Customer Churn: Case study and Analysis

By: Saumik Idnani

Problem Statement:

You are a data scientist working for a retail store application with a significant customer base. The company is interested in understanding customer churn patterns and segmenting its customer base to implement targeted marketing and retention strategies.

Introduction:

What is customer churn?
Customer churn, also known as customer attrition, refers to the loss of customers over a given period.

So, we will perform customer churn analysis to identify factors contributing to customer attrition and then conduct customer segmentation analysis to tailor marketing efforts to different customer segments. We have access to a Retail Store Customer Dataset, which includes customer information, order history, complaints raised, coupons used, and churn status.

Dataset: https://data.world/wayvytech/a-retail-store-customer-churn-dataset

Data Preprocessing:

The Dataset had missing values (null values) in them. The columns with null values are as follows:

- OrderAmountHikeFromlastYear 265
- CouponUsed 256
- OrderCount 258
- DaySinceLastOrder 307
- HourSpendOnApp 255
- WarehouseToHome 251
- Tenure 264

The following columns that had null values are filled the median of the values in their respective columns.

Fill missing values with the mean or median of the column. This is useful if the data is skewed (median).

Finally, The columns with text data were Label encoded to fit.

Analysis:

We performed three types of analysis:

1. Univariate Analysis - Univariate Analysis focuses on examining each variable individually to understand its distribution, central tendency, and dispersion. For categorical variables, the analysis involves visualizing the frequency distribution of each category. For numerical variables, histograms are used to visualize the distribution, central tendency (mean, median), and dispersion (variance, standard deviation).
2. Bivariate Analysis - Bivariate Analysis examines the relationship between two variables to identify patterns, correlations, and potential causal relationships. In the context of customer churn, it helps to understand how different variables are associated with the likelihood of a customer churning. By comparing the distribution of categorical variables with churn, we can identify which categories are more prone to churn. Box plots for numerical variables against churn can highlight differences in distribution and central tendency between churned and non-churned customers.
3. Clustering - Clustering is an unsupervised learning technique used to group similar data points together based on their characteristics. In the context of customer segmentation, clustering helps identify distinct groups of customers who exhibit similar behaviors and attributes.
    - Steps for Clustering
        i. Feature Selection: Select relevant features that capture important aspects of customer behavior and demographics.
        ii. Applying K-Means Clustering: Use the K-Means algorithm to partition customers into clusters.
        iii. Analyzing Segment Characteristics: Calculate the mean values of features for each segment to understand their characteristics.
        iv. Visualization: Plot the clusters to visualize how customers are grouped.

Modeling:

Predictive modeling involves using historical data to predict future customer behavior. By building a churn prediction model, businesses can proactively identify at-risk customers and take actions to prevent churn.

We will train a random forest model for evaluating customer churn.

Random Forest Model Performance

Overall Accuracy:

- The model achieved an accuracy of **96.36%**. This means that 96.36% of the predictions made by the model were correct.

Classification Report:

- **Precision**:
  - For class `0` (non-churners): 96%
  - For class `1` (churners): 99%
  - Precision indicates how many of the predicted positive cases were actually positive. A high precision for class `1` (churners) means the model is very good at correctly identifying churners.
- **Recall**:
  - For class `0`: 100%
  - For class `1`: 78%
  - Recall indicates how many of the actual positive cases were correctly identified by the model. A recall of 78% for class `1` means the model correctly identified 78% of the actual churners.
- **F1-Score**:
  - For class `0`: 98%
  - For class `1`: 88%
  - The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both. The high F1-score for both classes indicates that the model is performing well in distinguishing between churners and non-churners.
- **Support**:
  - The number of actual instances for each class:
    - Class `0` (non-churners): 941
    - Class `1` (churners): 185

Interpretation:

- The model performs exceptionally well in predicting non-churners (class `0`) with perfect recall and high precision.

- For churners (class 1), while the precision is very high, the recall is lower at 78%, indicating some churners are not being identified by the model.
- The overall high accuracy and balanced F1-scores suggest that the model is effective at distinguishing between churners and non-churners, though there may be room for improvement in identifying all churners accurately.

Insights:

The following insights can be derived from the clustered data.

Metric	Cluster 0	Cluster 1	Cluster 2
Churn Rate (%)	3.81	22.19	15.02
Customer Count	1314	3484	832
Average Tenure (months)	20.08	5.99	11.76
Avg Warehouse to Home Delivery Time (days)	16.3	15.34	15.38
Average Hours Spent on App	2.95	2.9	3.05
Average Number of Devices Registered	3.72	3.65	3.82
Average Satisfaction Score	3.04	3.06	3.14
Average Order Amount Hike from Last Year (%)	15.68	15.69	15.59
Average Order Count	2.41	1.84	8.52
Average Days Since Last Order	5.6	3.17	8.04
Average Cashback Amount	224.16	154.05	200.14

Churn Rates: Cluster 1 has the highest churn rate at 22.19%, indicating a significant portion of customers are leaving. Cluster 0 has the lowest churn rate at 3.81%, suggesting higher customer retention.

Customer Engagement: Cluster 0 has the longest average tenure (20.08 months) and the highest average cashback amount (224.16). Cluster 1 has the shortest average tenure (5.99 months) and the lowest average cashback amount (154.05).

Activity on the App: Cluster 2 customers spend the most time on the app (3.05 hours) and have the highest average order count (8.52).

Satisfaction and Delivery: All clusters have similar satisfaction scores around 3, with Cluster 2 slightly higher at 3.14. Average delivery times are relatively similar across clusters, with Cluster 0 having the highest at 16.30 days.

Cluster 0 seems to represent loyal customers with longer tenure and higher cashback amounts, whereas Cluster 1 indicates at-risk customers with high churn rates and shorter tenures. Cluster 2 appears to have moderately engaged customers who spend more time on the app and place more orders.

Retention Strategies:

Based on the insights from the customer churn data, here are tailored customer retention strategies for each cluster:

Cluster 0: Loyal Customers

- **Churn Rate**: 3.81%
- **Average Tenure**: 20.08 months
- **Average Cashback Amount**: 224.16

Retention Strategies:

1. **Loyalty Programs**: Enhance existing loyalty programs by offering exclusive benefits and rewards for long-term customers.
2. **Personalized Communication**: Send personalized thank-you messages, special offers, and birthday/anniversary greetings.
3. **Premium Services**: Introduce premium services or memberships that offer additional perks, such as faster delivery or early access to sales.
4. **Referral Programs**: Encourage these satisfied customers to refer friends and family by offering referral bonuses.

Cluster 1: At-Risk Customers

- **Churn Rate**: 22.19%
- **Average Tenure**: 5.99 months
- **Average Cashback Amount**: 154.05

Retention Strategies:

1. **Targeted Discounts**: Offer special discounts or cashback incentives to encourage continued usage and purchases.
2. **Engagement Campaigns**: Develop campaigns to re-engage these customers, such as personalized email marketing or in-app notifications highlighting new products or features.
3. **Customer Feedback**: Actively seek feedback from these customers to understand their pain points and address any issues promptly.
4. **Improved Onboarding**: Enhance the onboarding process to ensure new customers understand the value and features of the service, reducing early churn.

Cluster 2: Moderately Engaged Customers

- **Churn Rate**: 15.02%
- **Average Tenure**: 11.76 months
- **Average Order Count**: 8.52

Retention Strategies:

1. **Usage Incentives**: Provide incentives for increased usage, such as discounts for higher order volumes or frequency-based rewards.
2. **Content and Community**: Create engaging content, such as tutorials, reviews, and community forums, to keep these customers engaged and involved.
3. **Exclusive Previews**: Offer early access to new products or services to make these customers feel valued and appreciated.
4. **Proactive Support**: Implement proactive customer support strategies, such as regular check-ins and assistance, to prevent potential issues from causing churn.

Implementation Plan:

Here's an implementation plan for the suggested retention strategies, including a timeline to measure their success.

Phase 1: Preparation (Month 1)

- **Week 1-2**:
    - Form a retention strategy team.
    - Conduct detailed customer analysis for further segmentation.
    - Set specific goals for retention improvement.
- **Week 3-4**:
    - Develop detailed project plans for each strategy.
    - Allocate budget and resources.

Phase 2: Initial Rollout (Month 2-3)

- **Week 1-2**:
    - Launch loyalty programs and referral incentives for Cluster 0.
    - Start targeted discount campaigns for Cluster 1.
    - Implement usage incentives for Cluster 2.
    - Enhance onboarding processes for new customers.
- **Week 3-4**:
    - Begin personalized communication and engagement campaigns.
    - Introduce customer feedback mechanisms.
    - Establish proactive support strategies.

Phase 3: Full Implementation (Month 4-6)

- **Month 4**:
    - Launch premium services and exclusive previews.
    - Roll out content and community-building initiatives.
    - Implement multi-channel support systems.
    - Start proactive customer support outreach.
- **Month 5-6**:
    - Optimize and refine strategies based on initial feedback.
    - Conduct training for support and marketing teams.
    - Promote loyalty and referral programs through various channels.

Phase 4: Optimization and Monitoring (Month 7-12)

- **Month 7-8**:
    - Analyze initial results and feedback.
    - Adjust strategies based on performance data.
    - Intensify efforts on the most effective strategies.
- **Month 9-12**:
    - Continue monitoring and optimizing strategies.
    - Plan for long-term sustainability of retention initiatives.
    - Prepare a comprehensive report on the retention strategy effectiveness.