
CUSTOMER ACQUISITION
NEW CUSTOMERS
CHURN

OVERVIEW

This project aims to predict customer churn (i.e., whether a customer will leave the service) based on a variety of features related to their usage of the telecom services. The dataset contains customer-related data, such as their demographic information, service plans, and usage statistics. The goal is to build a predictive model to identify customers who are likely to churn so that the company can take preventive actions.

TABLE OF CONTENTS

OVERVIEW OF PRESENTATION STRUCTURE

BUSINESS UNDERSTANDING

- In the telecommunications industry, **customer churn** is a critical metric that directly impacts profitability. Churn refers to the loss of customers who decide to cancel or stop using the company's services. For telecom companies, understanding the factors that contribute to customer churn and being able to predict which customers are at risk of leaving can provide a significant competitive advantage. This project focuses on building a predictive model that helps the company identify customers likely to churn, enabling them to take proactive steps to retain these customers.

BUSINESS QUESTION

- What are the key factors contributing to customer churn?
- Which customer segments are most at risk of churn?
- How can customer retention be improved?
- What impact does churn have on the company’s revenue and growth?
- What is the effectiveness of current customer retention strategies?
- Which states or regions exhibit the highest churn rates, and why?
- Can predictive models help anticipate churn and reduce it proactively?

DATA UNDERSTANDING

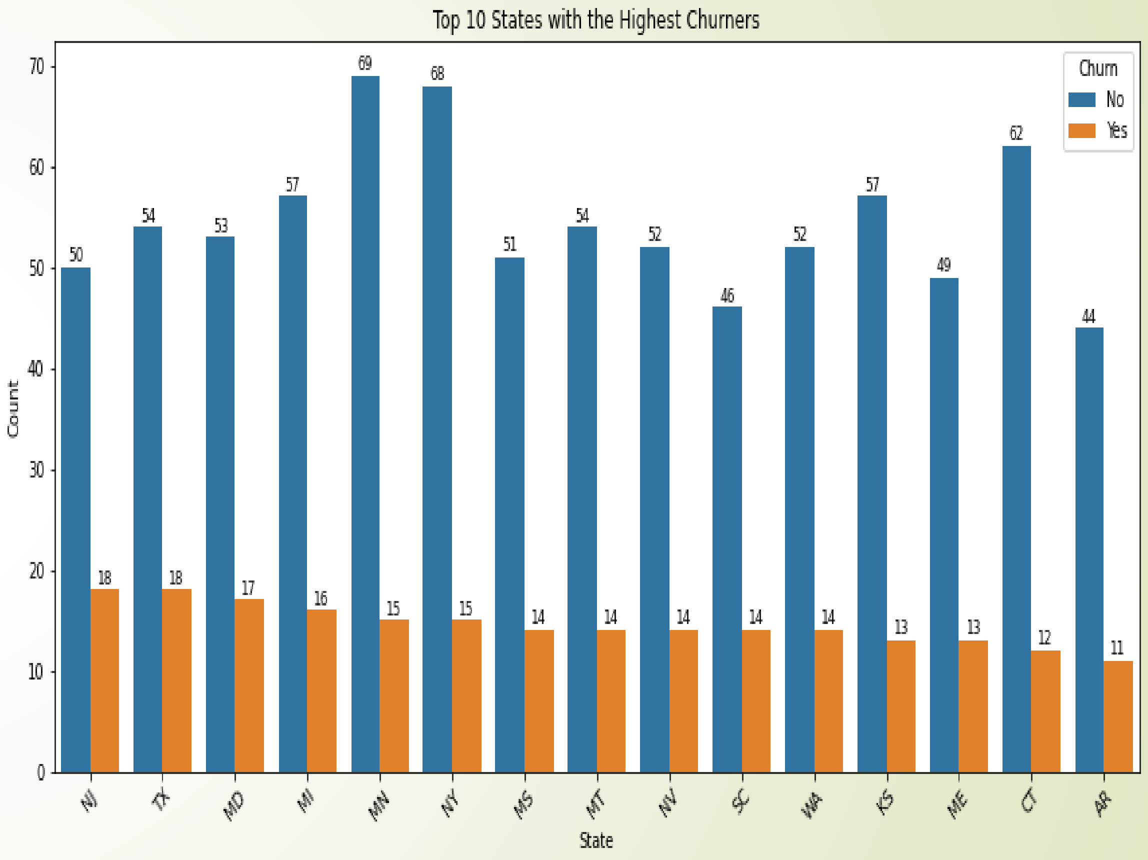
- The **Customer Churn Dataset** contains demographic and service usage data, with key features such as customer ID, gender, tenure, contract type, monthly charges, and service attributes like internet service, online security, and tech support. The target variable is **Churn**, indicating whether a customer has left the service. The dataset includes a mix of categorical, numerical, and binary features, with some missing values, especially in **TotalCharges**. It likely contains **class imbalance**, with more non-churned customers, and outliers in numerical features such as **MonthlyCharges**. Key insights reveal that shorter tenure, month-to-month contracts, and higher charges are associated with higher churn rates. Understanding these relationships is crucial for developing predictive models for customer retention.

DATA CLEANING

- The **data cleaning** process for the Customer Churn Dataset involved several key steps to ensure the data was ready for analysis and modeling. Initially, missing values were handled, particularly in the **TotalCharges** column, where empty cells were replaced with the median value or dropped if necessary. Categorical variables like **Contract**, **PaymentMethod**, and **InternetService** were encoded using one-hot encoding or label encoding to convert them into numerical format. Outliers in numerical columns, such as **MonthlyCharges** and **Tenure**, were assessed and handled appropriately, either through capping or removal. Duplicate records were also checked and removed to ensure the integrity of the data. Finally, feature scaling or normalization was applied where needed to ensure consistent ranges for numerical variables. The cleaned dataset was now suitable for exploratory data analysis (EDA) and building machine learning models.

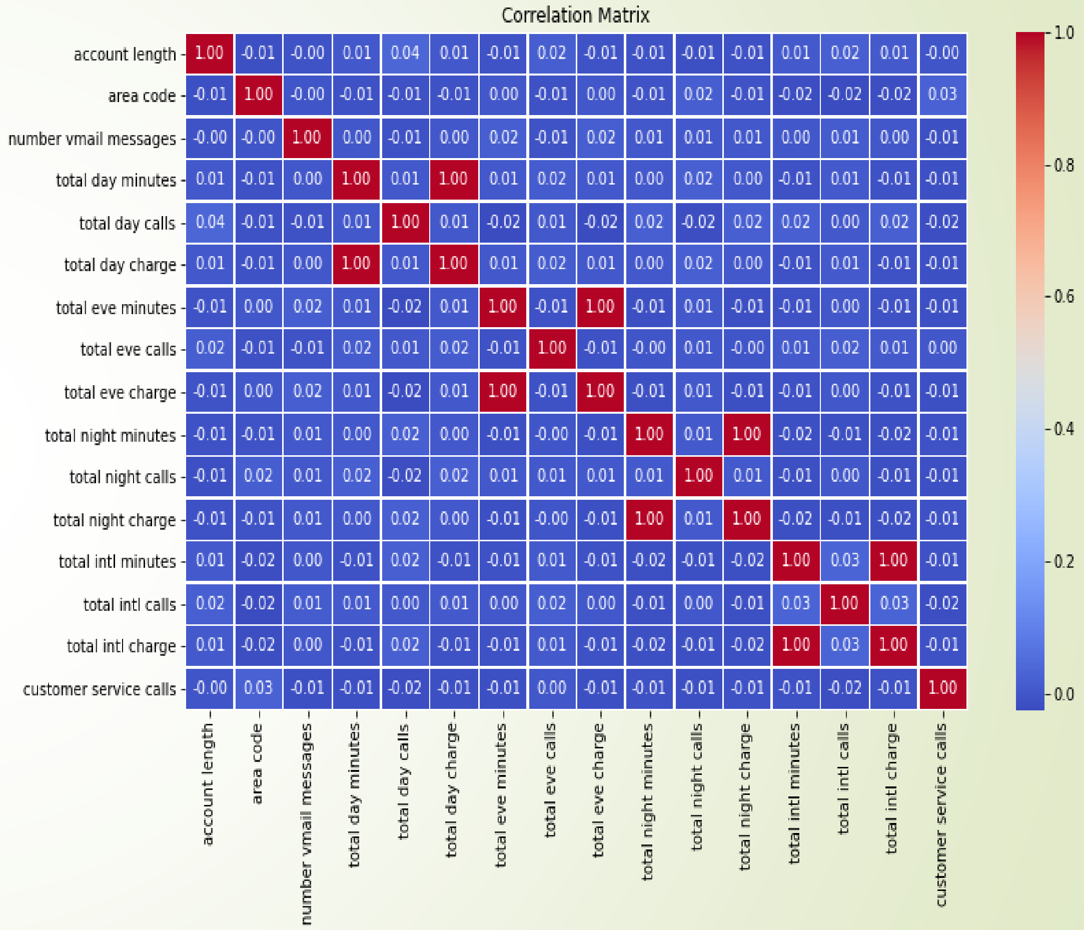
DATA ANALYSIS

- The **data analysis** focused on visualizing churn patterns across different U.S. states. By converting the 'churn' column to a string for compatibility with Seaborn, the analysis filtered the top 15 states with the highest number of churners. A **count plot** was created to compare the number of churned versus non-churned customers across these states, revealing geographic patterns in customer retention. The chart showed the distribution of churners and non-churners within each state, with clear distinctions highlighted by the color-coded bars. The added labels on each bar provided precise counts, enhancing interpretability. This analysis helped in identifying states where churn is more prevalent, guiding business strategies to address high churn regions.



DATA ANALYSIS

- The **correlation matrix analysis** aimed to explore the relationships between numerical variables in the dataset. By calculating the pairwise correlations, a heatmap was generated to visualize the strength and direction of these relationships. The heatmap, created using Seaborn, displayed correlation coefficients ranging from -1 to 1, with color coding (from blue to red) to indicate negative to positive correlations. The annotations on the heatmap provided exact correlation values for better clarity. This analysis helps identify highly correlated variables, guiding feature selection for predictive modeling and offering insights into which factors may influence customer churn.



RECOMMENDATION

- The business should implement a targeted retention strategy by focusing on key factors that contribute to churn, such as customer tenure, service type, and geographical location.
- Segmentation of high-risk customer groups, particularly in states with the highest churn rates, will allow for personalized interventions, such as tailored offers, loyalty programs, or customer service improvements.
- Additionally, leveraging predictive models can help identify at-risk customers early, enabling proactive retention efforts. The business should also consider adjusting pricing structures and contract terms to improve customer satisfaction and reduce churn. Monitoring and refining these strategies continuously will ensure sustainable customer loyalty and long-term business growth.

CONCLUSION

- In conclusion, the customer churn analysis project has provided valuable insights into the key factors influencing customer attrition. By identifying patterns in customer demographics, service types, and geographical locations, we gained a deeper understanding of the customer segments most at risk of leaving.
- The data-driven approach highlights the importance of focusing on high-churn states and offering targeted retention strategies to reduce churn rates. Furthermore, the use of correlation analysis and predictive modeling can assist in early identification of potential churners, allowing for more effective and timely interventions. By leveraging these findings, businesses can enhance customer retention efforts, improve satisfaction, and ultimately foster long-term growth and profitability.

THANK YOU!

- Should you have any queries or require further clarification , kindly do not hesitate to contact.

+254743080801

cynthiacheruto14@gmail.com

https://www.linkedin.com/in/cynthia-cheruto/