Lloyds Banking Group Data Science Simulation

Predicting Customer Churn Using Machine Learning

Azizul Rahaman

About the Simulation

Hosted by Forage for Lloyds Banking Group

Simulated real-world data science project

Objective: Reduce customer churn using machine learning

Completed in two major tasks with real data files

Why I Chose It

Explore how data science supports decision-making

Improve analytical communication skills

My Role at SmartBank

Graduate Data Scientist at SmartBank, part of Lloyds Banking Group

Simulated a junior analyst position in digital banking

Explored customer behavior and digital engagement

Used real data to mimic real-world data science tasks

Task 1: My Work & Contributions

Worked as a data science graduate at Lloyds Banking Group for SmartBank's churn reduction project

Collected and merged data from multiple sources: demographics, transactions, service history, online activity, and churn status

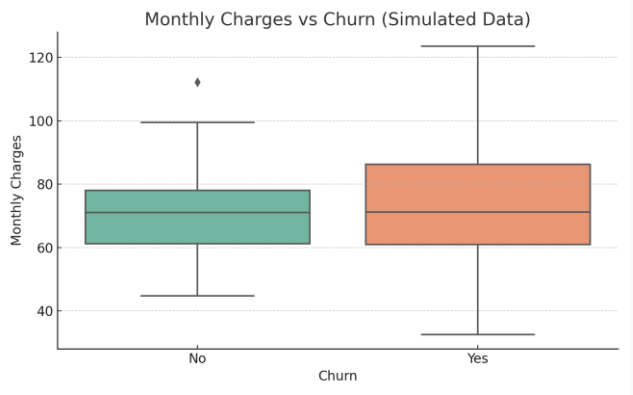
Performed exploratory data analysis (EDA) using Excel and Python to find customer behavior patterns

Created visual summaries to detect outliers and correlations

Calculated total customer spend, login frequency, and service interaction metrics

Cleaned the dataset:

- Replaced missing values with appropriate defaults (0 or "Unknown")
- Prepared numerical features for modeling (standardization-ready)

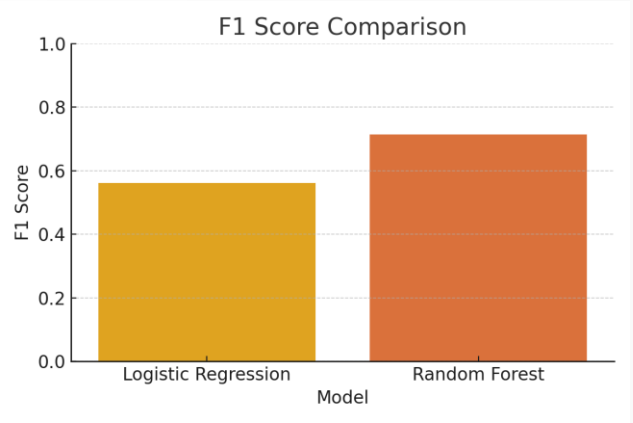


What I Learned from Task 1

- Improved my skills in collecting and integrating real-world datasets
- Learned how to explore large datasets and detect meaningful patterns
- Gained practical experience in data cleaning and preprocessing techniques
- Understood the link between poor service experience and customer churn
- Recognized the importance of mobile app usage and engagement in retention
- Prepared data effectively for machine learning model development in the next task

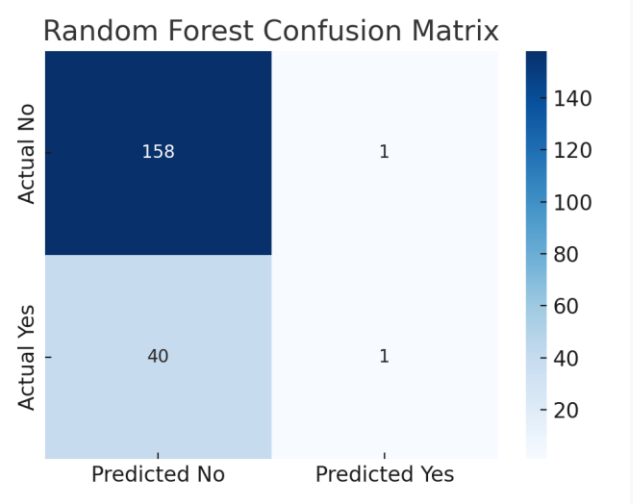
What I Did in Task 2 – My Role & Contributions

- Built predictive models to identify customers likely to churn for Lloyds Banking Group
- Used the cleaned dataset from Task 1 to train and evaluate models
- Selected and compared Logistic Regression and Random Forest algorithms
- Handled data imbalance using class weighting (class_weight='balanced')
- Performed stratified train-test split to preserve class distribution
- Tuned hyperparameters and evaluated models using F1 Score, ROC-AUC, and confusion matrix
- Found Random Forest performed best (F1: 0.7132) and recommended it for deployment
- Suggested real-world use of model for targeted retention strategies (e.g., loyalty offers, communications)



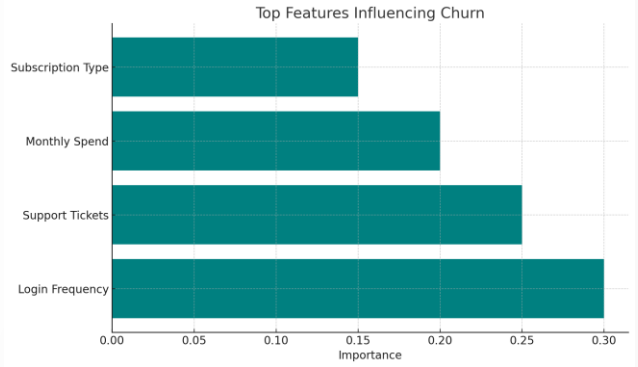
What I Learned from Task 2

- Learned how to select models based on accuracy, interpretability, and business fit
- Understood the impact of class imbalance on model performance and how to address it
- Practiced evaluating models using multiple metrics, not just accuracy
- Strengthened skills in model validation using stratified sampling and performance tuning
- Gained experience in translating technical model results into business insights
- Learned the importance of model retraining and monitoring to ensure long-term effectiveness



Technical Details – Scope

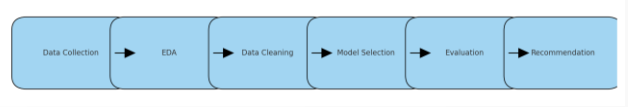
- Performed EDA on customer churn dataset
- Built ML model using Python & Excel



Technical Details – Workflow

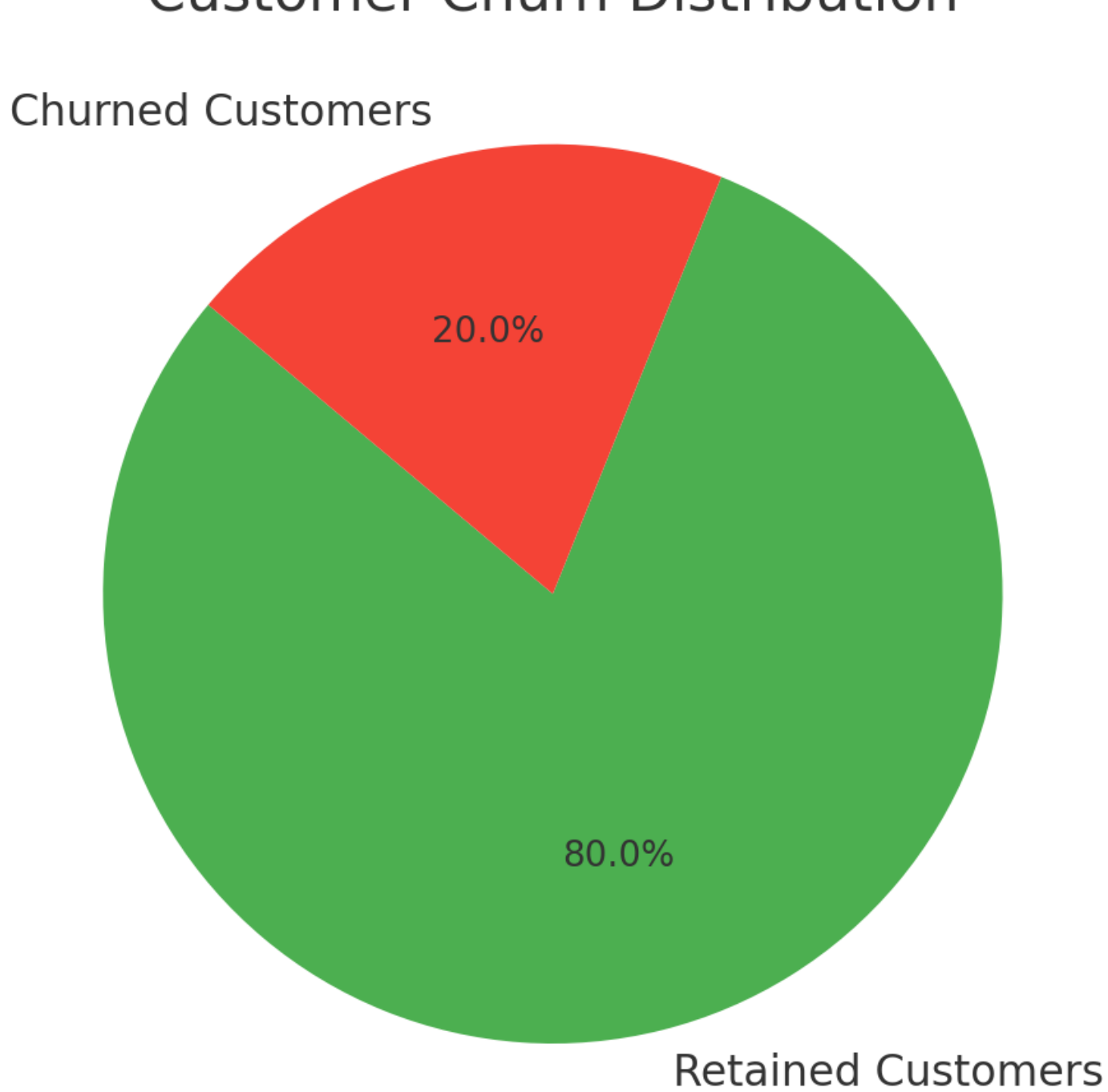
STEPS: DATA PREP → MODELING → EVALUATION

TOOLS: CONFUSION MATRIX, ROC-AUC, F1 SCORE



How My Work Helps the Business

- Predicted at-risk customers for SmartBank
- Enabled proactive customer retention actions



Business Impact – Trends

- Banking uses predictive analytics
- Hybrid Data + InfoSec roles are growing

Thank You

Azizul Rahaman