

CUSTOMER CHURN ANALYSIS

Introduction

Customer churn analysis is a critical business analytics process used to understand why customers discontinue a product or service. In highly competitive industries such as telecommunications, banking, subscription-based services, and e-commerce, retaining existing customers is significantly more cost-effective than acquiring new ones. Even a small increase in retention rates can result in substantial improvements in profitability. This project focuses on analyzing customer churn behavior using structured data analysis techniques to uncover patterns and risk factors associated with customer attrition. The analysis leverages both SQL and Python to explore customer demographics, service usage, contract types, billing patterns, and tenure. SQL is used for efficient data querying, aggregation, and churn rate calculation, while Python is employed for exploratory data analysis and visualization. By combining these tools, the project ensures both accuracy and interpretability of results. The ultimate goal is to provide actionable insights that organizations can use to proactively reduce churn and improve customer satisfaction.

Objective

The primary objective of this project is to analyze customer churn behavior and identify key factors that influence customers leaving the service. Specifically, the project aims to calculate overall churn rates, analyze churn across contract types, tenure groups, and billing patterns, and visualize churn trends for better decision-making. The insights derived from this analysis support business strategies such as targeted retention campaigns, pricing optimization, and contract restructuring to reduce churn risk and enhance long-term customer value.

Dataset Description

The dataset used in this project contains 705 customer records, with each record representing an individual customer. It includes demographic information such as gender and senior citizen status, service-related attributes such as internet service and contract type, billing details such as monthly and total charges, and customer tenure measured in months. The target variable, Churn, indicates whether a customer has discontinued the service. This dataset provides a comprehensive foundation for analyzing customer behavior and identifying churn patterns across multiple dimensions.

Tools & Technologies Used

MySQL was used for data storage and SQL-based analysis, enabling efficient aggregation and filtering of large datasets. Python was used for data analysis and visualization, with Pandas for data manipulation and Matplotlib for chart generation. Jupyter Notebook served as the development environment for Python-based exploratory data analysis.

Methodology

The project methodology followed a structured analytical approach. Initially, the CSV dataset was imported into a MySQL database and validated for schema consistency, data types, and missing values. SQL queries were then designed to compute churn metrics such as churn count, churn percentage, and churn rates across different customer segments. After completing SQL analysis, the same dataset was analyzed using Python to generate visualizations that highlight trends and patterns not easily observable in tabular data. This dual approach ensures both analytical rigor and visual clarity.

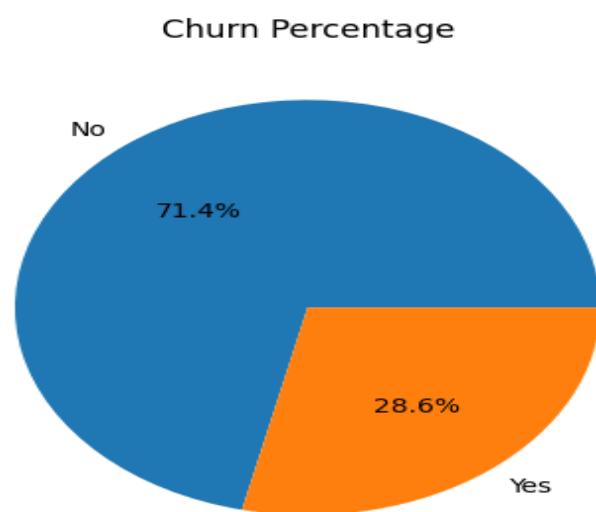
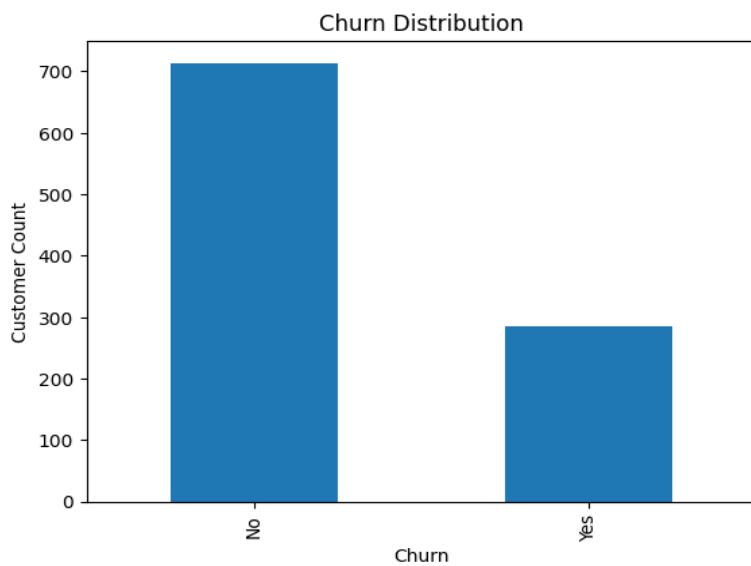
SQL Analysis

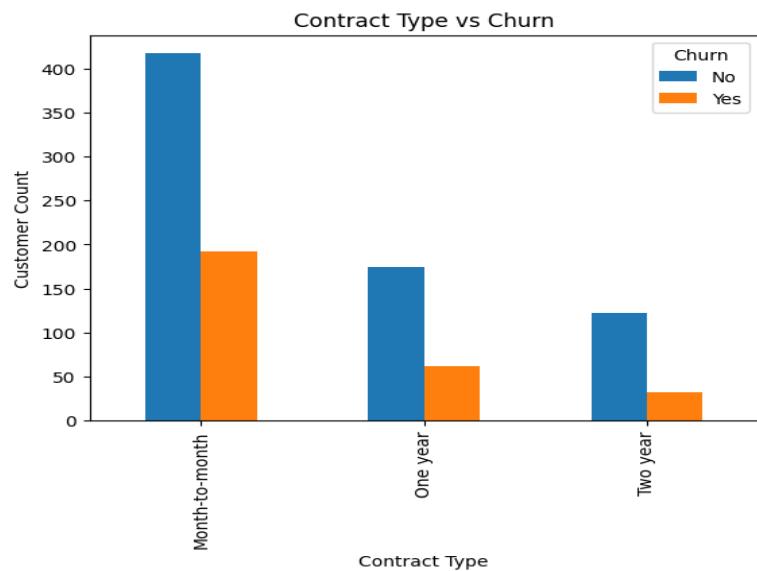
The following SQL queries were used to analyze churn patterns:

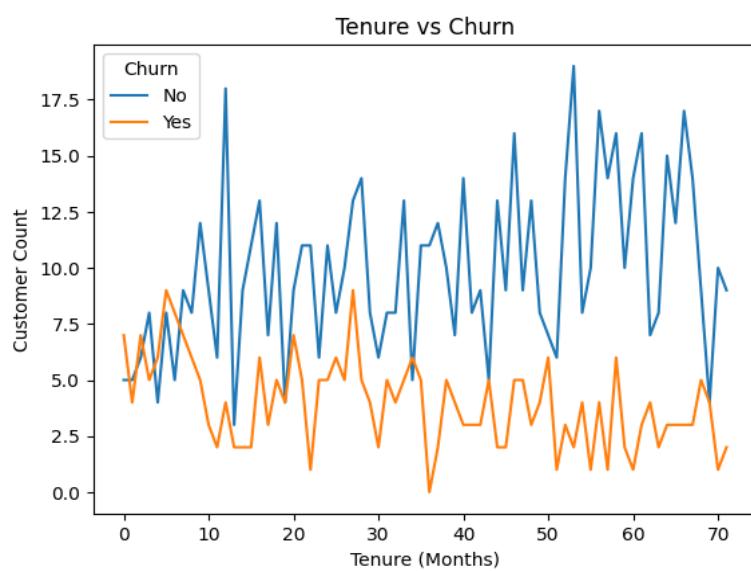
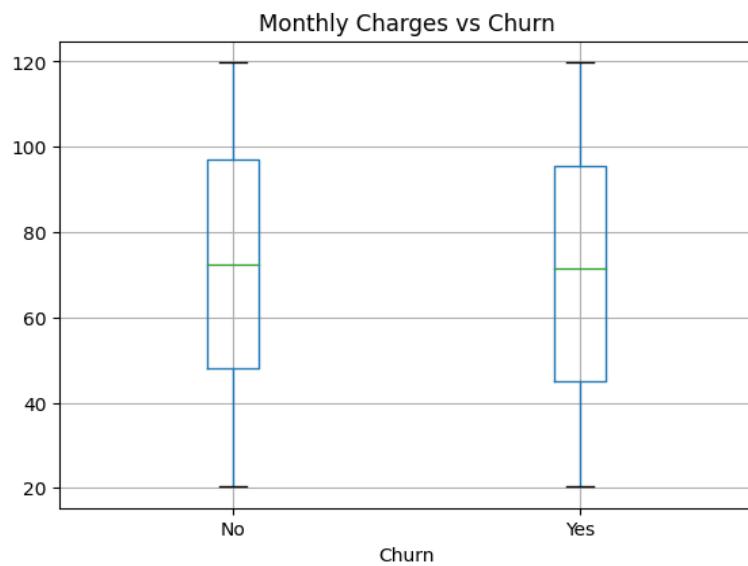
```
SELECT Churn, COUNT(*) FROM customer_churn GROUP BY Churn; SELECT Churn, ROUND(COUNT(*)*100.0/(SELECT COUNT(*) FROM customer_churn),2) FROM customer_churn GROUP BY Churn; SELECT Contract, COUNT(*) FROM customer_churn WHERE Churn='Yes' GROUP BY Contract; SELECT Contract, ROUND(SUM(CASE WHEN Churn='Yes' THEN 1 ELSE 0 END)*100.0/COUNT(*),2) FROM customer_churn GROUP BY Contract; SELECT Churn, ROUND(AVG(MonthlyCharges),2) FROM customer_churn GROUP BY Churn; Tenure-wise churn analysis using CASE statements.
```

Python Analysis & Visualization

Python was used to perform exploratory data analysis and generate visualizations that clearly communicate churn patterns. Bar charts were used to display churn distribution, pie charts to represent churn percentages, box plots to compare monthly charges between churned and retained customers, and line charts to illustrate churn behavior across tenure. These visualizations make complex patterns easily understandable for stakeholders.







Key Findings & Insights

Month-to-month contract customers exhibit the highest churn rate. Customers with shorter tenure are more likely to churn, while higher monthly charges increase churn probability. Contract type has a stronger influence on churn than gender or demographic factors.

Business Impact

The insights from this analysis enable organizations to identify high-risk customers early, design targeted retention strategies, optimize pricing models, and encourage long-term contracts. These actions directly contribute to improved customer lifetime value and revenue stability.

Conclusion

This project demonstrates how SQL and Python can be effectively combined to analyze real-world business problems. The churn analysis provides clear, actionable insights that support strategic decision-making and customer retention initiatives.

Future Enhancements

Future work may include building a predictive churn model using machine learning, developing an interactive Power BI dashboard, performing advanced customer segmentation, and integrating additional behavioral data for deeper analysis.