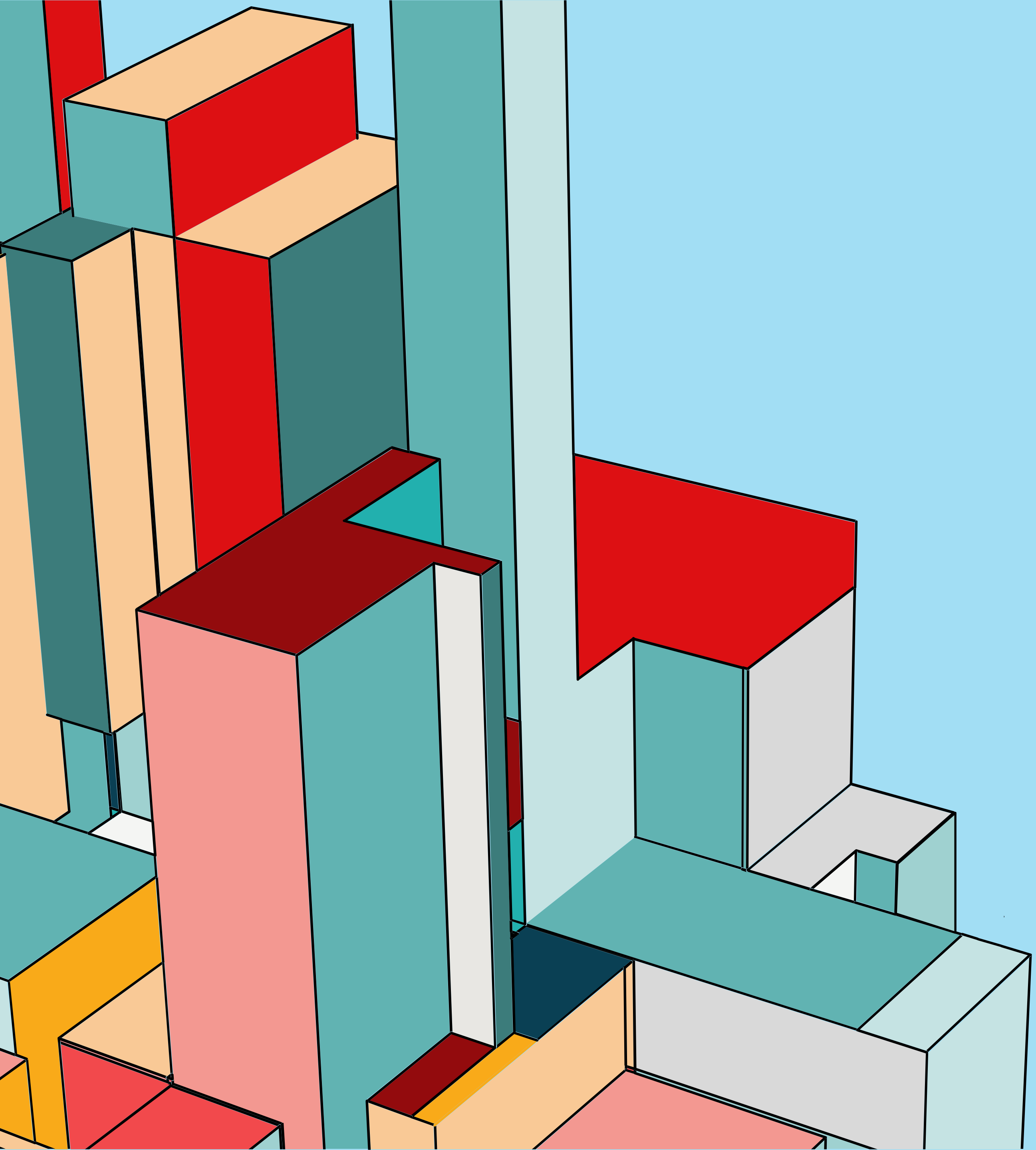


CUSTOMER CHURN ANALYSIS PROJECT

PREDICTING CUSTOMER CHURN FOR SYRIA TELECOM

NAME: BENSON MWIHIA

PHASE III DATA SCIENCE PROJECT

MORINGA SCHOOL

PROBLEM STATEMENT

What is churn?

The percentage of customers who stop using company's service during a certain time frame.

Why customer churn matters:

- Losing existing customers is costly
- Telecom companies face huge churn rates
- Understanding why customers leave can help retention strategies

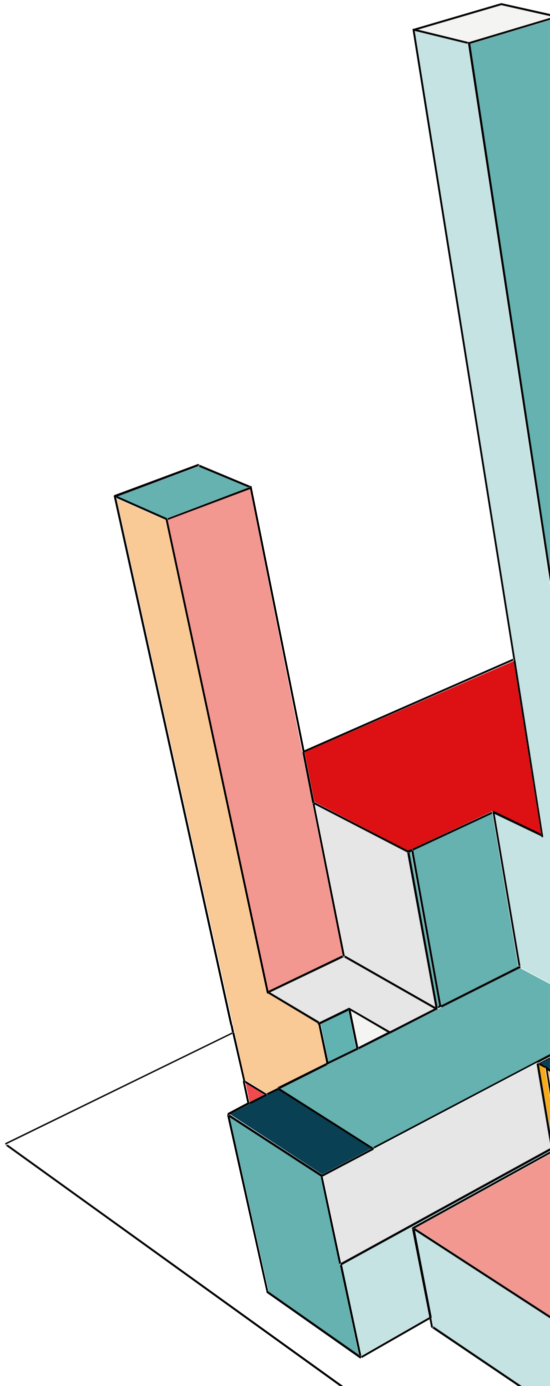
Goal:

- Predict which customers are likely ro churn and understand the drivers behind their decisions.

PROBLEM OBJECTIVES

Our Goals:

- ❖Perform data preprocessing and EDA
- ❖Explore customer data to understand churn behaviour
- ❖Build models to predict if a customer will churn
- ❖Compare models based on precision, recall and F1-Score
- ❖Tune the best model using GridSearchCV
- ❖Deliver insights & recommendations for the business



DATASET OVERVIEW

Data Description:

- 3,333 customers
- 20+ features: Usage, Service plans, demographics, etc.
- Target: Churn (True / False)

Feature Categories:

Feature Type	Examples
Demographics	State, Area Code, phone number
Services	International Plan, Voicemail Plan
Usage Patterns	Total minutes, charges (Day / Evening / Night / International
Customer Behaviour	Number of service calls

DATA PREPARATION

Cleaning and Preprocessing:

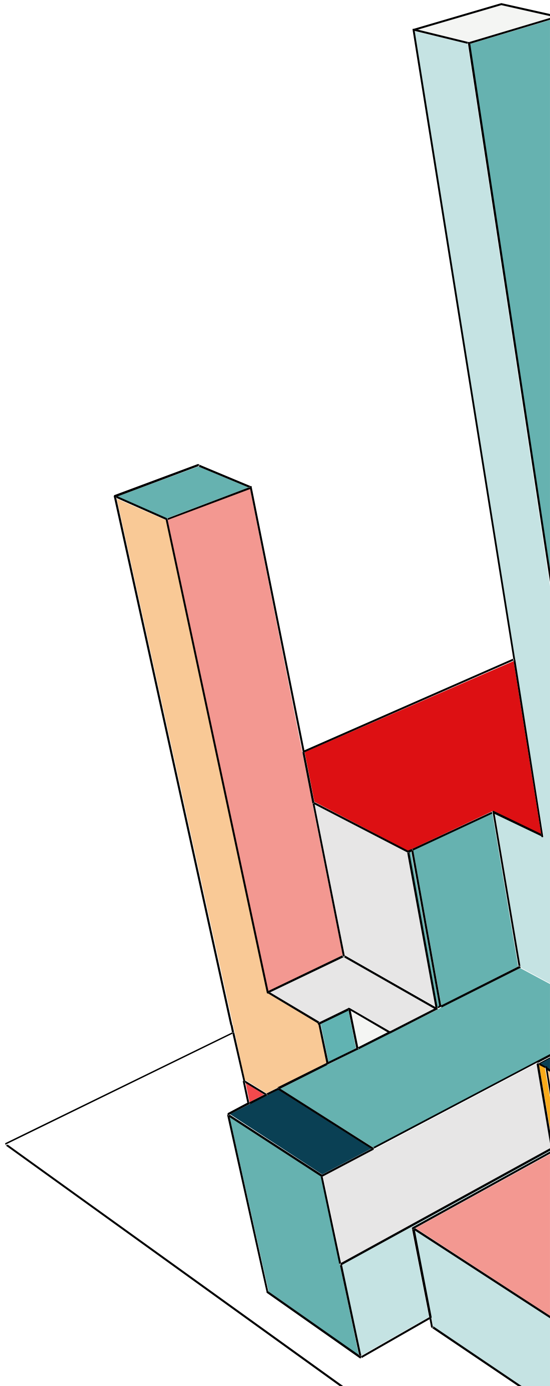
❖Dropped irrelevant fields (e.g., Phone numbers)

❖Encoded categorical variables using OneHotEncoding

❖Scaled numerical features with StandardScaler

❖Split into Train / Test sets (Stratified)

❖Class imbalance noted (Churn rate – 14.5%)



EXPLORATORY DATA ANALYSIS (EDA)

Key Visuals:

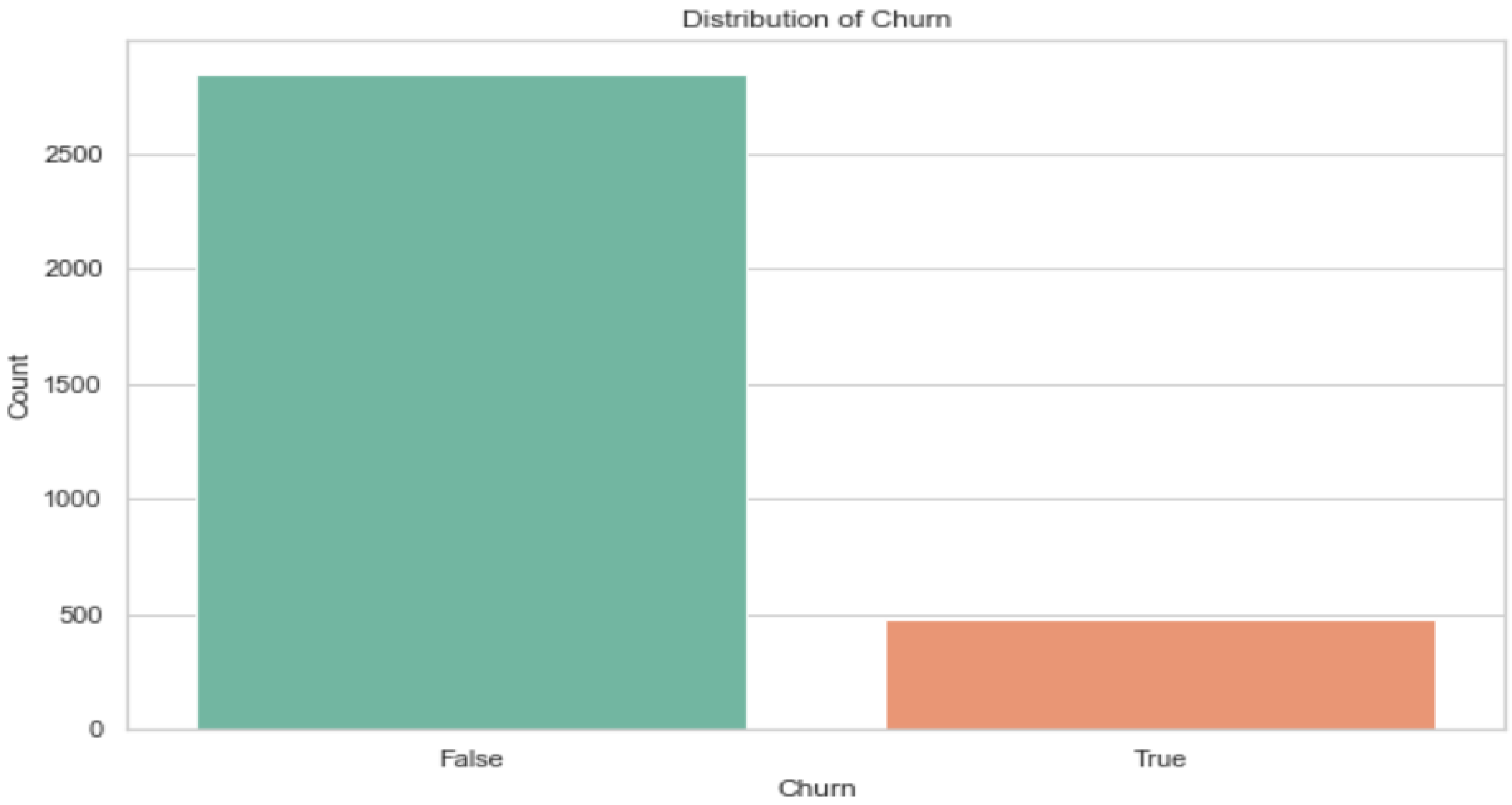
- Churn distribution plot
- Heatmap of feature correlations
- KDE plots for usage patters
- Countplots for categorical features

Key Insights:

- Customers with international plans churn more
- Churn increases with more than 3 customer service calls
- Usage merics (day / night calls) have a weak correlation with churn

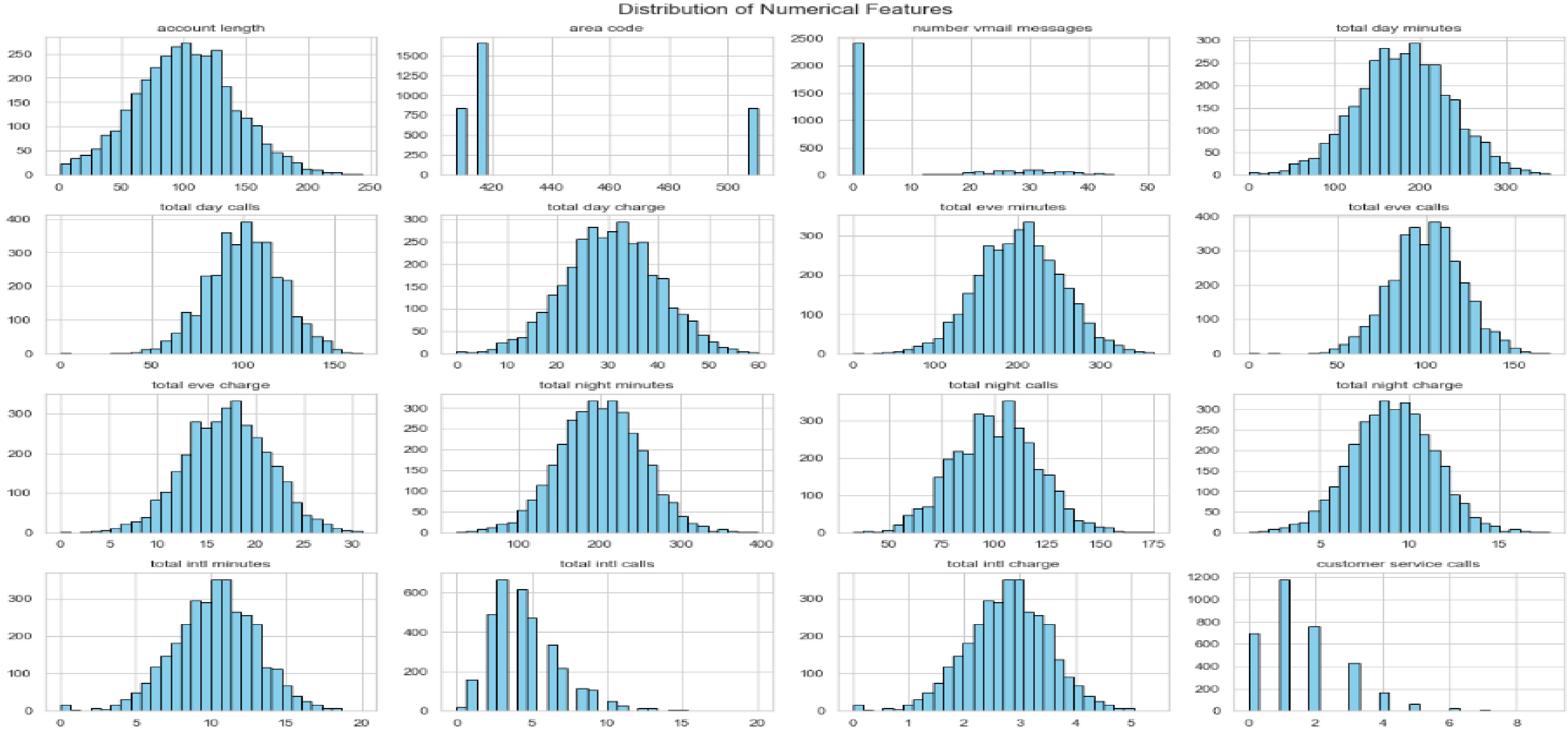
(Visuals) – Next slides

EDA – DISTRIBUTION OF CHURN



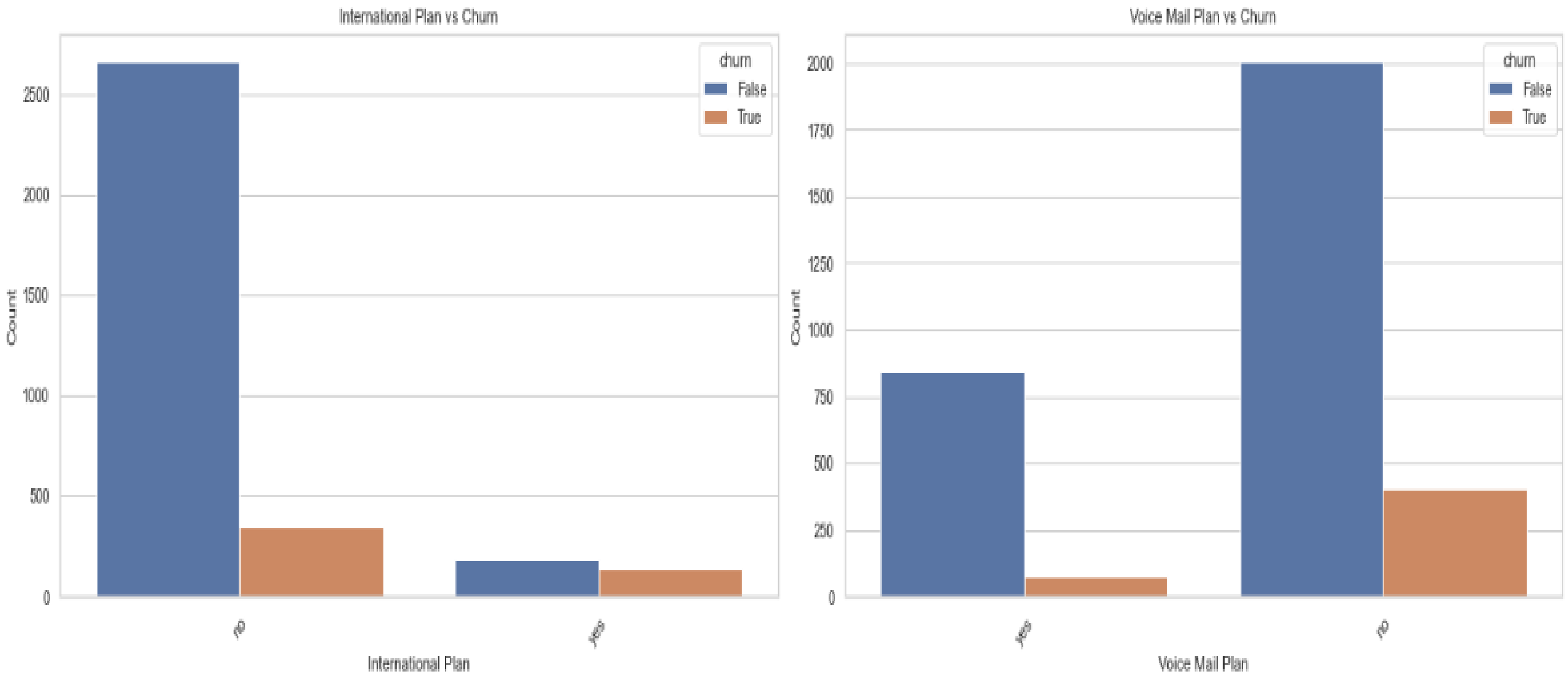
❖ Dataset is imbalanced with majority of customers not churning i.e., False – 86% and True – 14%.

EDA – NUMERICAL FEATURE DISTRIBUTIONS



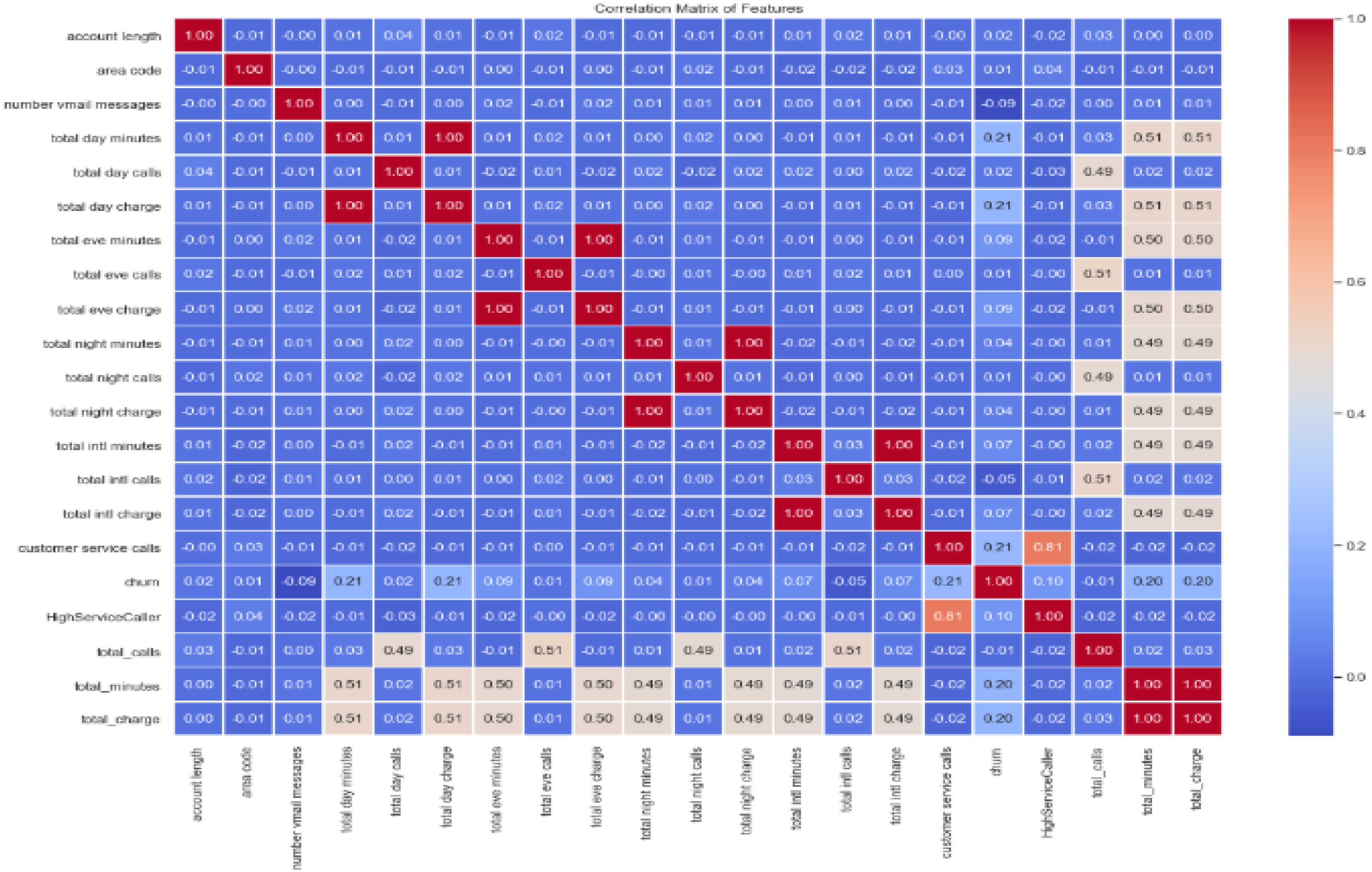
- ❖ Some features e.g. today day minutes, total intl charge) appear normally distributed.
- ❖ Number vmail messages and customer services calls may show skewness or zero inflation.

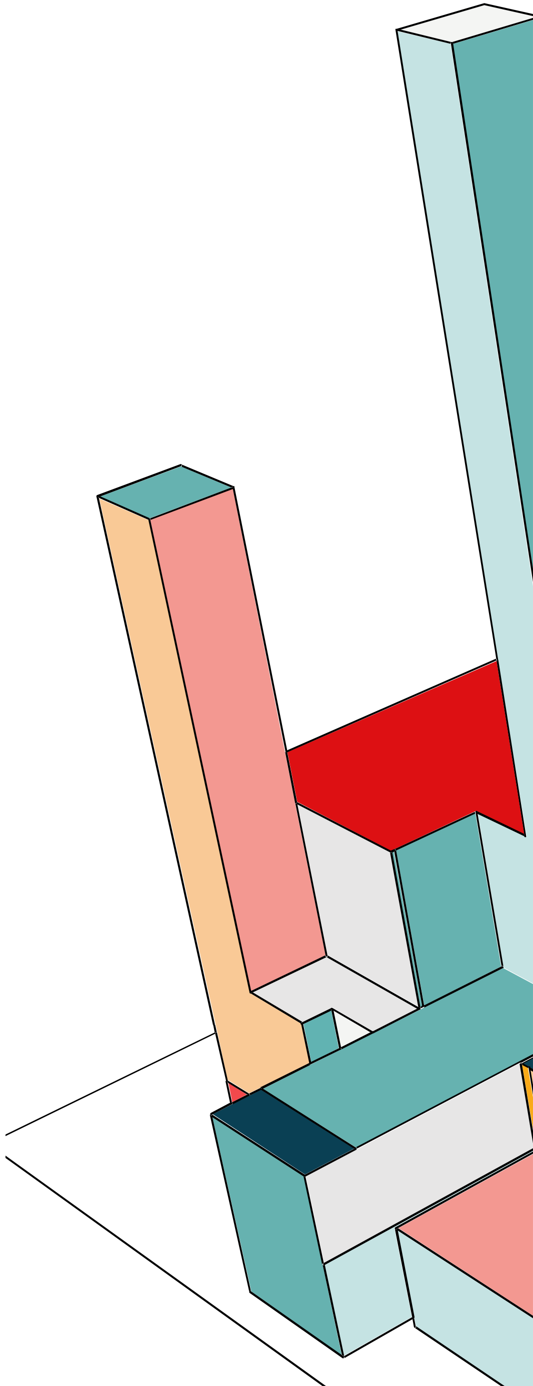
EDA – CATEGORICAL FEATURE DISTRIBUTION



❖ Visual of how each categorical feature is distributed and how it relates to the churn variable.

EDA – CORRELATION HEAT MAPS





MODELLING APROACH

Machine Learning Models Used:

❖Logistic Regression

❖Decision Tree

❖Random Forest

❖XGBoost classifier

Evaluation Metrics:

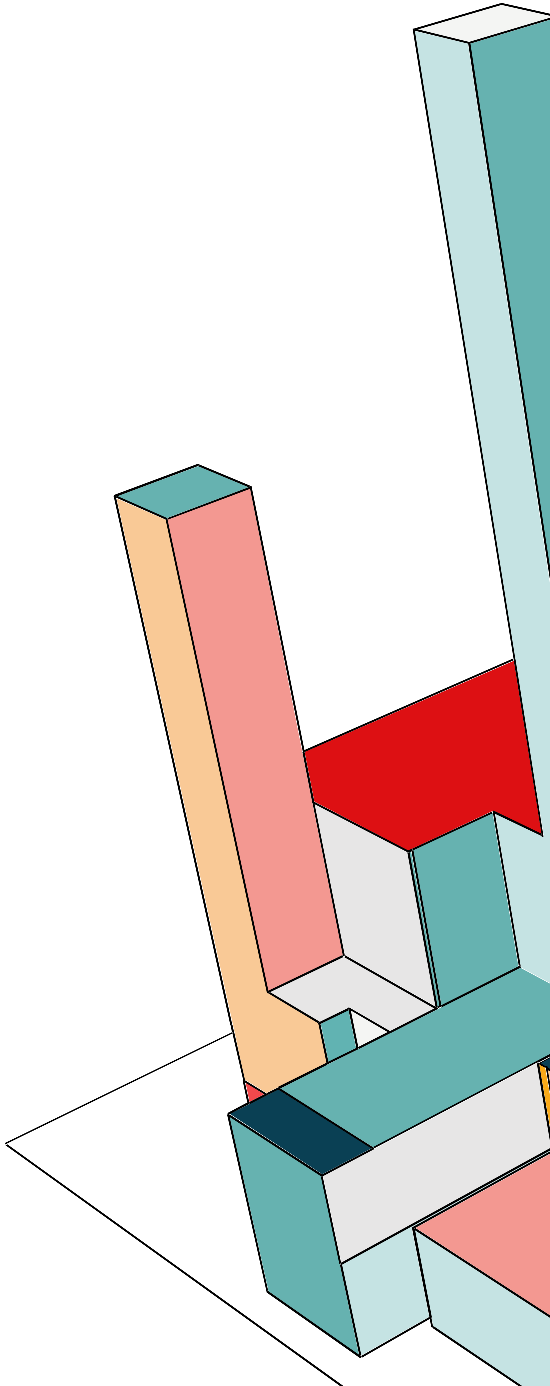
❖Accuracy

❖Precision

❖Recall

❖F1 – Score

❖Confusion Matrix



MODEL PERFORMANCE COMPARISON

Model Evaluation Results:

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.86	0.53	0.28	0.36
Decision Tree	0.93	0.82	0.63	0.71
Random Forest	0.91	0.95	0.38	0.54
XGBoost	0.95	0.90	0.72	0.80

XGBoost outperformed all other models on Recall and F1-score – critical for churn detection.

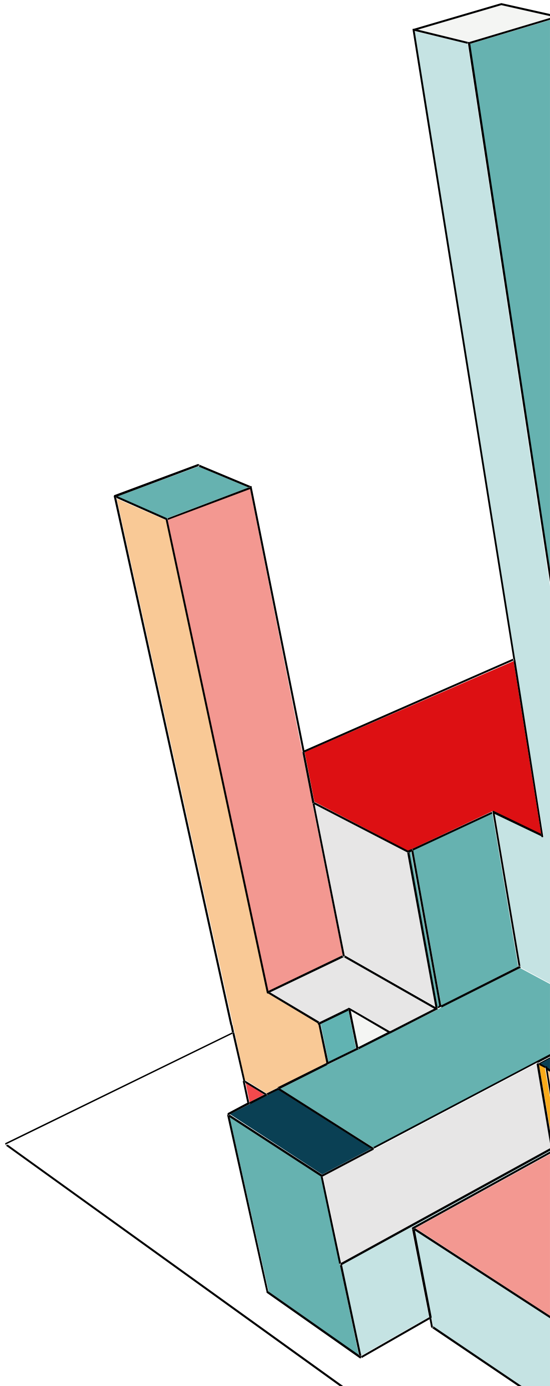
HYPERPARAMETER TUNING

Improving Random Forest with GridSearchCV

Best Parameters:

- n_estimators: 100, 200
- Max_depth: None, 10, 20
- Min_samples_split:2, 5
- Min_samples_leaf:1, 2

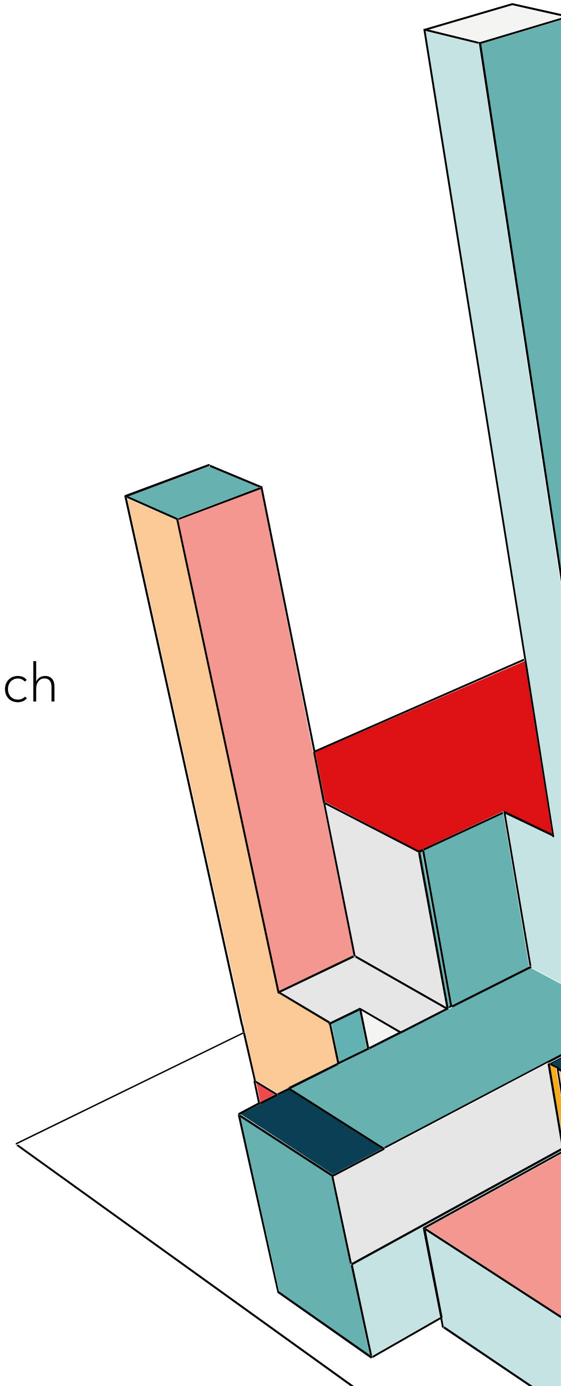
Despite tuning, XGBoost remained the top model.



KEY INSIGHTS

BUSINESS TAKEAWAYS

- ❖Customers with international plans are high risk – target them with loyalty incentives.
- ❖Frequent customer service calls are a churn signal – flag for retention outreach
- ❖Pure usage metrics alone are less predictive – focus on behaviour and complaints



RECOMMENDATIONS

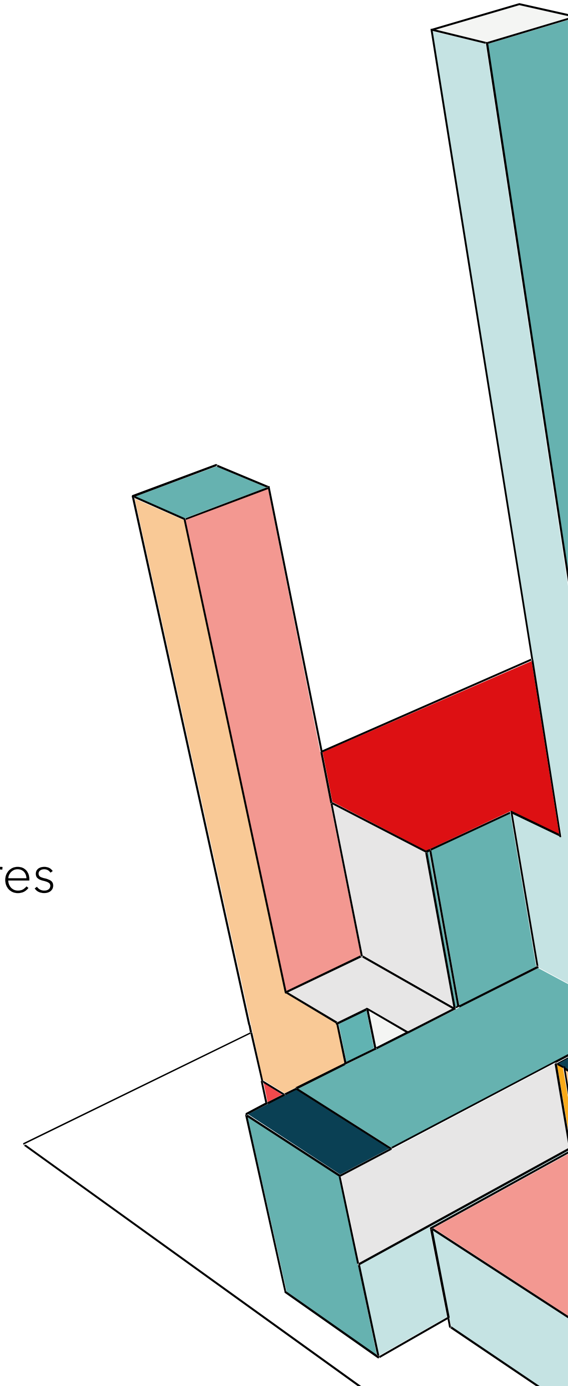
ACTIONABLE STRATEGIES

❖Retention Alerts: Flag customers with >3 support calls

❖Incentives: Offer loyalty rewards to internation plan users

❖Risk Scoring: Use XGBoost predictions to monitor high risk customers

❖Data Enrichment: Add contract type, billing history and satisfactions scores



CONCLUSION

- ❖Build a complete Machine Learning pipeline for EDA to deployment
- ❖Achieved 95% accuracy and 0.80 F1 – score with XGBoost
- ❖Provided real, actionable business recommendations

Company can now proactively reduce churn and save costs

THANK YOU

Name: Benson Mwihia

GitHub: https://github.com/BMwihia/Customer-Churn-Analysis

Email: bensonmwihia@gmail.com

LinkedIn: Benson Mwihia