

Reducing the Number of Customers Churned

Building a Powerful Model for Ted & Poppy's Subscription Model

BUSINFO 704 - 2025 Quarter 1
Master of Business Analytics



Group 4:
Shilpa Vijayan [schb362]
Pratik Bansal [pbm601]
Annie Li [sli471]
Xiaouri Wu [xwu784]



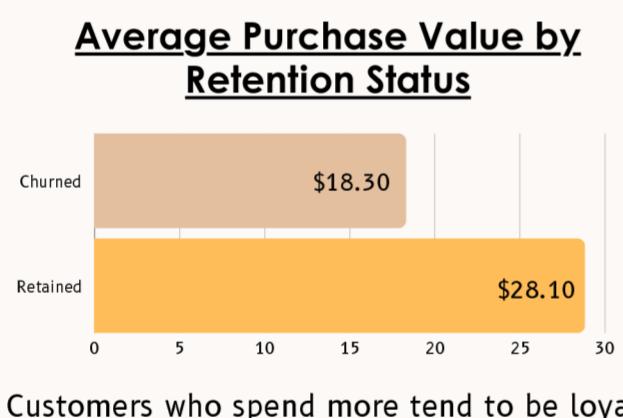
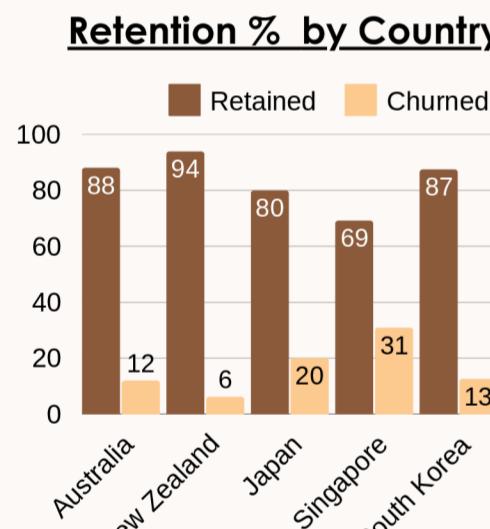
Background

Ted & Poppy, a rapidly growing international business, has built its success on the instincts and vision of its founders. Recently, Ted & Poppy has observed a concerning increase in customer churn but lacks clarity on the underlying causes. This project is designed to bridge that gap by implementing advanced analytics to uncover patterns, behaviors, and factors contributing to customer churn.

Data Overview

200,000 Observations
10 Variables

Customers Churned v/s Retained
16% Churned 84% Retained



Assumptions

- The dataset is representative of **customer behavior**.
- Downsampling** ensures balanced training data for churn classification.
- The patterns in the data remain **stationary** over time.
- The model is **scalable**.
- The model prioritizes key features.
- No multicollinearity** between variables.
- The classification **cut-off set at 0.5**.
- Leakage-free** between training and test sets.



Methods

1. Data Interpretation

- ✓ Data Inspection
- ✓ Data Transformation
- ✓ Exploring the Data through statistical analysis
- ✓ **Variable selection:**
Average Purchase Value, App Visits, Website Visits, Subscription Payment Problem, Subscription, etc.

2. Data Cleaning & Preprocessing

- 5 K-Fold Cross-Validation
Splitting data
75% train 25% test
- ✓ Enhance model performance
 - ✓ Remove redundancy
 - ✓ Balance the dataset

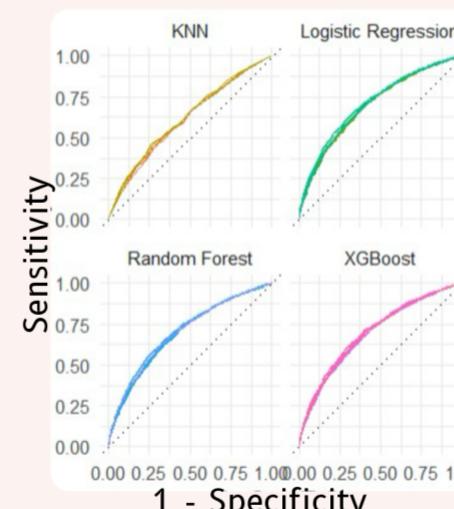
3. Modeling Approaches

💡 3.1 Training Dataset : Used for model development and learning.

- ✓ Define model
- ✓ Fit model to the folds
- ✓ Choose the best model
- ✓ Exam and compare results

	Accuracy	Sensitivity	ROC_AUC
Logistic Regression	0.708	0.670	0.778
Random Forest	0.702	0.684	0.770
XGBoost	0.704	0.686	0.772
KNN	0.614	0.621	0.685

💡 3.2 Testing Dataset : Used to evaluate model performance on unseen data.

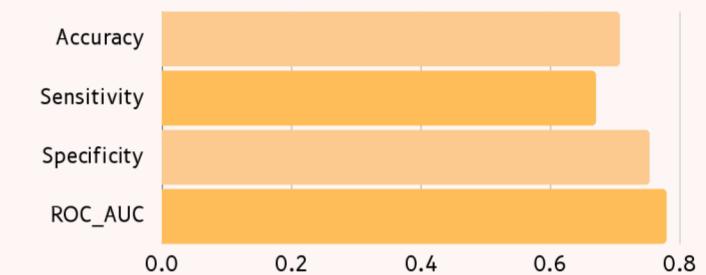


4. Interpreting Results

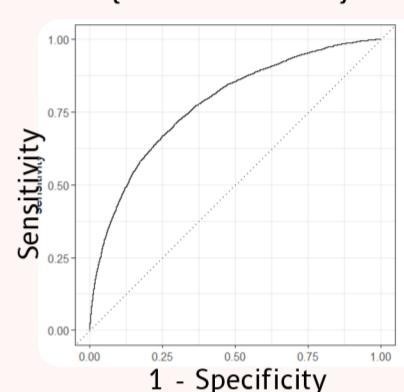
- Logistic Regression achieved 70.8% accuracy and a ROC AUC of 77.9%, meaning the model is reliable in distinguishing between retained and churned customers.
- The recall (67%) indicates that the model effectively captures customers at high risk of churning.

Results & Findings

Evaluation Metrics



ROC Curve - Logistic Regression [Churn Prediction]



Confusion Matrix

Actual		Churned	Retained
Prediction	Churned	2642	5195
Retained	1346	15818	

Key Insights from the Model

- The Model has strong reliability and performance.
- Payment problems for subscription are one of the strongest predictors.
- The model leans on the side of caution. This might lead to more promotion, in turn attracting new customers.

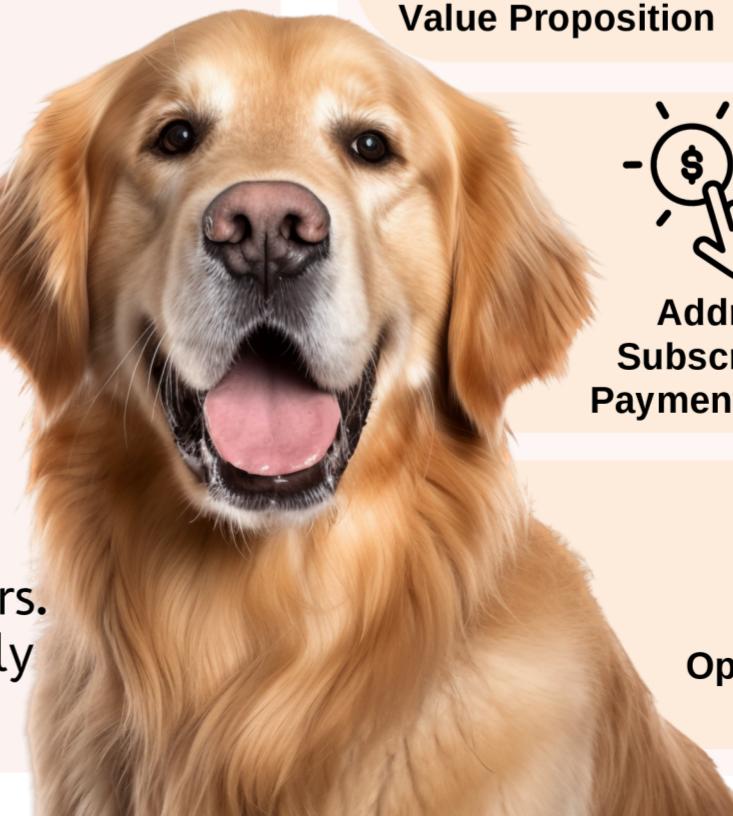
Conclusion & Recommendation

This model identified key factors influencing **customer churn**, and the findings suggest that **Logistic Regression** is the most effective model for predicting churn.



Optimize Pricing and Value Proposition

Conduct A/B testing on pricing and promotions, highlight each subscription's unique benefits, and explore bundling options to enhance premium tier.



Address Subscription Payment Issues

Improve payment processes, send proactive payment reminders, and provide flexible payment options to reduce churn.



Optimize Customer Engagement

Encourage higher app and website interaction through personalized content.