



# **Phase 3 Project – SyriaTel customer churn prediction**

---

By Collins Edwin Opiyo Ouko

# Overview



- For Telco companies it is key to attract new customers and at the same time avoid contract terminations (=churn) to grow their revenue generating base.
- **Churn Metric:** Percentage of customers canceling contracts/subscriptions within a specific period (typically monthly)
- This project is geared towards predicting and reducing customer churn for SyriaTel by analyzing customer behavior and applying machine learning models to identify high-risk customers

# Specific Objectives

1. Identify the factors that are most likely to lead to customer churn.
2. Develop a model that can accurately predict which customers are at risk of churning.
3. Take proactive steps to retain customers who are at risk of churning.

# PROJECT WORKFLOW

Insert Your Creative Idea

---



Business  
Understanding



Data  
Understanding



Modeling



Evaluation



Conclusion

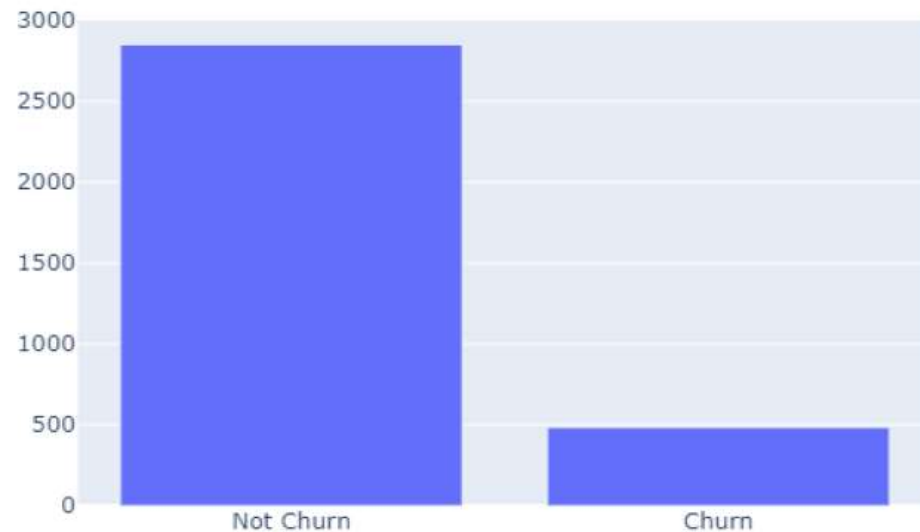


Recommendation

# DATA ANALYSIS

## How many customers have churned?

Churn Distribution



Total number of customers;  
**3,333**

Number of customers that churned;  
**483**

Churn Percentage;  
**14.5%**

# DATA UNDERSTANDING

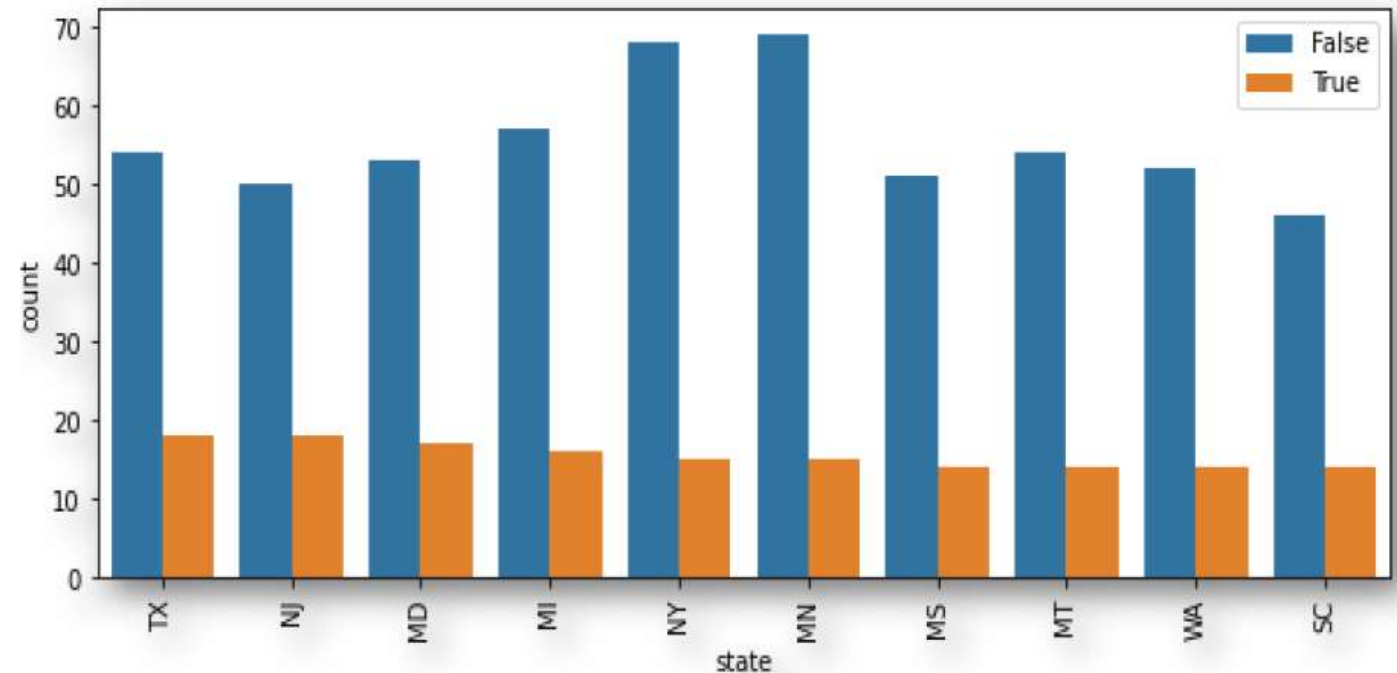


We worked with customer churn dataset provided as one of the project resources

It contains 3,333 records and 21 columns(5 categorical, 16 numerical).

# Which states have the highest churn rate?

Texas and New Jersey are leading followed by Maryland, Miami and New York.



# MODELLING



## Algorithms Used

- Logistic Regression
- Decision Tree

## Metrics Used

- ❖ Accuracy focuses on evaluating the overall performance of a classification model. It measures the proportion of correctly predicted instances
- ❖ Recall-focuses on the model's ability to correctly identify positive cases.



# EVALUATION

***Based on summary results table;***

Logistic Regression –0.869

Decision Tree –0.846

For accuracy

	accuracy	recall
classifiers		
LogisticRegression	0.869	0.258741
DecisionTreeClassifier	0.846	0.286713

# CONCLUSION

## **Models Evaluated:**

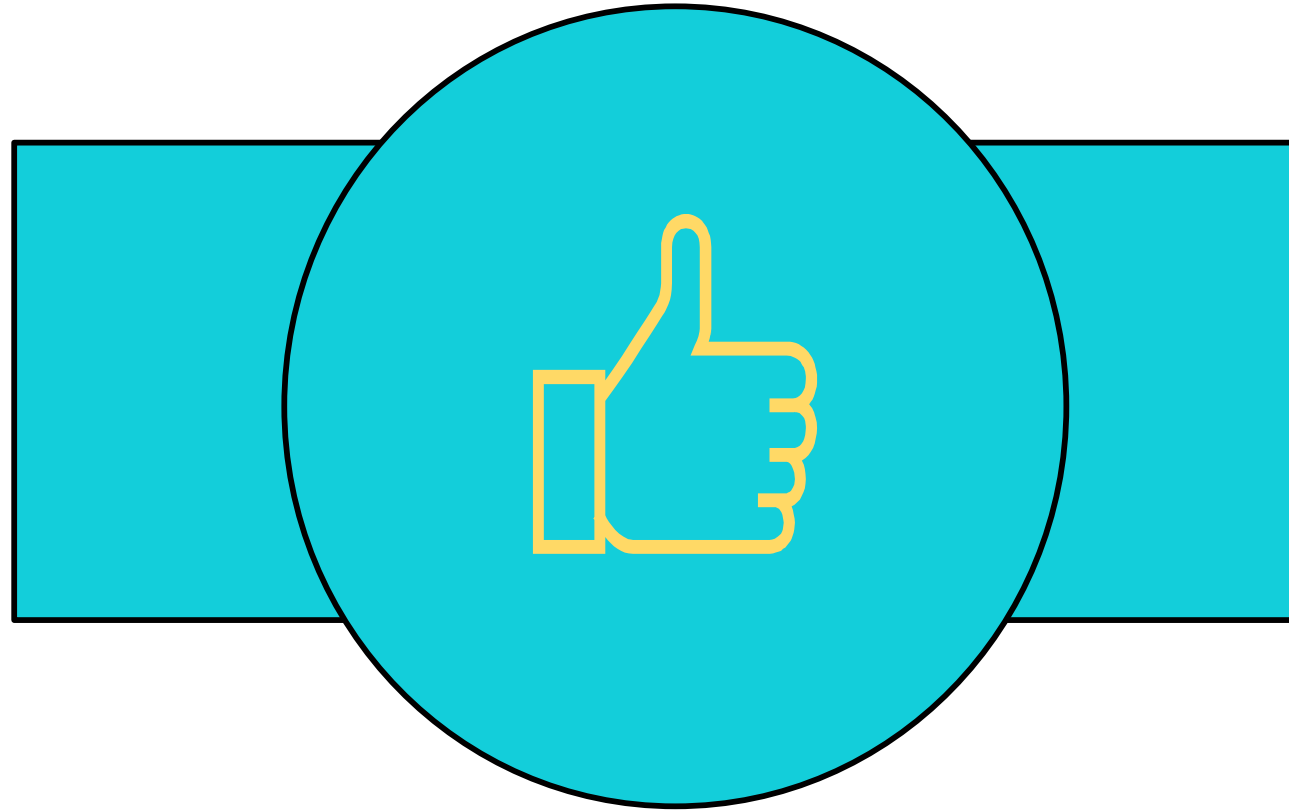
- Logistic Regression: Accuracy = 0.869, Recall = 0.259
- Decision Tree: Accuracy = 0.846, Recall = 0.287

## **Recommendation:**

- Opt for Logistic Regression due to its higher accuracy score.

## **Achievement:**

- Successfully predicted customer churn with a high accuracy score.



# Thank You