Project Report: Analyzing Customer Behavior for E-commerce Insights

Introduction

This project was undertaken to analyze customer behavior on an e-commerce platform with the goal of improving sales, enhancing customer engagement, and personalizing the shopping experience. The project involved the collection and analysis of customer data, the development of predictive models to anticipate customer churn, and the deployment of these models to provide actionable insights for the business. The final output is a predictive model that can be used to inform business strategies and optimize customer retention efforts.

Data Preprocessing

The dataset used in this project included various customer attributes such as demographics, browsing behavior, purchase history, and product interactions. The initial step involved cleaning the data by handling missing values and ensuring all variables were appropriately formatted for analysis.

For example, missing values in the `CouponUsed` column were addressed by analyzing the relationship between coupon usage and the number of orders a customer placed. This led to the implementation of a custom imputation strategy, where customers with more frequent orders were assigned a higher likelihood of coupon usage. Other numerical columns with missing data were imputed using the median, which is robust against outliers.

	count	mean	std	min	25%	50%	75%	max
Tenure	5366.0	10.189899	8.557241	0.0	2.00	9.00	16.0000	61.00
WarehouseToHome	5379.0	15.639896	8.531475	5.0	9.00	14.00	20.0000	127.00
HourSpendOnApp	5375.0	2.931535	0.721926	0.0	2.00	3.00	3.0000	5.00
OrderAmountHikeFromlastYear	5365.0	15.707922	3.675485	11.0	13.00	15.00	18.0000	26.00
CouponUsed	5374.0	1.751023	1.894621	0.0	1.00	1.00	2.0000	16.00
OrderCount	5372.0	3.008004	2.939680	1.0	1.00	2.00	3.0000	16.00
DaySinceLastOrder	5323.0	4.543491	3.654433	0.0	2.00	3.00	7.0000	46.00
CashbackAmount	5630.0	177.223030	49.207036	0.0	145.77	163.28	196.3925	324.99

Descriptive Statistics before cleaning

	count	mean	std	min	25%	50%	75%	max
Tenure	5630.0	10.134103	8.357951	0.0	3.00	9.00	15.0000	61.00
WarehouseToHome	5630.0	15.566785	8.345961	5.0	9.00	14.00	20.0000	127.00
HourSpendOnApp	5630.0	2.934636	0.705528	0.0	2.00	3.00	3.0000	5.00
OrderAmountHikeFromlastYear	5630.0	15.674600	3.591058	11.0	13.00	15.00	18.0000	26.00
CouponUsed	5630.0	1.758437	1.863501	0.0	1.00	1.00	2.0000	16.00
OrderCount	5630.0	2.961812	2.879248	1.0	1.00	2.00	3.0000	16.00
DaySinceLastOrder	5630.0	4.459325	3.570626	0.0	2.00	3.00	7.0000	46.00
CashbackAmount	5630.0	177.223030	49.207036	0.0	145.77	163.28	196.3925	324.99

Descriptive Statistics after cleaning

Exploratory Data Analysis (EDA)

Exploratory data analysis was conducted to uncover key patterns and trends in the data. This analysis provided valuable insights into customer behavior, which could be leveraged to improve customer engagement and sales. Some of the key findings include:

- **Customer Tenure:** The majority of customers had a short tenure on the platform, indicating that many users were relatively new. This suggests a need for strategies focused on early customer retention.
- **Order Frequency and Coupon Usage**: There was a notable correlation between the number of orders placed and the use of coupons, suggesting that promotional offers are an effective tool for encouraging repeat purchases.
- **Preferred Login Device:** A significant portion of customers preferred to access the platform via mobile devices, underscoring the importance of optimizing the mobile shopping experience.
- **Customer Engagement**: The data suggests that while most customers are relatively new (low tenure), they engage with the platform to a moderate extent, as seen in hours spent on the app and the number of orders placed. However, engagement drops off quickly for higher levels of interaction (e.g., order count, coupon usage).
- **Promotions and Incentives**: Cashback amounts and coupon usage show that while promotional incentives are being utilized, they are not widespread among all customers. There may be an opportunity to increase engagement by optimizing these promotions.
- **Logistics Impact**: The distance from the warehouse to home could impact delivery efficiency, and the distribution suggests that most customers are within a manageable range
- Mobile Phone and Laptop & Accessory categories have the highest number of customers, but they also exhibit the highest churn rates.

Fashion has a lower overall customer base, but the churn rate seems significant when compared proportionally.

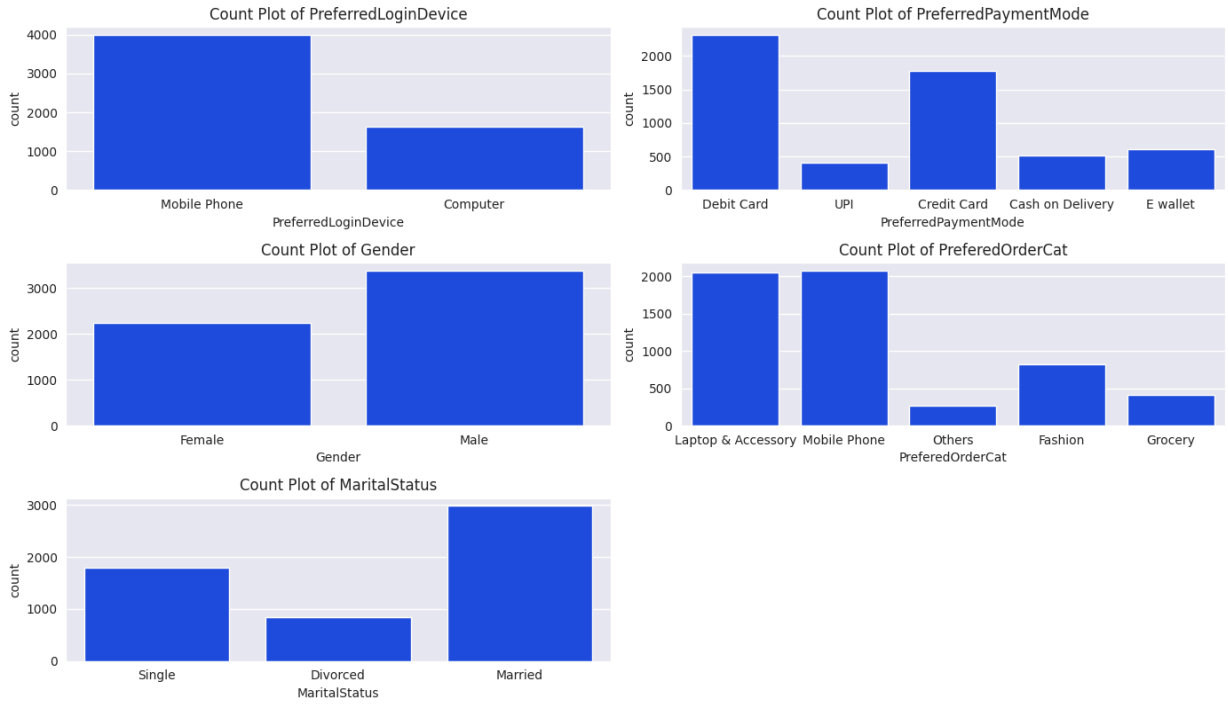
Grocery shows a low churn rate, which could indicate higher customer retention in this category.

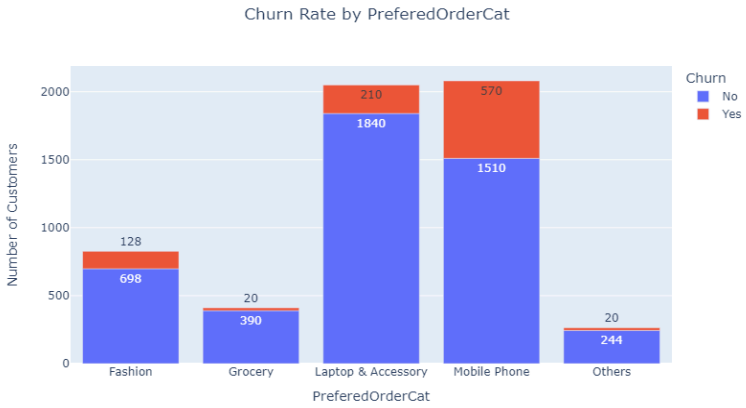
- A majority of customers prefer using Mobile Phones to log in, but this group also has a higher churn rate compared to those using Computers

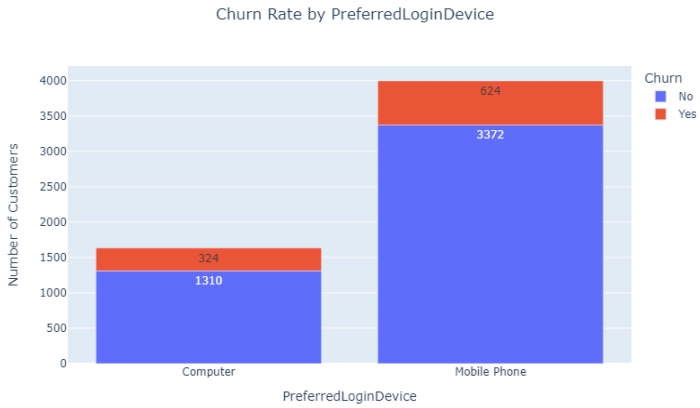
  The higher churn rate among mobile users could indicate issues related to the mobile experience,
- Customers with a lower tenure (less time with the company) are more likely to churn. This trend sharply decreases as tenure increases, suggesting that the longer customers stay, the less likely they are to churn.

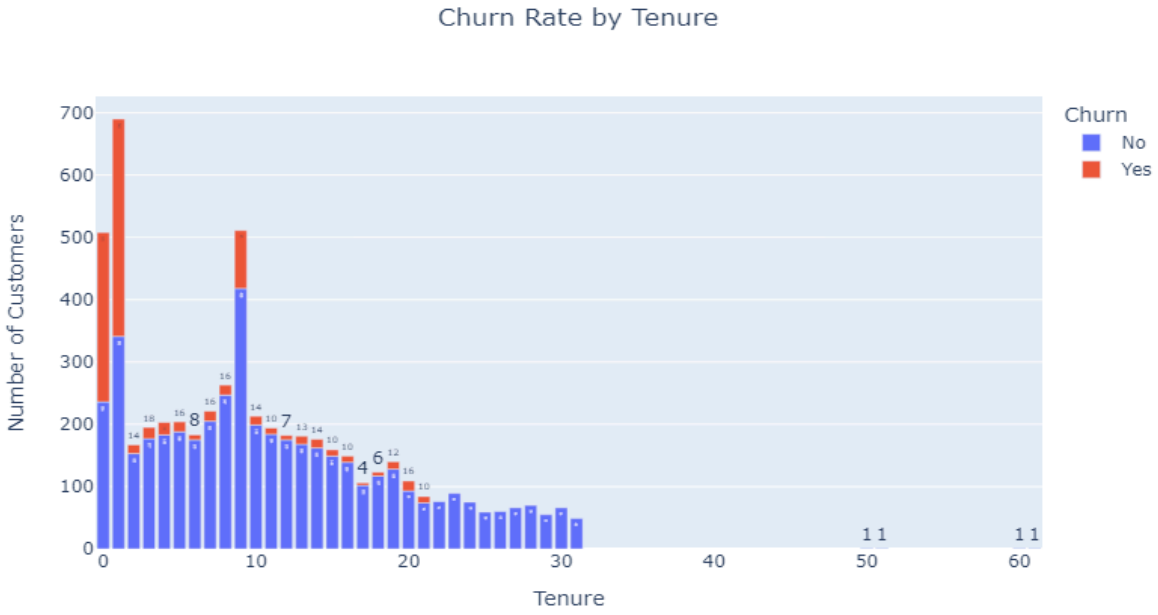
  Efforts to engage and retain customers in their early stages could reduce overall churn.
- Churn is negatively correlated with Tenure (`-0.35`), which aligns with the observation that customers with longer tenure are less likely to churn.
- Cashback Amount is moderately correlated with Tenure (`0.48`), indicating that longer-tenured customers tend to receive more cashback, possibly as part of retention strategies.
- Order Count and Coupon Used are strongly correlated (`0.74`), which suggests that customers who frequently order also use more coupons. This could indicate that coupons are an effective tool for encouraging repeat purchases.

Supporting charts




Churn Rate by PreferedOrderCat
Number of Customers
PreferedOrderCat
Fashion
Grocery
Laptop & Accessory
Mobile Phone
Others
128
698
20
390
210
1840
570
1510
20
244
Churn
No
Yes


Churn Rate by PreferredLoginDevice
Number of Customers
PreferredLoginDevice
Computer
Mobile Phone
324
1310
624
3372
Churn
No
Yes


Churn Rate by Tenure
Number of Customers
Tenure
Churn
No
Yes


WarehouseToHome
HourSpendOnApp
SatisfactionScore
Complain
OrderAmountHikeFromlastYear
CouponUsed
OrderCount
DaySinceLastOrder
CashbackAmount
Tenure
CityTier
WarehouseToHome
HourSpendOnApp
SatisfactionScore
Complain
OrderAmountHikeFromlastYear
CouponUsed
OrderCount
DaySinceLastOrder
CashbackAmount

Model Development and Evaluation

To predict customer churn, multiple machine learning models were developed and evaluated, including deep learning models, random forests, decision trees, and XGBoost. The XGBoost model emerged as the best performer, achieving high accuracy, precision, recall, and F1 scores, making it the ideal choice for deployment.

The chosen model was saved and integrated into a Flask API for real-time predictions. This allows the business to leverage the model in their operational systems to predict churn dynamically and take proactive measures to retain at-risk customers.

	Model	Accuracy	Precision	Recall	F1
3	XGB	0.986119	0.986128	0.986119	0.986119
1	RF	0.982381	0.982429	0.982381	0.982382
0	DNN	0.979178	0.979648	0.979178	0.979178
2	DT	0.951415	0.951544	0.951415	0.951418

Models and scores

Insights and Recommendations

The analysis and model predictions provided several actionable insights:

Early Customer Retention The analysis revealed that customers with shorter tenure were more likely to churn. To combat this, the business should focus on early engagement strategies, such as personalized onboarding experiences or targeted promotions, to enhance customer retention.

Mobile Optimization With a significant portion of the customer base preferring mobile devices, it's crucial to ensure the mobile experience is seamless and intuitive. Any issues with the mobile platform could lead to higher churn rates, so continuous improvement in this area is recommended.

Effective Use of Promotions The strong correlation between order frequency and coupon usage indicates that promotional strategies are effective in driving repeat purchases. The business should consider expanding and optimizing these promotional efforts to increase customer loyalty and sales.

Conclusion

This project successfully met its objective of providing actionable insights into customer behavior on an e-commerce platform. The developed predictive model for customer churn, particularly the XGBoost model, provides a robust tool for identifying at-risk customers and allows the business to implement targeted retention strategies. By focusing on improving

customer engagement, especially among new users, and optimizing the mobile shopping experience, the business can enhance customer satisfaction and drive sales growth. The deployment of the model via a Flask API ensures that these insights can be readily integrated into the business's day-to-day operations.