

E-commerce Sales Analysis And Customer Insights

2025



FOR EDUCATIONAL AND
ANALYTICS PURPOSE

PRESENTED BY
PHYLLIS WANGUI

Table of Contents

INTRODUCTION

**BUSINESS BUSINESS
UNDERSTANDING**

OBJECTIVES

**EXPLORATORY
DATA ANALYSIS**

MODELLING

**RECOMMENDATIO
NS**

CONCLUSION

Introduction

01

- This analysis examines sales performance and customer behavior to uncover revenue trends, top products, and purchasing patterns.
- Using transactional data, it highlights revenue distribution, customer segments, seasonal trends, and product performance.
- The goal is to provide actionable insights that support smarter marketing, inventory optimization, and long-term e-commerce growth.



Business Understanding

02



- Identify revenue drivers – determine which products, categories, and regions contribute most to overall sales.
- Understand customer behavior – analyze purchasing frequency, order value, and customer segments to highlight profitable groups.
- Track sales trends – uncover seasonal or time-based patterns that can inform marketing and inventory decisions.
- Improve decision-making – provide insights that help optimize pricing strategies, promotions, and customer retention efforts.



Objectives

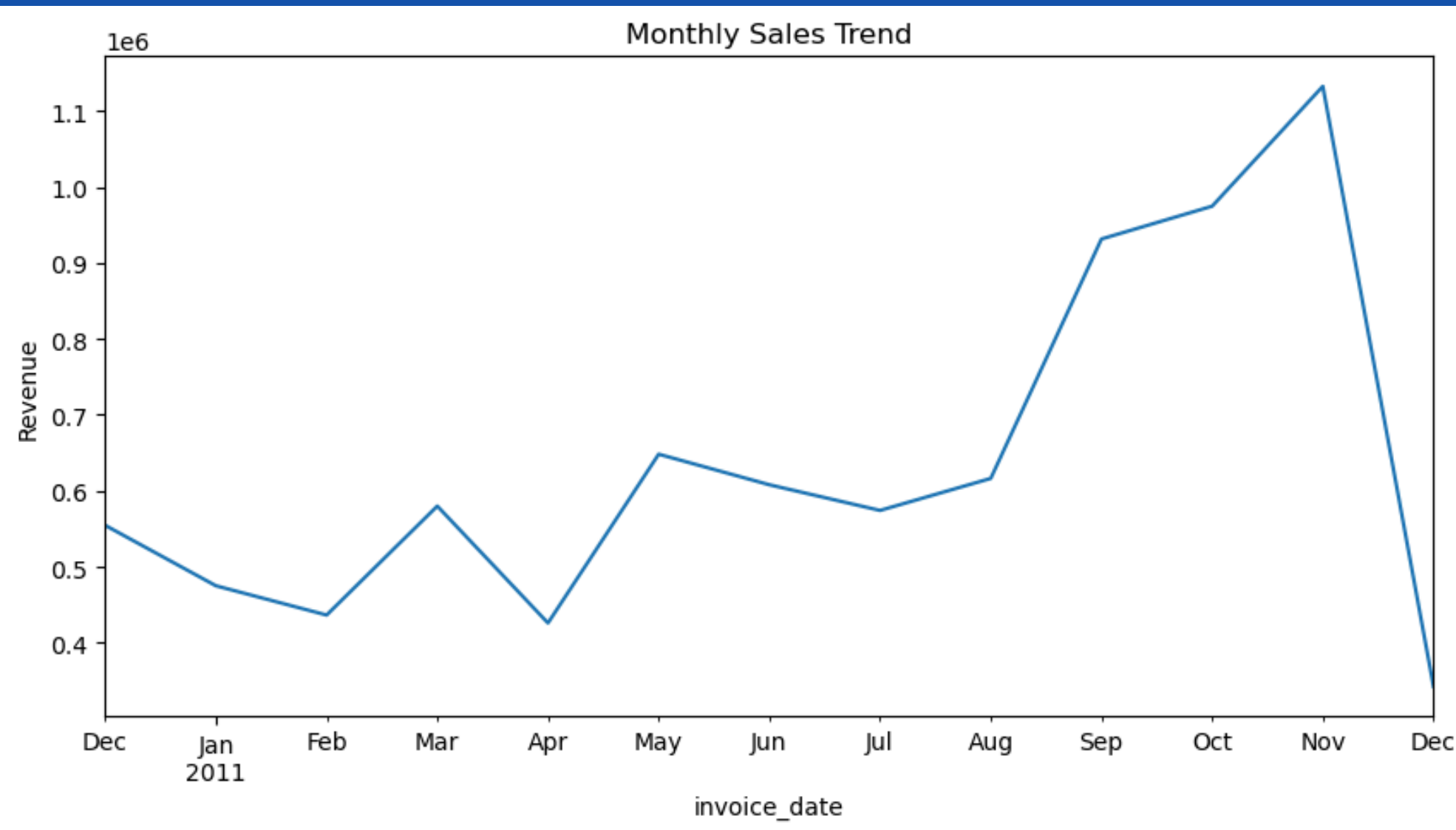
- Analyze sales to identify best-selling antiques and categories.
- Understand customer buying behavior and preferences.
- Detect seasonal or trend-based demand shifts.
- Provide insights to improve pricing, marketing, and inventory planning.



Exploratory Data Analysis

04

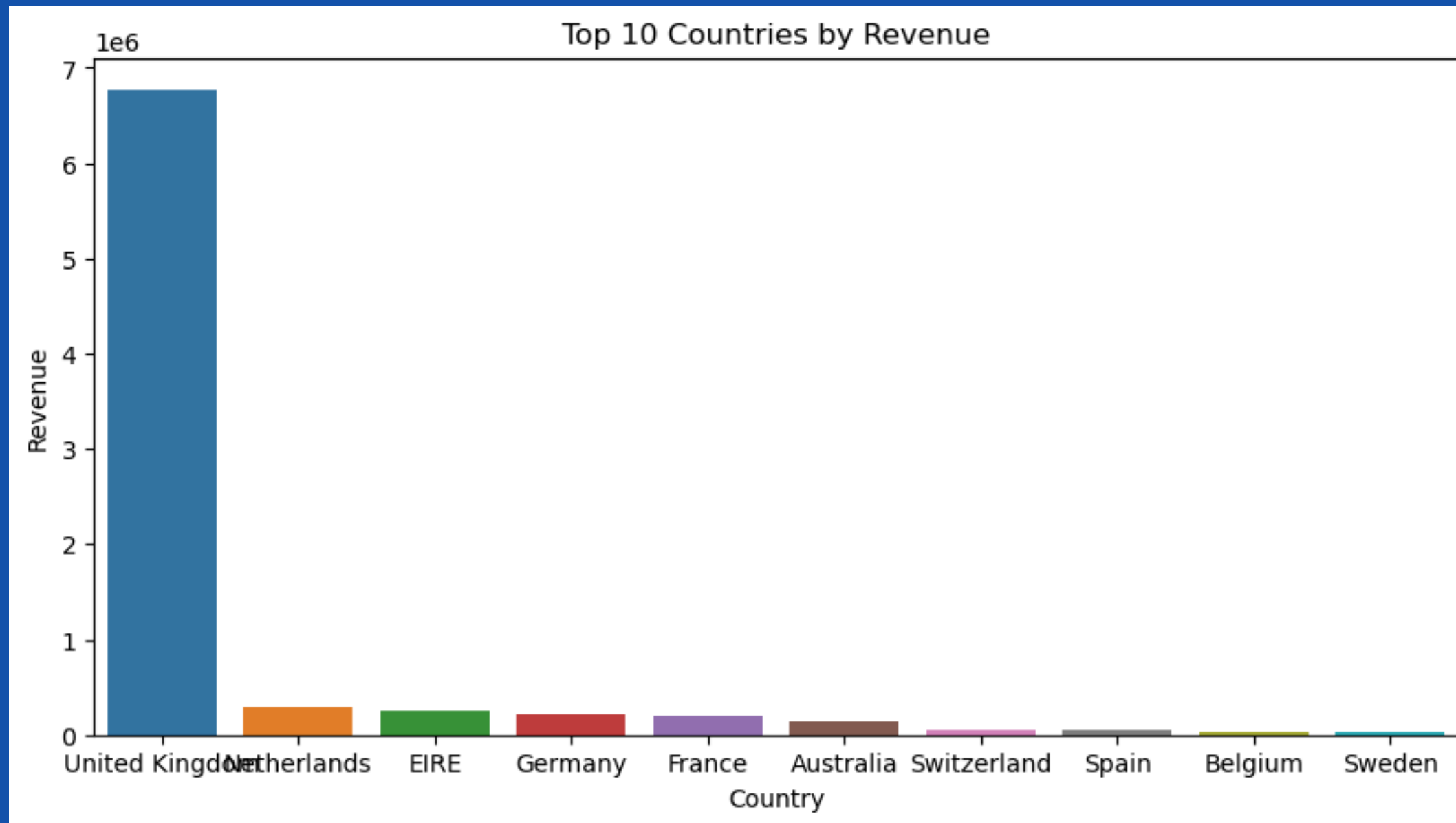
Monthly Sales Trend



Observation

The sales trend shows noticeable peaks during certain months, with higher revenue around the holiday season (e.g., November–December) and lower sales in the mid-year months. Demand is seasonal, likely tied to holiday shopping, so inventory and marketing should focus on peak months.

Top 10 countries by revenue

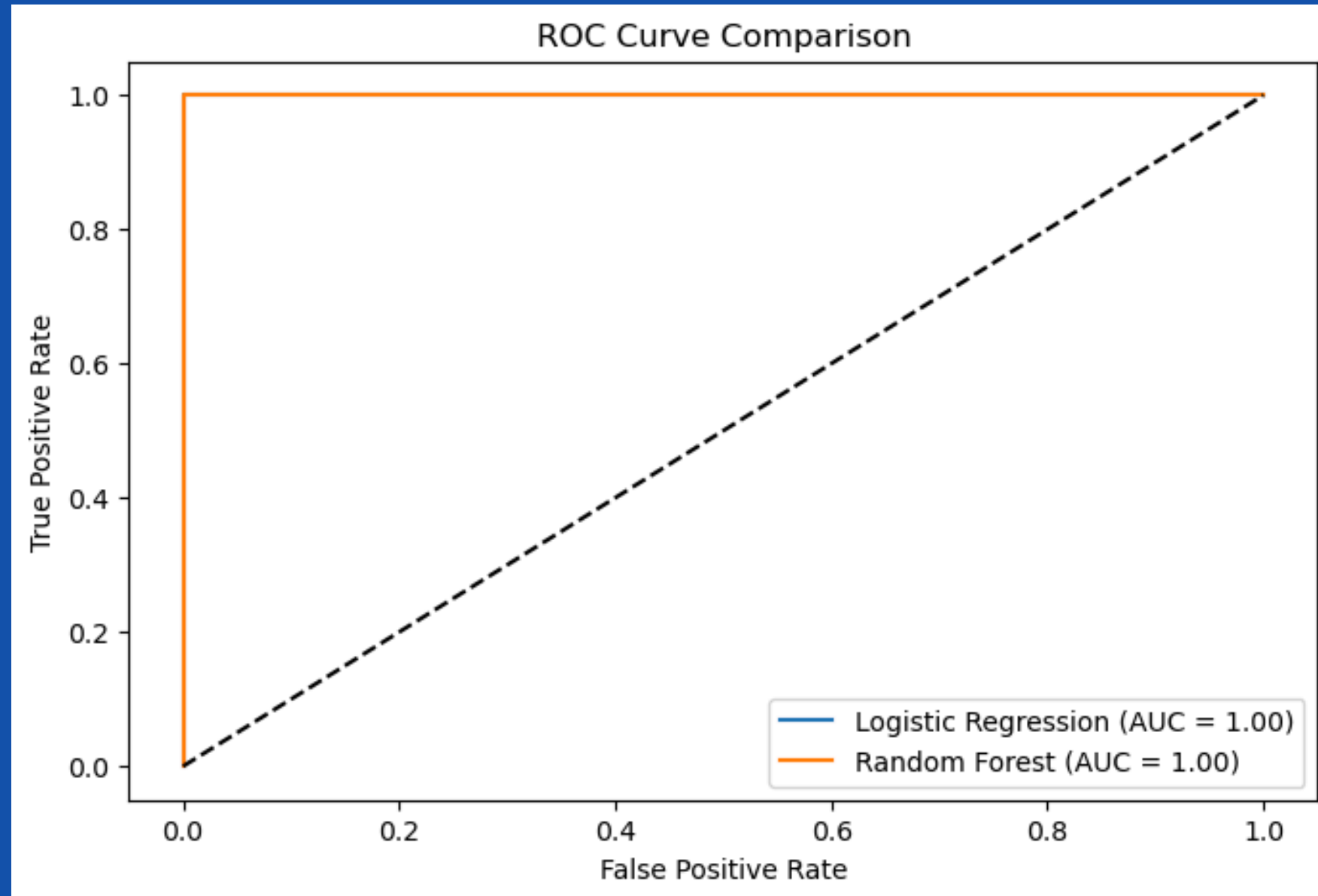


Observation

- Revenue is concentrated in a few countries, with one or two contributing the most, while others generate significantly less.
- The antique shop's customer base is not evenly distributed. Focusing marketing and logistics on top-performing countries could maximize returns, while exploring growth opportunities in lower-revenue regions could expand market reach.



ROC curve comparison



Observation

The Random Forest model has a higher AUC score than Logistic Regression, with its ROC curve staying closer to the top-left corner. Random Forest performs better at distinguishing between classes, making it more reliable for predicting customer behavior (e.g., purchase likelihood or churn) in the antique shop context.

10

RFM results into customer insights

CustomerID	Recency	Frequency	Monetary	Observation	Meaning
12346	326	2	0	Long time since last purchase, almost no spending	Inactive/lost customer, low likelihood of return
12347	2	7	4310	Very recent, frequent, and high-value purchases	Loyal & valuable customer, should be retained/rewarded
12348	75	4	1797.24	Moderately recent, mid-frequency, decent spending	Potential loyal customer, worth nurturing
12349	19	1	1757.55	Recent purchase but low frequency	New/occasional customer, may be converted into repeat buyer
12350	310	1	334.4	Old purchase, low frequency and low spend	One-time buyer, weak engagement

Modeling

11

The churn prediction model was trained to classify customers as Active (0) or Churned (1) based on RFM features (Recency, Frequency, Monetary). The evaluation results show:

- Accuracy: 100%
- Precision, Recall, F1-score: 1.0 for both classes
- Support: 1,058 active and 254 churned customers in the test set

The model perfectly classified all customers in the test set. This indicates strong predictive power, but such perfect performance may also suggest possible overfitting. Further validation with cross-validation or new unseen data is recommended to confirm generalization.

Recommendation

12

1. Focus retention efforts on customers with high Recency (i.e., those who haven't purchased recently), as Recency is the most important churn predictor.

2. Consider targeted marketing campaigns for customers with low Frequency and Monetary values to encourage repeat purchases and higher spending.

3. Use the trained Random Forest model to periodically score customers and proactively identify those at risk of churning.

4. Analyze top countries and products to tailor promotions and inventory for high-value segments.

5. Regularly update the RFM analysis and churn model as new data arrives to keep insights relevant.



Let's Make an Amazing Business Together!

2025

phylliswangui19@gmail.com

