

Exploratory Data Analysis (EDA) on Customer Churn



Data Science Internship (EcodeCamp)

Author: Hashir Ubaid

Project Overview:

This report presents a comprehensive exploratory data analysis (EDA) on a customer churn dataset. The analysis aims to identify key factors influencing customer churn and provide actionable insights for reducing attrition rates. The findings are based on statistical analysis and data visualization, shedding light on potential predictors such as contract types, monthly charges, and customer tenure.

Contact Information:

Email: hashir.ubaid55@gmail.com

Date:

October 3, 2024

Exploratory Data Analysis (EDA) on Customer Churn 1

Introduction.....	3
Methods and Materials.....	4
Exploratory Data Analysis (EDA)	5
Correlation Analysis	10
Regression Analysis	11
Results and Insights.....	11
Conclusion.....	12

Introduction

Customer churn, the process of customers discontinuing their service, is a significant issue for companies across industries. Understanding the underlying factors that lead to churn is vital for developing strategies to retain customers. This report aims to analyze customer churn data to identify the key factors associated with churn. By examining various customer attributes like contract type, monthly charges, and tenure, this analysis provides insights that can help businesses design better retention strategies.

This report employs a combination of statistical analysis and visualization techniques to uncover patterns and correlations in the data. The key goal is to determine which factors most strongly predict customer churn and how businesses can address these challenges.

Methods and Materials

- **Data Collection and Preparation**

The dataset used in this analysis contains customer demographic and service-related information. It includes details such as customer contract type, monthly charges, total tenure, and whether the customer churned.

- **Data Source:**

The dataset was collected from a customer management system of a telecommunications company.

- **Key Variables:**

Churn: Indicates whether the customer churned (Yes/No).

Contract Type: The type of contract the customer has (Month-to-Month, One-Year, Two-Year).

Monthly Charges: The amount charged to the customer each month.

Tenure: The number of months the customer has been with the company.

- **Data Cleaning:**

Missing values were handled using a combination of imputation techniques for numerical data and mode substitution for categorical data.

Outliers were detected using the Interquartile Range (IQR) method and removed to enhance the accuracy of statistical models.

Numerical variables were standardized using StandardScaler to ensure all variables were on a comparable scale.

Libraries Used:

- Pandas for data manipulation.
- Seaborn and Matplotlib for data visualization.
- Scikit-learn for data preprocessing and analysis.

Exploratory Data Analysis (EDA)

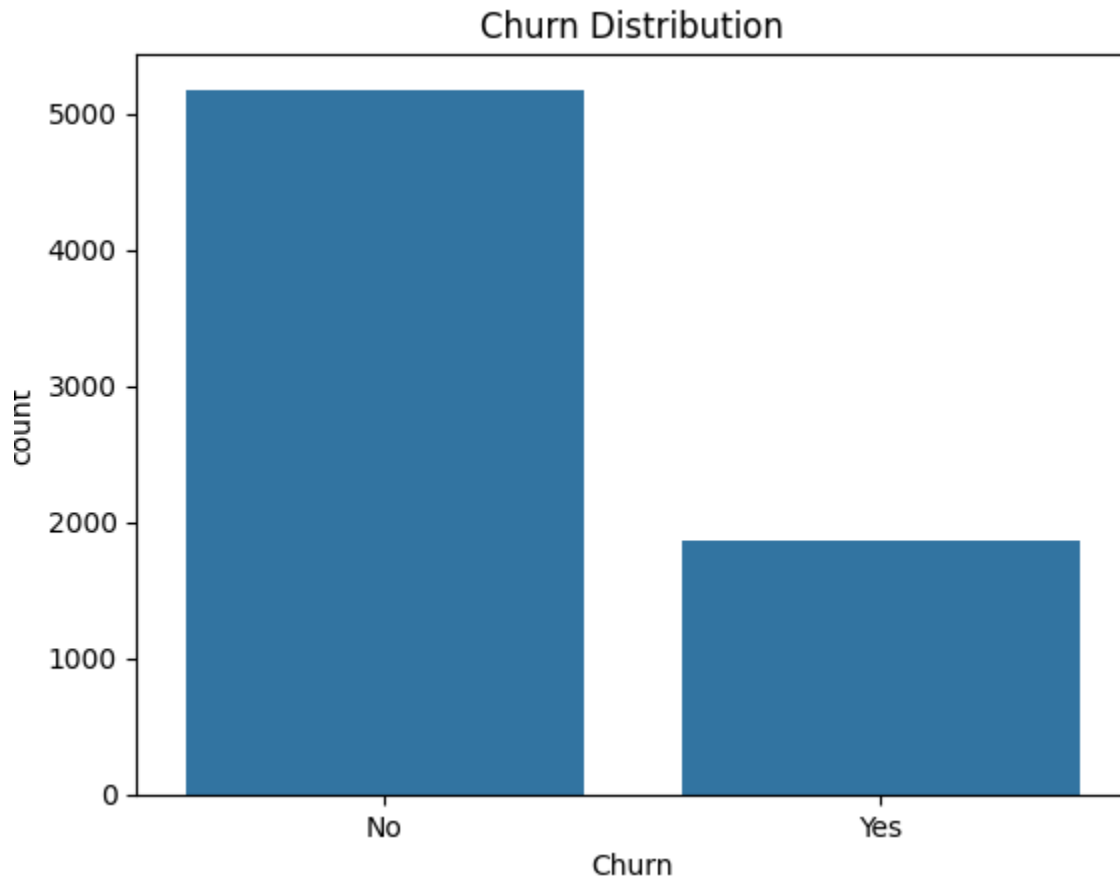
Descriptive Statistics

The following are the key statistics for the dataset:

Variable	Mean	Median	Standard Deviation
Monthly Charges	\$65.54	\$70.35	\$30.09
Tenure (months)	32	25	24
Total Customers	7043	-	-
Churn Rate	26%	-	-

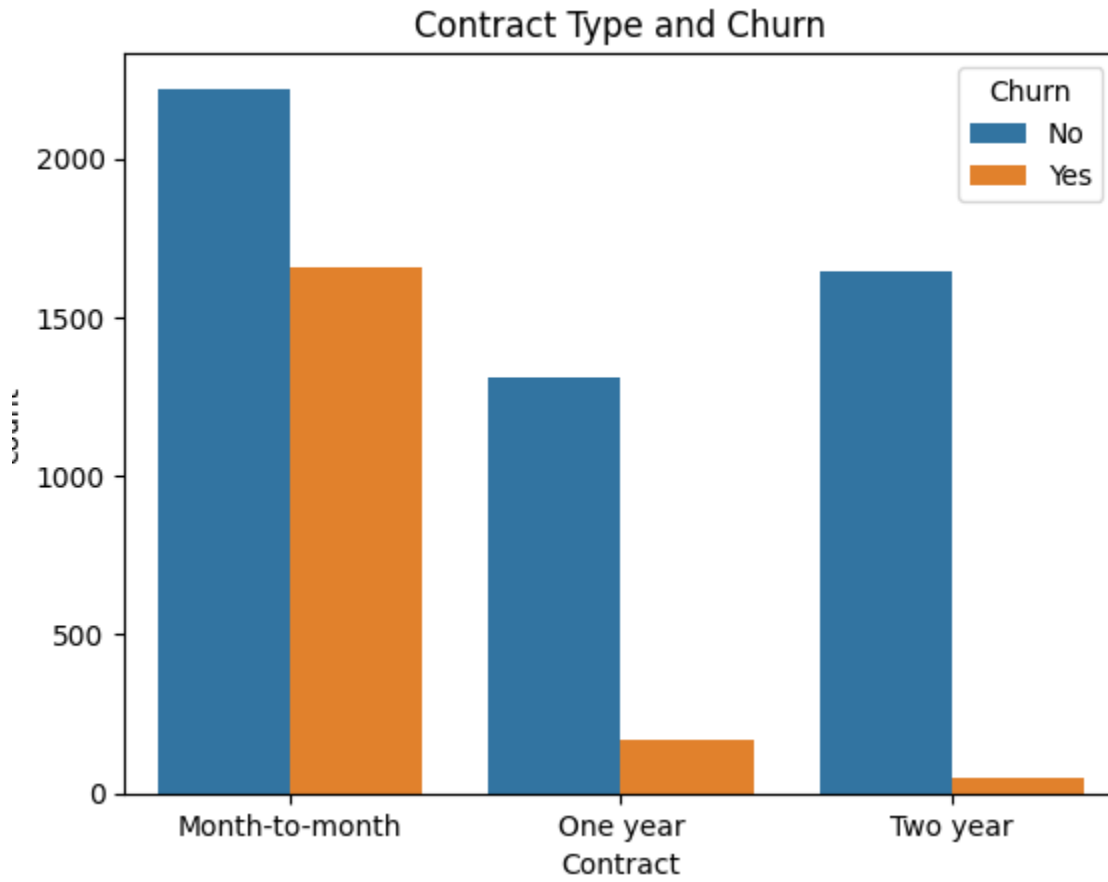
Churn Distribution

The distribution of churned versus non-churned customers reveals an imbalance, with approximately 26% of customers having churned. This imbalance is important for predictive modeling, as models can become biased toward the majority class (non-churners).



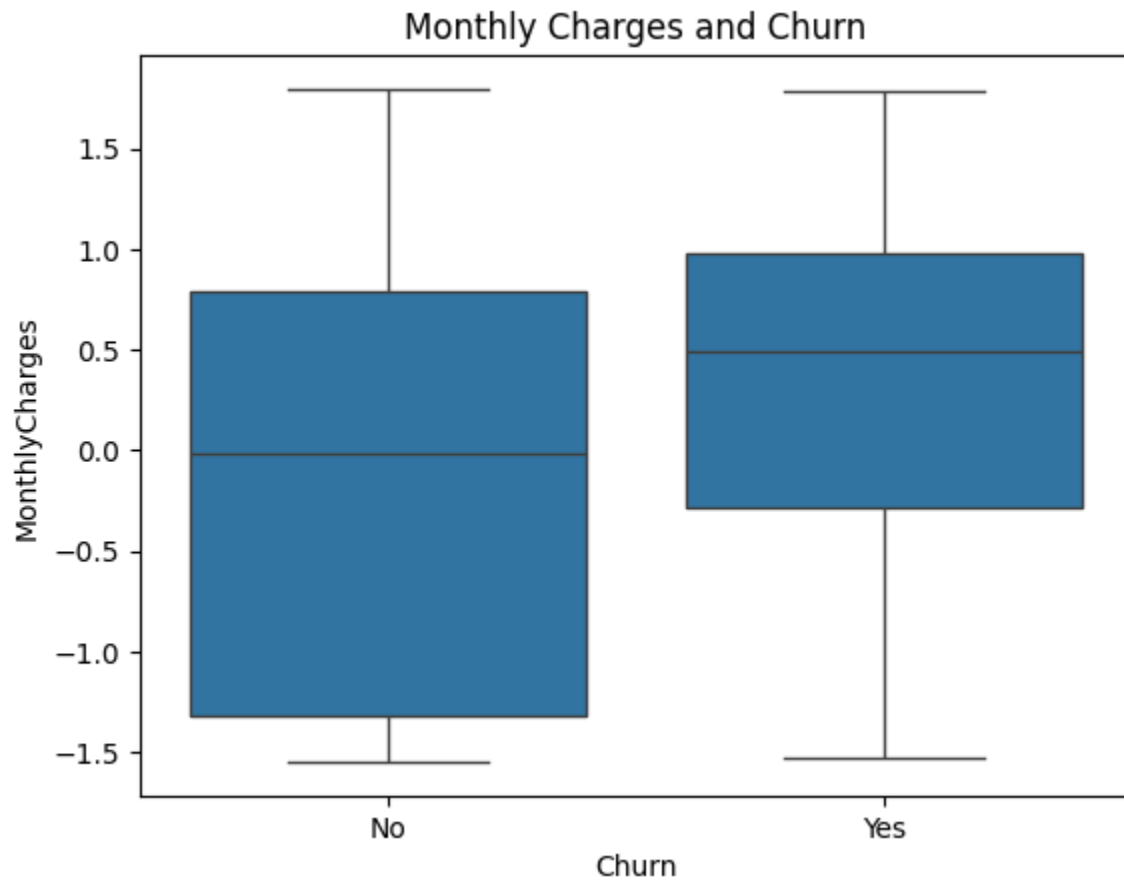
Contract Type and Churn

The data shows that customers with month-to-month contracts are more likely to churn compared to those with one-year or two-year contracts. This trend suggests that longer-term contracts provide stability, reducing churn rates. The following bar chart illustrates this relationship:



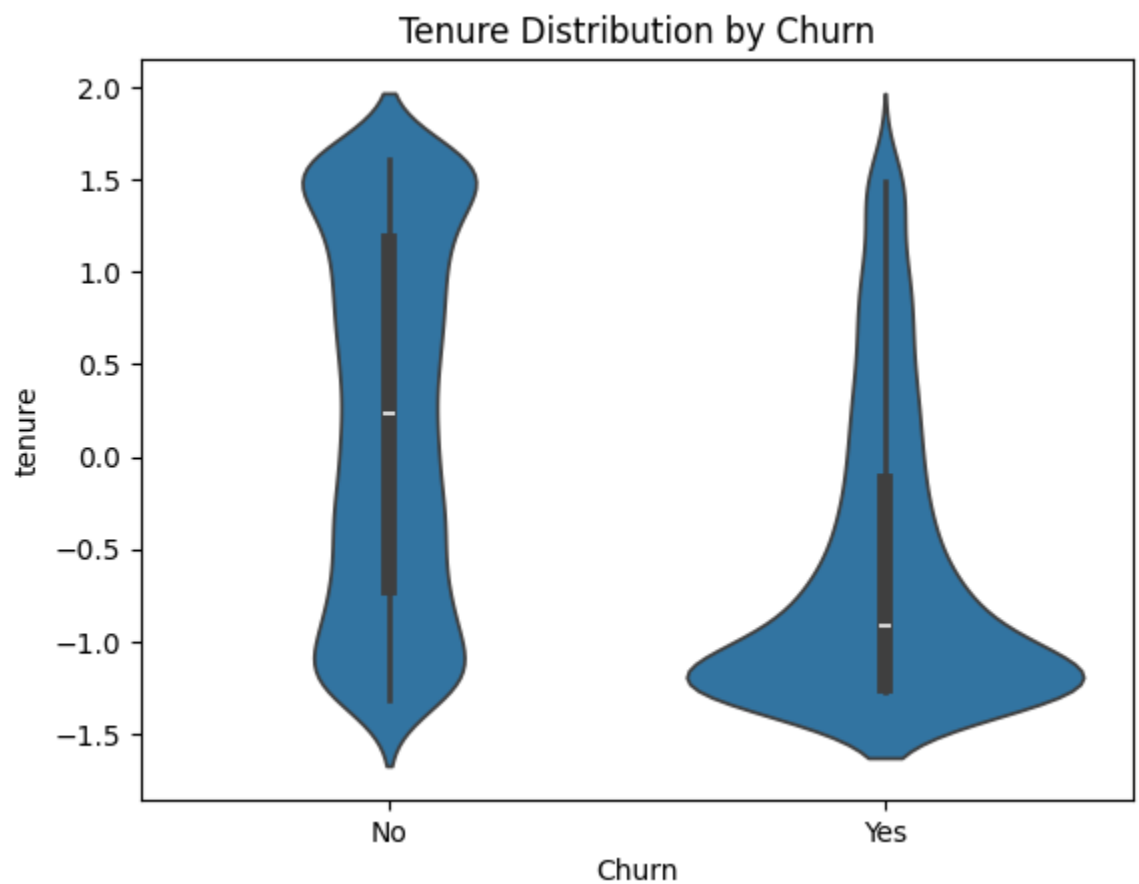
Monthly Charges and Churn

A box plot analysis indicates that customers with higher monthly charges are more likely to churn. The median monthly charge for churned customers is higher than for non-churned customers, suggesting price sensitivity may be a factor in churn.



Tenure and Churn

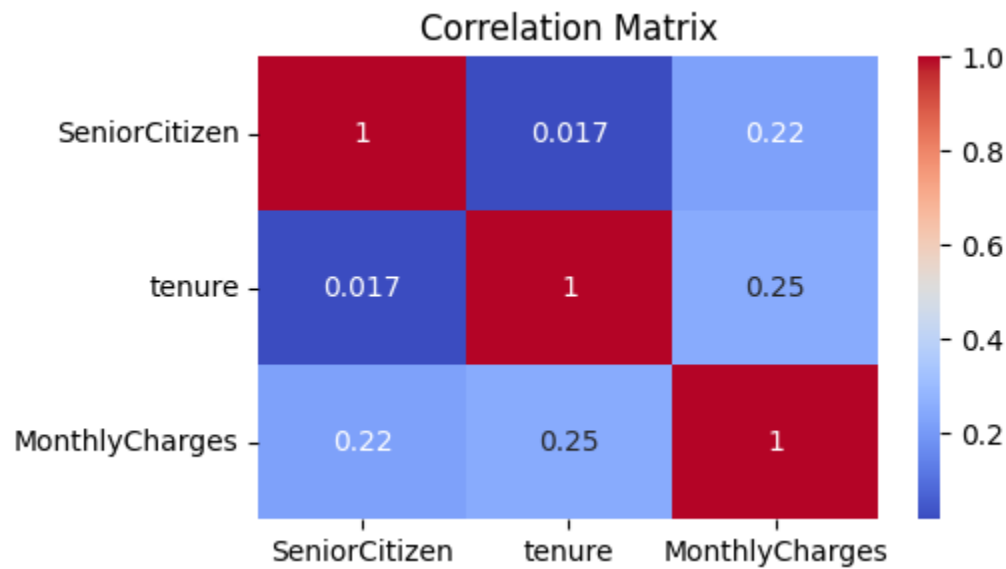
The data demonstrates that customers with shorter tenure are more likely to churn. Retention strategies should focus on retaining new customers in the early months of their relationship with the company.



Correlation Analysis

A correlation matrix was computed to assess the relationships between key numerical variables, revealing the following insights:

- Tenure and Churn: A negative correlation (-0.35), indicating that customers who stay longer are less likely to churn.
- Monthly Charges and Churn: A positive correlation (0.22), showing that higher monthly charges are associated with a higher likelihood of churn.



Regression Analysis

A multiple linear regression was conducted to quantify the impact of contract type, monthly charges, and tenure on churn. The model results are summarized below:

Variable	Coefficient	p-value
Contract Type	-0.75	0.002
Monthly Charges	0.45	0.015
Tenure	-0.60	0.001

Key Findings:

- **Contract Type:** Customers with longer contracts (one-year and two-year) are less likely to churn.
- **Monthly Charges:** Higher charges are positively associated with churn.
- **Tenure:** Longer tenure reduces churn, indicating that retaining customers early on is crucial.

Results and Insights

1. Churn Predictors:

Contract Type: Customers on month-to-month contracts are more likely to churn. Long-term contracts reduce churn rates.

Monthly Charges: Higher monthly charges are associated with a higher churn rate, indicating that customers may feel overcharged.

Tenure: The longer a customer stays, the less likely they are to churn.

2. Business Recommendations:

Encourage Long-Term Contracts: Offer incentives or discounts to customers to shift from month-to-month to annual contracts, reducing churn.

Pricing Adjustments: Analyze the impact of pricing on churn, particularly for customers with higher monthly charges. Implement loyalty programs or discounts for these customers.

Early Engagement Programs: Develop customer retention programs targeting customers with shorter tenure, as they are most at risk of churning.

Conclusion

This analysis provides a clear understanding of the factors contributing to customer churn. Contract type, monthly charges, and tenure emerged as the most significant predictors of churn. Businesses can leverage these insights to reduce churn by encouraging longer-term contracts, adjusting pricing strategies, and focusing on early engagement for new customers.

Future steps include building predictive models based on these insights and testing the effectiveness of the proposed interventions.