# Predicting Customer Churn at No-Churn Telecom

## A Data-Driven Strategy for Proactive Customer Retention

NO-CHURN TELECOM

NotebookLM

# The High Cost of Customer Churn

In a highly competitive market, customer retention is paramount. Our analysis reveals a significant challenge:

approximately **14%** of our customer base is at risk of churning.

## The Challenge

The telecom industry faces constant pressure from competitors, leading to high rates of customer churn.

## The Goal

Move from a reactive to a proactive retention model. Instead of waiting for customers to leave, we must identify them *before* they make the decision.

## The Project Objective

To develop a machine learning system that accurately predicts which customers are likely to leave, enabling targeted intervention.

NotebookLM

# The Foundation of Our Analysis: The Customer Dataset

Our predictions are built upon a comprehensive dataset of **4,617 customer records** (**#003366**), providing a detailed view of behaviour and interactions.

## Customer Profile

State

Account Length

Area Code

## Usage & Plan Metrics

Day, Evening, Night & International Usage (Mins, Calls, Charges)

International Plan & Voicemail Plan Subscriptions

## Customer Experience

Number of Customer Service Calls

Final Churn Status (The target for our prediction)

NotebookLM

# A Disciplined Path from Data to Decision

### 1. Data Preparation

Cleaning and structuring the raw data for analysis. Creating a clean `CHURN_FLAG` (1 for churn, 0 for no churn).

### 2. Feature Engineering

Creating new, more powerful signals for the model, such as `Total_mins` and `Total_charge`.

### 3. Exploratory Analysis

Investigating the data to uncover the hidden patterns and drivers behind customer churn.

### 4. Predictive Modelling

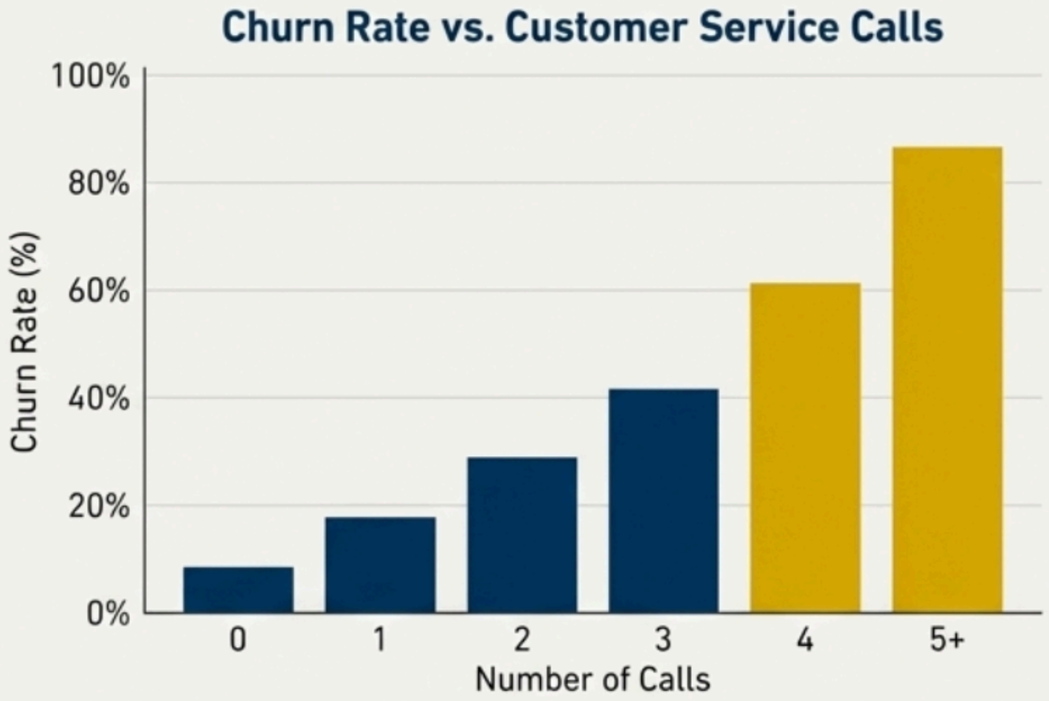
Building and testing machine learning models to generate a reliable churn prediction for every customer.

NotebookLM

# Key Insights: The Strongest Signals of Churn

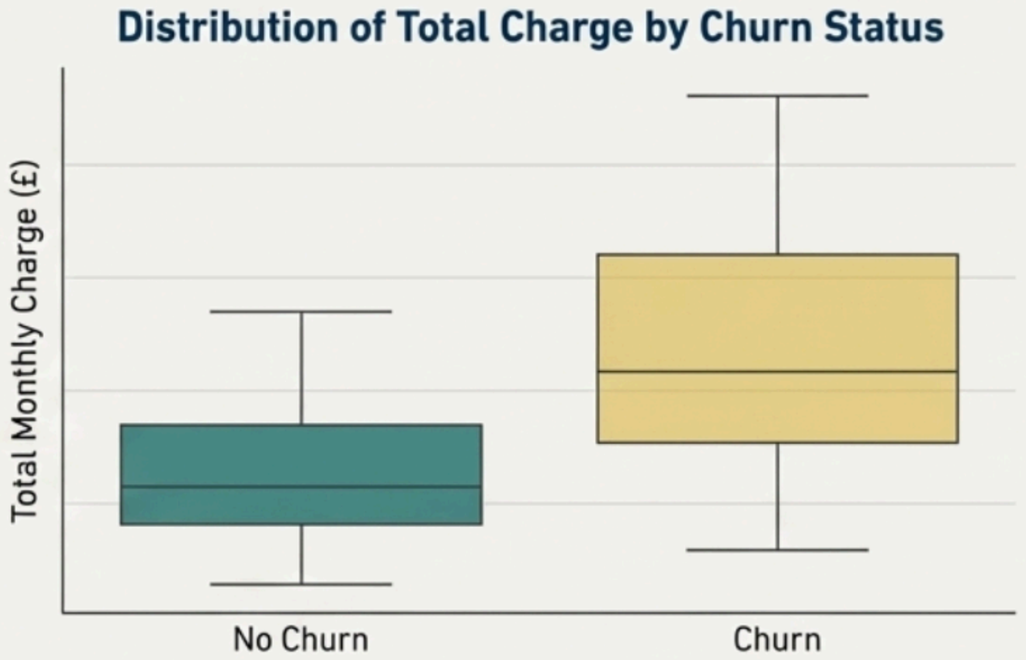
## The number of customer service calls is the most powerful predictor of churn.

The probability of a customer leaving increases dramatically with each call made to customer service, indicating that unresolved issues are a primary driver.



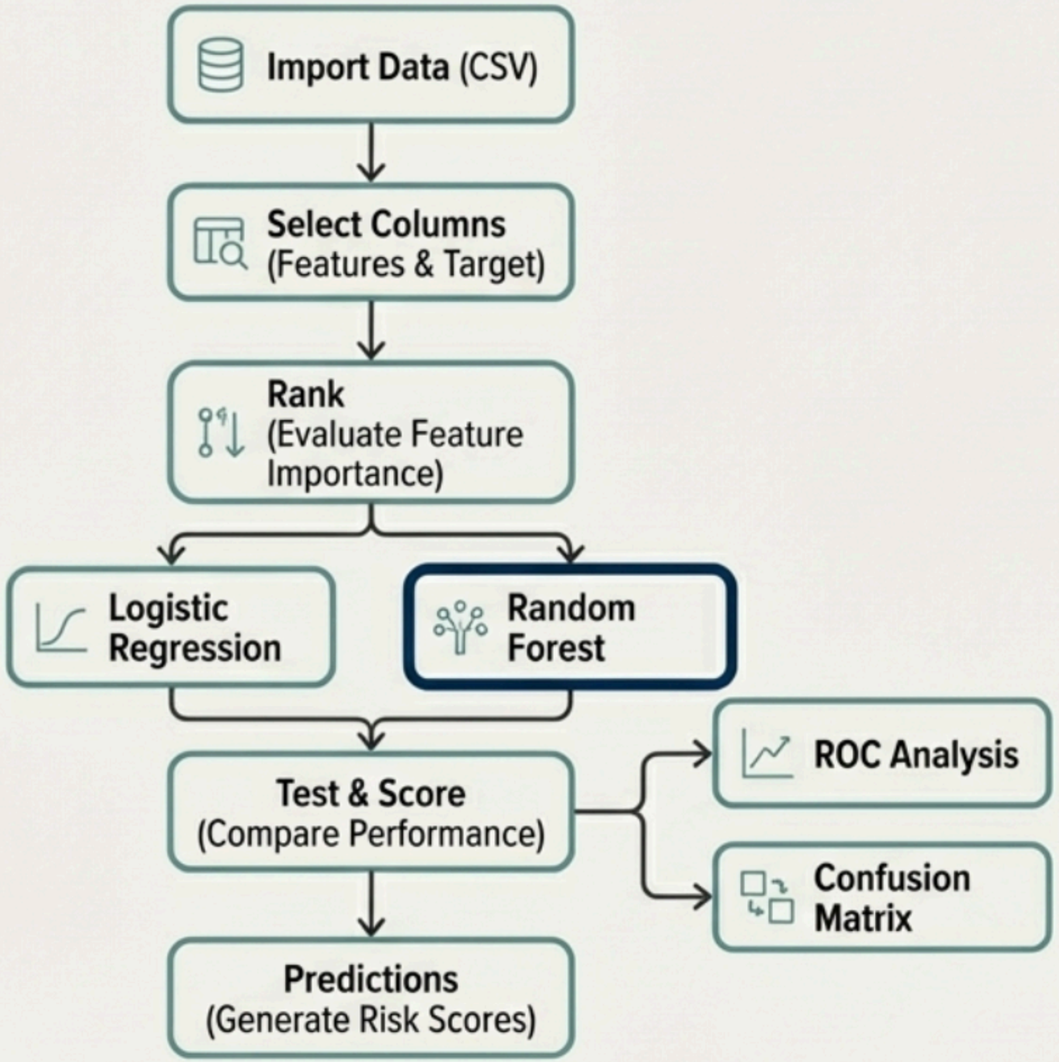

## Usage patterns and international plans are also significant factors.

High total charges and minutes are strongly correlated with churn. Customers on an international plan also exhibit distinct churn behaviour.




NotebookLM

# Building the Predictive Engine with Orange Data Mining




The analysis workflow begins by importing the cleaned telecom dataset into Orange. The data is passed through the **Select Columns** widget, where the feature variables and the target variable (CHURN_FLAG) are defined. Feature importance is evaluated using the **Rank** widget to understand which parameters influence churn the most.

Multiple models, including Logistic Regression and Random Forest, are connected to the **Test & Score** widget to compare their performance using key metrics such as AUC and Accuracy. The **ROC Analysis** widget visualises the performance curves, while the **Confusion Matrix** displays true and false predictions. Finally, the best-performing model is connected to the **Predictions** widget, which generates churn probabilities (risk scores) for each customer.

NotebookLM

# Random Forest: Our Chosen Model for Its Accuracy and Stability

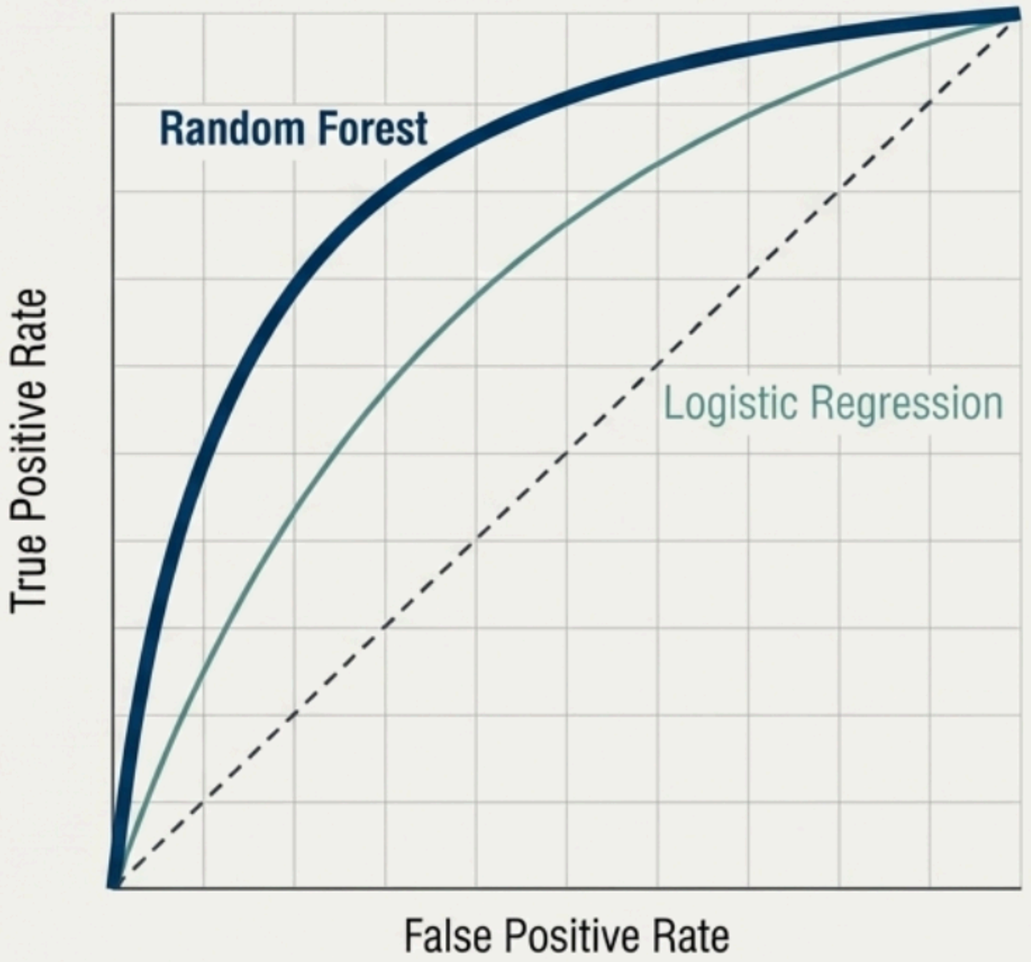
After evaluating multiple algorithms, the Random Forest model was selected as the most effective and reliable predictor of customer churn.

## Superior Performance

Consistently achieved a high Area Under the Curve (AUC), indicating an excellent ability to distinguish between churners and non-churners.

## Prediction Stability

Provided balanced performance, reliably identifying customers who were likely to churn without incorrectly flagging those who were not.




NotebookLM

# Measuring Predictive Accuracy: The Confusion Matrix Results

The model's performance demonstrates a high degree of reliability in identifying at-risk customers while minimising disruption to loyal ones.

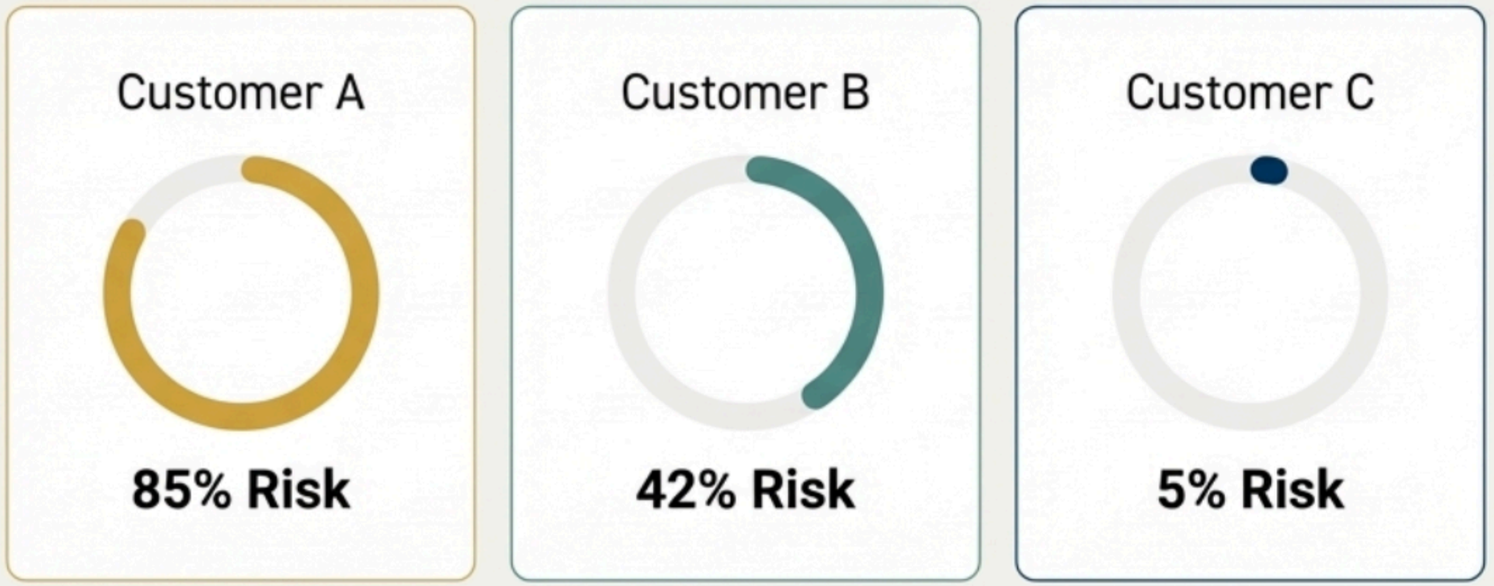
| | | Predicted: Staying | Predicted: Leaving |
|---|---|---|---|
| Actual | Staying | **3958**<br>Correctly identified as staying | **3**<br>Incorrectly identified as leaving |
| Actual | Leaving | **76**<br>Incorrectly identified as staying | **580**<br>Correctly identified as leaving |

- **High Success Rate:** We successfully identified **580** of the **656** actual churners.
- **Minimal False Alarms:** Only **3** non-churning customers were incorrectly flagged, ensuring retention efforts are focused efficiently.

NotebookLM

# From Prediction to Action: The Customer Risk Score

The Random Forest model assigns a precise churn probability—a **Risk Score**—to every customer in our database.




This score, ranging from 0% to 100%, represents the individual likelihood of a customer churning. A higher score indicates a greater risk. This allows us to move beyond a simple “yes/no” prediction and prioritise our retention efforts with surgical precision.

NotebookLM

# Activating the Data: Segmenting Customers by Risk Level

By applying simple thresholds to the Risk Score, we can segment our entire customer base into actionable tiers.

### High Risk (Score > 70)

These are the most vulnerable customers. They require immediete, personalised intervention: **Action:** Prioritise for proactive outreach and tailored retention risk.

### Medium Risk (Score 40–70)

These customers are showing signs of dissatisfaction. They should be monitored.
**Action:** Include in broad-based retention campaigns and monitor for escalating risk.

### Low Risk (Score < 40)

These are our loyal customers.
**Action:** Nurture with standard marketing; avoid disruptive retention offers.

# A Clear Path Forward: Strategic Recommendations

## Immediate Priorities

- Contact all **High-Risk** customers with proactive retention offers.
- Analyse their profiles to design personalised incentives.

## Operational Improvements

- Investigate and address the root causes of high customer service call volumes.
- Closely monitor customers who complain about usage and billing.

## Technical Governance

- Refresh the dataset and retrain the model on a quarterly basis to maintain accuracy.
- Continuously monitor model performance (AUC, Recall) and check for data drift.

NotebookLM

# The Value Unlocked: A Proactive Churn Reduction System

**This project delivers a powerful, data-driven capability that enables No-Churn Telecom to anticipate customer needs, mitigate churn, and protect revenue.**

## Identified

We now know the key drivers of churn, led by poor customer service experiences.

## Built

We have a reliable machine learning model that accurately identifies at-risk individuals.

## Enabled

Our actionable Risk Score allows for efficient, targeted retention strategies.

NotebookLM