GROUP 7

PHASE 3 PROJECT 2023

Customer Loyalty

Project Overview

This project aims to develop a classification model that predicts customer churn in Syria Tel by uncovering patterns and factors associated with it.

The company can therefore take proactive measures based on these insights to retain customers and ensure sustained growth.

Business Understanding

Syriatel Telecommunications is a company dedicated to connecting people through seamless communication.

In their relentless pursuit of excellence, they have encountered a challenge of customer churn. Each lost connection not only signifying a departure but also posing a threat to the company's financial growth and future.

By understanding factors and patterns associated with it and developing a classification model that predicts customer churn effectively, Syriatel can take targeted actions to prevent churn and ensure business continuity.

Objective & Performance Metrics

The main aim of this project is to create a model that would accurately predict whether customer would churn out or not.

As such the project will focus on optimizing the following metrics:

1. Recall, Precision & F1_score

2. AUC score from the ROC curve.

The main objective is to optimize recall.

Data Understanding

Source

- Consumer Data from Syria Telecommunication's company

Variables

- The Dependent variable is a binary. 1: Churn. 0: Not Churn
- The independent variable are features such as talk time, call rate etc

Shape

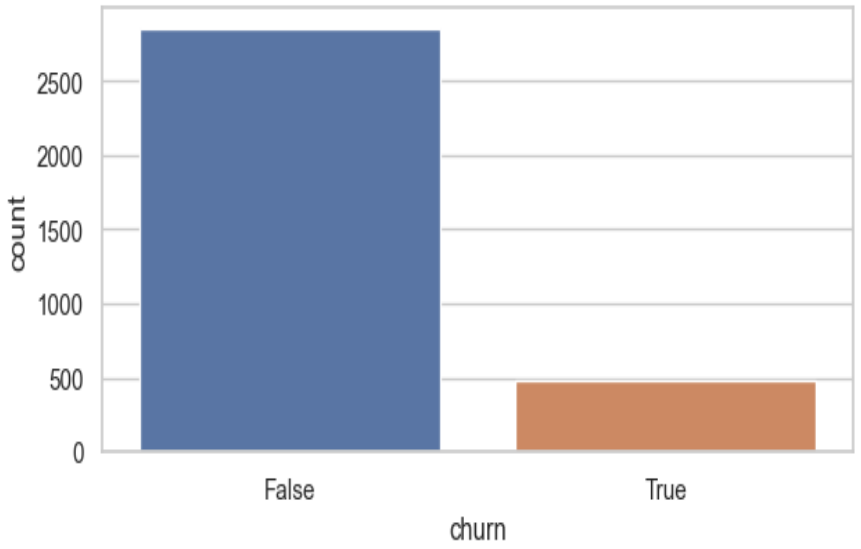
- The dataset contains 21 columns and 3333 rows. This means there are 21 different feature variables each with 3333 records.

Null & Duplicate values

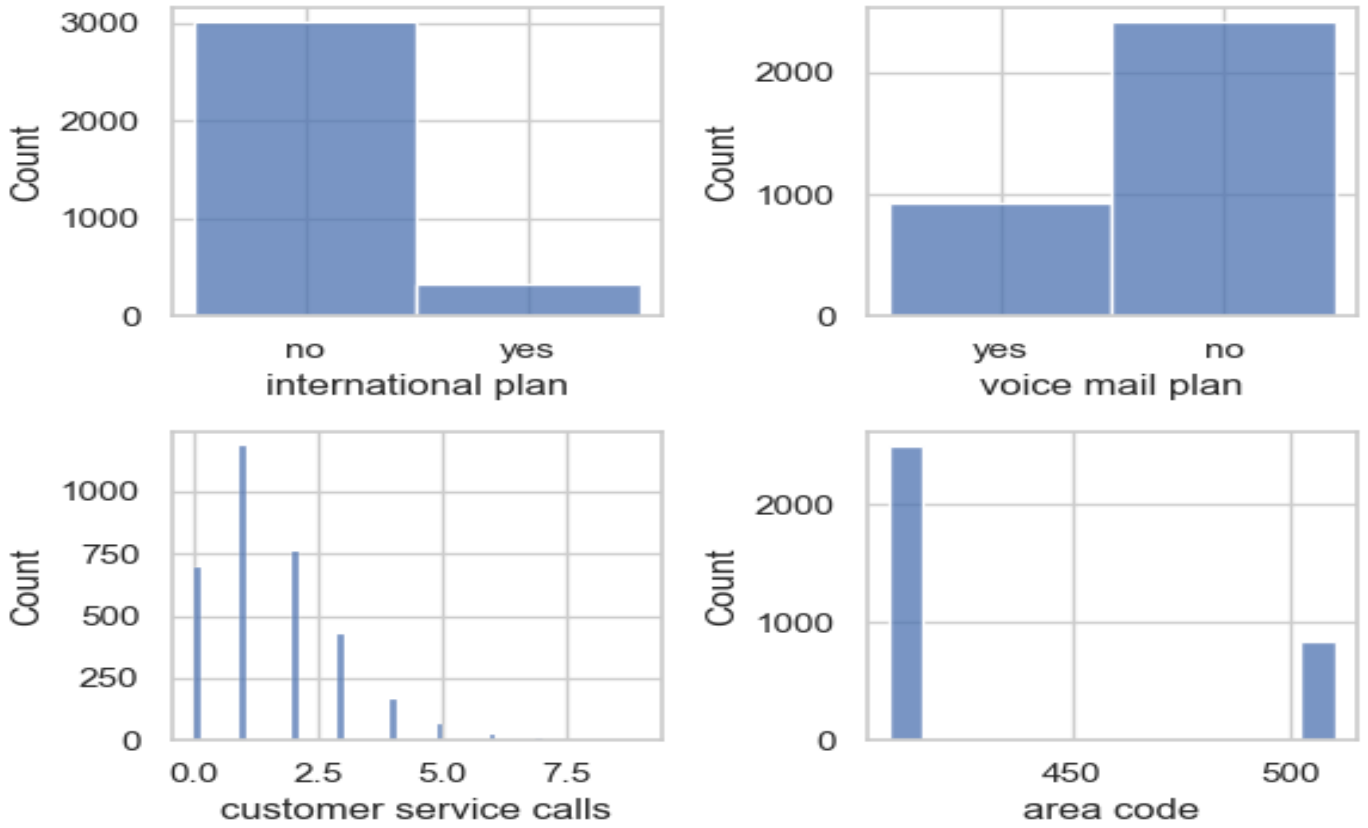
- There are no missing values.
- There are no duplicated rows in this dataset.

Explorative Data Analysis Univariate

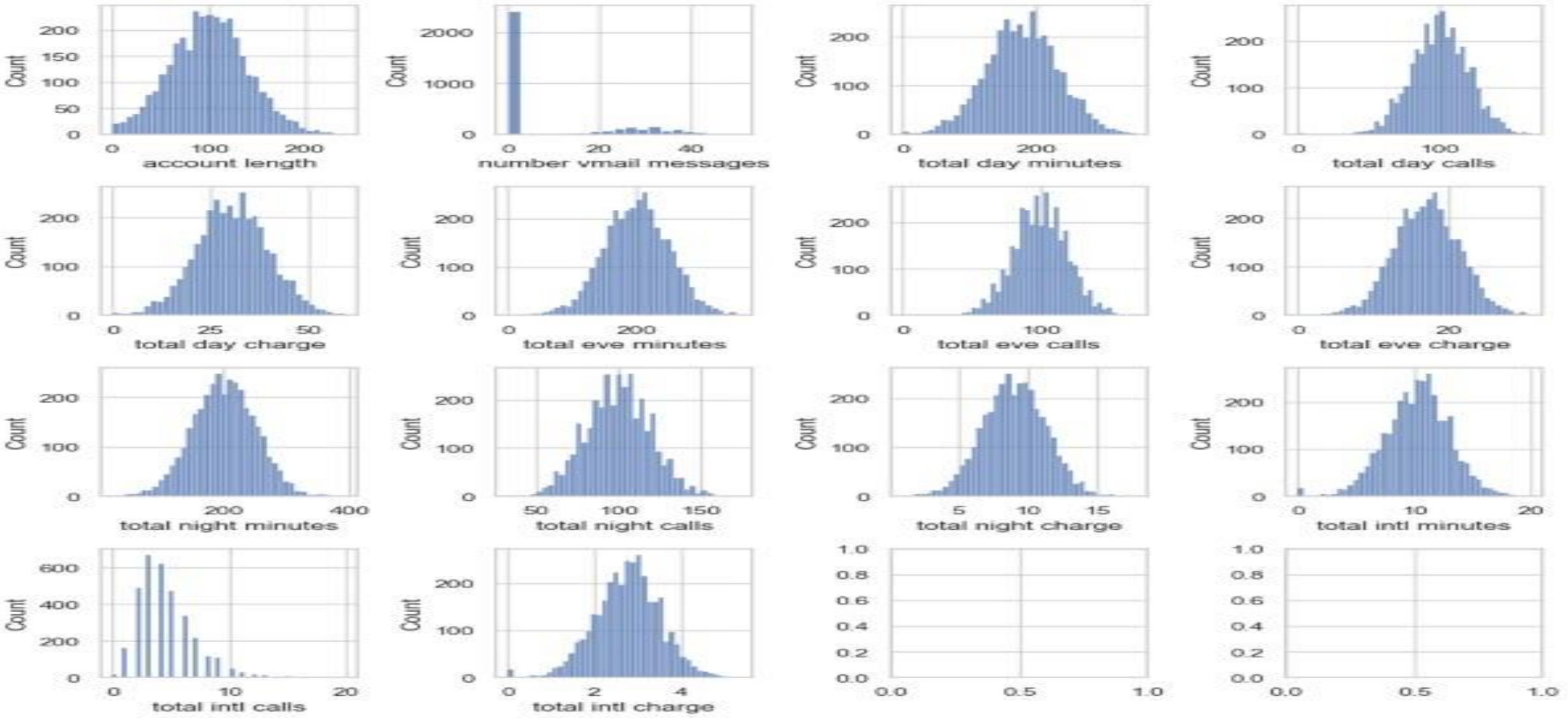
Dependent Variable Proportions



Categorial Independent Variable Distributions

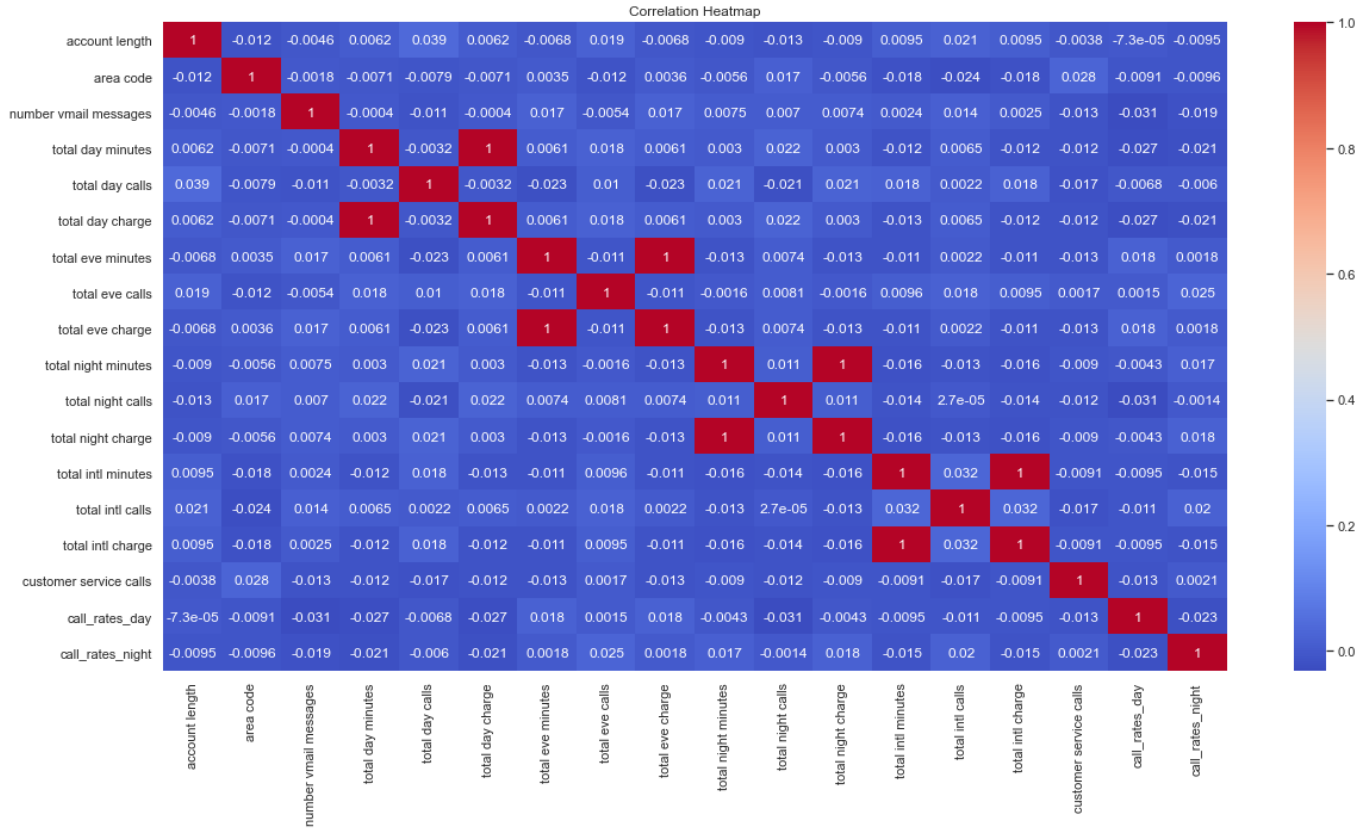


Explorative Data Analysis Univariate
Continuous Feature Variables: Histograms



EDA Multivariate

- From the Heat Map as shown, we have four pairs that have a very high correlation of 0.99 and above.
- In our model we dropped one of the features in each of the pairs



Baseline(Vanilla)Model

KNN Classifier

Accuracy = 0.90

Precision = 0.81

Recall = 0.317

F1_score = 0.45

Logistic Regression

Accuracy = 0.88

Precision = 0.57

Recall = 0.22

F1_score = 0.32

Decision Trees

Accuracy = 0.92

Precision = 0.67

Recall = 0.68

F1_score = 0.67

Hyper Parameter Tuned Models

KNN Classifier

Accuracy = 0.79

Precision = 0.35

Recall = 0.76

F1_score = 0.48

Best leaf size =1, p =1,
n_neighbors = 1

Logistic Regression

Accuracy = 0.79

Precision = 0.35

Recall = 0.76

F1_score = 0.47

Penalty =12, C =1 CW ={0:1, 1:5}

Decision Trees

Accuracy = 0.93

Precision = 0.67

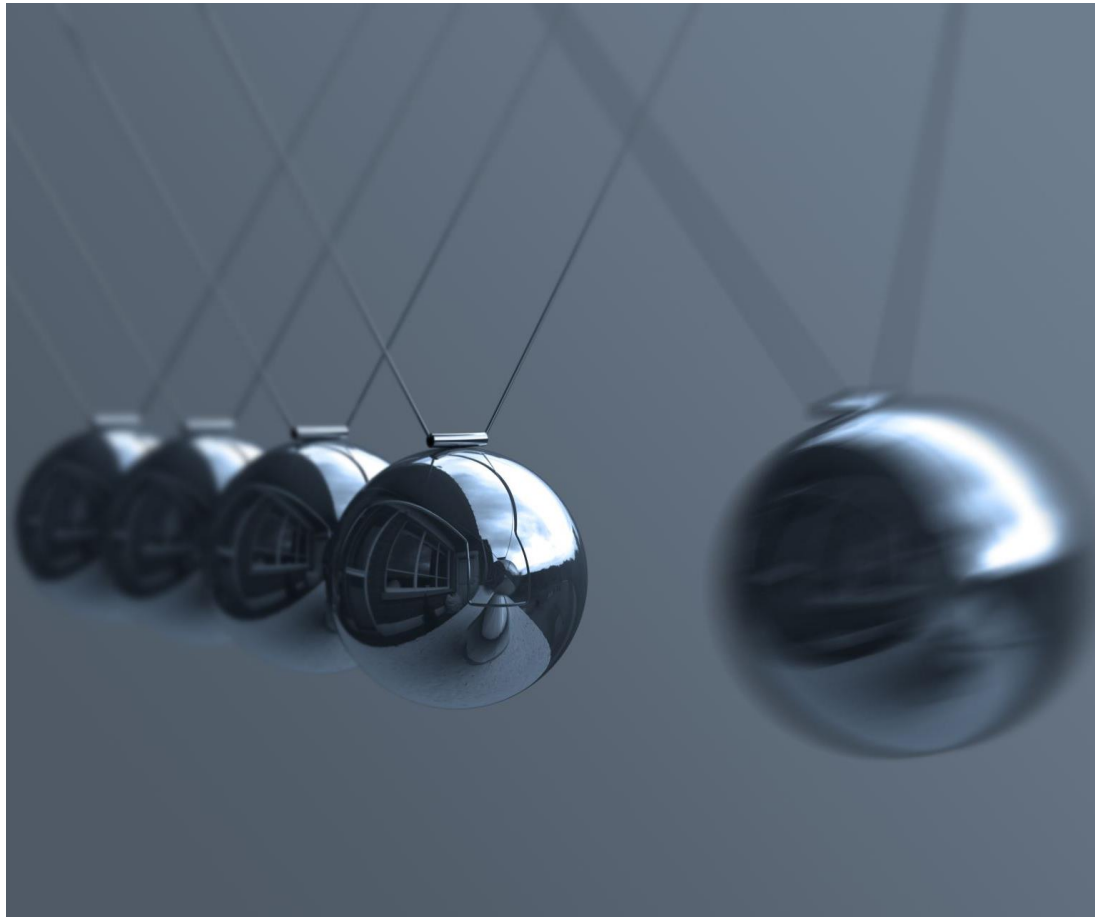
Recall = 0.8

F1_score = 0.73

Criterion = entropy,
max_depth =18,
Min_sample_split =2,
max_features=15,
min_samples_leaf =3

XGBoost Classifier: Accuracy: 0.967016491754123 Precision: 0.9090909090909091 Recall: 0.8235294117647058 F1-score: 0.8641975308641976

Comparison with ensemble methods



- **XG_BOOST**

XGBOOST CLASSIFIER

Accuracy = 0.97

Precision = 0.90

Recall = 0.82

F1_score = 0..86

Final Model: XGBOOST

For the Tuned Boost Classifier :

The recall score for the Tuned Boost Classifier at 0.8235 is the best one yet and aligns with our project objective, which is to identify as many churned customers as possible to gain insights into the reasons for churn. A high recall score of 0.8235 ensures that a 82.35& of actual churned customers are correctly identified by the model.

A high accuracy of, indicates that the model is making correct predictions with a high rate of success. while a precision score of 0.9091, which is very high and suggests that the model is effective in identifying customers who are likely to churn, with a high proportion of true positives.

An F1-score of 0.8642, shows a great balance between precision and recall.

Interpretation of the features

Patterns and Factors Associated with Customer Churn:

International Plan: Customers with international plans showed higher churn rates, suggesting that this feature is an essential factor in predicting churn.

Customers without voicemail plans had higher churn rates, indicating that the presence of voicemail plans could contribute to customer retention.

Customers with a high number of customer service calls were more likely to churn. This suggests that unsatisfactory customer service experiences could lead to churn.

Total Day Minutes and Total Day Charge: Churned customers tended to have higher total day minutes and charges.

Recommendations

-Utilize the improved XGBoost classifier as the primary model for customer churn prediction. It demonstrated the highest performance in accurately identifying churned customers.

-SyriaTel can focus on promoting and enhancing voicemail plan offerings to reduce churn.

-Improving customer service quality and addressing issues promptly can help reduce churn.

-Understanding the correlation between call usage and churn can help SyriaTel develop personalized plans and offers to keep customers engaged.

Way forward

-Develop customer retention strategies as recommended

-Segment customers according to their unique characteristics.

-Improve service quality to minimize rate of churn

-Establish a customer Feedback loop for sentiment analysis

-Regularly assess the effectiveness of implemented strategies and initiatives by tracking key metrics and comparing them against established benchmarks.

Contributed By:

Antony Kanai- TL

Joyce Muthike- ATL

Bryan Okwach

Sheila Kamanzi

Bahati Ndwiga

Lee Kimaita

Questions