NON - TECHNICAL DATA REPORT

1. BUSINESS UNDERSTANDING

Business Overview

Syriatel (Arabic: سيريتل) is a leading telecommunications company in Syria, known for its rapid growth and extensive market presence. With a robust network of 63 Points of Service across the country, Syriatel handles over 25,000 customer queries daily through its Call Centers and operates 2,783 radio base stations. The company proudly serves over 6 million customers, holding a 55% share of the Syrian market. Their skilled team is committed to delivering high-quality services and solutions, solidifying Syriatel's position as one of the region's fastest-growing telecom operators.

Problem Statement

As new customers begin using a product, each contributes to the growth rate of that product. However, over time, some customers may discontinue their usage or cancel their subscriptions for various reasons. Churn refers to the rate at which customers cancel or choose not to renew their subscriptions, and a high churn rate can significantly impacts revenue.

Syriatel has observed an increase in customer churn and is concerned about the financial losses associated with customers who discontinue their services prematurely.

Objectives

To Determine the features that serve as early indicators of customer churn.

To Analyze and identify the underlying reasons why customers discontinue their service.

To Build a Predictive Model that is capable of accurately predicting when a customer is likely to discontinue their service.

Success Criteron

This analysis aims to:

Identify Key Features: Determine at least five key features that strongly correlate with customer churn, providing actionable insights for Syriatel to monitor and address customer dissatisfaction effectively.

Develop a Predictive Model: Build a classifier model that achieves: At least 90% accuracy in predicting customer churn. A minimum precision of 75%, ensuring the model minimizes false positives and provides reliable predictions.

Support Business Decision-Making: Enable Syriatel to use the identified features and model predictions to implement targeted retention strategies, reducing churn and mitigating revenue loss.

2. DATA UNDERSTANDING

The Churn in Telecom's dataset was sourced from Kaggle. The dataset has 3,333 rows and 21 columns. Each column represents a customer and the columns represent the customer details. The customer details are as follows:

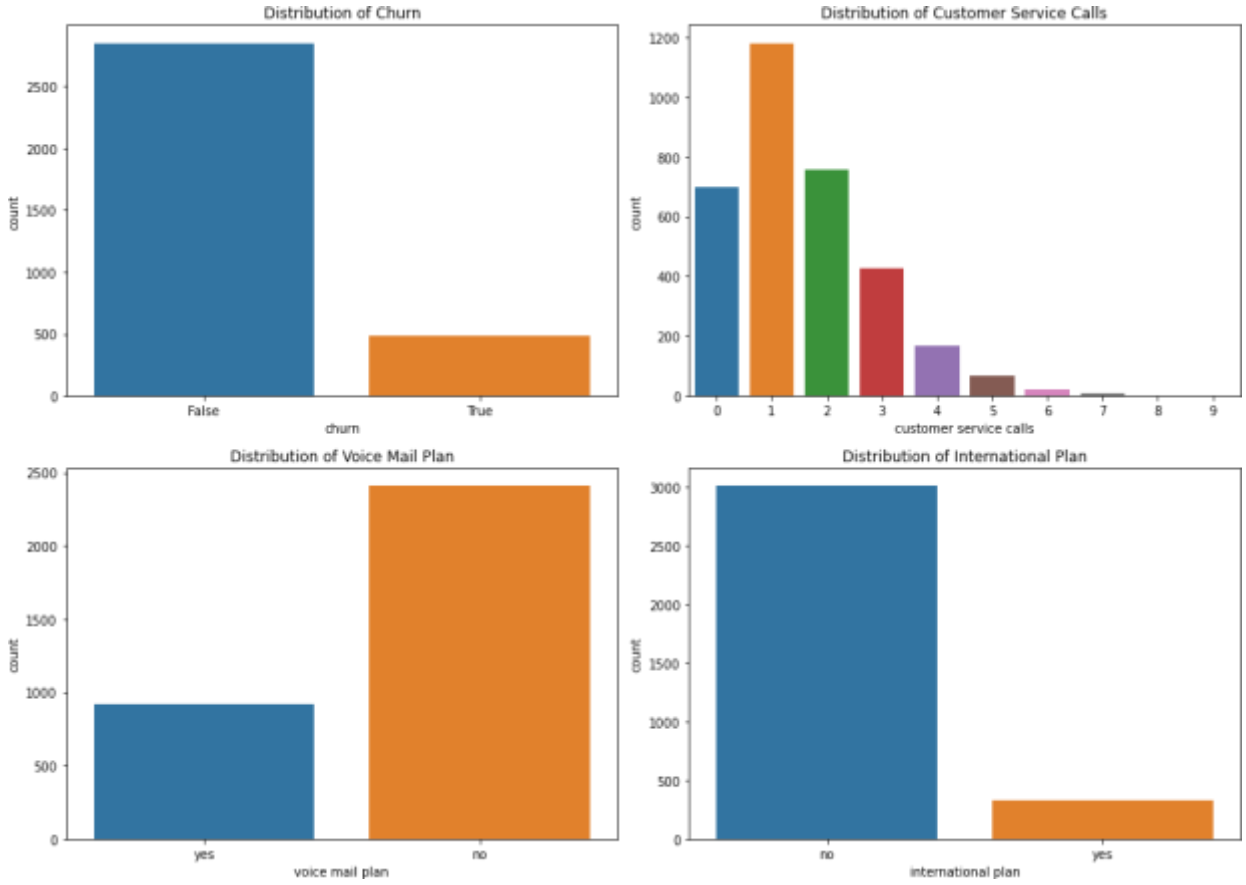
1. the state the cusomer lives in,
2. account length- the number of days the customer has had an account,
3. the area code of where the customer lives,
4. the customer's phone number,
5. international plan- true if the customer has the international plan, otherwise false,
6. voice mail plan- true if the customer has a voice mail plan, otherwise false,
7. number vmail messages- the number of voicemails the customer has sent,
8. total day minutes- the total number of minutes the customer has been on call during the day,
9. total day calls- total number of calls the user has done during the day,
10. total day charge- total amount of money the customer was charged for calls during the day,
11. total eve minutes- the total number of minutes the customer has been on call in the evening,
12. total eve calls- the total number of calls the customer has been on in the evening,
13. total eve charge - total amount of money the customer was charged calls during the evening,
14. total night minutes - the total number of minutes the customer has been on call at night,
15. total night calls- total number of calls the user has done at night,
16. total night charge- total amount of money the customer was charged for calls at night,
17. total intl minutes- the total number of minutes the customer has been on international calls,
18. total intl calls- total number of international calls the customer has done,
19. total intl charge- total amount of money the customer was charged for international calls,
20. customer service calls- number of calls the customer has made to customer service,

21. churn- true if the customer terminated their contract, otherwise false

3. DATA PREPARATION & ANALYSIS

Upon checking the data for duplicates, missing values and null values and there were none.

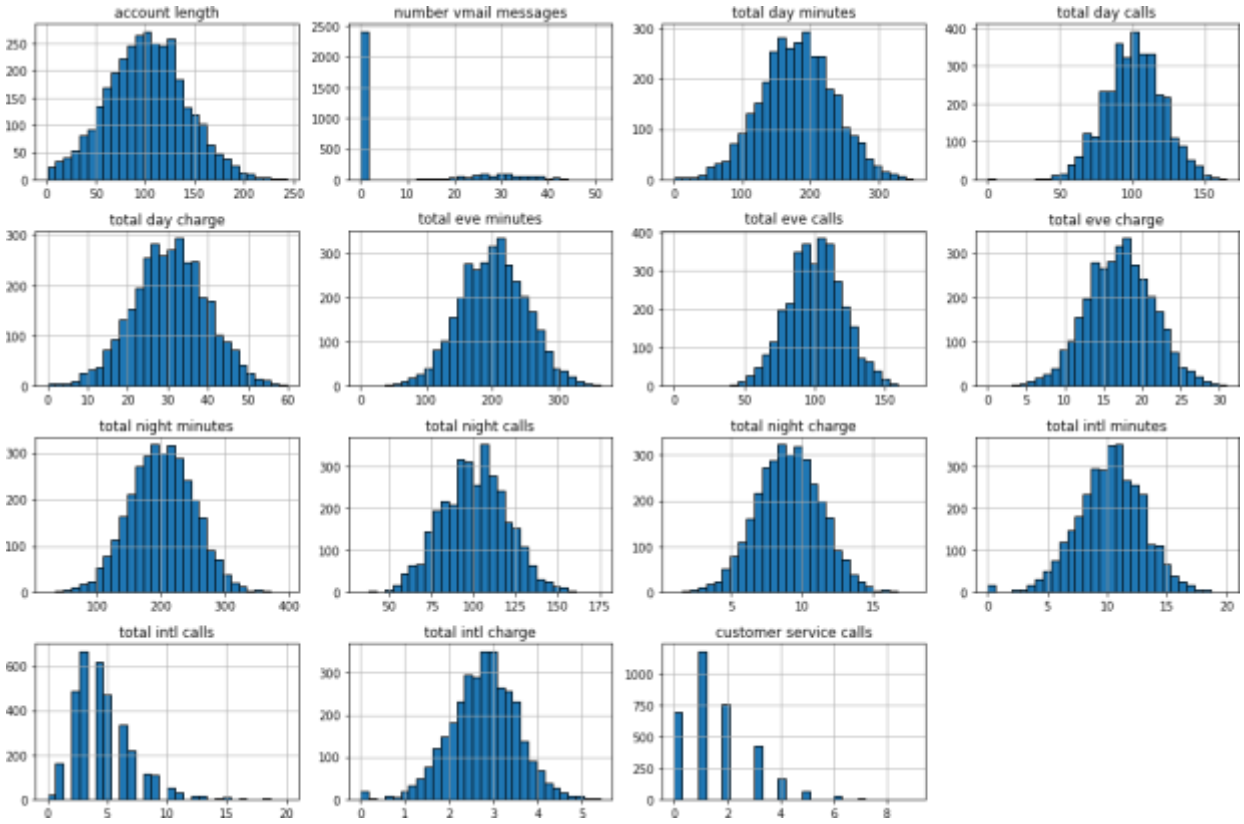
DISTRIBUTION OF CATEGORICAL VARIABLES..



As we can see from the graphs here, We can deduce the following;

1. They have retained most of their customers however the churn, though small could be significant in the longrun.
2. Most customer issues were solved by the 4th call and also a significant number of clients about 700 had no issues to reach out for.
3. Only about a third of their clients have a voicemail plan.
4. A very small amount of clients have an international plan and as such we can say that their services are mostly used domestically.

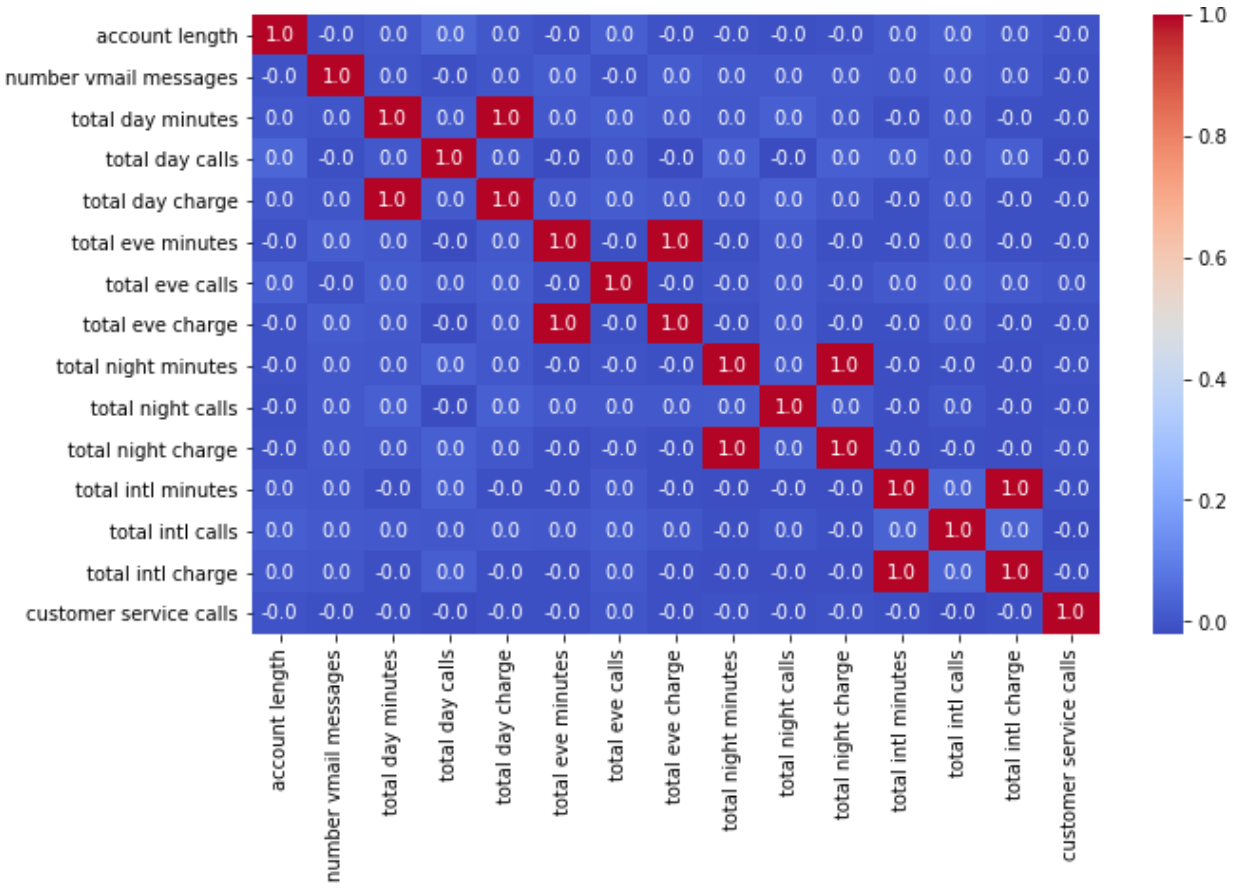
DISTRIBUTION OF NUMERIC VARIABLES.



As we can see from the graphs here, We can deduce the following;

1. The service usage is fairly normal, this is because we are seeing normal distribution especially on; day calls, evening calls, night calls and international calls.
2. The charges they've issued clientele match up o the usage and as such there are no mispricing issues.

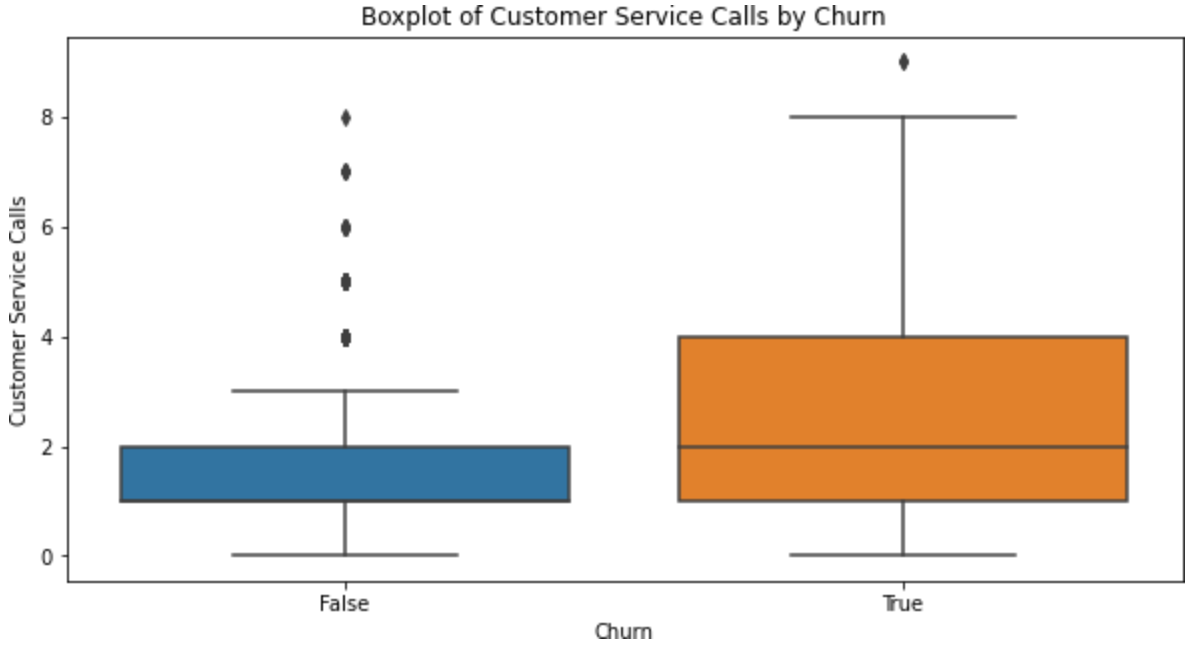
CORRELATION OF NUMERIC VARIABLES.



As we can see from the table here, We have done a correlation matrix for all the numeric variables. We can deduce the following;

1. There is very little relationship between the variables except for the charge and minutes.
2. The assumption made in the previous slide in regards to proper pricing of their products is further supported by this.

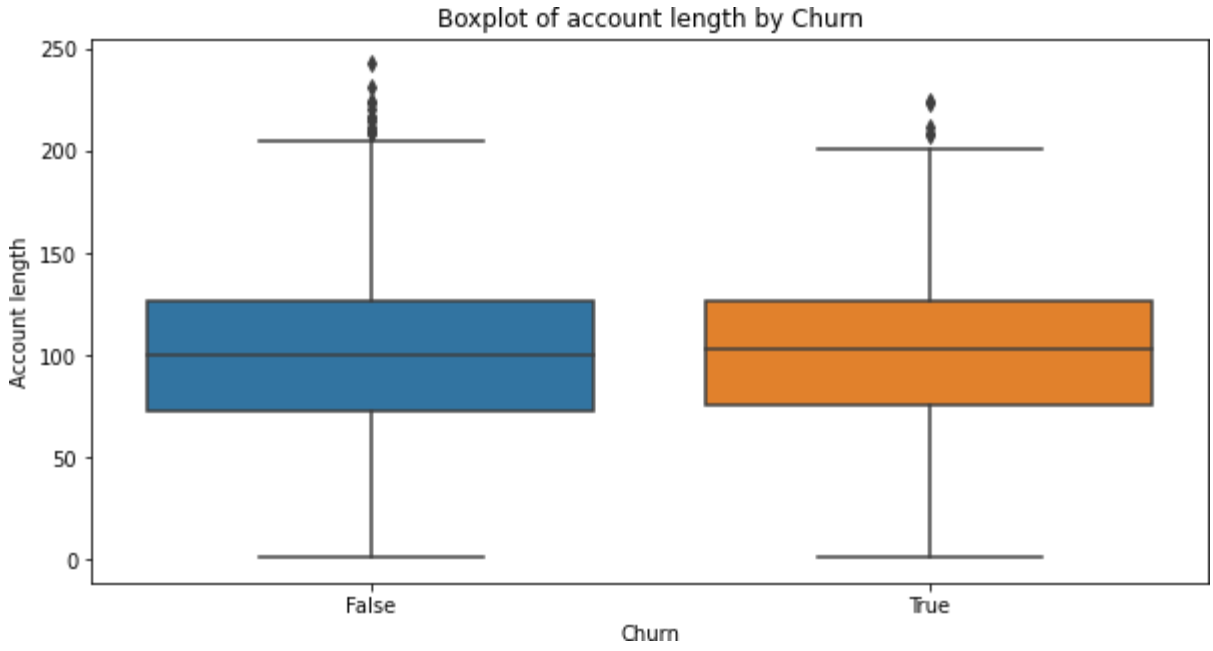
CROSS ANALYSIS



As we can see from the box plot here, The distribution of the customer service calls split between the churn values showing median, mean, quartiles and outliers. We can deduce the following;

1. There customers that left for the most part made attempts to reach out to customer service. This could mean the service may have had an issue for them that wasn't solved or wasn't sorted to their satisfaction.

CHURN VS CUSTOMER SERVICE CALLS.



As we can see from the box plot here, The distribution of the customer service calls split between the churn values showing median, mean, quartiles and outliers. We can deduce the following;
1. There customers that left for the most part made attempts to reach out to customer service. This could mean the service may have had an issue for them that wasn't solved or wasn't sorted to their satisfaction.

CLASSIFICATION MODELLING

MODEL / PARAMETER	LOGISTIC REGRESSION	RANDON FOREST CLASSIFIER	DECISION TREE CLASSIFIER
Accuracy	84%	95%	93%
Recall	11%	74%	69%
Precision	48%	94%	85%
F1- Score	19%	83%	76%

From the above table, we tried to implement and tune 3 different model types so as to see which one best fit the data and would act as a reliable classifier that Syriatel could use for analysis of their Churn.
1. Logistic Regression performed poorly and as such we don't recommend it for this purpose.

2. Random Forest Classifier performed the best yielding the most reliable predictions. We highly recommend this model.

3. Decision Tree Classifier was also came close but we shall not pick this one Definitions:
•Accuracy: The percentage of correctly predicted cases out of all cases.
•Precision: Out of all predicted positives, how many were actually positive.
•Recall: Out of all actual positives, how many were correctly predicted.
•F1 Score:A balance between precision and recall (useful for imbalanced data).

CONCLUSION, RECOMMENDATIONS.

The company should rely on our Random Forest Classifier when making analysis.

Also, they need to focus on items around customer service and analyze the issues raised so that they may be able to correct these along side implementing customer retention strategies and using the model to gauge effectiveness