CUSTOMER CHURN PREDICTION

By: Mathews odongo

Project Goal

This project aims to predict which customers are likely to leave using historical data.
By identifying these churn patterns early, Syriatel can take proactive steps to:

- Increase retention
- Enhance customer satisfaction
- Strengthen competitive advantage

BUSINESS & DATA UNDERSTANDING

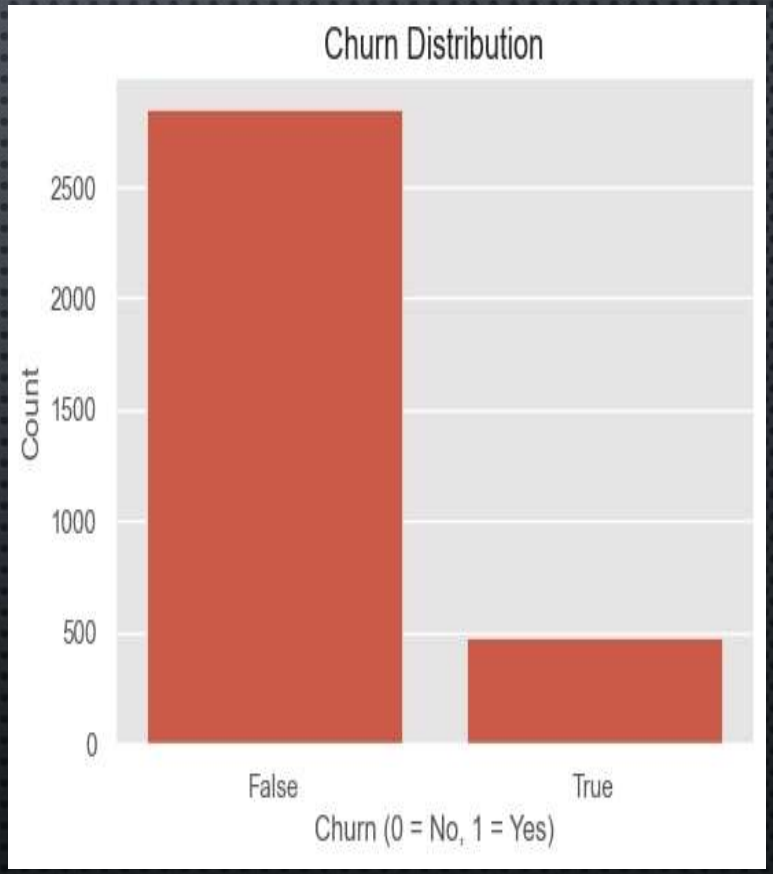
Dataset Overview

•: SyriaTel Telecom Dataset

•**Records**: 3,333 customers

•**Features**: 20 (numeric & categorical)

•**Target Variable**: Churn (Yes/No)



Data Preparation

- **What We Did:**
- Handled **missing values**
- Removed **duplicates** and **outliers**
- **Encoded** categorical variables
- Applied **SMOTE** to balance churn classes
- **Scaled** numerical features for modeling

MODELS TRAINED

- **Logistic Regression**
- Acts as the **baseline model**
- Focuses on **interpretability** and **recall**

- **Random Forest**
- Ensemble of decision trees
- Balances **precision and recall**
- Handles **non-linearity** well

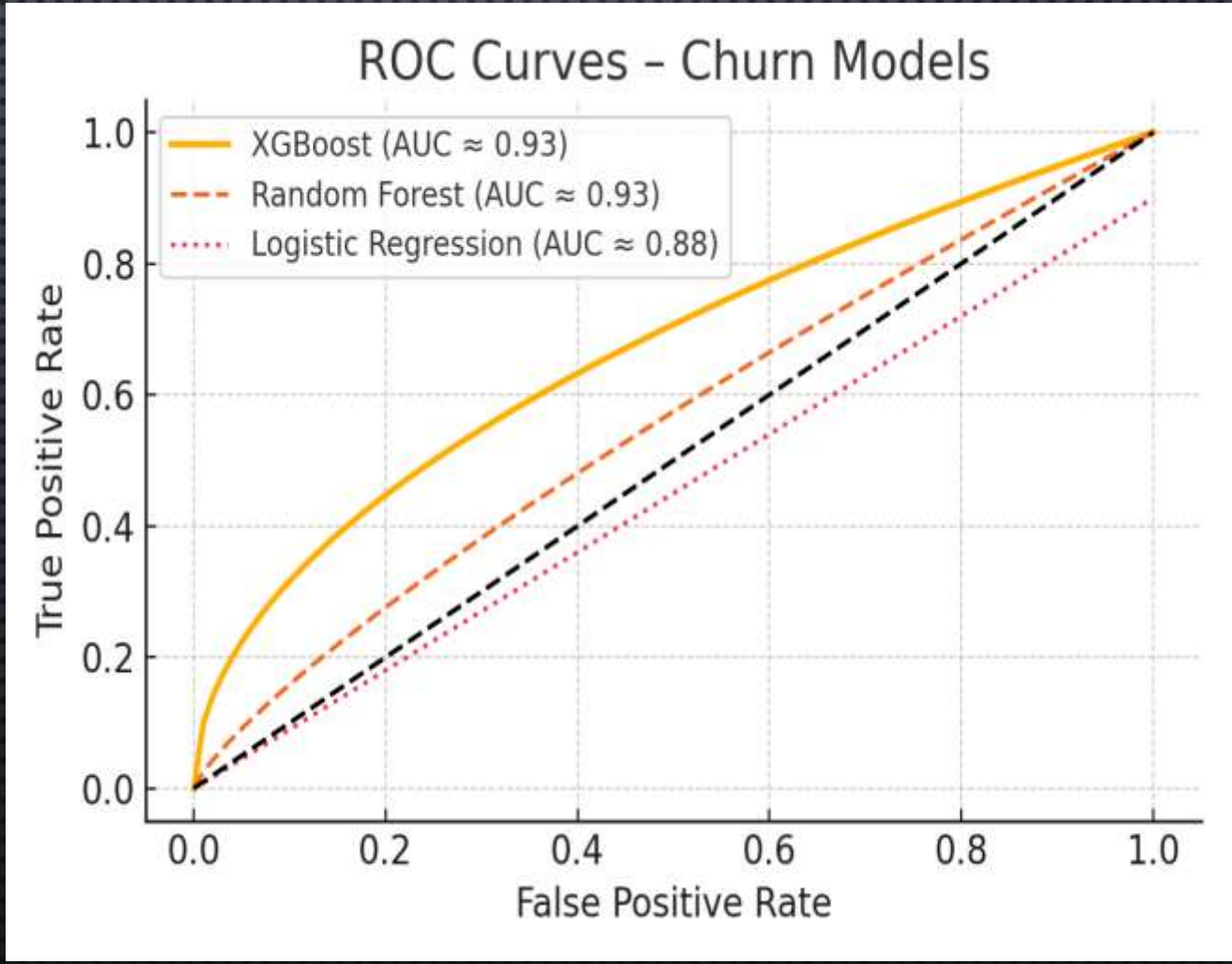
- **XGBoost**
- Advanced boosting technique
- Offers **highest accuracy** and **best ROC-AUC**
- Excels at **detecting complex churn patterns**

PERFORMANCE METRICS (KEY MODELS)

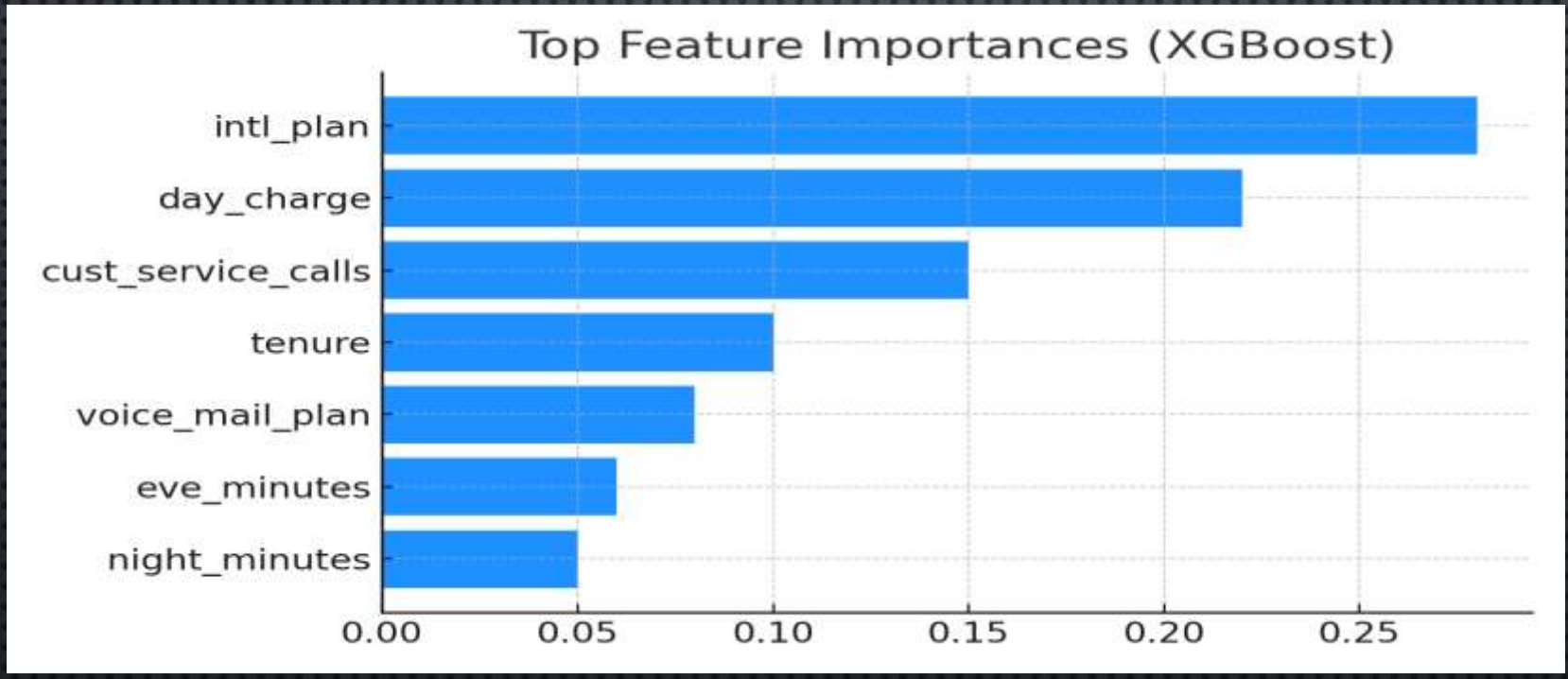
	Logistic Regression	Random Forest	XGBoost (Best)
Accuracy	~85%	~90%	**94%**
ROC-AUC	0.88	0.91	**0.93**
Precision (Churn)	0.51	**0.81**	0.77
Recall (Churn)	**0.86**	0.73	0.85
F1-Score (Churn)	0.64	0.77	**0.81**

MODEL COMPARISON

- XGBoost: F1 = 0.81, AUC = 0.93 (Best)
- Random Forest: Balanced and interpretable
- Logistic Regression: High recall, low precision



FEATURE IMPORTANCES (XGBOOST)



- Top drivers: Intl. Plan, Day Charges, Support Calls
- Also important: Tenure, Area Code, Voicemail Plan

RECOMMENDATIONS

- Based on model evaluation, **XGBoost** emerged as the top-performing model. Its ability to accurately identify customers most likely to churn — while minimizing false positives — makes it highly effective for **targeted retention strategies**. It ensures resources are focused on customers who truly need intervention, maximizing ROI on loyalty campaigns.

- **Random Forest** also delivered excellent performance with a strong balance between recall and precision. It's a reliable alternative, especially when aiming for a **broad churn prevention strategy** that captures both at-risk and potentially dissatisfied customers. It works well even with class imbalance and is more interpretable than XGBoost.

- While **Logistic Regression** offers speed and interpretability, its lower precision resulted in many false positives. It may still be useful for **quick insights or early detection**, but it's not ideal for guiding business-critical retention decisions on its own.

Conclusion

- Predicting churn isn't just about accuracy—it's about **retaining revenue** and **protecting customer relationships**.
- Our models, especially **Random Forest** and **XGBoost**, empower SyriaTel to act **before customers leave**, driving smarter decisions and long-term growth.

NEXT STEPS

- **Deploy** the XGBoost model to production systems
- **Monitor** model performance and update with fresh data
- **Collaborate** with customer service teams to act on churn predictions
- **Explore** further segmentation to tailor retention strategies
- **Evaluate** potential ROI of targeted interventions

THANK YOU

EMAIL: WANDERA59@GMAIL.COM

GITHUB: MATHEWS-ODONGO