BBA Semester – VI
Research Project – Interim Report

Name	Mr. Ukesh Kumar
USN	212VBBR01351
Elective	Data Science & Analytics
Date of Submission	20-09-2024

A study on "Prediction of Customer Churn and Strategy to Retain Customers for an E Commerce/DTH Company "

Research Project submitted to Jain Online (Deemed-to-be University)

In partial fulfillment of the requirements for the award of:

Bachelor of Business Administration

Submitted by:

Mr. Ukesh Kumar

USN:

212VBBR01351

Under the guidance of:

Mr. Milind Desai

(Faculty-JAIN Online)

Jain Online (Deemed-to-be University)

Bangalore

2023-24

DECLARATION

I, *Mr. Ukesh Kumar,* hereby declare that the Research Project Report titled *"**Prediction of Customer Churn and Strategy to Retain Customers for an E Commerce/DTH Company***" *has been* prepared by me under the guidance of the *Mr. Milind Desai.* I declare that this Project work is towards the partial fulfillment of the University Regulations for the award of the degree of Bachelor of Business Administration by Jain University, Bengaluru. I have undergone a project for a period of Eight Weeks. I further declare that this Project is based on the original study undertaken by me and has not been submitted for the award of any degree/diploma from any other University / Institution.

Place: Chennai

Date: 20-09-2024

_____

Mr. Ukesh Kumar
USN:212VBBR01351

CERTIFICATE

This is to certify that the Research Project report submitted by Mr. Ukesh Kumar (212VBBR01351) bearing on the title "Predicting Customer Churn and Designing Retention Strategies for E-Commerce/DTH Provide" is a record of project work done by him/ her during the academic year 2023-24 under my guidance and supervision in partial fulfillment of Bachelor of Business Administration.

Place:

Date: Mr. Milind Desai

EXECUTIVE SUMMARY

In E-commerce or DTH (Direct to home) Markets, Customer acquisition and retention play a crucial role on deciding the present, past and the future of the business. Lots of companies in this market focus on optimizing both acquisition and retention to scale the business from one level to the next. Customer churn - where a customer stops doing business with a company and discontinue their relationship or subscription, often lead the company to lose their customer base which leads to a significant loss in revenue as well.

The primary objective of this project is to develop a churn prediction model to identify at-risk accounts in e-commerce and DTH companies, while proposing targeted, cost effective campaigns to retain the customers.

In this project we will identify potential churners and provide some tailored strategies to retain the customers, by analyzing customer behavior and transaction data from the collected data of multiple users of the company.

This Project employs a four stages of data analysis, Prepare, Model, Visualize and Analyze. The historical data on customer is collected which has various customer attributes including account information, behavior, demographics, Transactional metrics, customer care interactions, is prepared. During preparation process raw data is turned into information that is trusted and understandable while using Exploratory Data Analysis (EDA), key insights from the data is understood and the story from the data becomes clearer. Once we connect with the story, it will become easier to transform and model the data for churn prediction.

Using the Prepared Data, a semantic model is developed with various methods like logistic regression, decision tree, random forest and gradient boosting, after evaluation of the model using various metrics like accuracy, precision, recall, and F1 Score. The model is then deployed to predict churn of new accounts in real time. The model is also tested and can be changed in real time with different metrics in future depends on the change patterns or trends or any necessities of the company.

In conclusion, through this project a comprehensive churn prediction model is prepared for an e-commerce and DTH company to effectively identify and retain at-risk customers, while on the other hand drive revenue growth and customer loyalty. The proposed campaigns and offers are unique, cost-effective solution that can be tailored to different customer segments ensures a definite path to customer acquisition and retention.

TABLE OF CONTENTS

Title	Page Nos.
Executive Summary	5
List of Tables	8
List of Graphs	8
Chapter 1: Introduction and Background	9
Chapter 2: Data Report	17
Chapter 3: Exploratory Data Analysis & Business Insights from EDA	19
Chapter 4: Findings , Recommendations and Conclusion	32

CHAPTER 1

INTRODUCTION AND BACKGROUND

INTRODUCTION AND BACKGROUND

1.1 Introduction

The e-commerce and DTH industries are currently facing a high competition by an influx of new entrants and also changes in customer preferences. As a result of this the retention of customers has become a significant challenge for companies in this sector. Customer churn - where a customer stops doing business with a company and discontinue their relationship or subscription poses a threat to growth and sustainability of the companies in this sector.

1.2 Definition of Problem Statement

By Defining the problem statement given, the company in question here also recognizes that account churn in a detrimental pattern, as every single account the company has holds tremendous value. Therefore, losing of several customers can amplify the impact on revenue and market share. The challenge is further compounded by the fact that there a numerous other companies in the same space at their disposal, which makes it easier for the customers to switch if their needs are not met with current company or if they perceive better value elsewhere.

1.3 Need for the Study

The cost of customer acquisition is comparatively higher than the retention cost. So this project aims to develop a customer churn prediction model that can identify accounts at-risk of churn. By leveraging data analysis, this model will analyze various customer related variables provided, such as tenure, service score, complain history and payment preferences to predict which account are likely to disengage. This predictive capability will enable the company in question to take data driven decisions to target potential churners.

1.4 Understanding Opportunities

1.4.1 Business Opportunities

1. **Retention as a revenue driver**: Retaining existing customers is much cost effective than acquiring new ones. By minimizing churn using the prediction model, the company can maintain a more stable revenue pool and reduce customer acquisition costs. The company can focus resources on growth and overall profitability.

2. **Enhancement in customer insights**: Developing this model allows the company to gather better insights into customer preference and behavior patterns. Using this knowledge company can allocate resources on product development, marketing strategy and improved customer service initiatives, leading to customer satisfaction and a loyal base.

3. **Targeted Marketing**: As Discussed above using the knowledge of customer behavior, preference pattern and also the analysis of potential at-risk churners from the model, the company can structure its marketing campaign that resonates better with the customers. This type of targeted campaign provides effective usage of marketing budgets and higher conversion rates for retention.

4. **Competitive Advantage**: In current competitive market scenario, companies that can effectively manage customer relationship and demonstrate high retention rates have competitive edge. By investing in churn prediction and structured retention strategies, the company can differentiate itself from the competitors.

5. **Data Utilization for decision making**: By Implementing a churn prediction model leverages the company to take data driven decision. The Analysis from the model provides strategic advantage which can be applied not only in just churn prediction but also in other departments like customer service and sales forecasting.

1.4.2   Social Opportunities

1. **Improving Customer Experience:** This model helps the company to implement strategies to reduce customer churn, which shows the customers that the company is committed to listening to customer feedback and address their needs. This helps in enhancing overall customer satisfaction and experience, making the company more trustworthy and reliable for customers.

2. **Building Brand Loyalty and Community base:** By Fostering strong relationship with consumers contributes to a sense of community. satisfied customers will in turn become advocates of the company to drive more brand visibility, through word of mouth referral and social media. This can impact positively on brand reputation and community engagement.

3. **Enhanced Accessibility:** By Analyzing data regularly, the company may identify gaps in the service or area to great extent. This insight can be leveraged to improved service offerings that can cater to underrepresented demographics or cities or customer base, fostering inclusivity and expansion of customer base.

4. **Sustainable Business Practice:** By Prioritizing customer retention by addressing current customer needs through ethical practices contributes to company's social responsibility. Through this demonstration the company can position itself as a socially responsible and customer centric company. Improvement in public perception and stakeholder's trust can be observed as well.

5. **Long Term Customer Relationship:** A long Term customer relationship initiative taken by the company shows that the company has long term vision for its business. This fosters a more stable environment for the business and helps create a culture of loyalty, which can positively impact on society by promoting customer focused interactions and reduce service or goods related frustrations of the customers.

CHAPTER 2

Data Report

Data Report

2.1 Data Collection

2.1.1   Time

1. **Historical data:** Data Collected Spans a historical period, specifically focusing on last 12 months for the most variables. This can be used to analyze recent customer behavior in terms of trends and patterns which are relevant for predicting customer churn.

2. **Comparative Analysis:** Some variables like rev_growth_yoy involve comparison between different time periods last 12 months' vs the previous 12 months. This data can help in understanding trends and changes in customer behavior over time.

3. **Real-Time Data:** The Dataset does not contain any real-time data.

2.1.2   Frequency

1. **Monthly Average:** Variables such as rev_per_month and cashback_l12m are calculated on a monthly basis. Allows a granular view of customer activity and financial performance, facilitating a trend analysis.

2. **Contact frequency:** Variable like CC_Contacted_L12m indicates the frequency of customer care contact in the last 12 months. This frequency metric may help in assessing customer engagement and satisfaction levels, and also critical key indicators for potential customer churn.

3. **Complaint Tracking:** Complain_l12m tracks any complaints raised by the account in the last 12 months. This trend can also be a critical key indicator for potential lead to customer churn.

4. Day since last customer care contact: Day_Since_CC_connect variable measures number of days since no customers in the account has contacted the customer care. This metric indicates customer engagement levels, if there has been a long gap without contact may also potentially lead to churn due to customer dissatisfaction.

2.1.3 Methodology

1. **Surveys and feedbacks:** Service_Score and CC_Agenct_Score is likely collected by surveys or feedback forms/links. This metric may help in gauging into the customer's perception of the company's service and agent's interaction to identify areas of improvement.

2. **Transactional Data:** rev_per_month, cashback_l12m and Payment are derived from transactional data, which is records customer purchase and payment behaviors makes it essential for understanding revenue generation and customer spending patterns.

3. **Customer Interactions:** CC_Contacted_L12m, Day_Since_CC_connect, complain_l12m, Account_user_count, Login_device is typically collected through CRM (Customer Relationship Management) Systems. Helps to Track Engagement and Service quality**.**

4. **Demographic Data Collection:** Gender, Marital_Status, City_Tier, account_segment is collected through account registration process or profiles. It is crucial for segmenting customers and tailoring retention strategies.

5. Coupon Tracking system: coupon_used-l12m tracks how many times customers have used coupons for payments, this data gives insight into pricing preference, purchase behavior and responsiveness to promotions. Higher responsiveness indicates pricing issue on the product, lower indicates either disinterest on promotion or comfort in pricing.

2.2 Visual Inspection of Data:

Rows (Records)	Columns(Variables)
11260	19

2.2.1 Description of Data

Variable	Data Type	Description
AccountID	String	account unique identifier
Churn	Binary	account churn flag (Target)
Tenure	Integer	Tenure of account
City_Tier	Integer	Tier of primary customer's city
CC_Contacted_LY	Integer	How many times all the customers of the account has contacted customer care in last 12months
Payment	Categorical	Preferred Payment mode of the customers in the account
Gender	Categorical	Gender of the primary customer of the account
Service_Score	Integer	Satisfaction score given by customers of the account on service provided by company
Account_user_count	Integer	Number of customers tagged with this account
account_segment	Categorical	Account segmentation on the basis of spend
CC_Agent_Score	Integer	Satisfaction score given by customers of the account on customer care service provided by company
Marital_Status	Categorical	Marital status of the primary customer of the account
rev_per_month	Float	Monthly average revenue generated by account in last 12 months
Complain_ly	Binary	Any complaints has been raised by account in last 12 months
rev_growth_yoy	Float	revenue growth percentage of the account (last 12 months vs last 24 to 13 month)
coupon_used_l12m	Integer	How many times customers have used coupons to do the payment in last 12 months
Day_Since_CC_connect	Integer	Number of days since no customers in the account has contacted the customer care
cashback_l12m	Float	Monthly average cashback generated by account in last 12 months
Login_device	Categorical	Preferred login device of the customers in the account

CHAPTER 3

Exploratory Data Analysis
&
Business Insights from EDA

Exploratory Data Analysis

3.1. Objectives

To understand the underlying patterns, relationships and overall story of the data which may influence a customer to churn/discontinue the service provided by the E-commerce company or DTH provider.

3.2. Data Analysis Tools (EDA)

Python, numpy, pandas, Matplotlib, Seaborn are used in this stage.

3.3. Exploratory Data Analysis

3.3.1. Univariate Analysis

In univariate analysis, we focus on each metric individually to understand its distribution, central tendency, and variability.
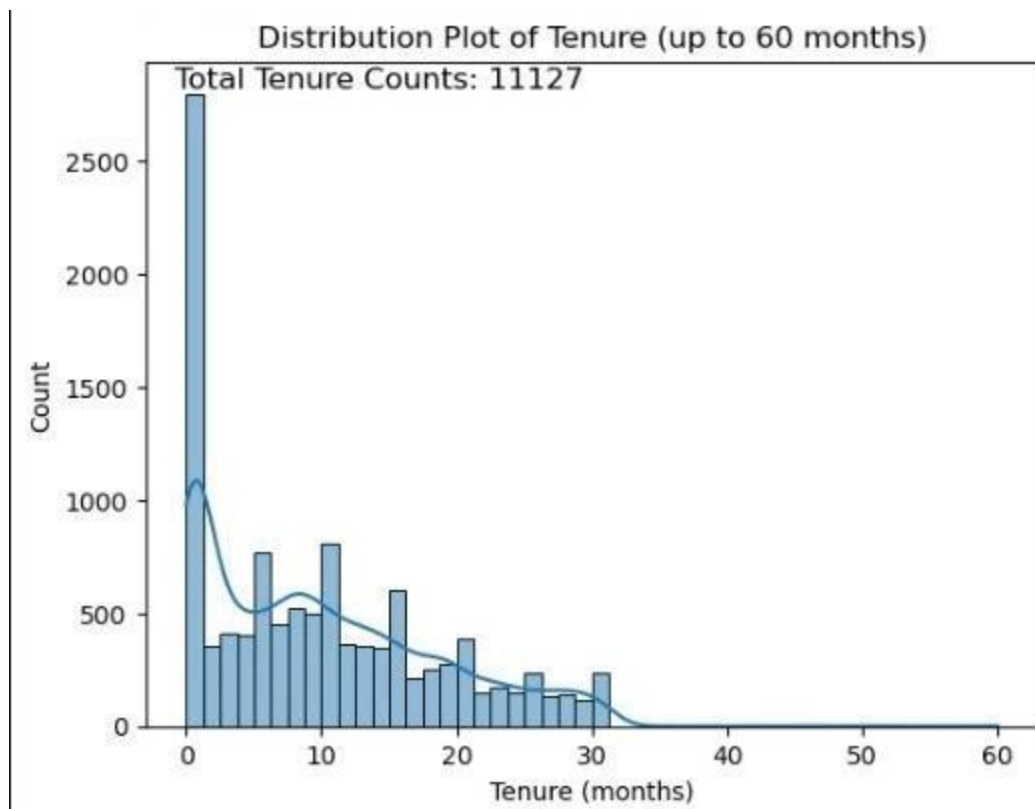
Metrics which are most indicative of potential churners

1. Churn Flag:

- **Analysis Method:** Frequency distribution of churned vs. non-churned accounts
- **Impression:** From the data we can observe that 16.83% i.e. around 1896 accounts have churned.
- **Business Insights:** By Understanding this metric the company should make decisions and allocate resources to reduce the churn of accounts

Frequency of Churn Vs Non Churn Accounts

Account Status
— Non-Churn: 9364
— Churn: 1896
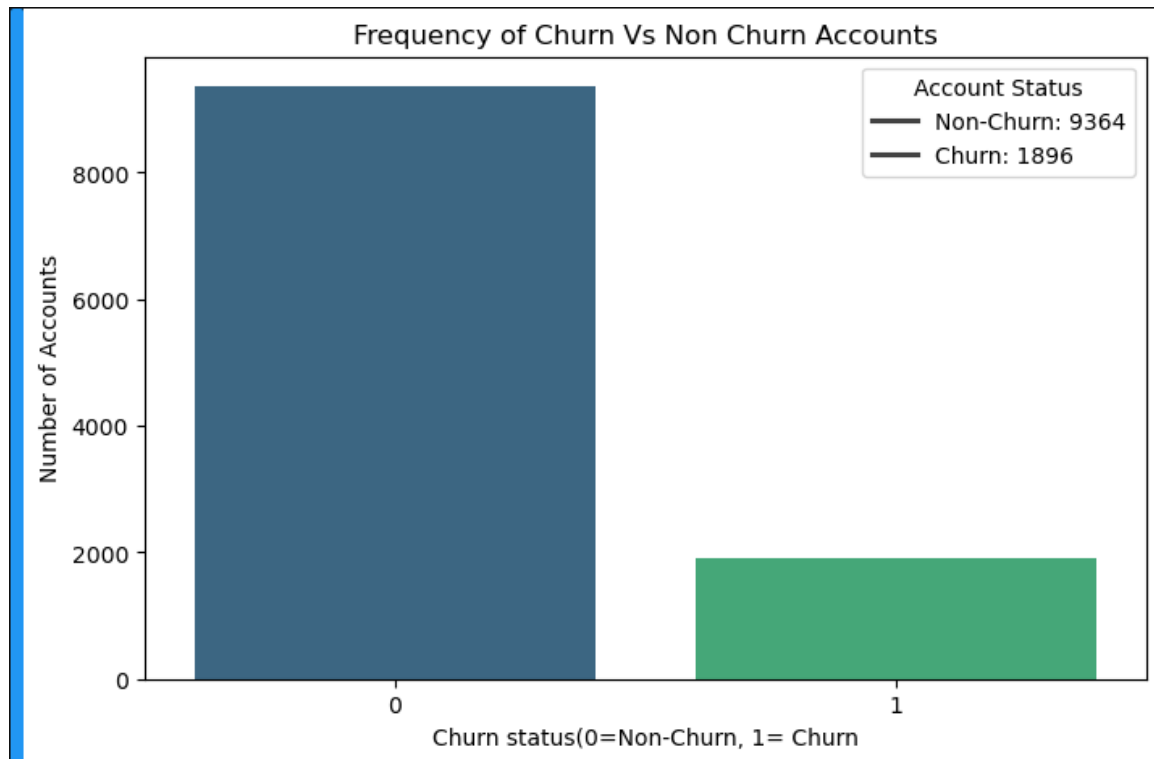
3.3.1.1 & 2 Churn observation & Tenure Plot

Distribution Plot of Tenure (up to 60 months)

Total Tenure Counts: 11127

2. Tenure:

- **Analysis Method:** Descriptive statistics (mean, median, mode) to understand the average duration of accounts using Histograms or box plots to visualize the distribution.
- **Observation:** The chart 3.3.1.2 Tenure Plot is right skewed which indicates there are chances of churn by non-churners if the business doesn't take proactive measures to retain customers.
- **Business insights:** Taking measures to retain customers in a long term is much recommended

3. CC_Contacted_LY:

- **Analysis Method:** Overview of the number of contacts customers made to customer care in the last 12 months, focusing on frequency distribution.
- **Observation:** The chart 3.3.1.3 indicates high frequency of customer care contact in the last 12 months, indicating customers are facing issues with the service provided.
- **Business Insights:** Customer dissatisfaction may lead to churn, understanding the pain points of the customers and taking measures to address concerns will help in retaining the customers

Frequency Distribution of Customer Contacts in Last 12 Months
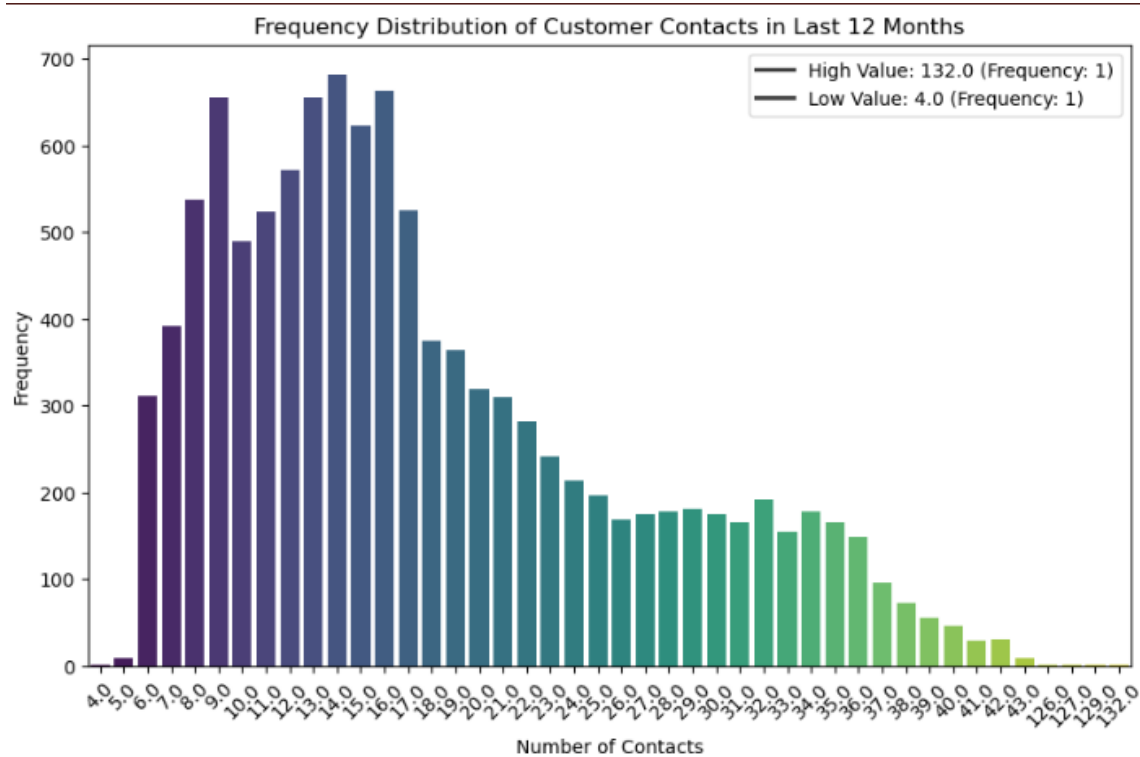
3.3.1.3 & 4 CC_Contacted_LY and Service Score

Distribution of Service Satisfaction Scores (0-5)
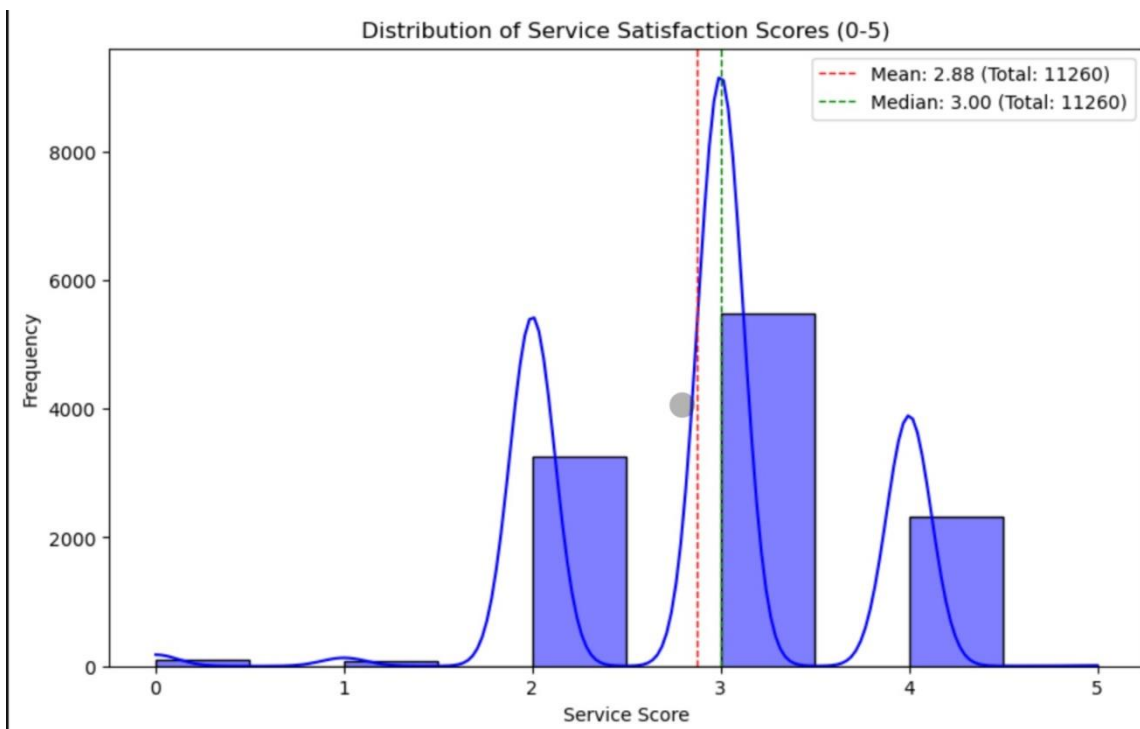
4. Service_Score:

- **Analysis Method:** Descriptive analysis of service satisfaction scores to identify trends in customer satisfaction over time.
- **Observation:** As shown in chart 3.3.1.4 Mean: 2.88 indicates customer are not happy with the service.
- **Business insights:** Training sales professional with focus on delivering best service may help in retaining the customers.

5. CC_Agent_Score:

- **Analysis Method:** Distribution and summary statistics of customer care satisfaction scores to gauge overall experience with customer service.
- **Observation:** As shown in chart 3.3.1.5 Mean: 3.03 indicates customer are not happy in the interaction they had with customer care professionals.
- **Business Insights:** Re-Training of existing and new customer care professional will drive the satisfaction score to a better scale, which in turn will lead to customer retention.

6. Complain_ly:

- **Analysis Method:** Frequency of complaints raised within the last 12 months, including average complaints per customer.
- **Observation:** As shown in 3.3.1.6 Total Complaints Raised in a year is 3111 i.e. 27.6%
- **Business Insights:** By Improving services and customer care professionals a business can reduce the complaints raised by the customers.
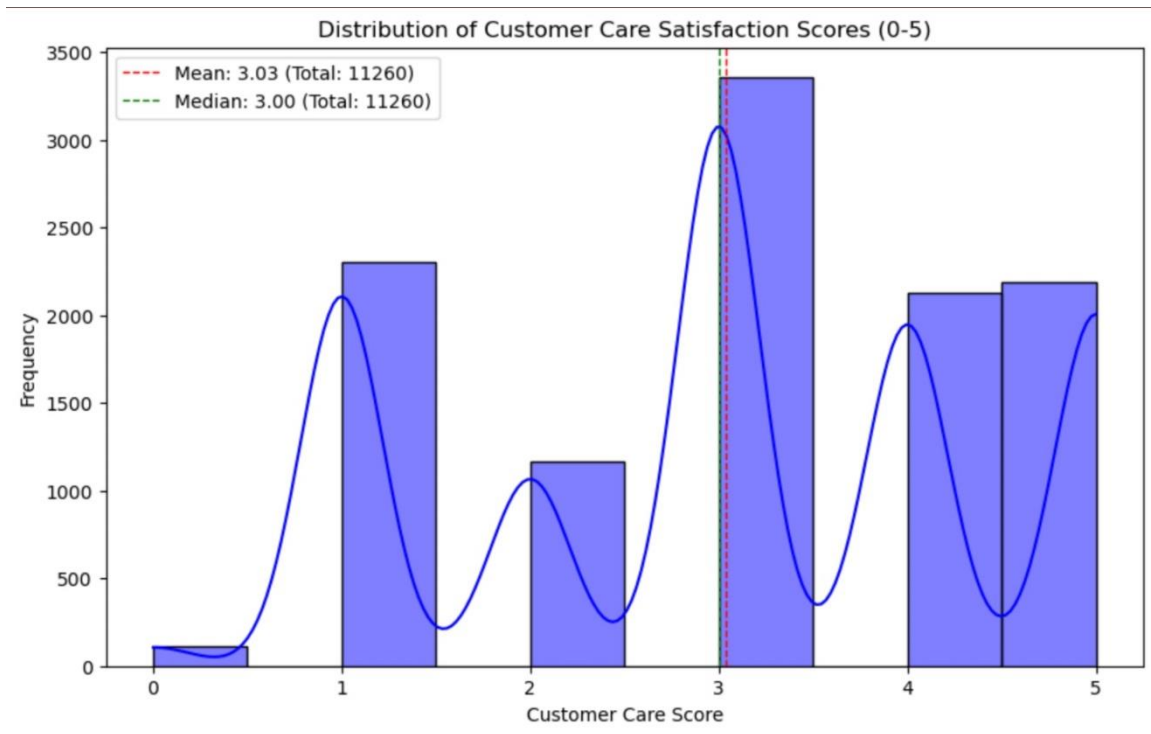
3.3.1.5 Customer Care Satisfaction Score

3.3.1.6 Frequency of Complaints raised in the Last 12 Months
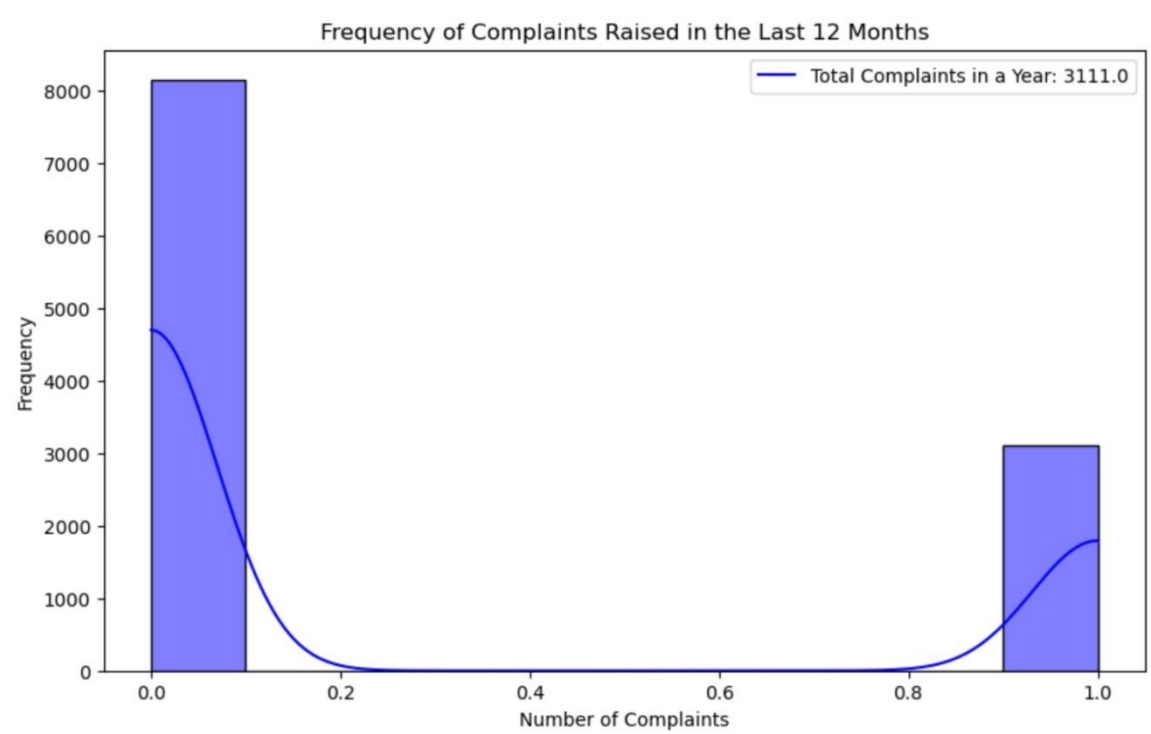
3.3.2. Bivariate Analysis

In bivariate analysis, we explore the relationships between pairs of variables to identify potential correlations or patterns.

1. Churn Flag and Gender:
- **Analysis Method:** Examine the relationship between gender and churn status e.g., using a bar chart.
- **Observation:** Chart 3.3.2.1 Indicates Male customer are churning more comparing to female customers**.**
- **Business Insights:** The company should focus more on retaining more male customers by targeted market campaign.

2. Complaint_LY and CC_Contacted_L12m:
- **Analysis method:** Analyze the correlation between the number of contacts with customer care contacted and complaint raised.
- **Observation:** This observation from chart 3.3.2.2 helps us in gauging the insights of how many customers call with a complaint regarding the services provided**.**
- **Business insights:** This statistic will help a business in making data driven decision to solve the complaint promptly which will in turn help the company receive better service score and customer satisfaction**.**
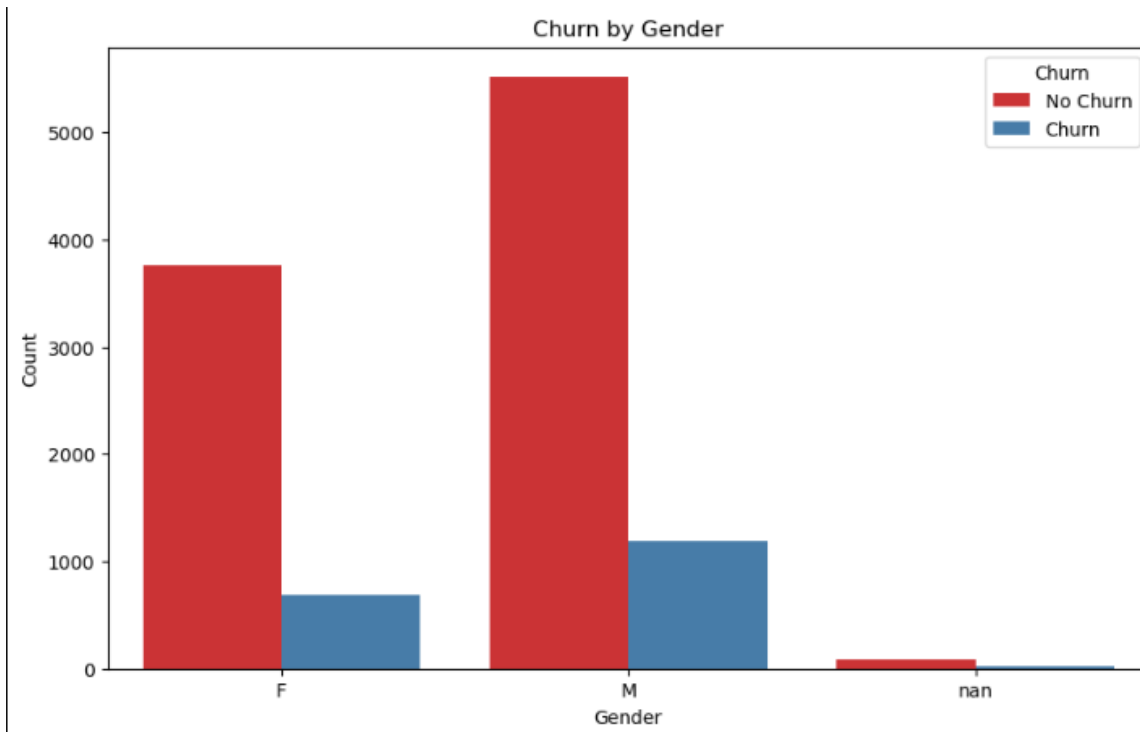
3.3.2.1 Churn by Gender

**3.3.2.2 Total Complaints last year vs cc connect span**
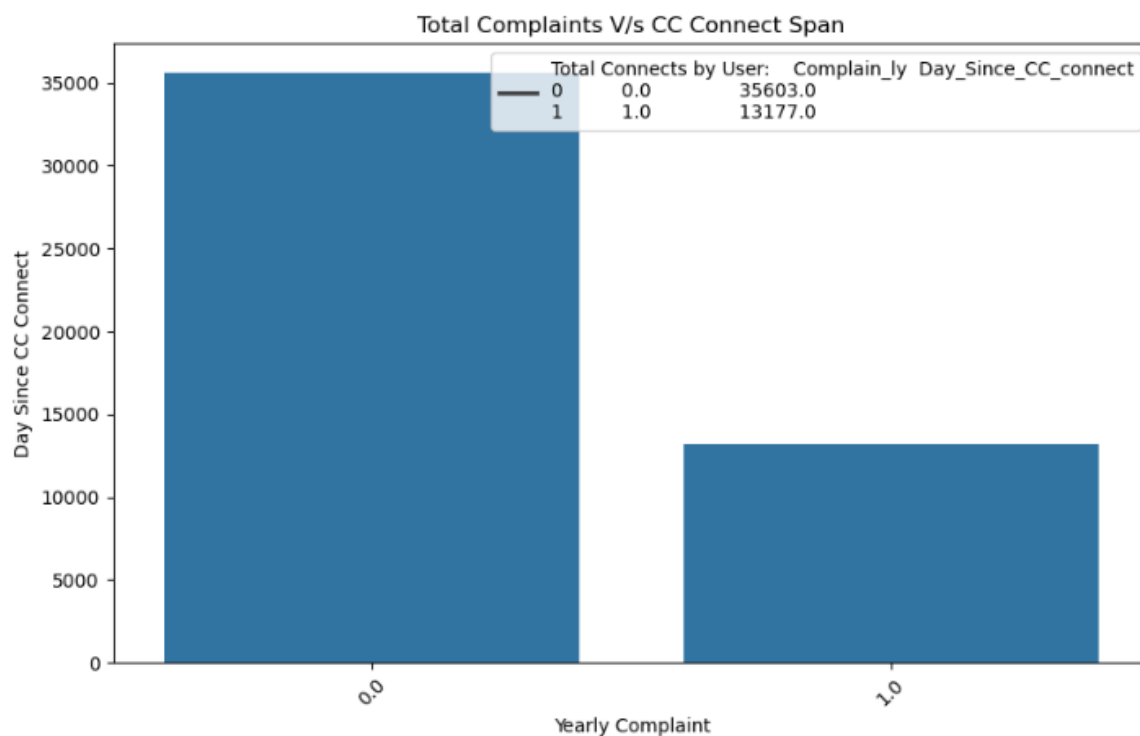
3. Coupon Used and Account Segment

- **Analysis method:** Understanding correlation of coupon usage between account segments.

- **Observation:** From Chart 3.3.2.3 we can observe Super and Regular plus are using the coupon more comparing to all other segments

- **Business Insights:** Focusing on right pricing will help company recover losses on coupon issued. Another insight is that accounts like regular plus and super can be up selling point for sales and customer care professional to scale this segment of accounts.

4. City Tier and Account Segment

- **Analysis Method:** Analyze the relationship between City_Tier and account segments using chart and heatmap for correlation analysis.

- **Observation:** From chart 3.3.2.4 summarizes the count of different account segments across the city tier (1.0, 2.0, 3.0). It also gives a better understanding at the no. of distribution in each segments of customers in each city tier.

- **Business Insights:** Regular Plus and super appears to be the most common across all city tiers. By making the common segments of accounts more accessible in all city tier by a targeted marketing efforts will increase the customer base at a better scale.
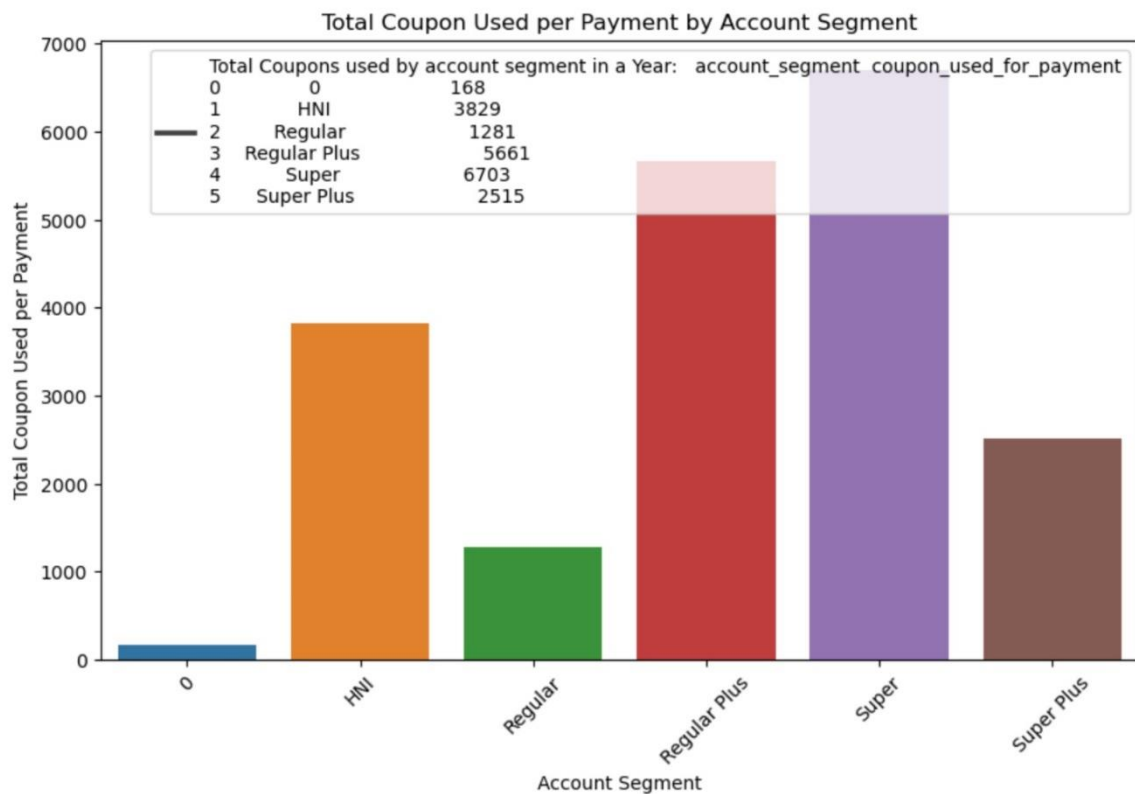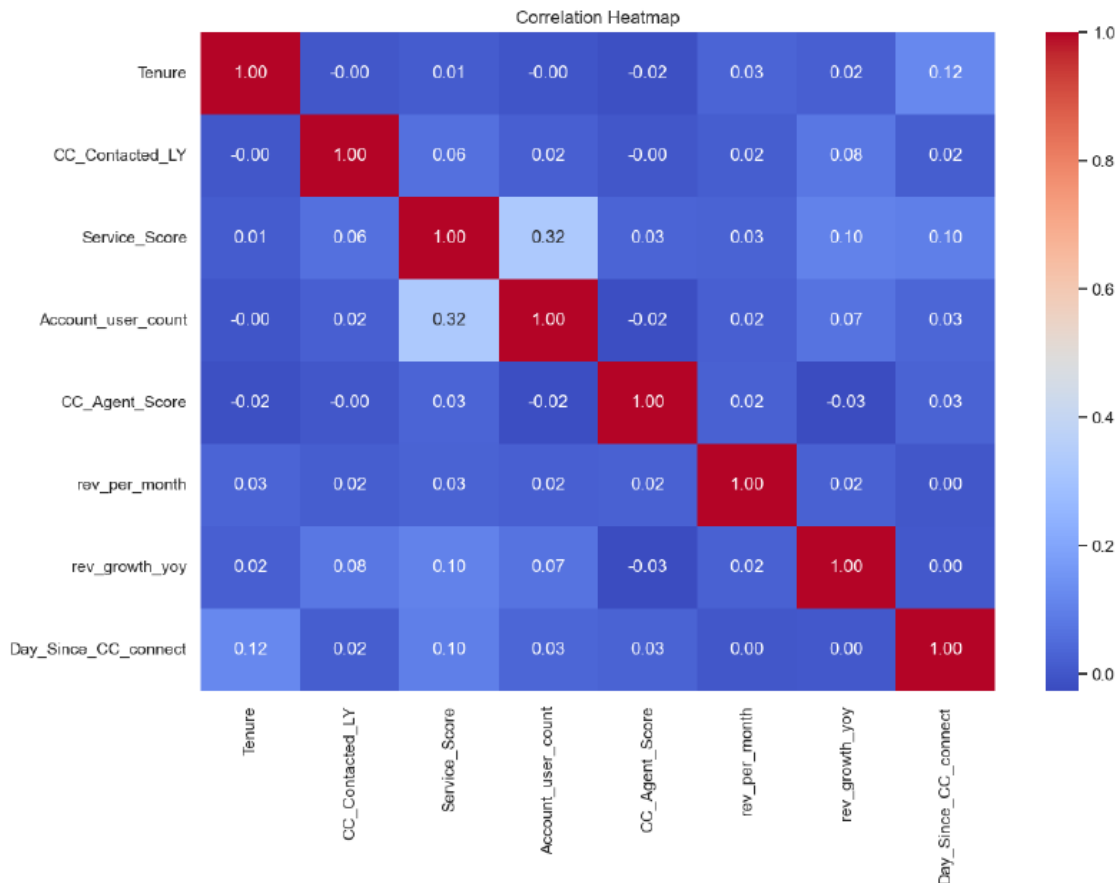
3.3.2.3 & 4 – Total Coupon Used per payment by account segment and Comparison of City Tier and Account Segment
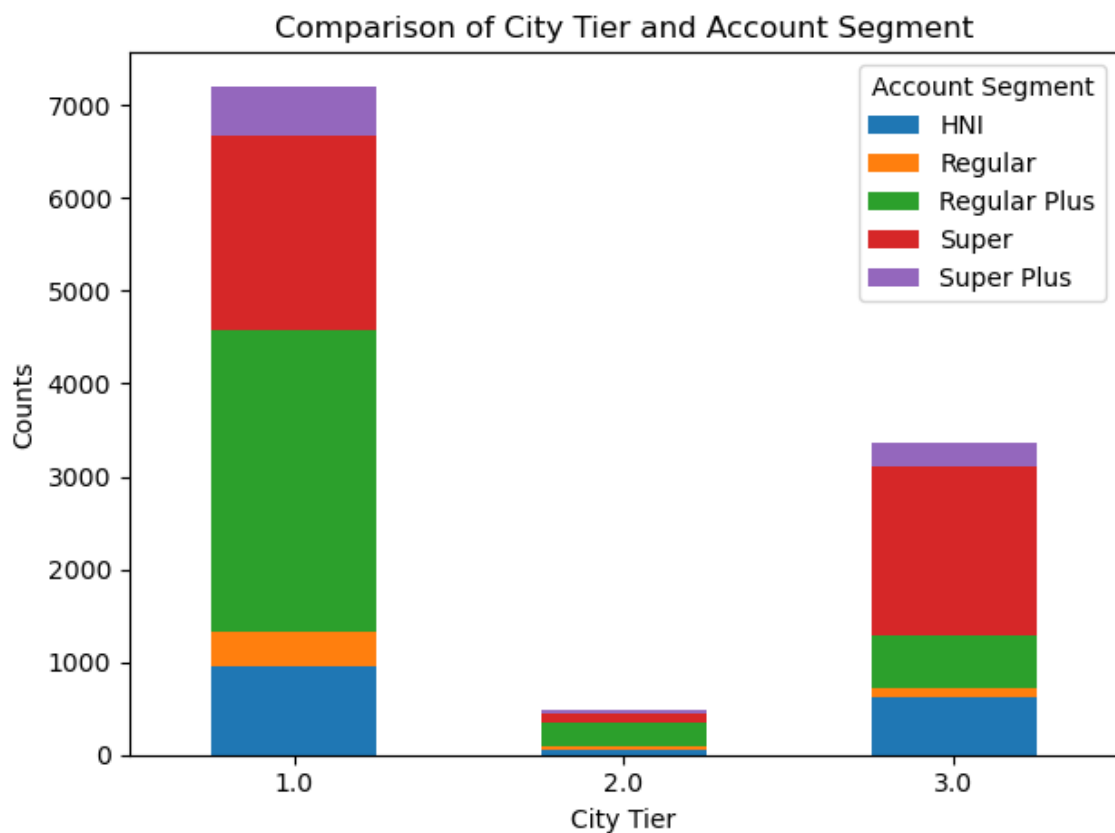
3.3.3. Correlation Analysis

- **Analysis Method: A Correlation matrix to display relationship between various metrics like Tenure, CC_Contacted_LY, Service_Score and other transactional metrics.**

3.3.3 Heat map of correlation matrix

Correlation Heatmap

	Tenure	CC_Contacted_LY	Service_Score	Account_user_count	CC_Agent_Score	rev_per_month	rev_growth_yoy	Day_Since_CC_connect
Tenure	1.00	-0.00	0.01	-0.00	-0.02	0.03	0.02	0.12
CC_Contacted_LY	-0.00	1.00	0.06	0.02	-0.00	0.02	0.08	0.02
Service_Score	0.01	0.06	1.00	0.32	0.03	0.03	0.10	0.10
Account_user_count	-0.00	0.02	0.32	1.00	-0.02	0.02	0.07	0.03
CC_Agent_Score	-0.02	-0.00	0.03	-0.02	1.00	0.02	-0.03	0.03
rev_per_month	0.03	0.02	0.03	0.02	0.02	1.00	0.02	0.00
rev_growth_yoy	0.02	0.08	0.10	0.07	-0.03	0.02	1.00	0.00
Day_Since_CC_connect	0.12	0.02	0.10	0.03	0.03	0.00	0.00	1.00

- **Observation:** We can derive at a key insight by looking at the matrix above which shows a positive correlation between Service_Score and Account_user_count. This correlation is a clear indication that increase in service score will in turn result in increase in user count.

- **Business Insights:** The Company Should focus on Service Driven Growth Strategy to retain their customer base.

Chapter 4

Findings, Recommendations and Conclusion

Findings, Recommendations and Conclusion

4.1 Findings Based on the Observations

Campaign Recommendations:

Segmentation: Segment potential churners based on their churn probability and other relevant factors like account tenure, revenue, and customer demographics from the observations of the metrics.

Campaign Offers:

- **High-Value Customers (High Churn Probability, High Revenue):**
    - **Personalized Loyalty Program:** Offer exclusive discounts, early access to new products, and personalized recommendations based on past purchase history.
    - **Dedicated Customer Support:** Provide dedicated account managers for high-value customers to address their concerns and offer personalized solutions.
    - **Exclusive Events and Promotions:** Invite them to exclusive events, product launches, and offer special discounts on premium products.
- **Mid-Value Customers (Moderate Churn Probability, Moderate Revenue):**
    - **Targeted Discounts and Promotions:** Offer discounts on specific product categories or brands based on their past purchase behavior.
    - **Free Shipping and Returns:** (Exclusive for E-Commerce) Provide free shipping and easy returns to incentivize repeat purchases.
    - **Referral Programs:** Offer rewards for referring new customers to the platform.

- **Low-Value Customers (Low Churn Probability, Low Revenue):**
  - **Targeted Email Marketing:** Send personalized emails with relevant product recommendations and special offers.
  - **Free Trial Offers:** Offer free trials of premium services or subscriptions to encourage engagement.
  - **Bundled Offers:** Offer bundled deals with complementary products to increase average order value.

Campaign Implementation:

- **A/B Testing:** Conduct A/B testing to compare the effectiveness of different campaign offers and optimize the campaign strategy.
- **Real-time Personalization:** Use the churn prediction model to personalize offers and communication based on individual customer behavior and preferences.
- **Continuous Monitoring and Evaluation:** Monitor the campaign performance and make adjustments as needed to ensure its effectiveness.

Revenue Assurance:

- **Cost-Benefit Analysis:** Conduct a cost-benefit analysis to ensure that the campaign offers are profitable and do not result in significant revenue loss.
- **Transparent Reporting:** Provide clear and transparent reporting to the revenue assurance team on the campaign performance, including cost, revenue, and customer retention metrics.

Unique Campaign Offer:

"Loyalty Rewards Program"

- **Tiered Rewards:** Customers are assigned to different tiers based on their spending and loyalty. Each tier offers increasing benefits like discounts, early access, and exclusive perks.

- **Points System:** Customers earn points for every purchase, which can be redeemed for discounts, free products, or other rewards.

- **Personalized Recommendations:** Customers receive personalized product recommendations based on their purchase history and preferences.

- **Exclusive Events:** High-tier customers are invited to exclusive events, product launches, and meet-and-greets with brand ambassadors.

Conclusion:

By implementing a comprehensive churn prediction model created in this project and targeted retention campaigns, the E-commerce or DTH company can effectively identify and retain at-risk customers, ultimately driving revenue growth and customer loyalty. The proposed "Loyalty Rewards Program" offers a unique and cost-effective solution that can be tailored to different customer segments, ensuring its approval by the revenue assurance team. Regular testing, refining of the feature engineering, expansion of data sources and personalization of the model is required for accurate performance of this "Customer Churn Prediction Model".

ANNEXURE (if any)

The questionnaires, financial statements and any other relevant document can be put here. The annexures have to be numbered in case there are more than one annexure.