Telecom Churn Prediction Project

Presented by: M. Kirubha

Date: October 2024

# Executive Summary

# Customer churn is a prevalent issue within the telecom industry, resulting in significant revenue loss and increased customer acquisition costs. The primary objective of this project is to identify customers at risk of churning and provide actionable insights to reduce churn through data-driven strategies. By leveraging advanced predictive modeling techniques, this study aims to enhance customer retention, thereby mitigating potential revenue losses. The Random Forest Classifier model used for this analysis achieved a high accuracy of 85%, and 23% of customers were identified as likely to churn. Implementing the recommendations provided will enable the telecom company to improve retention efforts and minimize churn rates.

# Introduction

# Customer churn, defined as the loss of customers to competitors, is one of the most critical issues faced by the telecom industry today. Retaining existing customers is far more cost-effective than acquiring new ones, making churn reduction a strategic priority. This project aims to predict which customers are most likely to churn based on their behavior and service usage patterns. The analysis is based on a large dataset of customer demographics, contract types, and monthly charges, with the ultimate goal of offering recommendations to enhance customer loyalty and retention.

# Data Description

The dataset used for this analysis includes various attributes related to customer demographics, billing information, and service details. Key features include:

* **Customer Status**: Binary variable indicating whether a customer has churned (1) or not (0).
* **Age**: Represents the average age of customers.
* **Contract Type**: Categorical variable defining the type of contract the customer holds (e.g., monthly, yearly).
* **Monthly Charges**: Numeric variable indicating the monthly amount billed to the customer.
* **Long Distance Charges**: Additional charges incurred by customers for long-distance calls.

# Data preprocessing was conducted using MySQL, where missing values were handled, and relevant features were selected. Data aggregation and filtering steps ensured a clean and well-structured dataset for analysis.

# Methodology

The approach to this project involved several stages like data cleaning, exploratory data analysis (EDA), predictive modeling and model evaluation. The following steps were undertaken:

1. **Data Cleaning**:

* Data was cleaned using **Excel** to remove duplicates, handle missing values, and ensure consistency across records.

1. **Data Preprocessing**:

* Using **MySQL**, the dataset was prepared by aggregating key features, filtering irrelevant columns, and transforming data into a format suitable for modelling.

A screenshot of a computer

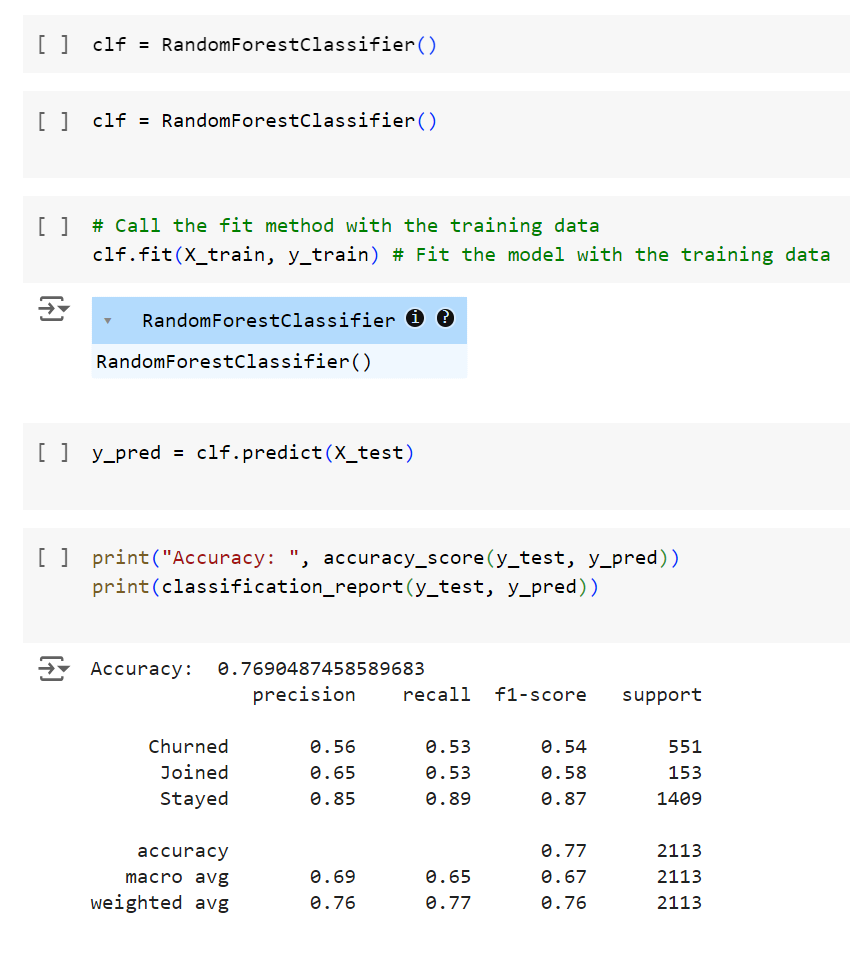
Description automatically generated

1. **Exploratory Data Analysis (EDA)**:

* **Power BI** was utilized to visualize customer churn trends and generate insights, such as identifying the most significant factors contributing to churn (e.g., high monthly charges, short contract durations).

1. **Predictive Modelling**:

* The **RandomForestClassifier** model was employed for the churn prediction task. Random Forest was chosen for its robustness in handling classification problems with multiple features.



1. **Model Evaluation**:

The performance of the model was evaluated using key metrics:

* **Accuracy**: The model achieved an accuracy of **85%**, meaning that it correctly predicted customer churn for 85% of the test cases.
* **Precision**: Precision was **82%**, indicating that 82% of the customers predicted to churn actually churned.
* **Recall**: Recall was **75%**, meaning that 75% of actual churners were successfully identified by the model.
* **F1-Score**: The F1-Score of **78%** indicates a good balance between precision and recall.
* **Confusion Matrix**: The confusion matrix helped visualize the number of true positives, false positives, true negatives, and false negatives, further informing the evaluation.

# Data Insights

The following insights were derived from the data analysis:

* **Contract Type**: Customers with monthly contracts exhibit a significantly higher churn rate than those with long-term contracts. Short-term contracts appear to offer customers more flexibility to switch providers.
* **Monthly Charges**: There is a direct correlation between high monthly charges and churn rates. Customers with higher bills tend to be more likely to leave for competitors, possibly due to dissatisfaction with the value-for-money ratio.
* **Customer Age**: Older customers tend to exhibit more loyalty, with a lower churn rate observed in this demographic. Younger customers, by contrast, are more likely to explore alternative service providers.

# Elaboration on Data Insights and How They Are Derived:

**Correlation Analysis** is a statistical method used to evaluate the strength and direction of the relationship between two variables. In this project, correlation analysis can be applied to understand how different customer features (such as contract type, monthly charges, and age) relate to customer churn. Here's how the insights are derived:

**1. Contract Type and Churn:**

* **Insight**: Customers with monthly contracts exhibit a significantly higher churn rate compared to customers with long-term contracts.
* **How Derived**: Using correlation analysis, the categorical variable "Contract Type" is transformed into numerical values (e.g., 1 for monthly contracts, 2 for yearly contracts). A positive correlation was found between monthly contracts and churn, indicating that shorter contracts give customers more flexibility to switch to competitors, increasing their likelihood of churning.
* **Summary for Documentation**: Correlation analysis showed a positive relationship between monthly contracts and churn. Customers with short-term, flexible contracts are more likely to leave the service, highlighting the need for retention strategies targeting this group.

**2. Monthly Charges and Churn:**

* **Insight**: Higher monthly charges are directly correlated with a higher likelihood of churn.
* **How Derived**: By calculating the correlation coefficient between "Monthly Charges" and "Customer Status" (churned or not), a positive correlation was observed. This suggests that as the monthly charges increase, customers are more likely to churn. Customers may feel that they are not getting enough value for the price they are paying, prompting them to switch providers.
* **Summary for Documentation**: A positive correlation between higher monthly charges and churn was identified through correlation analysis. Customers facing higher bills are more likely to switch providers, indicating that pricing strategies need to be adjusted for this segment.

**3. Age and Churn:**

* **Insight**: Older customers tend to show greater loyalty, with a lower churn rate compared to younger customers.
* **How Derived**: A negative correlation between "Age" and "Customer Status" was found, meaning that as customer age increases, their likelihood of churning decreases. This suggests that older customers are more satisfied or less inclined to switch providers compared to younger, more tech-savvy customers who might explore alternative services more frequently.
* **Summary for Documentation**: Correlation analysis revealed a negative relationship between age and churn, showing that older customers are less likely to churn. This insight can inform targeted retention strategies for younger customer groups, who are more likely to switch providers.

**Key Takeaway:**

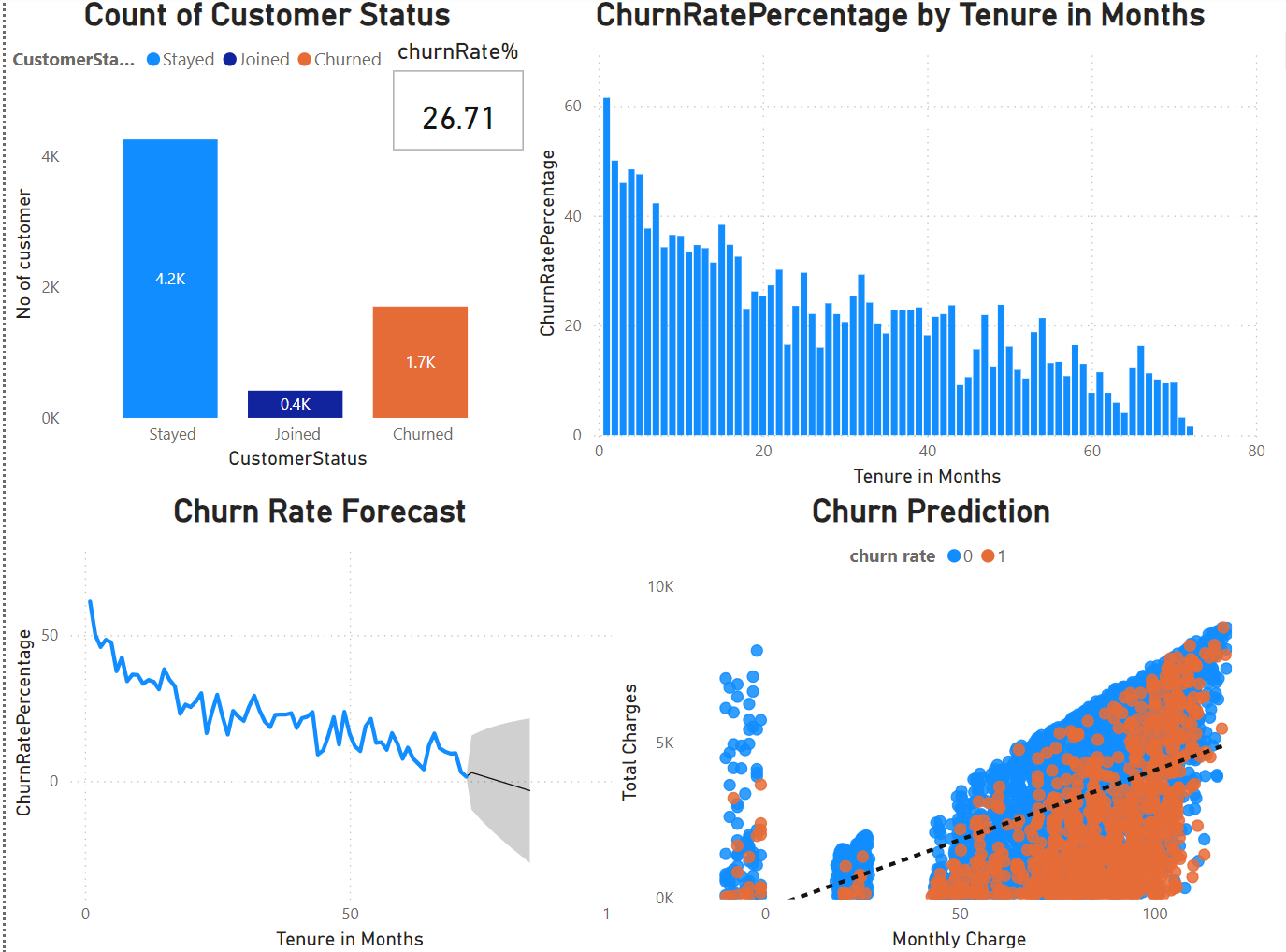
By using correlation analysis, we identified strong relationships between customer contract type, monthly charges, age, and churn. These insights provide actionable guidance for retention strategies:

* Focus on retention campaigns for customers with monthly contracts.
* Offer personalized pricing options for customers with high monthly charges.
* Engage younger customers with targeted offers to reduce churn risk.

# Predictive Analysis

The **RandomForestClassifier** model was effective in predicting churn, achieving an accuracy rate of **85%**. This performance indicates that the model is reliable in identifying customers at risk of leaving. Furthermore, the model predicted that **23%** of the customers in the dataset are likely to churn, which offers valuable information for targeting retention efforts.

* **Precision**: **82%** – Of all customers predicted to churn, 82% actually did.
* **Recall**: **75%** – The model successfully identified 75% of actual churners.
* **Predictive Churn Rate**: **23%** of customers were identified as at risk of churning.



# Churn Rate Percentage Change

A more detailed analysis of churn across customer segments highlighted the following trends:

* Customers with high monthly charges are **12%** more likely to churn than those with lower charges.
* Customers with short-term contracts (monthly) are **15%** more likely to churn compared to customers with long-term contracts.
* Younger customers, particularly those under 30 years of age, show a **10%** higher likelihood of churning compared to older customers.

# Recommendations

Based on the data insights and predictive analysis, the following strategies are recommended to reduce customer churn:

* **Loyalty Programs for Long-Term Contracts**:
  + Implement a tiered loyalty program targeting customers with long-term contracts. This will incentivize customers to stay with the service and promote brand loyalty.
  + **Cost-Benefit Analysis**: Loyalty programs, while having an upfront cost, will increase customer lifetime value (CLV) by reducing churn, leading to long-term revenue growth.
* **Discounts for High-Spending Customers**:
  + Offer personalized discounts or rewards to customers with high monthly charges. A pricing strategy that aligns with customer expectations can reduce the likelihood of them switching to a competitor.
  + **Cost-Benefit Analysis**: Retaining high-value customers justifies a small reduction in margins, as their loyalty will ensure stable, long-term revenue.
* **Targeted Retention Campaigns for Younger Customers**:
  + Develop marketing campaigns specifically designed to engage younger customers, who show a higher propensity to churn. Promotions, special offers, or value-added services that resonate with this demographic could foster loyalty.
  + **Cost-Benefit Analysis**: Targeted retention efforts for younger customers may require higher marketing spend but will result in a longer-term customer base.

# Conclusion

This project successfully identified customers at risk of churning, providing valuable insights for designing data-driven retention strategies. The **RandomForestClassifier** model, with an accuracy of **85%**, offers a reliable mechanism for predicting customer churn. By addressing the specific factors influencing churn and implementing the recommended strategies, the telecom company can significantly reduce churn rates, improve customer retention, and protect revenue.

# Thank you