaving the product and services.

```
                       Feature  Importance
0                       Tenure    0.226162
11                    cashback    0.086077
7                 Complain_ly     0.070317
2                CC_Contacted_LY  0.066875
10          Day_Since_CC_connect  0.066659
8               rev_growth_yoy    0.060849
6                rev_per_month    0.058058
5               CC_Agent_Score    0.056902
4           Account_user_count    0.036386
9      coupon_used_for_payment    0.030066
1                    City_Tier    0.027358
22        Marital_Status_Single   0.027164
18  account_segment_Regular Plus  0.021976
16                 Gender_Male    0.021485
3                Service_Score    0.021064
23         Login_device_Mobile    0.021048
12         Payment_Credit Card    0.020393
21       Marital_Status_Married   0.018240
13          Payment_Debit Card    0.015217
19        account_segment_Super   0.015063
14            Payment_E wallet    0.013700
15                 Payment_UPI    0.008319
20   account_segment_Super Plus   0.004080
17      account_segment_Regular   0.003608
24         Login_device_Others    0.002933
```

Recommendations:

- By implementing this churn prediction model, the company can significantly reduce account churn without excessive financial risk.
- The structured approach ensures that **high-value customers are retained**, and **low-value churners are not over-incentivized**.
- We need to **segment customers** based on their churn risk and tailor retention offers **without excessive cost**.

- **High Risk** (Likely to Churn) customers with low engagement and frequent complaints like City Tier 3 customers and customers segmented at Super Plus and Regular Plus. We can offer them **Retention Discount** (10-15% off for the next 3 months) IF they commit to a longer subscription. Provide **personalized support** to resolve service issues. Increases commitment while reducing churn risk.
- Encouraging long-term commitment without giving away free services.
- **Medium Risk** (Uncertain) customers with reduced engagement like City Tier 2 Customers and customers segmented at HNI and Super Plus. We can offer them **Personalized Upsell** (Bundle upgrade at a discount for limited time e.g., extra channels for DTH, priority delivery for E-Commerce). Provide a **personalized loyalty program**. This Increases perceived value while ensuring revenue growth.
- **Low Risk** (Loyal Customers) with regular payments and high usage like City Tier 1 customers and the customers who fall under the Super category. We can offer them **Exclusive Early Access or Rewards** (early product access, priority support). Provide **exclusive perks** (e.g., faster deliveries, priority support). No discount, just added benefits to maintain loyalty.
- **Discounts only for long-term commitments** (e.g., **6-month lock-in**).
- **Targeted offers** based on **predicted churn probability.**
- Providing comprehensive self-service tools empowers customers to resolve issues independently, leading to increased satisfaction and reduced workload for agents.
- Developing a robust knowledge base with FAQs, tutorials, and troubleshooting guides can significantly improve the customer experience.
- Enabling online account management for tasks like billing and service modifications enhances convenience.
- Personalization fosters a stronger connection between the customer and the service provider.
- Utilizing customer data to tailor interactions and offers can enhance satisfaction and loyalty.
- Invest in Comprehensive Agent Trainings and Implement Quality Assurance Programs
- Training programs should cover product knowledge, communication skills, problem-solving techniques, and the use of customer service technologies. Continuous education keeps agents updated on new products and policies, enhancing their performance.

THE END