Telka Customer Churn Analysis

A deep dive into behavioural, demographic and service data to understand customer churn at Telka.

Timothy Kabucho - 08/06/2025

Business & Data Understanding

Business Understanding

Operational Context & Business Problem

- Telka is a national telecommunications company offering internet, phone, and bundled services to a wide customer base.
- In recent months, the company has observed a steady increase in customer churn especially among month-to-month subscribers and users paying via electronic methods like e-checks.
- With an ever saturating telecom market and rising customer acquisition costs, Telka needs to turn churn into a strategic priority—and fast.

Business Understanding

Project Goal and Aim

- We aim to build a **classification model** that predicts churn using customer demographics, billing details, and service usage.

- By understanding which factors contribute most to churn, Telka can:

  1. Prioritise at-risk customers for intervention

  2. Refine service offerings and contract structures

  3. Increase overall customer retention and lifetime value

Modelling

Modelling

Data preprocessing

- **Binary Encoding:** Converted binary features like Partner, Dependents, and PhoneService into numerical 0/1 format.
- **One-Hot Encoding:** Categorical features with more than two classes (e.g., InternetService, Contract, PaymentMethod) were transformed using OneHotEncoder to ensure no ordinal relationships were inferred.
- **Feature Scaling:** Applied **StandardScaler** to numerical features (tenure, MonthlyCharges, TotalCharges) for models sensitive to magnitude (e.g., SVM, Logistic Regression).
- **Multicollinearity Handling:** Used **Variance Inflation Factor (VIF)** to assess collinearity. TotalCharges was flagged (VIF = 8.08) and treated cautiously due to its derivation from tenure × monthly charges.

Modelling

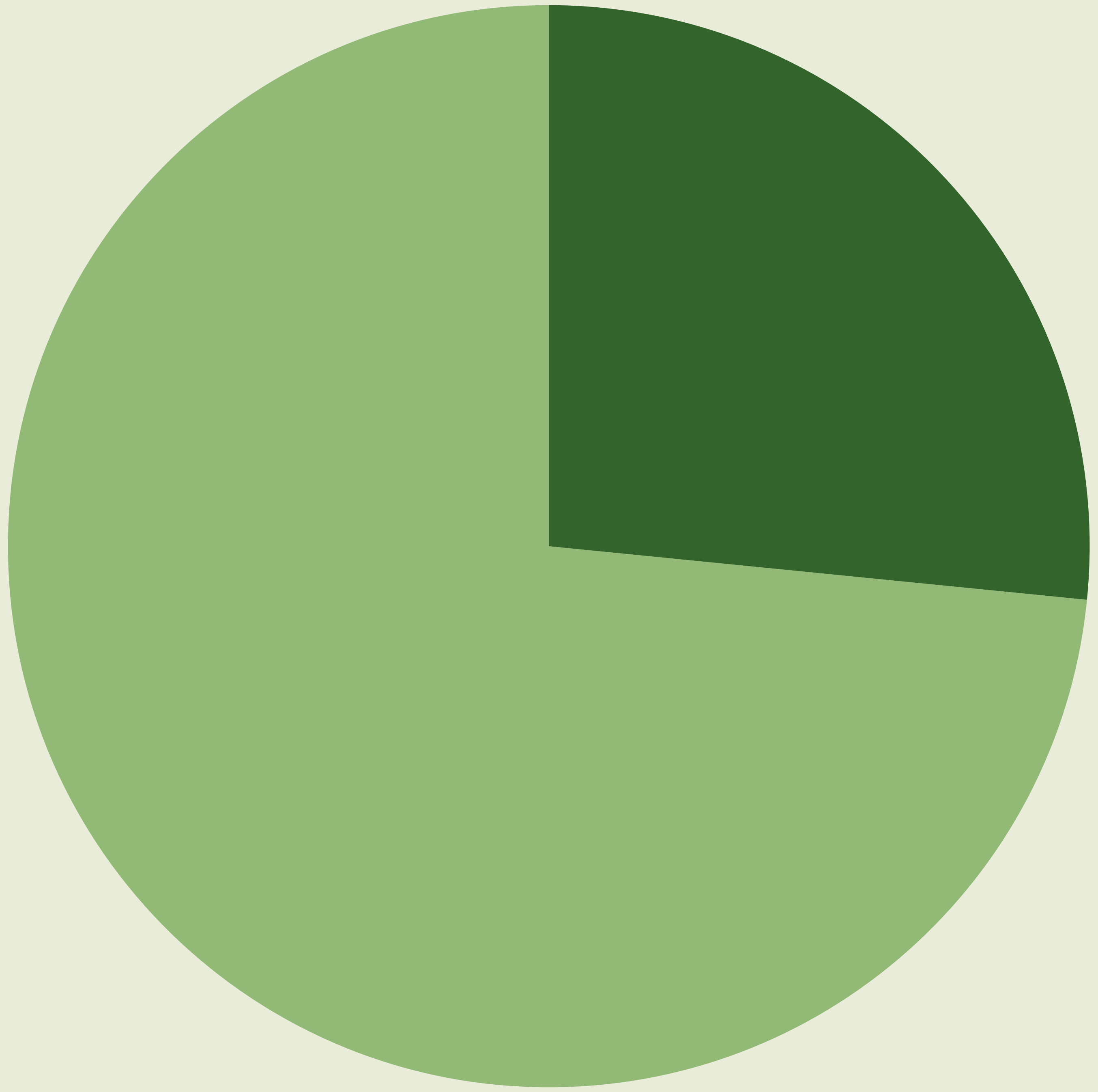
Evaluated Models

- **Logistic Regression:** A baseline linear model for binary classification. Fast, interpretable, and great for benchmarking.
- **Support Vector Machine (SVM):** Effective for high-dimensional spaces. Used RBF and linear kernels with tuned regularisation.
- **Random Forest:** A powerful ensemble method. Handles nonlinearities, feature importance analysis, and is robust to multicollinearity.

Modelling

Imbalance in Target Variable

- Dataset had **~27% churners** and **~73% non-churners**.
- Applied **SMOTE (Synthetic Minority Oversampling Technique)** *only on training data* to generate synthetic churn examples and balance the classes.



Modelling

Model Optimisation

- **Hyperparameter Tuning:** GridSearchCV with 5-fold cross-validation was used to optimize model parameters (e.g., C, penalty for LR; kernel, gamma for SVM; n_estimators, max_depth for RF).
- **Pipeline Workflow:** All models were trained using **imblearn Pipelines**, chaining SMOTE, preprocessing, and model training steps—keeping it clean, consistent, and leakage-free.

Evaluation

Evaluation

Evaluation Metrics

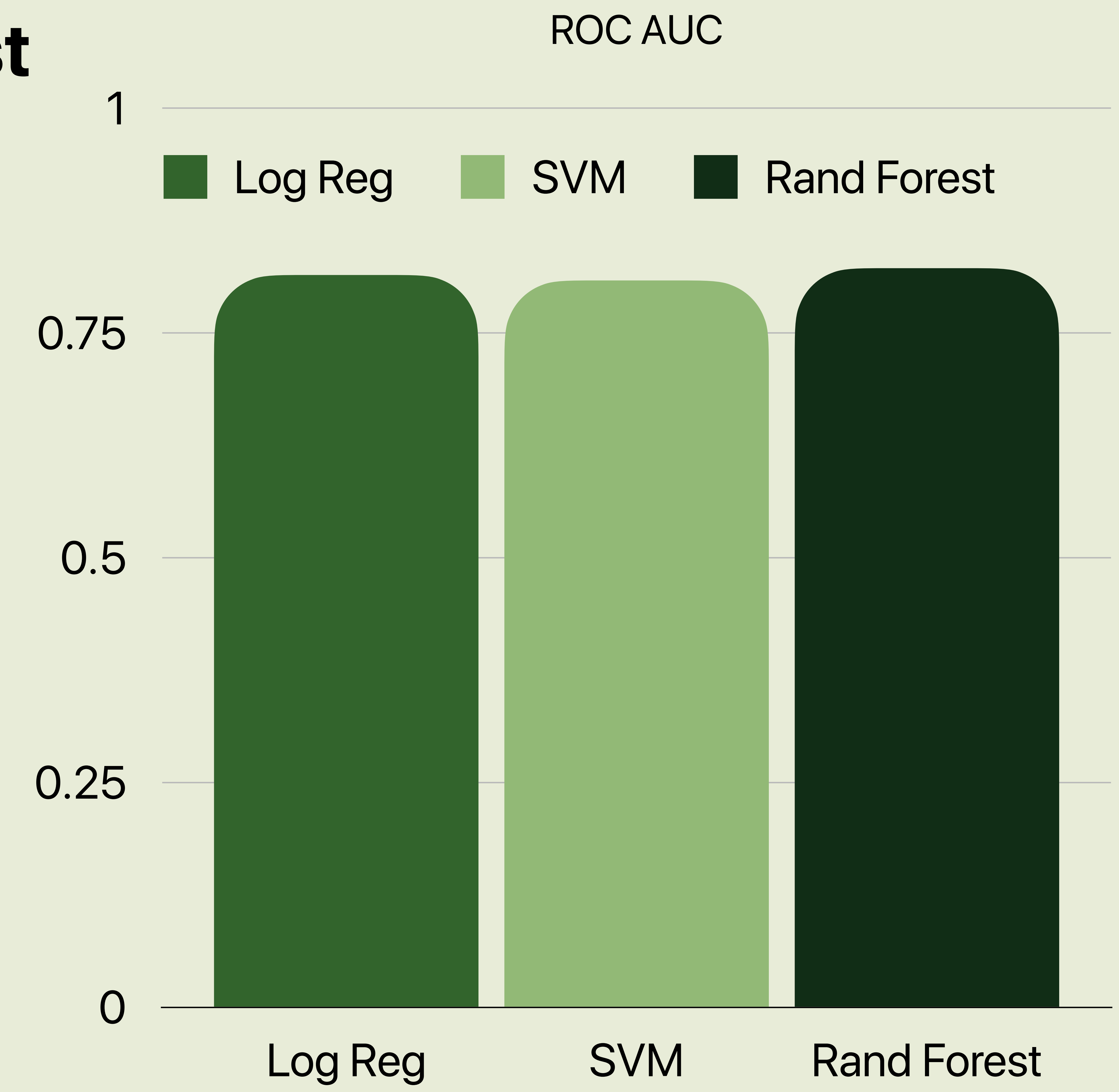
To rigorously assess model performance—especially on an imbalanced dataset—we used:

- **Accuracy:** Overall proportion of correct predictions. Good for general performance, but can be misleading with imbalanced classes.
- **Precision:** Out of all predicted churners, how many were actually churners? (Helps reduce false positives.)
- **Recall:** Out of all actual churners, how many were correctly identified? (Crucial for retention strategies!)
- **F1-Score:** Harmonic mean of precision and recall. A balanced measure, especially important when tradeoffs matter.
- **ROC AUC:** Measures the model's ability to distinguish between churners and non-churners across all classification thresholds.

Evaluation

Best Performing Model: Random Forest

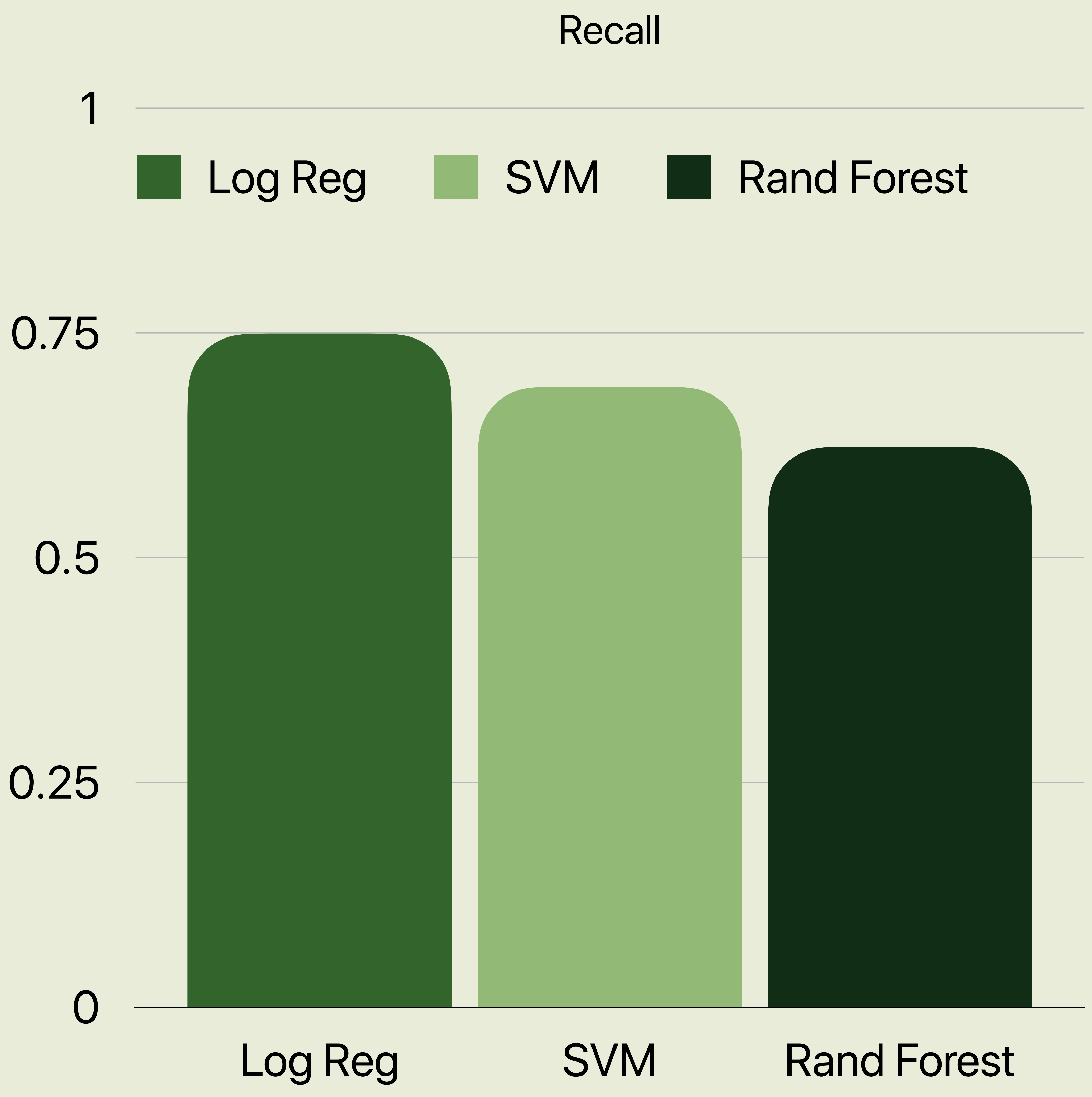
- **Top Accuracy:** 0.77
- **Highest ROC AUC: 0.82**, making it the best at separating churners from loyal customers.
- **Strong Precision:** 0.57, meaning fewer false alarms on predicted churners.
- **Good balance across all metrics**, and it handled feature interactions + non-linearities like a champ.



Evaluation

Other Evaluated Models

- **Logistic Regression** excelled in **Recall (0.75)** – great at flagging *most* churners, ideal when missing a churner is costly.
- **SVM** showed balanced performance across all metrics – a solid all-rounder, especially when interpretability isn't the top priority.



Recommendations

Recommendations

Adopt Random Forest for Production

- Among all tested models, **Random Forest** delivered the best performance across key metrics (Accuracy, Precision, ROC AUC).
- It’s especially effective at handling non-linear interactions and provides feature importance insights — a big win for business interpretability.

Recommendations

Monitor Key Churn Indicators

- **Contract Type:** Customers on month-to-month plans churn significantly more — push them toward longer contracts.
- **Online Security:** Lack of security services correlates with higher churn — upsell these as retention tools.
- **Tenure:** The shorter the tenure, the higher the risk. Target new customers early.
- **MonthlyCharges:** High monthly bills = higher churn risk. Flag high spenders for value check-ins.

Recommendations

Retention Timing is Everything

- The **first 6 months** are critical. Churn is concentrated here.
- Launch onboarding sequences, loyalty perks, and personalized check-ins during this window.

Recommendations

Offer Targeted Incentives

- Use churn predictions to **proactively engage at-risk customers** with:
  - Discounts
  - Loyalty bundles
  - Personalised offers
- Focus especially on high-value customers and those lacking supportive services (e.g., no TechSupport or DeviceProtection).

Next Steps

Next Steps

Deploy Model into Production

- Package the **Random Forest** model with all preprocessing steps into a pipeline.
- Integrate it into the company's **CRM system** to score new and existing customers in real time.
- Use predictions to prioritize retention workflows and personalize customer engagement.

Next Steps

Setup Churn Risk Alerts

- Implement automatic alerts for customers flagged as **high-risk**.
- Trigger workflows like:
  - Retention team follow-ups
  - Automated offers
  - Escalation for top-tier clients
- Ensure that the business responds before churn happens — not after.

Thank You