Technical Report: Customer Churn Analysis

1. Introduction

Customer churn is a critical business problem for subscription-based organizations. When customers discontinue their services, businesses face direct revenue loss and increased costs associated with acquiring new customers. Understanding why customers churn is essential for improving retention strategies and ensuring sustainable growth.

This technical report presents an end-to-end data analysis of a customer churn dataset. The analysis applies data cleaning, exploratory data analysis (EDA), and statistical techniques to identify key drivers of churn and translate them into actionable business insights.

2. Business Problem Statement

The business is experiencing customer attrition, particularly among certain customer segments. The goal of this analysis is to answer the following questions:

- What proportion of customers are churning?
- Which customer attributes are most strongly associated with churn?
- Are the observed differences statistically significant?
- What actions can the business take to reduce churn?

3. Dataset Description

The dataset used for this project is customer_churn.csv, which contains customer-level information related to demographics, billing, contract details, and churn status.

Key Variables:

- **Customer ID**: Unique identifier for each customer
- **Tenure**: Number of months the customer has stayed with the company
- **Monthly Charges**: Monthly service charges
- **Total Charges**: Total charges paid by the customer
- **Contract**: Contract type (Month-to-month, One year, Two year)
- **Payment Method**: Payment method used by the customer
- **Senior Citizen**: Indicator of senior citizen status
- **Churn**: Target variable indicating whether the customer churned (Yes/No)

The dataset contains 500 customer records and was cleaned before analysis.

4. Data Cleaning and Preparation

To ensure reliable analysis, several data preprocessing steps were performed:

- Standardized column names by converting them to lowercase and replacing spaces with underscores.
- Converted numerical columns stored as text (e.g., Total Charges) into numeric format.
- Handled missing values using:
    - Median imputation for numerical variables
    - Mode imputation for categorical variables
- Removed duplicate records.
- Validated data types and ensured consistency across columns.

The cleaned dataset was saved as cleaned_customer_churn.csv and used for all subsequent analyses.

5. Exploratory Data Analysis (EDA)

EDA was conducted to understand customer behavior and identify potential churn drivers.

5.1 Churn Distribution

- The churn rate was calculated to determine the scale of the problem. Approximately 10% of customers in the dataset have churned, indicating a moderate but significant retention issue.

5.2 Monthly Charges vs Churn

- Box plot analysis shows that customers who churn tend to have higher monthly charges compared to those who remain. This suggests that pricing may play a role in customer dissatisfaction.

5.3 Tenure vs Churn

- Customers with shorter tenure are significantly more likely to churn. Long-term customers exhibit stronger loyalty and lower churn rates.

5.4 Contract Type vs Churn

- Customers on month-to-month contracts display the highest churn rates, while one-year and two-year contracts are associated with much lower churn.

5.5 Correlation Analysis

- Correlation analysis among numerical variables indicates:
- A negative relationship between tenure and churn

- A positive relationship between monthly charges and churn
- These exploratory findings guided the statistical analysis phase.

6. Statistical Analysis

Statistical tests were applied to validate whether observed differences are significant.

6.1 Hypothesis Testing: Monthly Charges

- An independent t-test was conducted to compare monthly charges between churned and non-churned customers.
- **Null Hypothesis (H_0)**: No difference in monthly charges
- **Alternative Hypothesis (H_1)**: Churned customers pay higher monthly charges
- The test results show a statistically significant difference ($p < 0.05$), confirming that higher monthly charges are associated with churn.

6.2 Hypothesis Testing: Tenure

- A t-test comparing tenure between churned and non-churned customers also showed a statistically significant difference. Churned customers have significantly shorter tenure.

6.3 Chi-Square Test: Contract Type

- A chi-square test was performed to analyze the relationship between contract type and churn.
- The results indicate a strong and statistically significant association between contract type and churn ($p < 0.05$).

7. Customer Segmentation

Based on the analysis, customers were segmented into risk groups:

- **High Risk**: Customers with tenure less than 12 months and high monthly charges
- **Low Risk**: Long-tenure customers with moderate or low charges

This segmentation helps identify customers who require immediate retention interventions.

8. Key Findings

The main findings of the analysis are:

- High monthly charges significantly increase churn risk
- New customers are the most vulnerable to churn
- Month-to-month contracts are a major churn driver
- Long-term contracts improve customer retention

All major findings were supported by statistical evidence.

9. Business Recommendations

Based on the analysis, the following recommendations are proposed:

- Encourage long-term contracts through discounts and incentives
- Provide special retention offers for new customers
- Review pricing strategies for customers with high monthly charges
- Improve onboarding and customer engagement during the first year

10. Limitations

- The dataset size is limited to 500 records
- No time-series analysis was performed
- Customer satisfaction metrics were not available

Future analysis with larger and more detailed datasets could provide deeper insights.

11. Conclusion

This technical report demonstrates how data cleaning, exploratory analysis, and statistical testing can be combined to understand customer churn. The findings highlight clear drivers of churn and provide actionable recommendations that can help reduce customer attrition and improve long-term business performance.