Ecommerce Customer Churn Analysis and Prediction

~ NUPOOR LONDHE

INDEX

Overview

This project focuses on analyzing an e-commerce customer dataset to uncover key insights into churn drivers, behavioral patterns, and emerging trends. The Power BI dashboard enables the identification of potential data anomalies and manipulation patterns through interactive visualizations. Following the exploratory analysis, the labeled data is used to build and train predictive machine learning models. Both supervised algorithms—such as Random Forest and Logistic Regression—and unsupervised algorithms—such as KNN and SVM—are applied. Model performance is then evaluated based on accuracy metrics to determine the most effective predictive model.

Key Concepts

What is customer Churn ?

In terms of business, churn is the measure of how many customers stop using a company's product, platform or service during a specific time period.

Why are we concerned ?

As a business owner, understanding customer behavior is crucial because it directly influences:

- The growth and long-term sustainability of your busines
- The market viability of your products or services
- Consistent revenue generation
- Competitive positioning and advantage
- The cost and efficiency of customer acquisition
- The ability to identify and address customer dissatisfaction

Acquiring new customers is typically far more expensive than retaining existing ones often estimated to be five times (or more) costlier. This is because new customer acquisition requires extensive marketing and advertising efforts, whereas customer retention relies on more cost-effective strategies such as improved customer service, personalized offers, and loyalty programs.

How is Churn Calculated ?

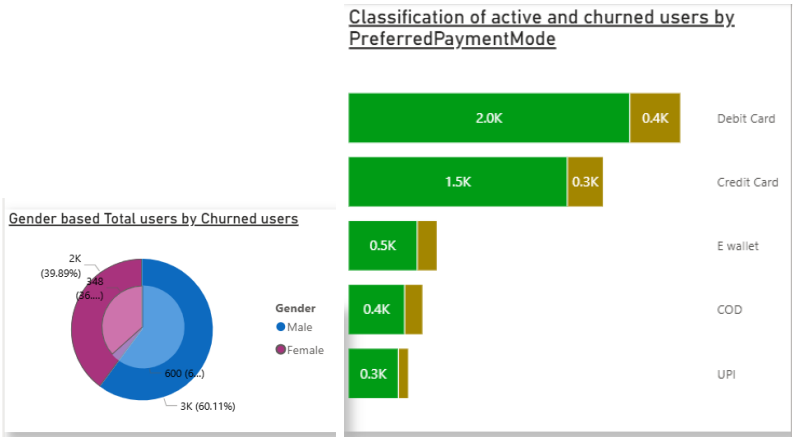
$$\text{Churn Rate} = \left(\frac{\text{Lost Customers}}{\text{Total Customers at the Start of the Time Period}}\right) \times 100$$

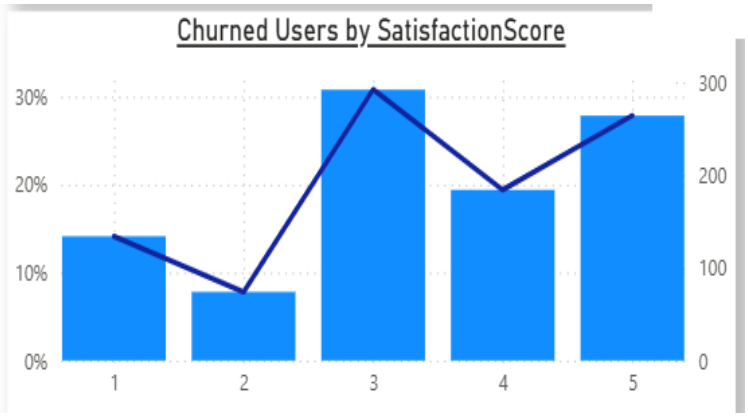
Dataset Description

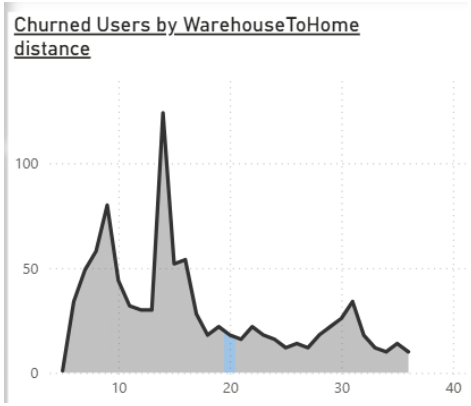
E Comm CustomerID: Unique customer ID
E Comm Churn: Churn Flag
E Comm Tenure: Tenure of customer in organization
E Comm PreferredLoginDevice: Preferred login device of customer
E Comm CityTie:r City tier
E Comm WarehouseToHome: Distance in between warehouse to home of customer
E Comm PreferredPaymentMode: Preferred payment method of customer
E Comm Gender: Gender of customer
E Comm HourSpendOnApp: Number of hours spend on mobile application or website
E Comm NumberOfDeviceRegistered: Total number of deceives is registered on particular customer
E Comm PreferedOrderCat: Preferred order category of customer in last month
E Comm SatisfactionScore: Satisfactory score of customer on service
E Comm MaritalStatus: Marital status of customer
E Comm NumberOfAddress: Total number of added added on particular customer
E Comm Complain: Any complaint has been raised in last month
E Comm OrderAmountHikeFromlastYear: Percentage increases in order from last year
E Comm CouponUsed: Total number of coupon has been used in last month
E Comm OrderCount: Total number of orders has been places in last month
E Comm DaySinceLastOrder: Day Since last order by customer
E Comm CashbackAmount: Average cashback in last month
https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction

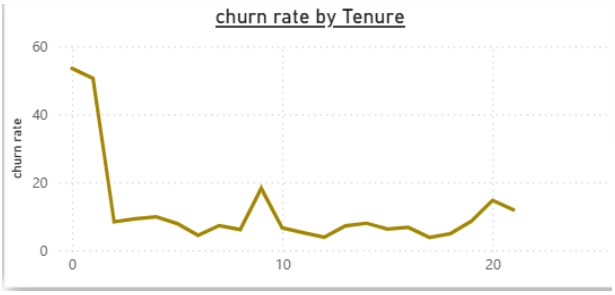
Tableau Dashboard

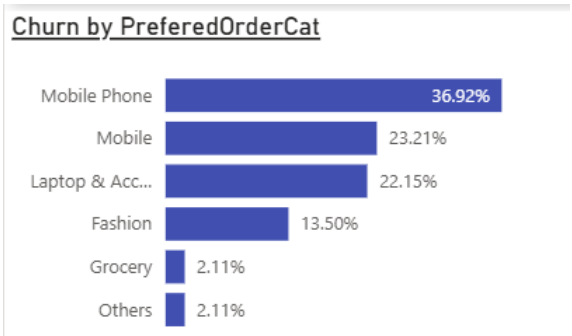
Snapshots of the dashboard







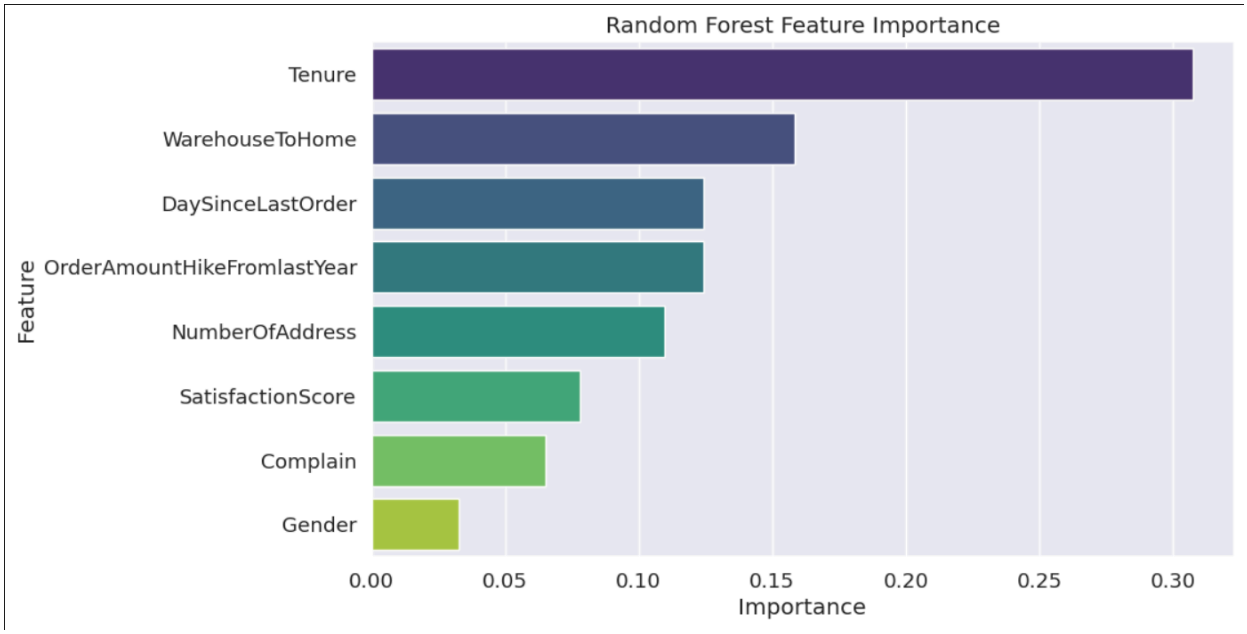


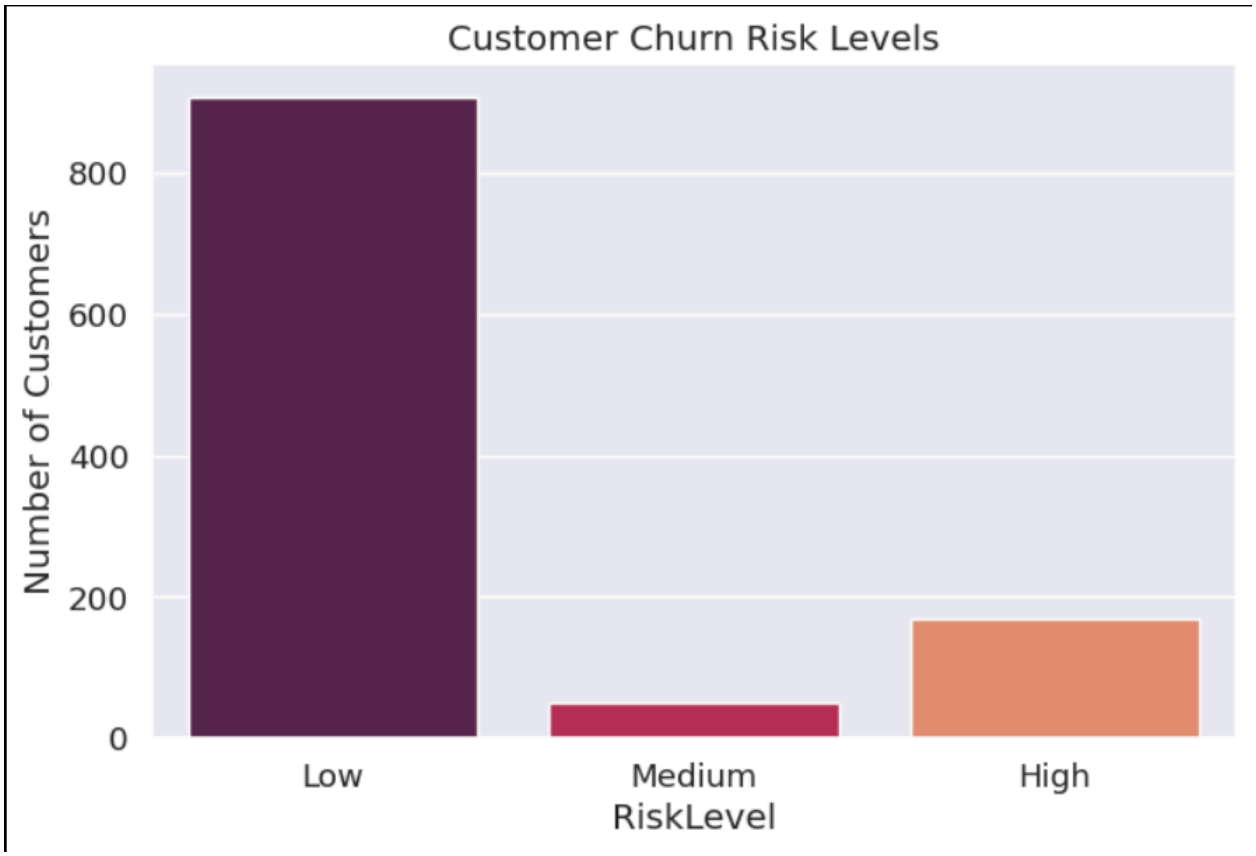


Predictive Machine learning Models

Logistic Regression
Random Forest
SVM
KNN

	Accuracy	Precision	Recall	F1-score	AUC
Random Forest	96.36%	89.01%	89.47%	89.24%	98.98%
Support Vector Machine	93.78%	85.29%	76.32%	80.56%	95.59%
K-Nearest Neighbors	93.87%	94.81%	67.37%	78.77%	93.44%
Logistic Regression	86.94%	66.67%	45.26%	53.92%	85.59%


Random Forest Feature Importance
Tenure
WarehouseToHome
DaySinceLastOrder
OrderAmountHikeFromlastYear
NumberOfAddress
SatisfactionScore
Complain
Gender
Feature
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Importance


Customer Churn Risk Levels
Number of Customers
800
600
400
200
0
Low
Medium
High
RiskLevel

Impact

- Understanding Churn Triggers: By identifying what prompts customers to churn, you can gain a deeper understanding of their motivations and pain points. These triggers could range from dissatisfaction with your product or service to more attractive offers from competitors, poor customer service experiences, or unmet expectations. Recognizing these triggers allows you to address them directly, thereby improving customer satisfaction and loyalty.
- Identifying Weak Points in Your Product or Service: Churn analysis can reveal areas where your product or service may not be meeting customer expectations. Addressing these weak points can help reduce churn.
- Predicting and Preventing Churn: One of the most powerful outcomes of churn analysis is the ability to predict which customers are likely to churn. By developing predictive models based on historical data and identified churn triggers, you can proactively reach out to at-risk customers with tailored interventions, such as special offers, personalized support, or loyalty programs designed to re-engage and retain them.

Future Scope

The insights derived from this project can play a vital role in shaping data-driven retention strategies and business decisions. By identifying the key drivers of customer churn, organizations can proactively implement targeted interventions—such as personalized marketing campaigns, loyalty programs, or improved customer service—to reduce attrition rates and enhance customer satisfaction.

In the future, the predictive models developed in this project can be integrated into real-time decision-making systems. This would enable businesses to monitor customer behavior continuously and automatically flag customers at high risk of churn, allowing timely and customized retention actions.

Additionally, with more diverse datasets and advanced analytical tools, the scope can be expanded to include sentiment analysis, customer lifetime value prediction, and recommendation systems. Such enhancements would strengthen strategic planning, improve business sustainability, and help organizations maintain a competitive edge in the evolving e-commerce landscape.