SYRIATEL CUSTOMER CHURN ANALYSIS

By: Gloria Mutheu Tisnanga

Business Understanding for Syriatel Churn Analysis

Syriatel's Business and the Impact of Emerging Technologies

- Syriatel is a leading telecommunications company in Syria, providing a range of services including mobile and fixed-line telephony, internet access, and data services. As a state-owned enterprise, Syriatel plays a crucial role in the country's communication infrastructure.
- Understanding the Problem:
- Churn, or customer attrition, is a significant concern for telecommunications companies like Syriatel. When customers leave, it directly impacts revenue, brand reputation, and overall business health. Therefore, understanding the factors driving churn is crucial for developing effective retention strategies.

Understanding the Problem:

The goal is to predict which Syriatel customers are likely to churn and identify the underlying reasons for this behavior. This information can be used to develop targeted retention strategies and improve overall customer satisfaction.

Key Questions and Corresponding Classification Tasks:

1.Which customer segments are most likely to churn?

2.What are the primary reasons for customer churn?

3. What is the cost of customer churn to Syriatel?

Business Objectives:

- Reduce customer churn rate: Implement strategies to retain existing customers.
- Improve customer satisfaction: Enhance the overall customer experience.
- Optimize pricing and promotional offers: Develop pricing plans that meet customer needs and incentivize loyalty.
- Enhance customer support: Provide timely and effective assistance to address customer concerns.
- Identify at-risk customers: Proactively reach out to customers who are likely to churn.

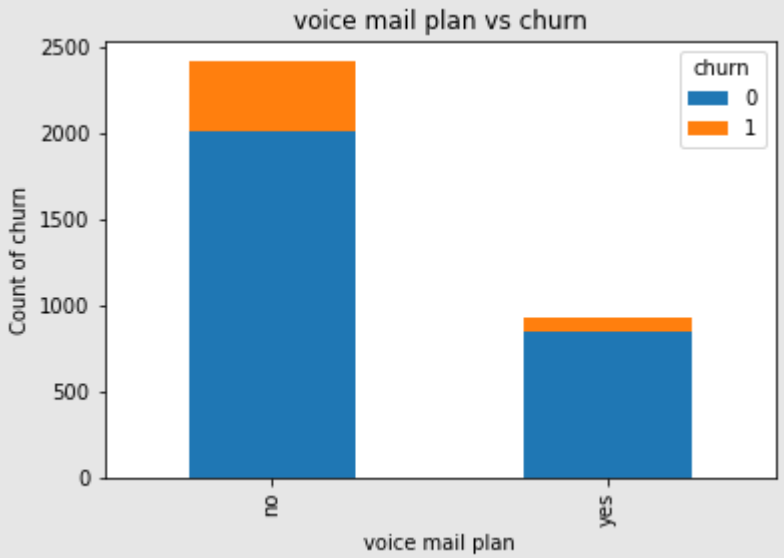
By addressing these questions and objectives through data analysis, Syriatel can gain valuable insights into customer behavior, develop effective retention strategies, and improve overall business performance.

Data Understanding

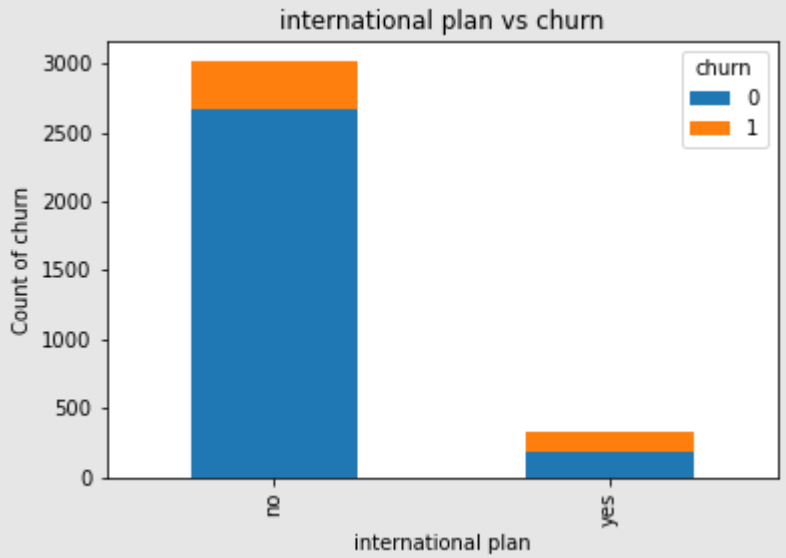
- The SyriaTel dataset from Kaggle is composed of a comprehensive set of customer features, providing multifaceted information about customer usage behaviors, preferences, and interactions.
- The dataset contains 3333 entries and 21 columns.
- The total memory usage of the dataset is approximately 524.2 KB.
- The columns represent various customer attributes, including state, account length, area code, phone number, international plan, voice mail plan, number of voice mail messages, call durations and charges for different time periods and international calls, customer service calls, and churn status.
- The dataset does not have any missing values, as indicated by the non-null counts.
- The data types of the columns include bool, float64, int64, and object.
- The bool column represents the churn status, indicating whether a customer discontinued the service (True) or not (False).
- The float64 columns represent numerical values for call durations and charges.
- The int64 columns represent numerical values for account length, area code, number of voice mail messages, call counts, and customer service calls.
- The object columns include state, phone number, international plan, and voice mail plan, which are categorical variables.
- By understanding these features and their implications, we can conduct in-depth analyses and predictive modeling to tackle the issue of customer churn.

Explorative Data Analysis

Voice male plan vs churn

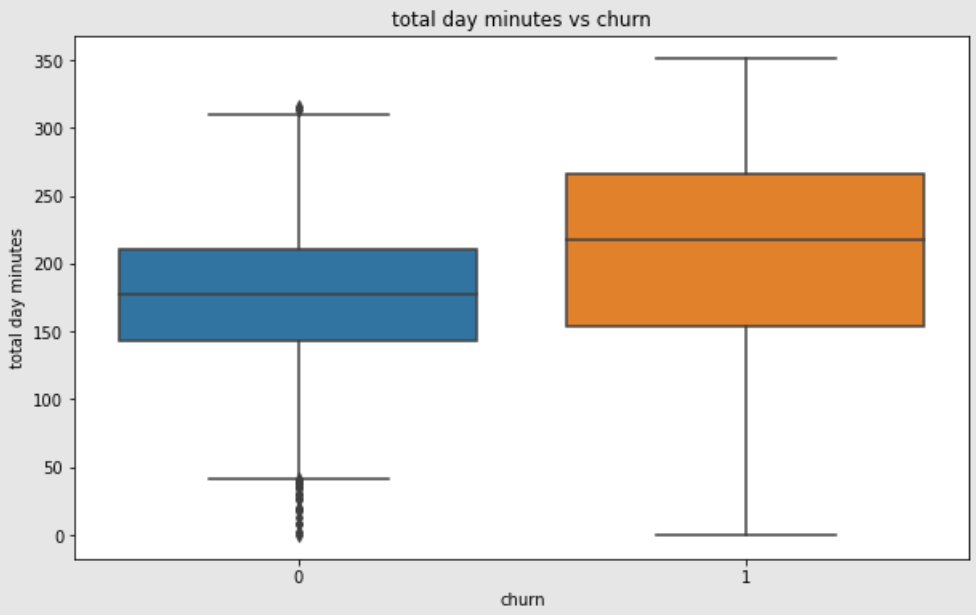


International Plan vs churn

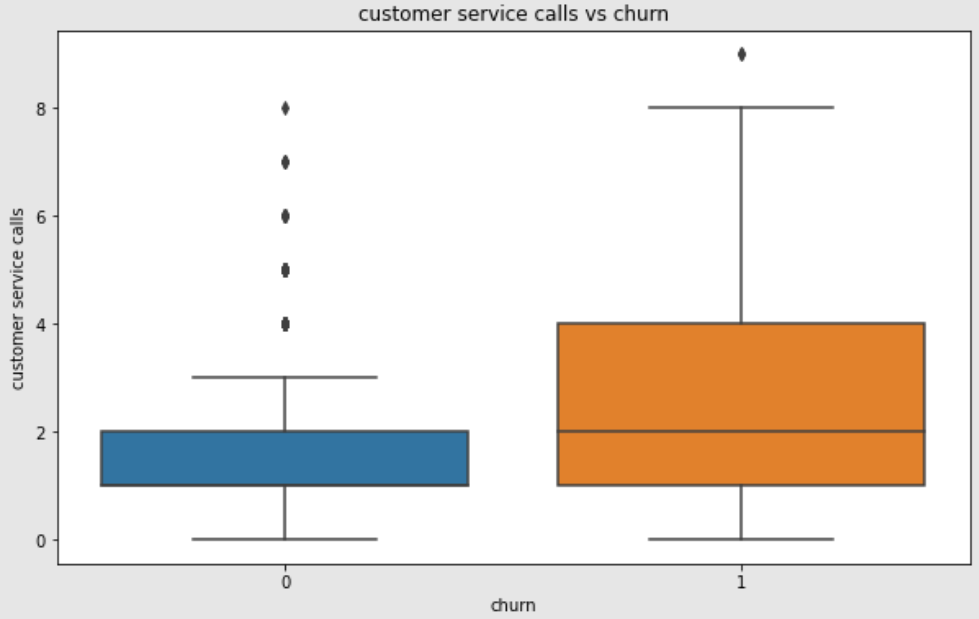


Explorative Data Analysis

Total day minutes vs churn



Customer service calls vs churn



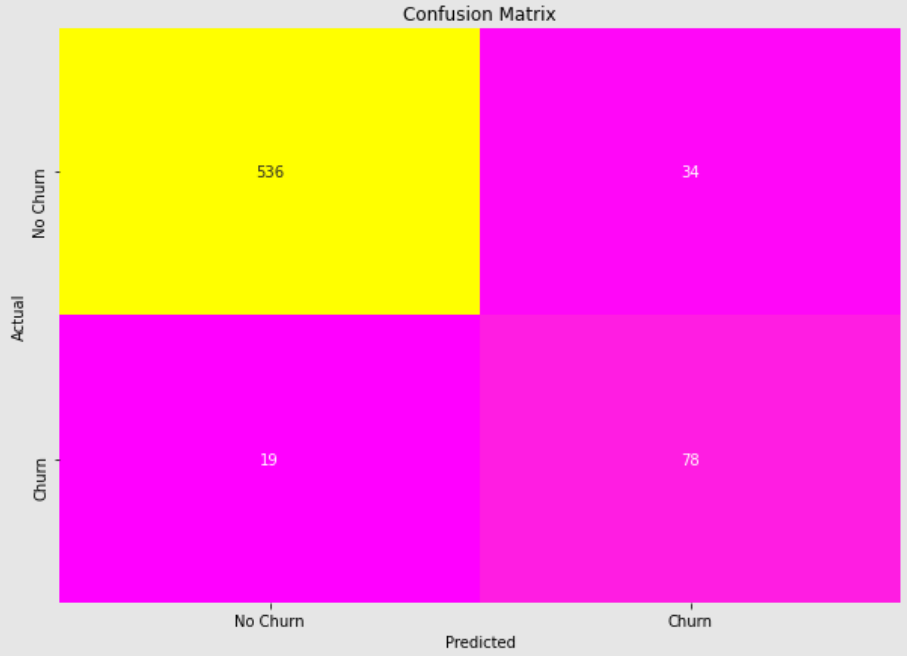
Modelling Approach

Model Performance Metrics

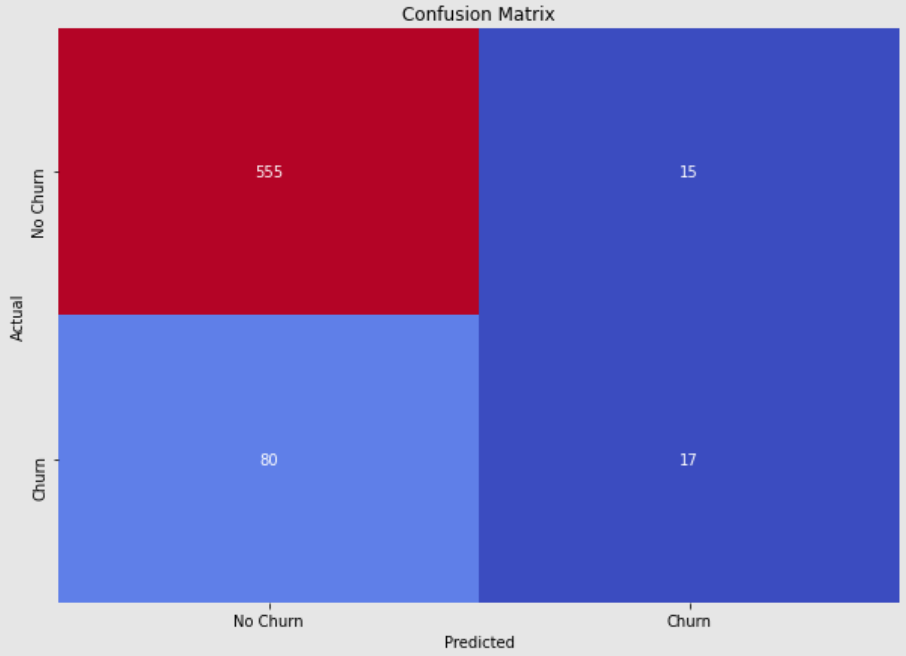
	Model	Accuracy (%)	ROC AUC	Precision (Class 0)	Precision (Class 1)	Recall (Class 0)	Recall (Class 1)	F1-score (Class 0)	F1-score (Class 1)
Decision Tree	Decision Tree	96.85	0.8918	0.96	1.0	1.0	0.78	0.98	0.88
Logistic Regression	Logistic Regress	86.36	0.6036	0.88	1.0	0.97	0.57	0.92	0.24
Random Forest	Random Fores	91.3	0.7053	0.91	1.0	1.0	0.98	0.95	0.41
Gradient Boosting	Gradient Boosti	96.85	0.8918	0.96	1.0	1.0	0.78	0.98	0.88

Model Perfomance Evaluation

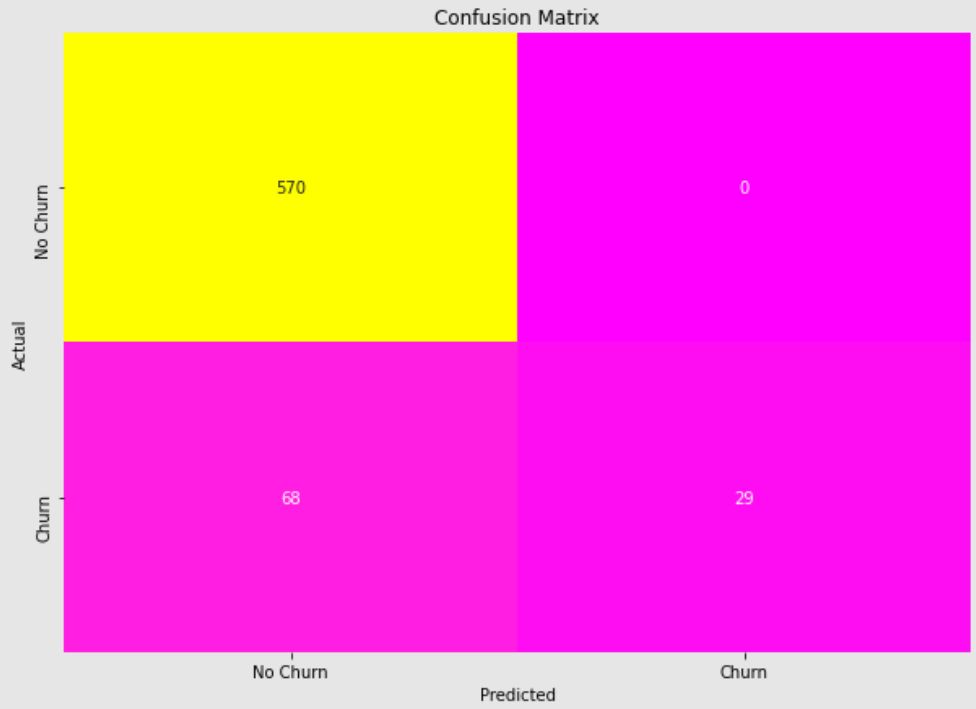
ConfusionMatrix: Decision Tree

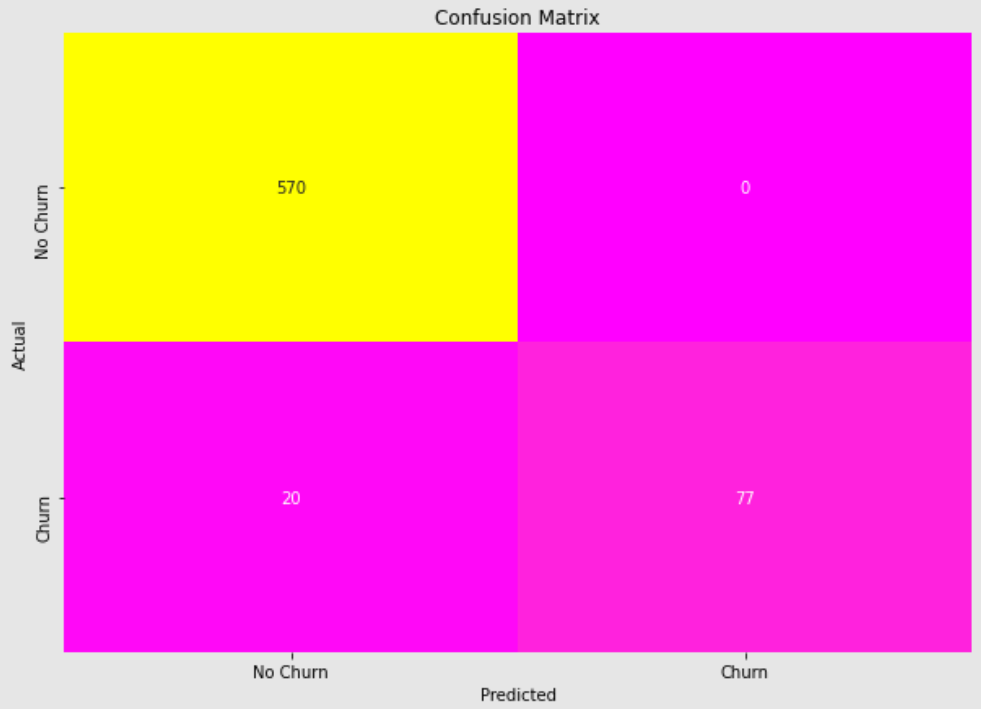


Logistic Regression



Confusion Matrix: Random Forest and Gradient Boost

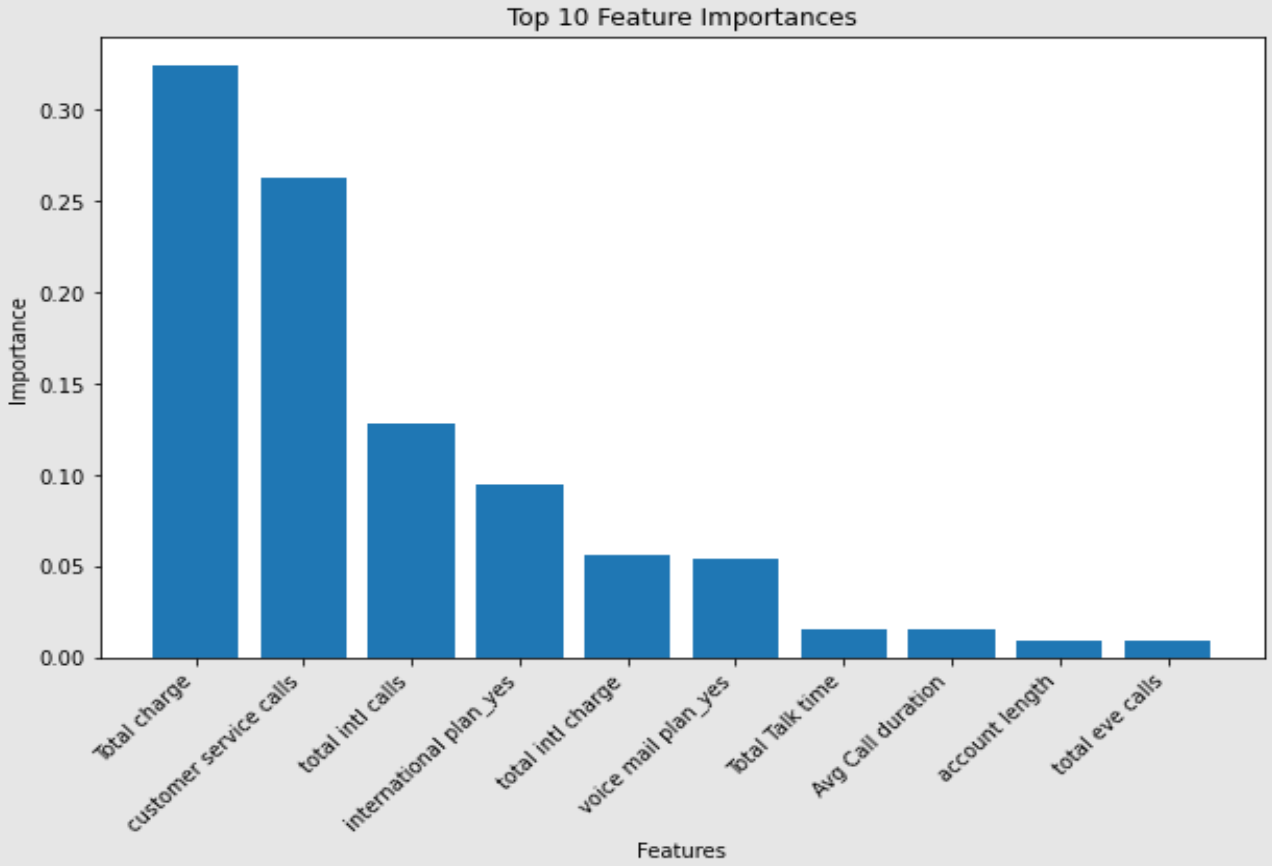




Tuning Findings

- Decision Tree: Hyperparameter tuning reduced the accuracy slightly (0.9685 to 0.9205) but might have improved performance on specific classes based on the F1-score in the classification report.
- Random Forest: Tuning resulted in a slight improvement in accuracy (0.9130 to 0.8981) but a significant increase in ROC AUC (0.7053 to 0.9207), indicating a better ability to distinguish between churners and non-churners. However, recall for class 1 (churners) remains low (0.30).
- Gradient Boosting: Tuning maintained the high accuracy (0.9700) while slightly improving the ROC AUC (0.8918 to 0.9281).
- Logistic Regression: Tuning had minimal impact on performance (accuracy remained around 0.86).
- Comparing Model Performance:
- Overall Accuracy: Gradient Boosting remains the most accurate model (0.9700).
- ROC AUC: Random Forest has the highest ROC AUC (0.9207), indicating the best ability to distinguish churners from non-churners.
- Precision and Recall:
- Precision: Gradient Boosting and Decision Tree have high precision for class 0 (non-churners). Random Forest has high precision for class 0 but struggles with class 1.
- Recall: Gradient Boosting achieves perfect recall for class 0. Both Decision Tree and Gradient Boosting have improved recall for class 1 compared to the untuned models.

Feature Importance



Conclusion

Recommendations

- **Customer Service Enhancement**
  - Improve service quality to reduce call frequency.
  - Collect feedback from frequent callers.
- **Usage Patterns Analysis**
  - Tailor plans for high call duration or total talk time.
  - Monitor and ensure positive experiences during long calls.
- **Plan-Specific Strategies**
  - Offer promotions and support for international and voicemail plans.
- **Billing and Charges**
  - Review and adjust billing practices; consider loyalty programs.
  - Align billing with high engagement call volumes.
- **Communication & Retention Programs**
  - Create targeted campaigns using insights.
  - Implement churn prevention strategies.

Conclusion

Future Work

- **Feature Engineering**
  - Explore new features and create interaction terms.
- **Model Improvement**
  - Refine features using advanced selection techniques.
  - Continue hyperparameter tuning.
- **Model Comparison**
  - Test additional models like XGBoost, LightGBM.
  - Use cross-validation for robustness.
- **Customer Segmentation**
  - Perform segmentation and cluster analysis for targeted strategies.
- **Monitoring & Feedback Loop**
  - Implement feedback loops and A/B testing for strategy optimization.

Thank You.