Predicting Customer Churn at SyriaTel

Agenda

- Business Context
- Data Understanding
- Key Insights
- Business Impact & Recommendation
- Next steps & Improvement

Business Understanding

What is Customer Churn and Why Should We Care?

- Customer churn occurs when a customer cancels their services or switches to another provider. For Syriatel, this represents a serious threat to profitability and long-term sustainability. In the highly competitive telecommunications industry, where churn rates typically range from 15% to 25%, churn is one of the most critical performance metrics to monitor and reduce.

- Retaining customers is a central part of Syriatel's strategy. This is because it is significantly more cost-effective to keep existing customers than to acquire new ones. Research shows that acquiring a new customer can be 5 to 10 times more expensive than retaining an existing one.

- Syriatel is one of Syria's leading telecommunications providers, offering mobile and data services to millions of customers across the country. However, due to the competitive nature of the industry, retaining customers is increasingly challenging. To maintain its market position and continue delivering high-quality services, Syriatel must find effective ways to understand and reduce customer churn.

- The objective of this project is to analyse customer behaviour and predict churn using machine learning techniques. By identifying customers at high risk of churning, Syriatel can apply targeted retention strategies to improve customer satisfaction and loyalty.

Stakeholders:

Who this helps?

- Syriatel Management ⟶ To strategize & Implement Customer retention programsSeamless integration
- Marketing ⟶ Targeted retention campaigns
- Customer Service ⟶ Improve support experience
- Finance ⟶ Revenue forecasting & budget allocation to customer retention

Business Objectives

- Predicting whether a customer will stay or churn
- Understanding the key reasons behind churn
- Recommending actionable solutions

Data Understanding:

- The dataset used for this project is from Kaggle
- it has 3333 rows which represents customers and 21 column which captures features that influence their decision to stay (Not churned) or leave (Churned) the service.

Feature Overview:

- Customer Info: State, Area Code, Phone number
- Tenure: Account Length
- Service Plans: Voice mail & International Plan
- Usage Patterns: Day/Night/Evening/International calls & minutes
- Billing: Charges by time period (Day, Night, Evening, International)
- Support: Number of Customer Service Calls

Data Limitation:

- Class Imbalance: about 85.5% of the customers did not churn, while only 14.5% did. This imbalance can affect model performance by making it biased toward predicting the majority class.

Exploratory Data Analysis (EDA):

Data Preparation :

- Explored dataset structure and key statistics
- Handled missing values and removed duplicates
- Dropped irrelevant features (e.g., phone numbers)
- Corrected data types and addressed outliers
- Encoded categorical variables
- Normalized numerical features for modelling

Key Data Insight:

- To better understand what drives customers to leave, I grouped the variables into thematic areas and explored their impact on churn. These factors include:

a) Tenure (Account Length) - to understand if longer-term customers are more or less likely to churn.

b) Geographical Factors(State) - to detect regional or location-based trends in churn.

c) Service Plan Subscriptions (International Plan & Voice Mail Plan) — Does having specific plans influences churn likelihood

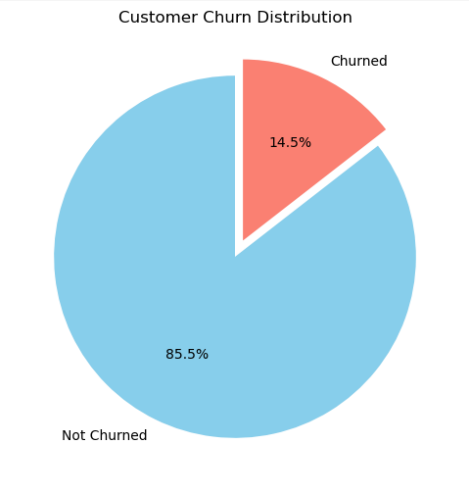
d) Usage Behaviour (Call & Minute Usage during all period of time) — Does high or low usage patterns are linked to churn.

e) Financial Impact (Charges): to evaluate if higher billing is associated with customer dissatisfaction and churn.

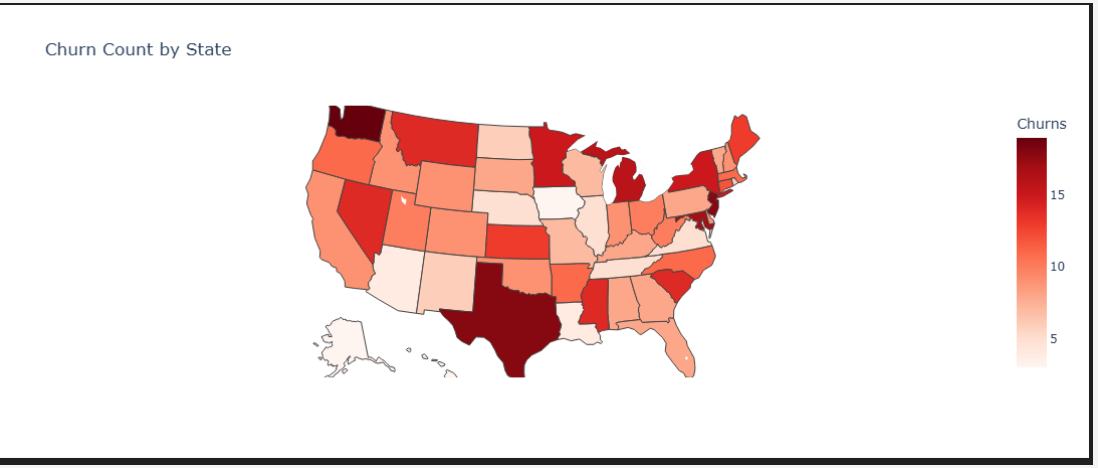
f) Customer Service Interaction (Number of Customer Service Calls) — Does frequent service contact signals dissatisfaction.

Understand the key reasons behind churn:

a) Churn Distribution : we observe that 85.5% of the customers did not churn while only 14.5% churned. This shows a significant class imbalance.

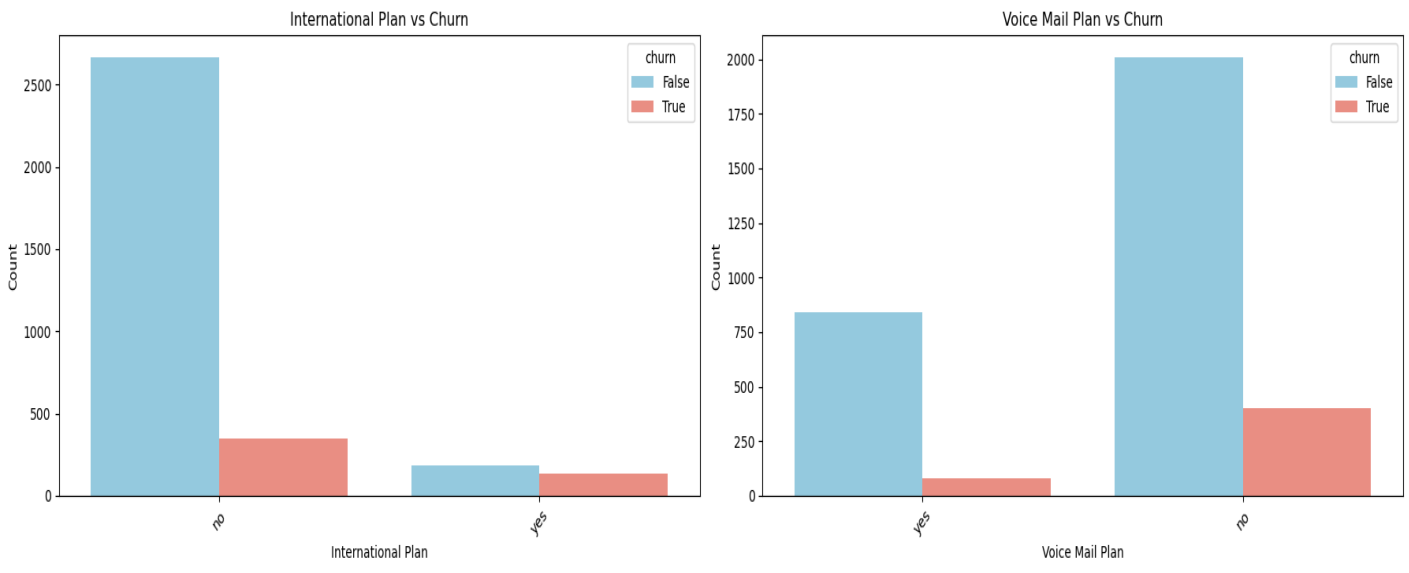


b) Geographical analysis : To support the company's location-based retention strategy, I identified which states have the highest number of churned customers.



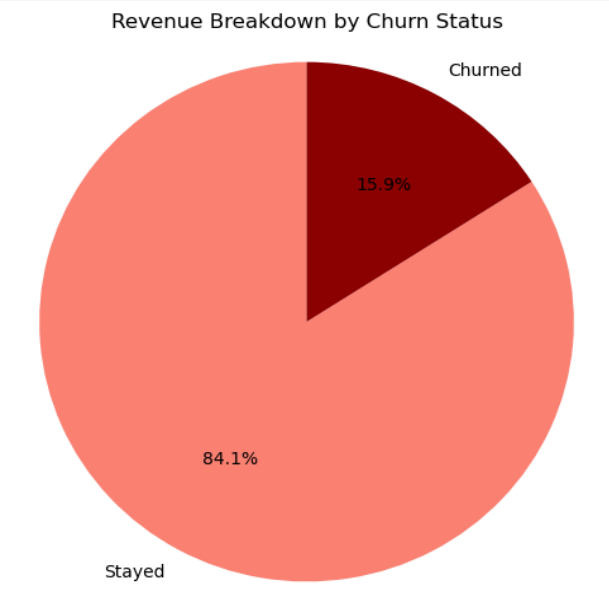
- I visualized the churn intensity to helps the company to prioritize outreach efforts and marketing campaigns in high-churn states and design targeted interventions by region.
- Washington, New jersey & Texas have the highest customers who have churned
- Majority of other states reflect a moderate level of churn and shouldn't be overlooked when designing region-specific retention strategies

c) Service Plan : To understand the impact of service subscriptions on churn, I examined whether having an international plan or a voice mail plan made customers more or less likely to churn.



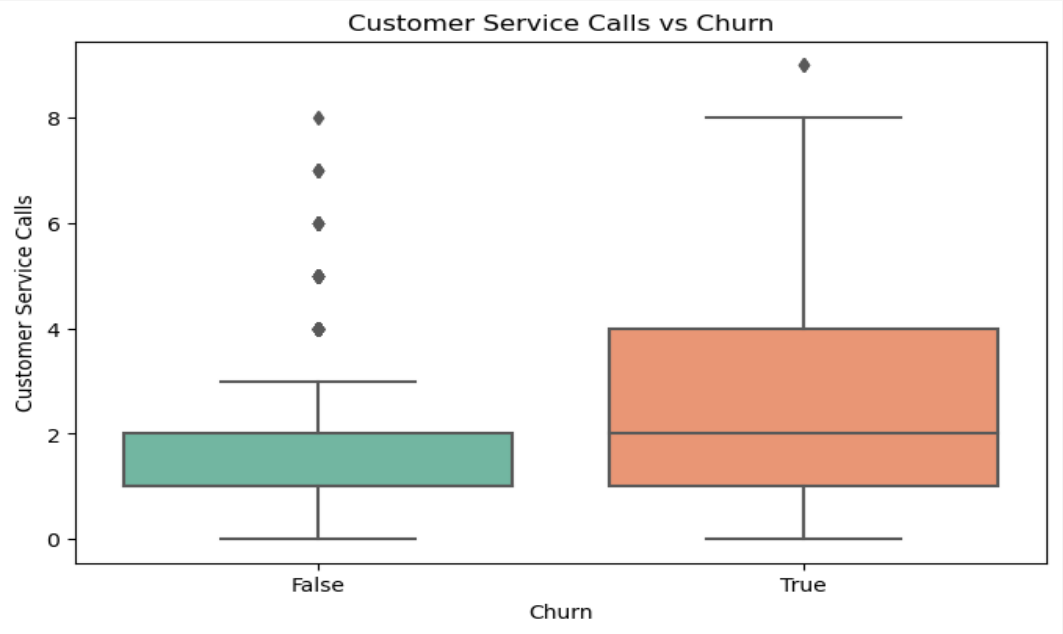
- Customers with an international plan are significantly more likely to churn compared to those without.
- Customers with a voice mail plan are less likely to churn.
- The highest risk group is international plan customers with no voicemail Services and the company should have immediate action for these group of customers

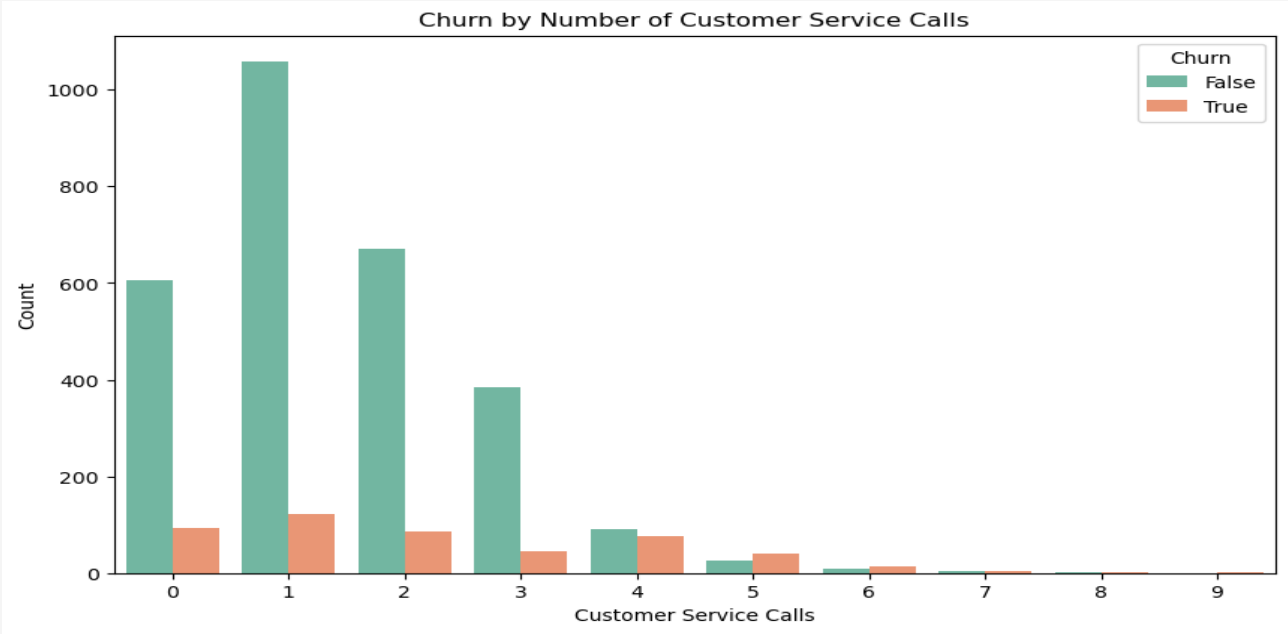
d) Financial Impact: I compared the total revenue from churned vs. non-churned customers and shows what percentage of overall revenue each group contributes.



- Churned customers account for 15.9% of the company's total revenue loss. On average, each customer who leaves contributes approximately 1.4% to that lost revenue.

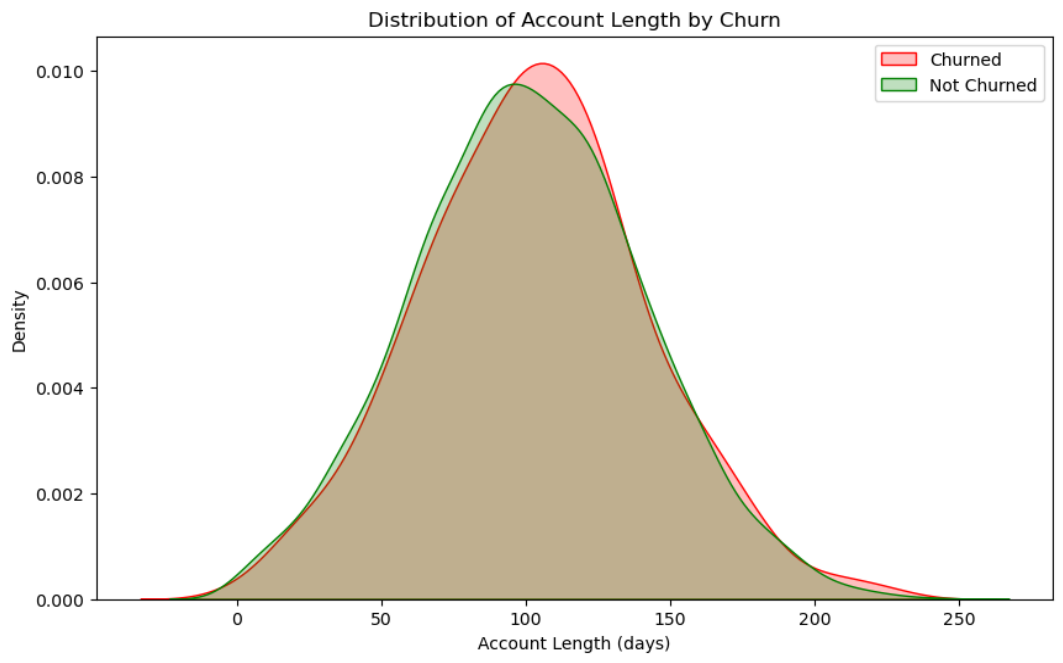
e) Customer Service Calls : To assess whether customer dissatisfaction drives churn, I examined the number of customer service calls made, as repeated contact without support may indicate unresolved issues or frustration.





- Average call made by customers who churned was 2.23 while those who stayed made an average of 1.45 calls.
- The difference of just one extra call on average suggests a thin line between staying and churning, even a small increase in service issues can tip a customer toward leaving.
- The churn rate increases sharply with the number of service calls therefore the company should work on the first contact resolution
- Once customers hit 5 or more service calls, churn shoots up (above 50%+) indicating rising dissastifaction.

e) Tenure: To understand how customer longevity influences churning, I analyzed how long each customer has been with the company (in days).



- The spread and median of tenure for churned and non-churned customers look very similar.
- Both groups have a median around 100 days.
- Churned customers do not appear to have dramatically shorter or longer tenures compared to loyal ones.

Modelling:

Objective:

- The goal is to predict customer churn for SyriaTel using available customer data. This helps the company proactively retain customers likely to stop using the service.

Models used

- Logistic Regression (No resampling): checking how well logistic regression works without doing anything about imbalanced data (if some classes appear much more than others).
- Logistic Regression + SMOTE : Using SMOTE, which artificially creates more examples of the minority class to improve fairness of the model. This helps the model learn both classes more equally.
- Random Forest Classifier: Performed significantly better with high accuracy and good recall for churners, making it suitable for deployment.

Model Comparison:

Model	Confusion Matrix	Accuracy	Recall	F1- Score
Logistic (No Smote)	[[552, 22], [75, 18]]	0.85	0.19	0.27
Logistic + Smote	[[402, 172], [18, 75]]	0.72	0.81	0.44
Random Forest + Smote	[[543, 31], [17, 76]]	0.93	0.82	0.76

Modelling:

Model Interpretation:

- Logistic no resampling: has high accuracy, but very poor recall and F1 for the minority class. It misses many potential churners, making it less suitable for identifying at risk customers.
- Logistic + SMOTE improves recall for churners a lot, but loses accuracy by catching churners, but misclassifies many non-churners.
- Random Forest + SMOTE gives the best of both worlds because it has the highest overall accuracy, recall and F1- score for churners

Model Selection

- I selected Random Forest model combined with SMOTE resampling as the final model for deployment as it achieves:
  a) Highest accuracy (93%)
  b) Strong recall (82%) for identifying churners
  c) Best F1-score (76%) which balances precision and recall
- Class imbalance is addressed using SMOTE, improving the model's ability to detect minority cases.
- Random Forest captures complex relationships in customer behaviour better than linear models like logistic regression.

Modelling:

Business Implication for SyriaTel

- This model enables early and accurate detection of customers likely to churn, which directly supports SyriaTel's goal of proactive customer retention. With this insight, the business can:

  a) Target at-risk customers with personalized offers or service improvements

  b) Reduce churn rates which protects recurring revenue

  c) Optimize marketing spend focusing only on customers who truly need retention efforts

  d) Improve customer satisfaction by intervening before dissatisfaction leads to loss

Areas for Enhancement:

- While the current analysis provided a strong starting point for churn prediction, there are several areas that could be explored further to improve model performance and business impact, given more time or deeper expertise.

  a) Model Refinement - Tuning the model parameters and explore more advanced models (Gradient Boosting)

  b) Model Deployment & Monitoring - Integrating model into CRM and set up real-time performance tracking

Recommendation:

- Run localized campaigns in high-churn states (Washington, Texas) with deeper investigation into region specific issues like service quality, network coverage or billing concerns.
- Develop onboarding programs for new customers and loyalty program for long term customers to reduce early drop-offs and late disengagement.
- Proactively monitor customers with 3+ surpport calls and prioritize them for resolution. Train customer service team to resolve issues on the first contact to prevent frustration.
- Reassess the value proposition of the international plan. This could involve improving call quality, reducing costs, or bundling with other perks to increase satisfaction.
- Consider making the voice mail plan a default offering or promote it more actively, given its positive association with retention.
- Introduce spending caps or usage notifications for customers who pay more especially during daytime & International calls to help manage expectations and reduce bill shock as these users are more likely to churn.

Summary

- In this project, I analyzed SyriaTel's customer data to understand why customers churn and built a machine learning model to predict those at risk of leaving. The analysis explored key churn drivers, including international plan subscriptions, frequent customer service calls, and high call charges.
- Using a Random Forest model with SMOTE to handle class imbalance, the model achieved 93% accuracy and strong recall for churners. This model enables the company to proactively target at-risk customers, reduce revenue loss, and improve retention strategies.
- With further refinement and deployment, this solution can be integrated into SyriaTel's CRM to support real-time, data-driven decision-making.

Thank you

Hafsa M.Aden

Hafsa.mohammed.aden@gmail.com