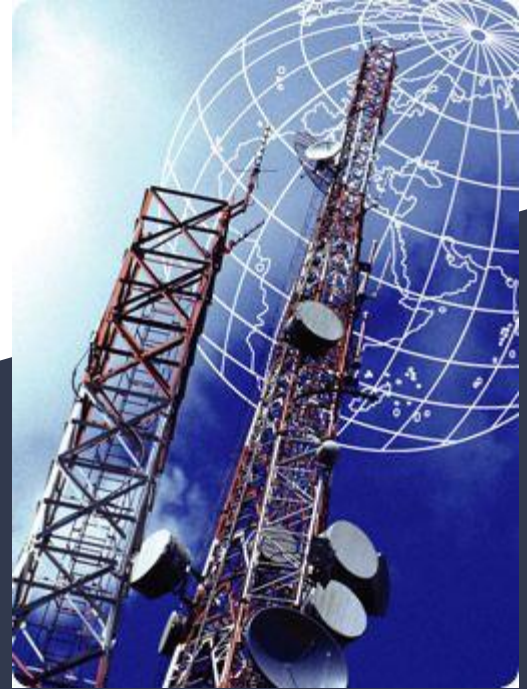


SYRIA TEL ANALYSIS

Presented by: Mualuko Janet Mumo.



OVERVIEW.

1. Business Understanding.
2. Data Understanding
3. Model Evaluation
4. Recommendations
5. Modeling
6. Strategy.

Business Understanding.

Customer churn refers to the losing of customers over a period of time. It impacts the company's revenue heavily and leads to losses.

This study aims to investigate the factors that contribute to this problem and strategize for ways to mitigate this problem.



Data Understanding.

We sourced the data from the company .(Churn Dataset).Below is the summary to explain the key features in the dataset.

Customer Demographics.(State and Area Code.), which explain where the customers resides.

Account length which describes the duration of the customer's relationship with the company

Service Usage.(Total Day Minutes, Calls, Charge. Similar metrics for the evening and night and International Calls,Minutes, and Charge.

Service Plans. International Plan and Voce Mail Plans. Whether a customer has or does not have the plans.

Customer Service Calls. The number a of times a customer has made to the company

Churn. Shows whether the customer contract ended or whether the customer is still using the company's services.



Objectives

The main objective is to use a predictive model to analyse the data and get insights of the number on the churners.

Other Objectives.

Develop strategies to retain customers

Investigate which features contribute to the churn challenge.

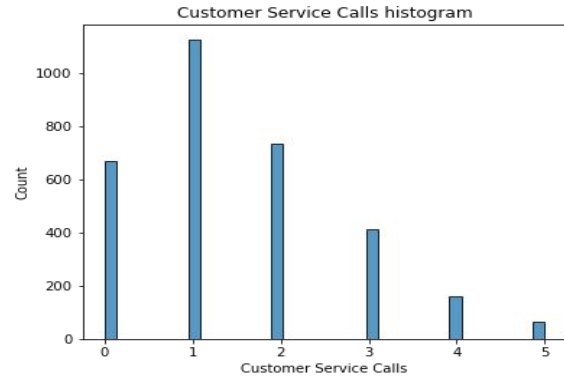
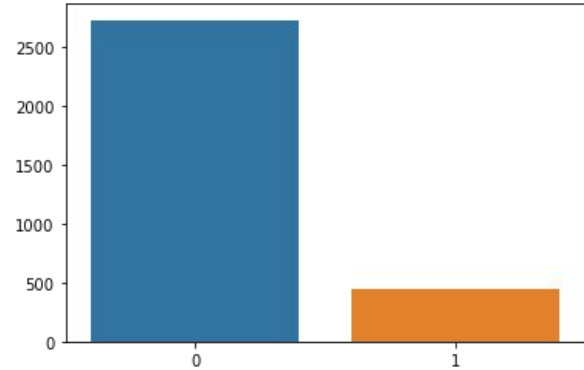
Modeling.

We decided to use machine learning since it is designed to forecast future outcomes based on the historical data. Unlike simple data analysis, which focuses mainly only on understanding historical data trends. We first , cleaned and prepared the data. We then explored the data to check for the features characteristics and then went ahead to preprocess the data to help improve the models predictive power.

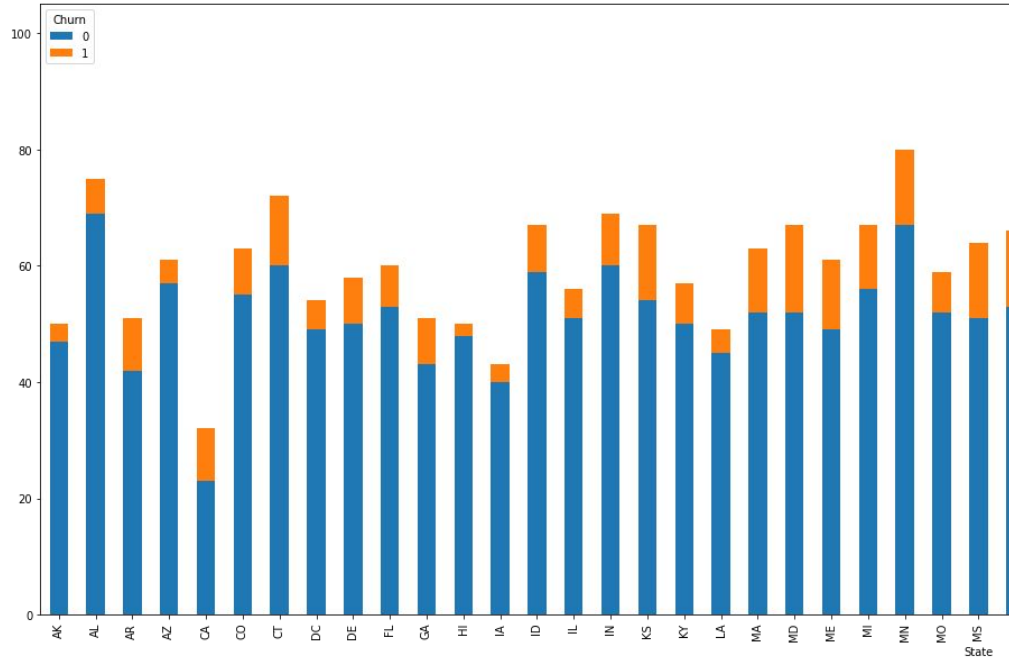
Shows the distribution of the customers who are still using the company's products and those that churned.

This shows the counts of the n

umber of customer service calls. More than 1000 customers made one customer service call. This might indicate there might be one common problem that the customers are facing

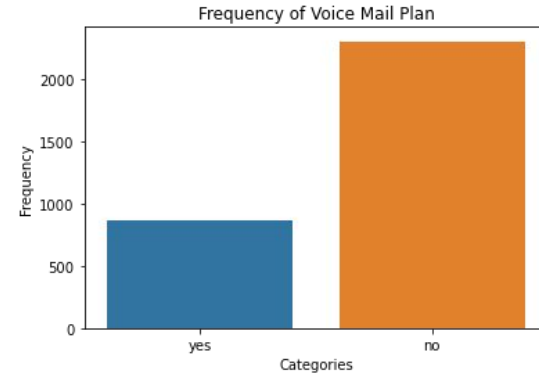
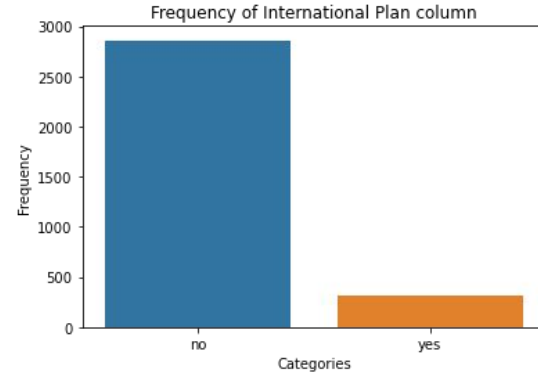


The graph shows the distribution of the churn and non-churn customers by state. The company should focus on the states with a high churn rate to mitigate the challenge.

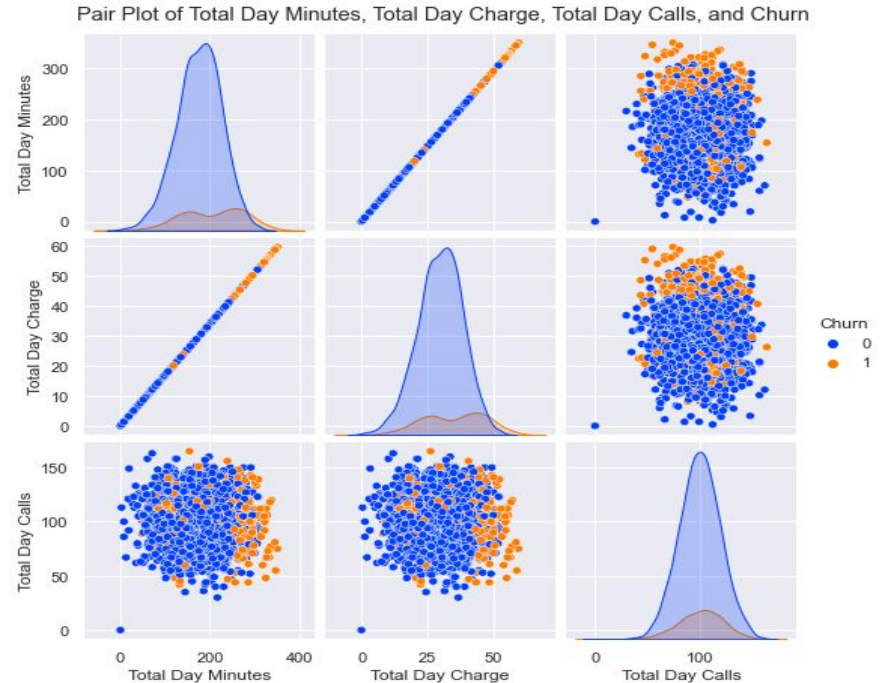


Both customers with or without the International Plan tend to churn but those with an International have a lower number of customers who churn

Similar to the above graph, the pattern is the same..

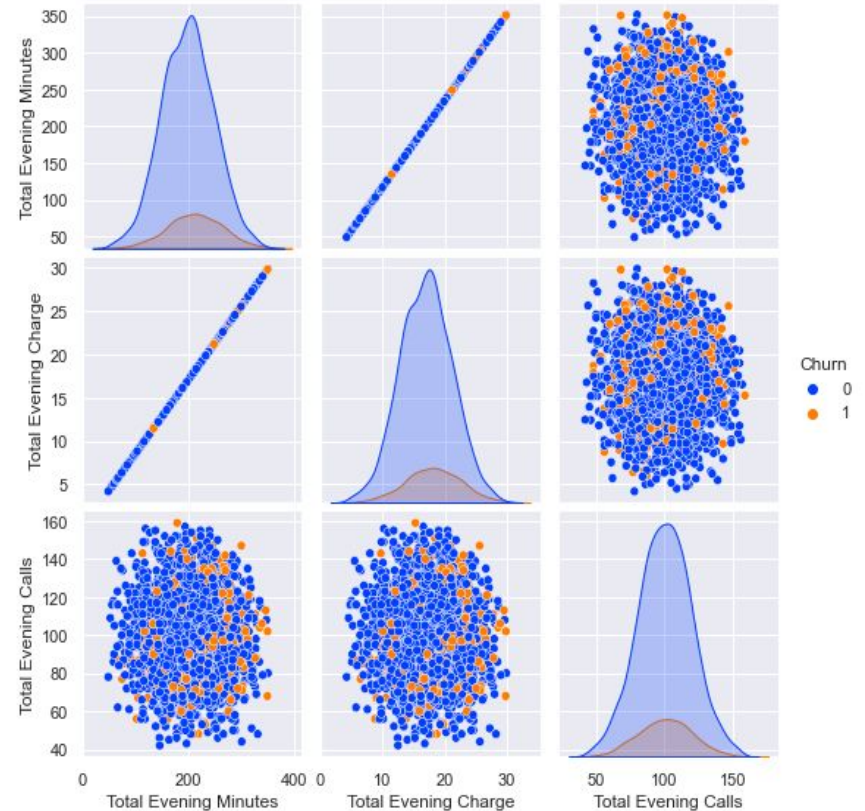


Here, we can see linear relationships between the different features. As one increases, the other increases. The cluster points saints show as the variables increase, the churn rate also increases.

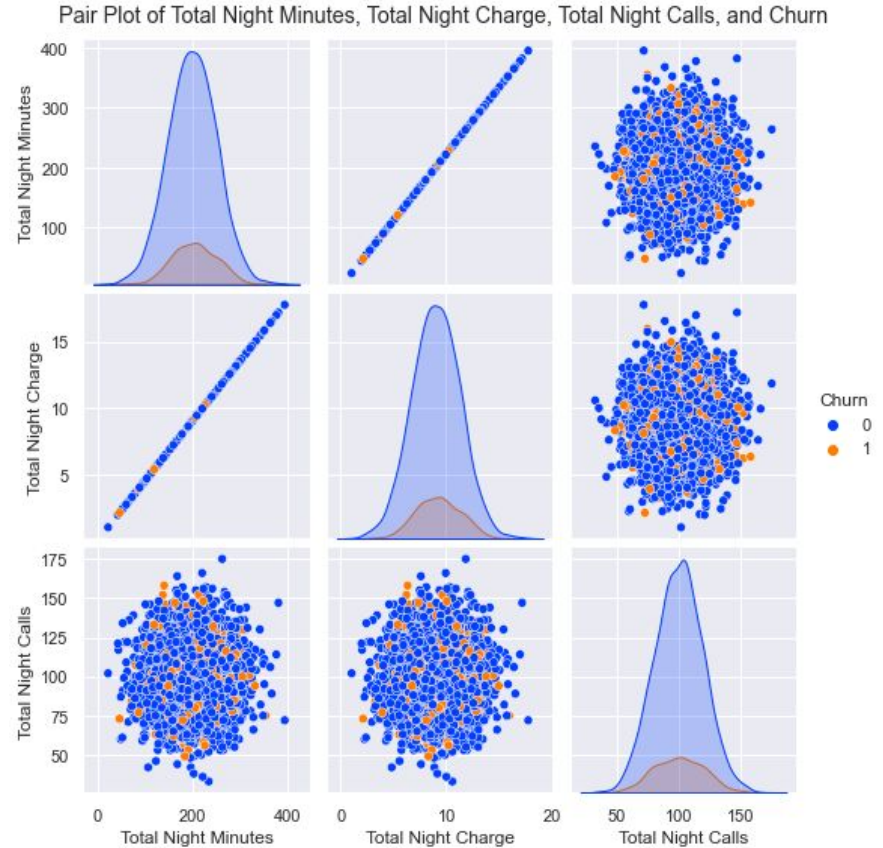


Similarly, there is also a linear relationship between the variables but the cluster points are spread unevenly.

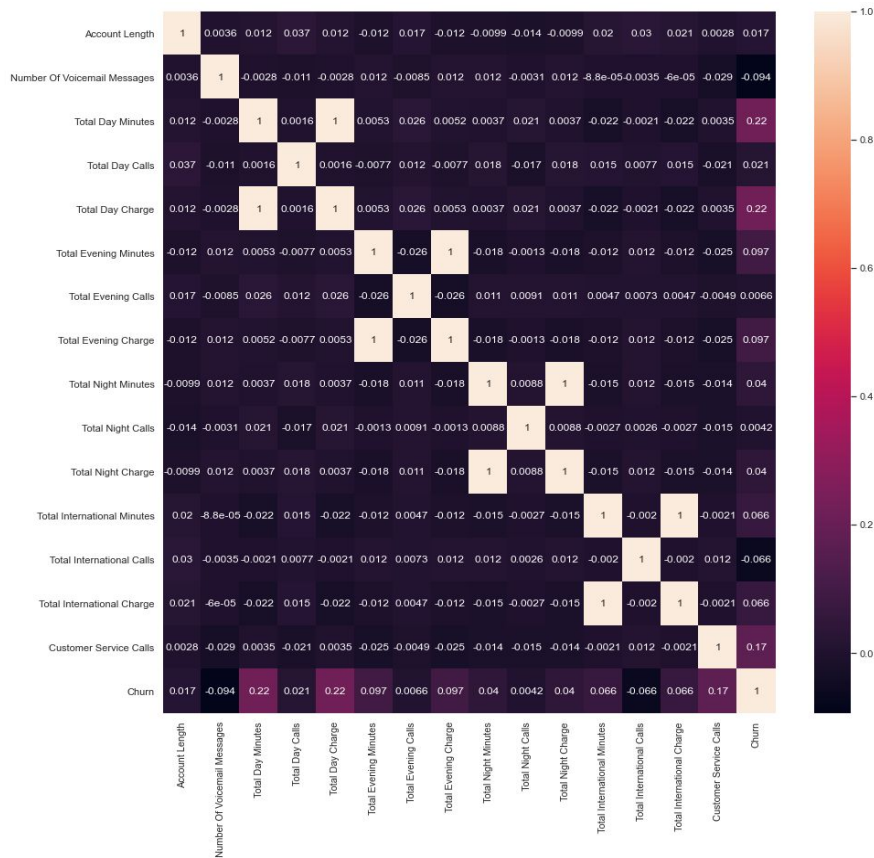
Pair Plot of Total Evening Minutes, Total Evening Charges, Total Evening Calls and Churn



There's a linear relationship, the cluster points are spread unevenly but are very few showing that the churn rate is not heavily impacted by the variables.



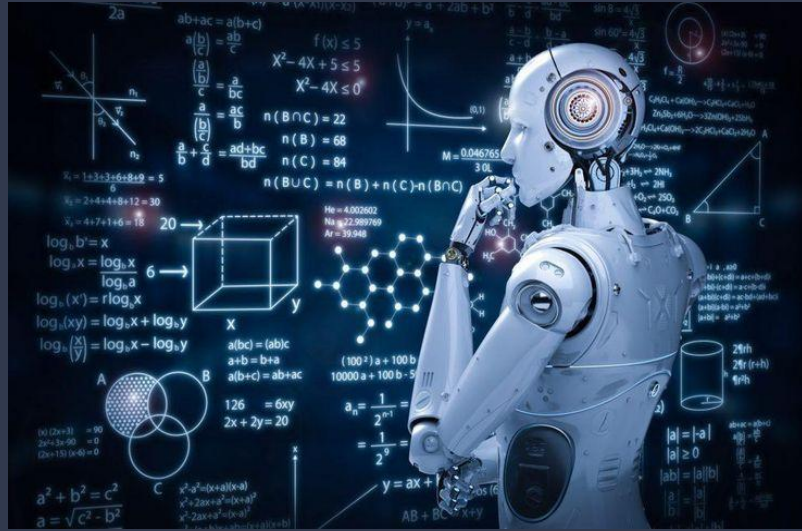
This shows how different features are correlated with each other. As our target variable is churn, these are the features that are highly correlated with Customer Service Calls, total day minutes and total day charge.



Model Evaluation.

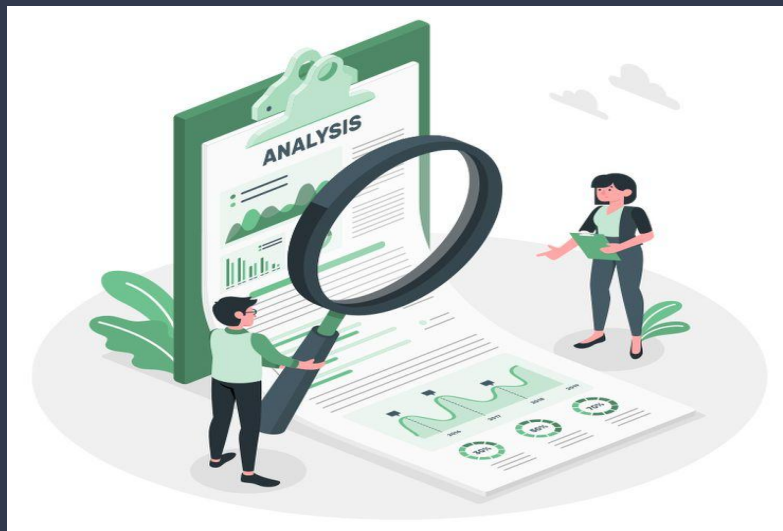
In machine learning, there are different supervised learning techniques. For this project, we used classification models i.e Logistic Regression model and K-Nearest Neighbours model to train data and predict unseen data.

There are some few classification metrics that we used to compare the different models. (Precision and Recall). Precision measures how often predictions for the positive class are correct. Recall measures how well the model finds all positive instances in the dataset. We used the ROC and AUC curves to evaluate the models which shall be explained.



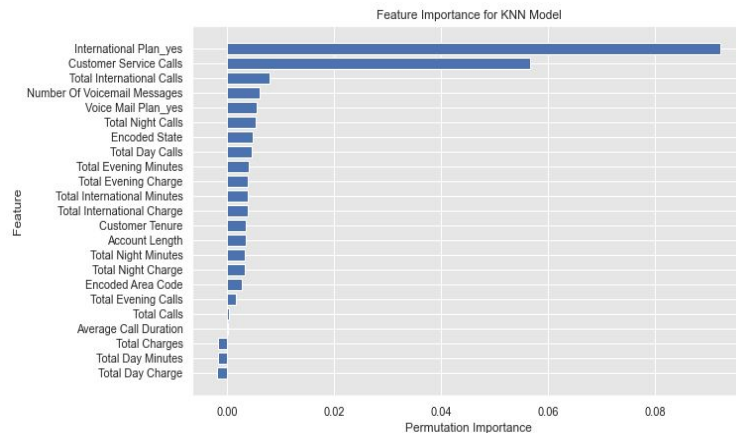
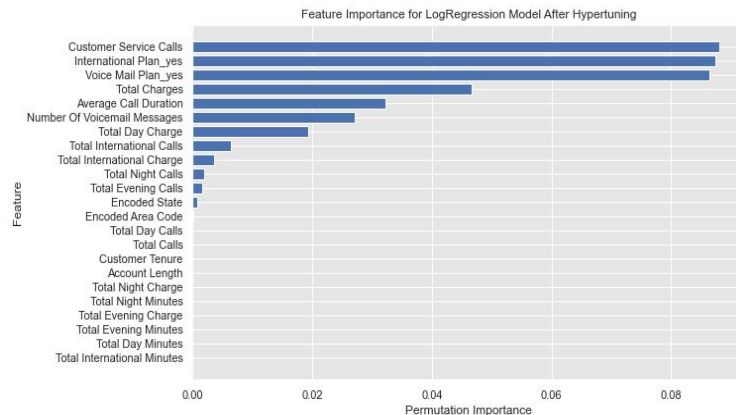
From the models, we that the baseline metrics had the highest classification scores compared to all the models. K-Nearest Neighbours had a precision score of 0.42 which means only 42% of the predicted values were classified correctly.

The Logistic Model after balancing the dataset had the highest recall score of 0.84, while after tweaking the parameters it had a recall score of 0.83 which is a very small difference. Using classification metrics, the Logistic Regression models are the best for predicting unseen data.



The ROC and AUC helps compare the quality of models that might output different ranges of predicted probabilities. From the results, the Logistic REgression model after hypertuning has an AUC of 0.8004 which is an shows it generalizes well on the unseen data.

We then performed feature importance to show which features should be considered in the feature to reduce the churn rate from both the logistic regression and KNN models. The top 4 were Customer Service Calls, International Plan, Voice Plan and Total International Calls.



Conclusions.

In conclusion, machine learning was chosen over simple data analysis due to its superior ability to handle complex data. The best models to predict if a customer is likely to churn are the Logistic Regression after balancing and after hypertuning. Reducing the churn will have a positive impact on the company's revenue and in addition customers will be more satisfied. It is very cost effective for the company to retain its customers.

Recommendations.



We would recommend for the company to develop strategies to retain customers like:

1. Improve the packages i.e improve on the Voice Mail Plan and International Plan.
2. Offer discounts to the customers who have a huge customer tenure in the company.
3. Moderate on the call charges and have customers fill in feedback forms to help the company improve their services.

∞ The end ∞