CUSTOMER CHURN ANALYSIS FOR SYRIATEL - COMPANY

AUTHOR: BEATRICE NDANU KIILU
DATE: 23RD JULY 2025

INTRODUCTION

- The telecommunications industry is quite competitive, and customer retention has become increasingly crucial.
- This project is centered on analyzing **customer churn**, which occurs when clients discontinue their service. By studying historical data, we aim to discover underlying patterns and identify the key factors contributing to churn.
- These insights will enable the company to develop effective strategies for improving customer loyalty and minimizing churn rates.

BUSINESS UNDERSTANDING

- Understanding the rate at which customers are likely to churn will enable us come-up with better customer retention plans.

- This will in the long run increase profits and enable the organization plan better

- It is therefore important to predict the rate of customer loss

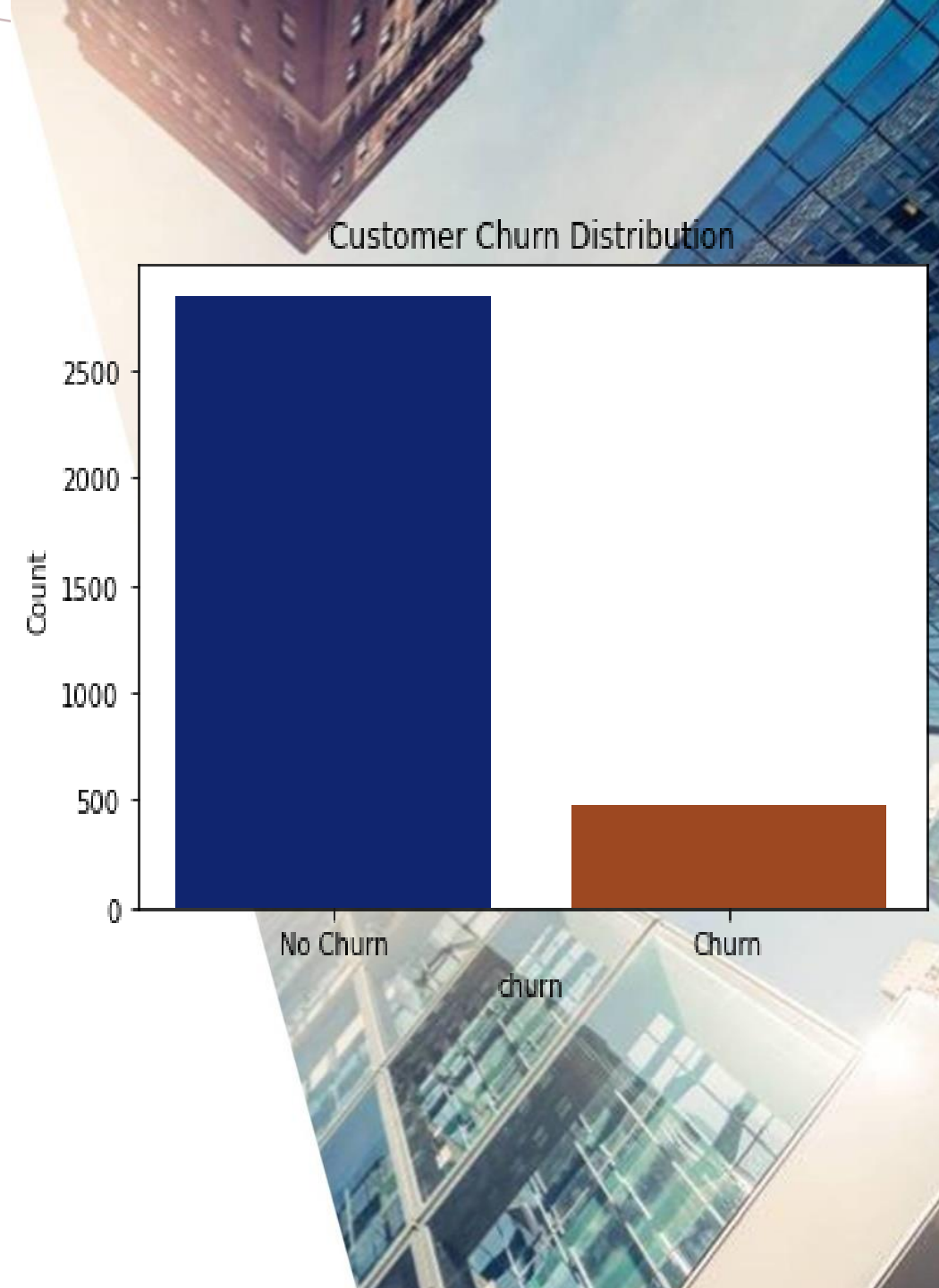
DATA UNDERSTANDING

This slide outlines the key features within our dataset, crucial for understanding customer behavior and, most importantly, identifying patterns related to **customer churn**.

I worked on the following Key features:-**Total Day, Evening, and Night Calls, Charges, and Minutes:**

Detailed records of usage across different times of the day. These metrics are vital for understanding customer engagement levels and their associated costs, which often directly impact satisfaction and churn.

As illustrated in this bar graph, the distribution of churned vs. non-churned customers provides an initial insight into the scale of the churn problem and serves as a baseline for further in-depth analysis. This visual representation helps us understand the proportion of customers leaving the service.



MODELING AND EVALUATION

- Features prepared by one-hot encoding categorical variables (e.g., international plan), scaling numeric features; target variable is churn (True/False).
- Dataset split into training (80%) and test (20%) sets ensuring unbiased evaluation.
- Models evaluated:
  - **Random Forest Classifier:** handles nonlinearities and interactions between features well.
  - **Logistic Regression:** simple, interpretable baseline model providing coefficients reflecting feature influence.

EVALUATION

- Confusion matrices demonstrate Random Forest's superior performance in identifying churners with fewer false negatives.
- Logistic Regression useful for understanding feature impacts but less suited for complex churn patterns.

Model	Accuracy	Precision (Churn)	Recall (Churn)	F1-Score (Churn)	Notes
Random Forest	94%	97%	60%	74%	Strong predictive power, handles complexity, robust to noise.
Logistic Regression	86%	58%	21%	31%	Highly interpretable but less effective with complex data.

RECOMMENDATIONS

Customers with more than 3 service calls are very likely to churn — consider proactive service quality improvement.

- International plan users churn more — analyze pricing or satisfaction issues related to this offering.

- High daytime usage correlates with churn — these may be high-value but high-risk customers.

- Use this model to flag high-risk customers and apply retention campaigns or offers.

The business model needs to change

From the analysis, it is evident that the cost of night calls is significantly higher than other call types, despite having the highest number number of hours spent on them. This suggests that the business model may need to be adjusted to better accommodate customer needs and preferences, particularly in terms of attracting the most clients.

This indicates a potential misalignment between customer usage patterns and the pricing strategy, which could lead to lost revenue opportunities for the business.

NEXT STEPS

- **Reassess night call pricing model** to align cost with usage patterns and improve customer satisfaction.
- **Improve customer service experience** for frequent callers to prevent churn escalation. Consider proactive outreach for customers with multiple service calls.
- **Review international plan offerings and pricing**, ensuring competitive, transparent, and value-aligned options.
- Deploy **machine learning models in CRM systems** to flag at-risk customers early and trigger retention campaigns.
- Promote **voice mail plans** and other value-added services shown to correlate with lower churn.
- Explore personalized plan options addressing customer-specific usage to reduce churn related to plan mismatch.

CONCLUSION

- Analysis combined detailed call usage analytics and customer plan data to uncover churn drivers.
- Predictive modeling with Random Forest demonstrated strong capability to identify churn risk with 94% accuracy, outperforming logistic regression.
- Insights emphasize the need for coordinated pricing, service improvement, and targeted customer engagement strategies.
- A data-driven churn prediction approach provides scalable support for retention efforts, vital in a highly competitive telecom industry with annual churn up to 25%
- Ongoing monitoring and model refinement recommended for sustained performance.

THANK YOU

Beatrice Ndanu Kiilu

beatricekiilu716@gmail.com