Online Retail Store

Driving Business Decisions with Customer Analytics

RFM Segmentation | Cohort Analysis | Churn Prediction | CLV Modelling

Online Retail Store

table of content

Online Retail Store

Online Retail Store provides unique giftware for all occasions

This dataset, sourced from Kaggle, contains transaction records from a UK-based, registered online retail business that operated without a physical storefront between December 1, 2009, and December 9, 2011. The company leveraged a digital-first approach to efficiently serve its customer base, emphasizing the advantages of e-commerce in reaching a broad audience and streamlining operations.

Why This Analysis Matters?

Goals

The main objective of this analysis is to identify and segment high-value customers of the online retail store. By doing so, it aims to provide actionable insights for the business development team to implement targeted marketing strategies that enhance customer retention and maximize long-term value.

Problem

Although the store has over 1 million transactions in two years, it struggles with low customer retention. Many buyers make only one purchase, and the store lacks insight into which customers are most valuable—hindering efforts to build effective retention strategies.

Increase Customer Loyalty

Online Retail Store

To achieve the goals of this analysis, key features will be used in a structured process to measure customer retention, segment customers by value and purchase frequency, and predict potential churn. The insights will support data driven marketing strategies to boost loyalty and retention.

Data Understanding & Analysis Overview

Data Understanding

Dataset

The dataset, sourced from Kaggle.com, includes key features such as Invoice Number, Quantity, Invoice Date, Unit Price, Customer ID, and Country.

Analysis Overview

RFM Analysis

This analysis segments customers based on how recenty they purchased (Recency), how often they buy (Frequency), and how much they spend in total (Monetary).

Cohort Analysis

This analysis helps us understand customer retention behavior over time by examining their purchase history across different months.

Churn Prediction Model

This model leverages machine learning techniques to predict the likelihood of customer churn based on their behavior patterns.

CLV Analysis

CLV identifies high-value customers, enabling focus on long-term relationships, which will be combined with churn prediction to target at-risk, high-value customers.

Online Retail Store

“80% of Sales Come From 20% of Customers”

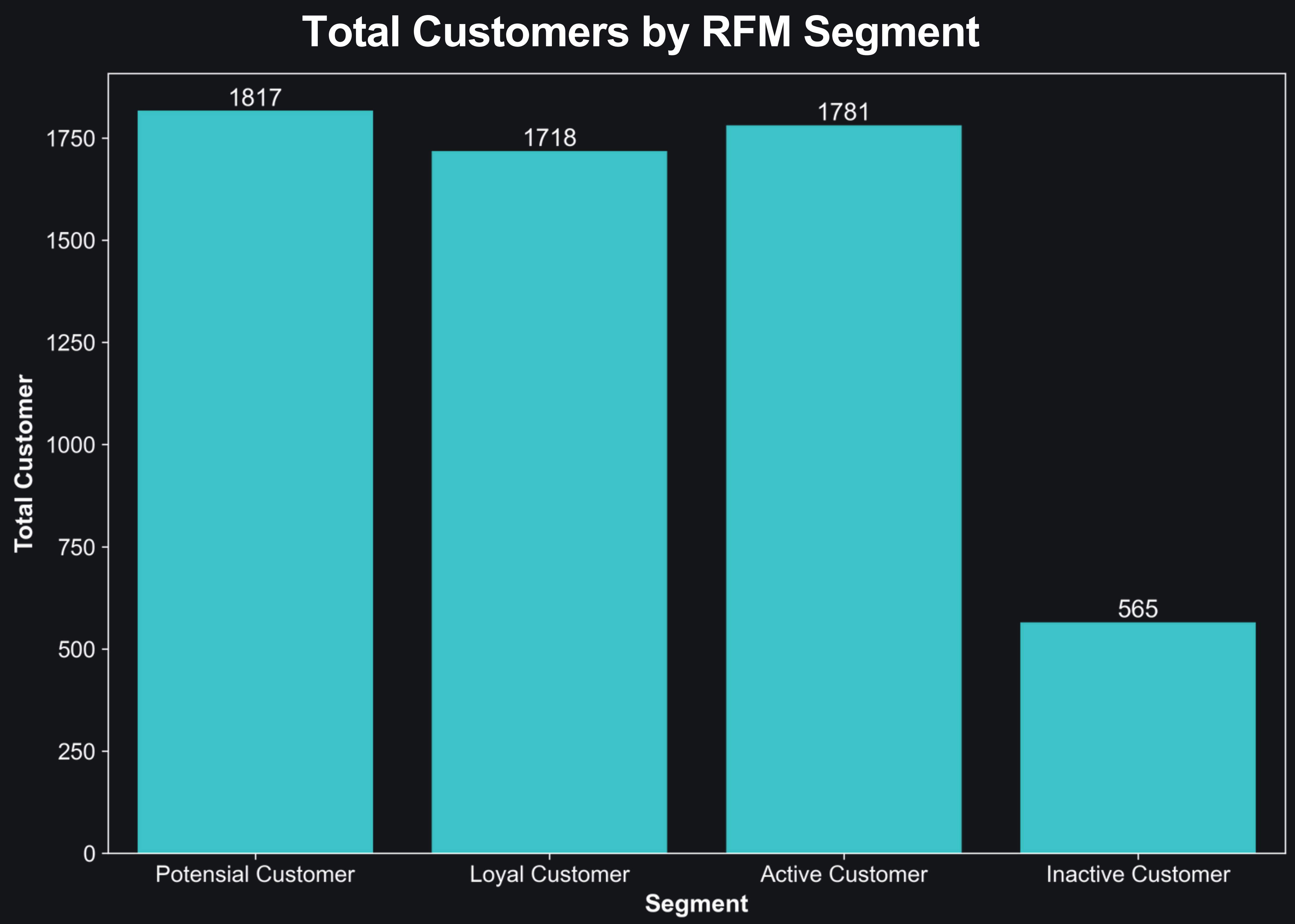
The *Pareto Principle* shows that a small group of customers contributes most of the revenue. So, **Who Are They?** To find out, we need to understand their purchasing behavior. RFM analysis helps us identify loyal, high value customers as well as those with low engagement.

RFM Analysis

Key Features

- Recency
- Frequency
- Monetary

Online Retail Store



This analysis applied the RFM method for customer segmentation by dividing each metric into quatiles and assigned scores from 1 to 4. A higher recency score indicates more recent activity, while higher frequency and monetary scores such as R=4, F=4, M=4 are categorized as top customers. Then summed the RFM scores for each customer and classified them into segments ranging from loyal to inactive customers.

1. Loyal Customer
These customers have an RFM score of >= 10, indicating high purchase frequency, high spending, and strong engagement with store.

2. Potential Customer
These customers have an RFM score of 7-9, indicating higher value from past purchases than average customers, making them more likely to be loyal.

3. Active Customer
These customers have an RFM score of 4-6, indicating average customers and have decent purchase amount, decent frequency, and a tendency to swtich between stores.

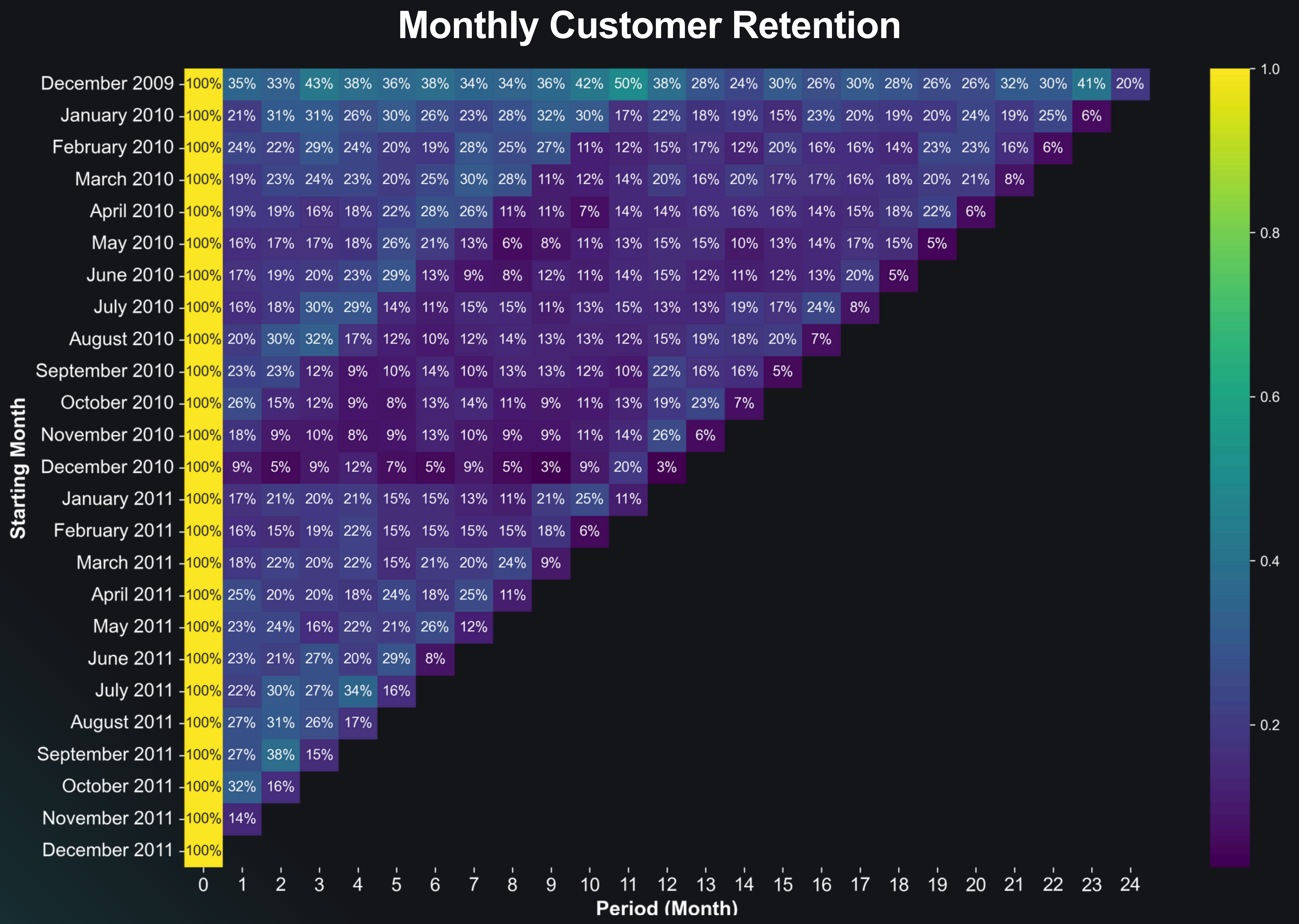
4. Inactive Customer
These customers have an RFM score of below 4, indicating they make infrequent purchases from the store, often only buying occasionally or after long periods of inactivity.

Online Retail Store

Big Question:
Who Stays? and Who Leaves?

Online stores often experience the highest customer turnover due to the abundance of choices and attractive offers available to consumers. With a large portion of both active and inactive customers, it’s essential to understand their retention patterns. Through cohort analysis, we can identify the behavior of customers from the time of their first purchase and observe whether they continue to make repeat purchases in the following months.

Online Retail Store



Overall, customer retention from the first purchase to the following months shows a fluctuating trend, mostly staying below 40%. Howeveer, there is one notable exeption: customers who made their first purchase in December 2009 maintained a stable and consistent retention rate over time.

This cohort chart reveals that most customers do not shop frequently on this platform. A likely explanation lies in the nature of the products giftware items that are typically purchased for specific occasions, rather than for regular use. As a result, repeat purchases tend to be seasonal or event-driven, rather than habitual.

Online Retail Store

Spotting the Next Goodbye Before It's Happens!

Churn Key Indicators

Declining Purchase Frequency

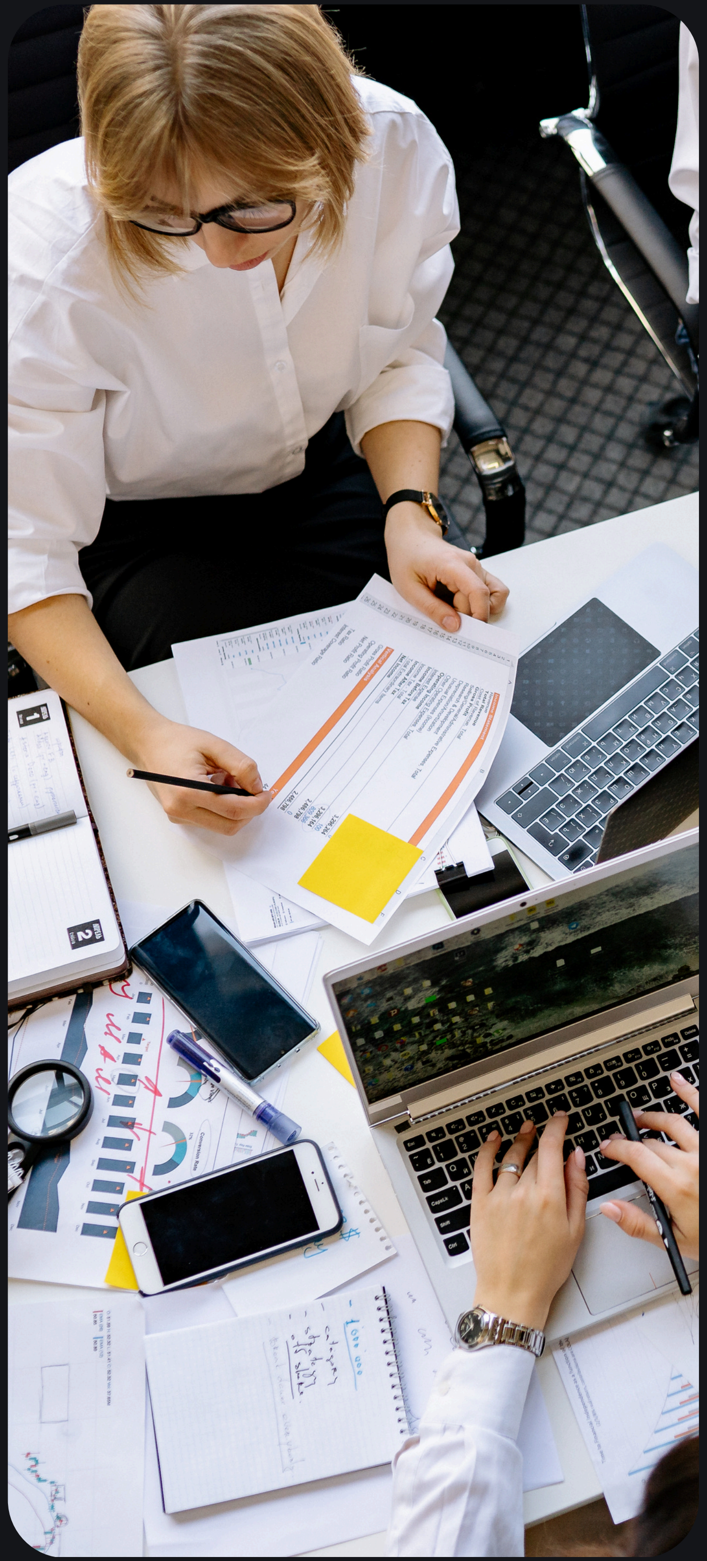
The Customers who previously made frequent purchases but have reduced or stopped their transactions over a certain period of time.

Decreased Activity on Platform

This could refer to reduced usage or fewer transactions in particular product category, indicating that customers are losing interest.

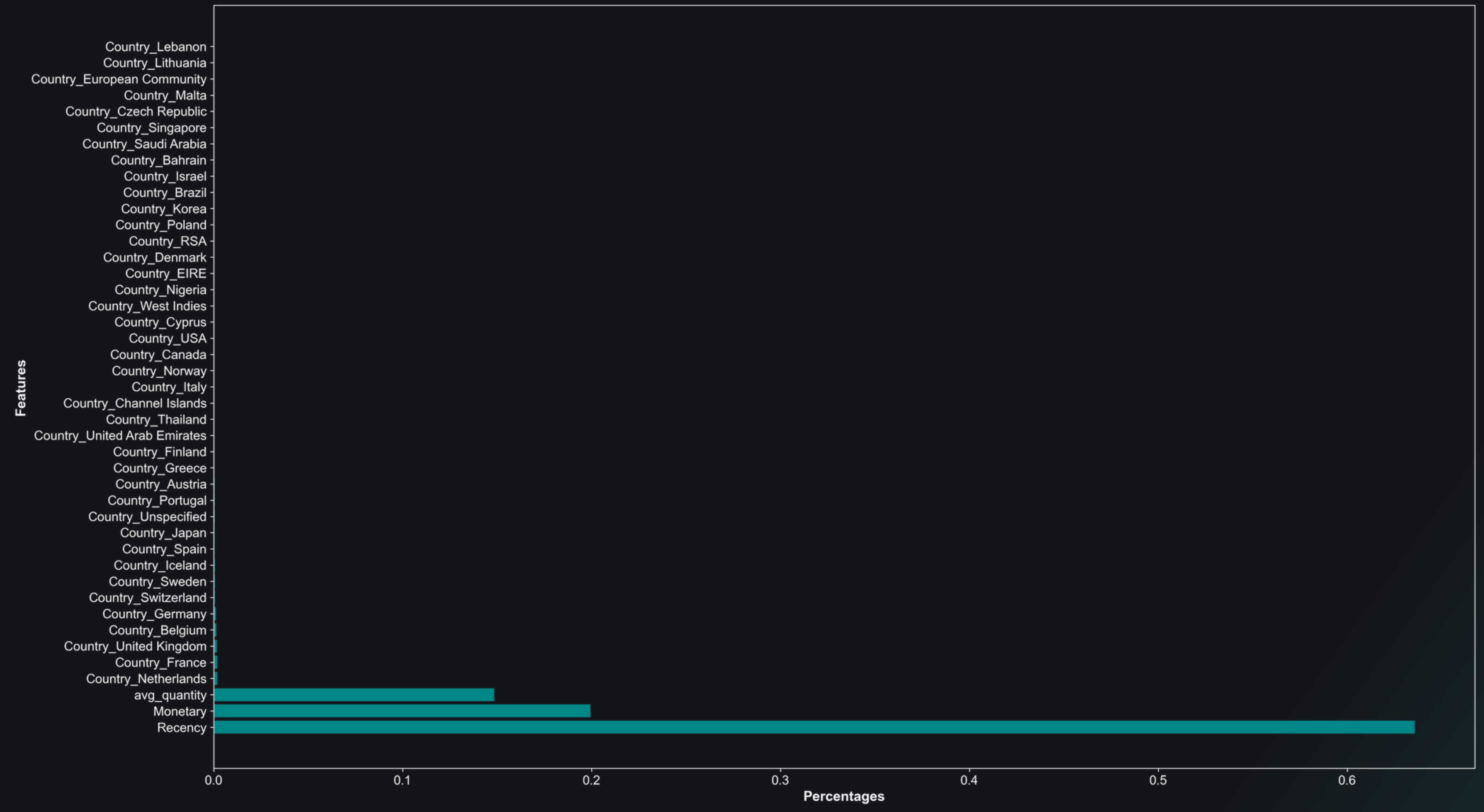
Spending pattern shift

Customers who discontinue payments entirely may indicate a risk of churn.



Online Retail Store

Key Features Indicator for Churn Predictive Model



To address this challenge, applied machine learning to identify patterns in customer behavior using the dataset. By implementing a Random Forest Classification model, we can predict the likelihood of churn based on each customer's purchase behavior as a percentage. On the right, you'll see the key indicators used by the model to estimate churn probability for every customers.

Insights

1. Key Features

This model analyzes potential customer churn by learning from behavioral patterns such as recency, monetary value, and avergae purchase quantity. by examining how recently they buy, the model identifies which customers are at higher risk of leaving.

2. Accuracy

To evaluate the model's performance, we used tools like a confusion matrix, classification report, and accuracy score. The results show that the model correctly identified 774 customers who are not likely to churn, and 266 who are. It made 39 false alarms and missed 98 actual churn cases. Overall, the model reached an accuracy of 88%, which indicates strong performance in predicting customer churn.

3. Room for improvement

Although its performance is great for running a predictive job, this model can be further upgraded by incorporating additional relevant features such as age, gender, customer engagement level, frequency of purchases, and customer satisfaction scores. By including these factors, the model's ability to accurately predict churn rates can be enhanced, leading to more precise insights and better retention strategies.

Online Retail Store

Let’s Talk Value:
Who’s Really Worth Retaining?

We can’t stop everyone from leaving, but we can choose who’s truly worth keeping around. There are many ways to determine which customers deserve our focus, but in this case, we can use Customer Lifetime Value (CLV) analysis to estimate how much revenue each customer is expected to generate over the course of their relationship with the company.

In conclusion, combining CLV analysis with churn prediction allows businesses to shift from trying to retain every customer to focusing on those who truly drive long-term growth.

Online Retail Store

CLV analysis allows us to segment customers by their projected revenue contribution. when paired with churn risk predictions, we gain a clear view of who brings the most value and who we risk losing.

Insights

1. CLV Segmentation

Customers are segmented based on their expected long-term value to the business. This value is calculated by combining each customer's monetary contribution and estimated lifespan in years.

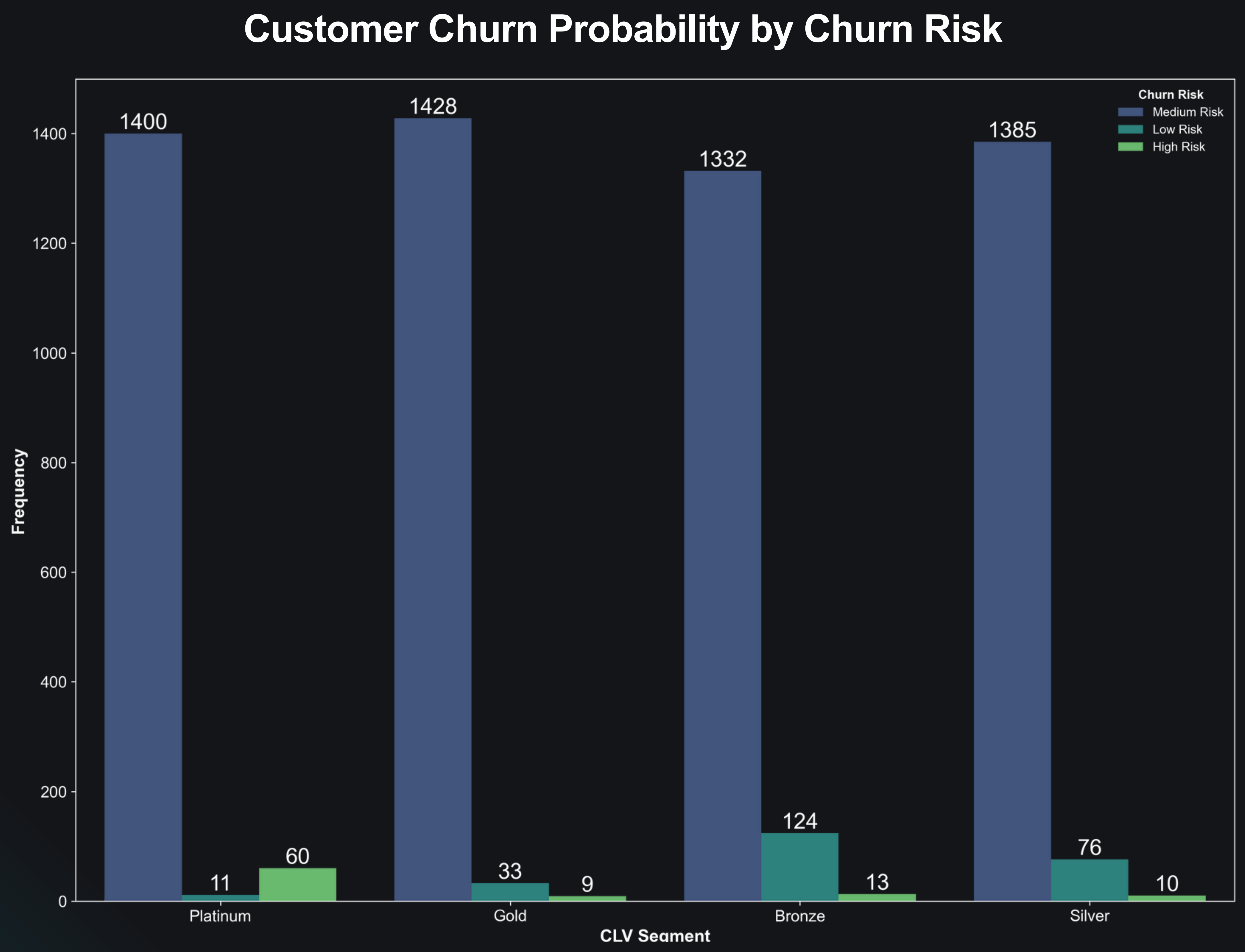
The segments are as follows:

- Bronze Customers: ≤ €290
- Silver Customers: ≤ €1,010
- Gold Customers: ≤ €3,458
- Platinum ≥ €3,459

2. Churn Risk

In this section, customers are classified based on their likelihood to churn, as predicted by a machine learning model trained on behavioral and transactional data. The segmentation thresholds are determined using data quantiles to reflect different levels of risk:

- High Risk: ≥ 76%
- Medium Risk: ≤ 75%
- Low Risk: ≤ 55%



Online Retail Store

Summary

To summarize the analysis conducted, here are the key takeaways and insights derived from the data and visualizations. These findings highlight customer behavior patterns, potential risks, and strategic opportunities form improving retention and revenue.

RFM Analysis

The majority of customers fall into the potential, active, and loyal segments, showing a strong foundation for growth and retention. With potential customers as the largest group, there's a key opportunity to convert them into loyal buyers through targeted engagement. Active customers, while consistent, may be at risk of switching-suggesting a need for loyalty incentives. Meanwhile, the smaller inactive segment presents a chance for reactivation through personalized win back campaigns. This segmentation highlights where to focus efforts to maximize customer lifetime value.

Cohort Analysis

The cohort analysis shows that most customers have low retention after their first purchase, likely due to the seasonal nature of giftware products. Retention rates generally stay below 40%, with a notable exception in the December 2009 cohort, which maintained consistent engagement likely driven by holiday demand. This suggests that timing and occasion based campaigns are key to boosting repeat purchases.

Churn Prediction Model

The churn prediction model, built using Random Forest Classification, achieved 88% accuracy in identifying at-risk customers based on behavioral patterns like recency, monetary value, and purchase quantity. It effectively distinguishes loyal vs. churn-prone users, offering actionable insights for retention. While already strong, the model can be enhanced with additional features like age, engagement, and satisfaction to boost precision and drive smarter churn prevention strategies.

CLV Analysis

Based on the CLV analysis integrated with churn probability, it is observed that customers across all CLV segments fall into a medium churn risk category. The distribution of customers is relatively even across each CLV segment. This insight highlights the opportunity to prioritize retention strategies for high value customers in the Platinum and Gold segments to proactively reduce their churn probability and enhance long-term customer royalty.

Online Retail Store

What's Our Next Move?

We have already segmented our customers based on their CLV values. To drive significant results, our retention strategies will focus primarily on the Platinum and Gold Segments (The 'Gold High Risk' segment shows null values due to the very small number of customers in that segment), which exhibit medium to high churn probabilities. The following marketing actions are recommended to increase customer retention and reduce churn risk.

1. 10% Retention - Email Marketing

A study by Dr. Somanchi in The British Journal of Administrative Management found that 68.3% of customers trust email marketing, and 78.3% find it convenient. Personalized and consistent email campaigns can improve customer engagement and boost retention by up to 10%.

2. 20% Retention - Social Media Engagement

Research by Rootman & Cupp emphasizes a strong link between social media activity and customer retention. Sharing relevant, credible content and interacting regularly with customers can reduce churn by up to 20%.

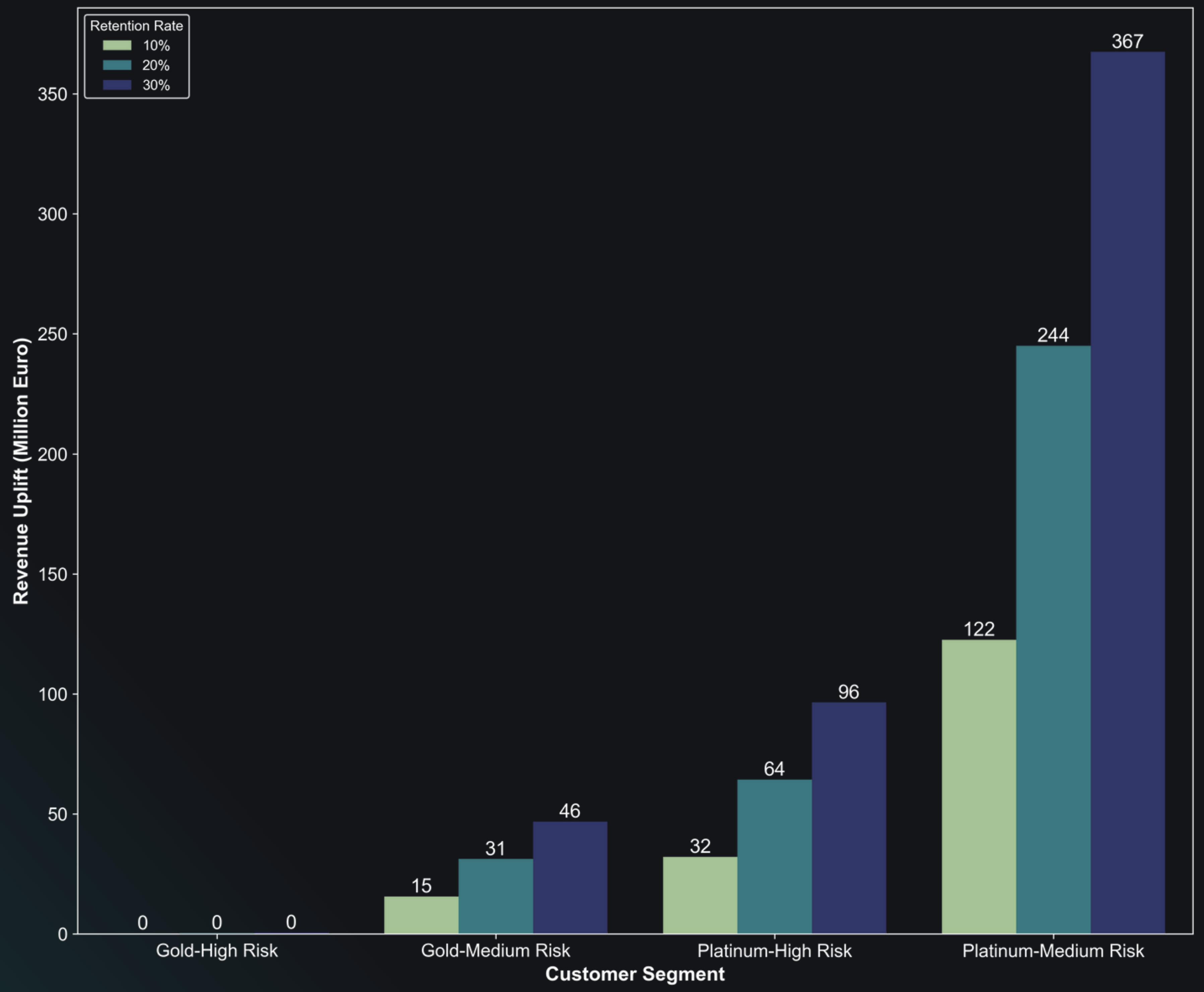
3. 30% Retention - Loyalty Programs

According to McKinsey's report Next in Loyalty, offering reward points or exclusive discounts to frequent buyers significantly improves loyalty. These programs can reduce churn by up to 30%, especially among high-value customers.

Room For Improvement

While these percentages are supported by research, it's essential to validate their effectiveness in our unique business context. To achieve more reliable and tailored results, conducting A/B testing is recommended to evaluate the performance of each strategy. This will help identify the most effective combination to maximize retention and minimize churn.

Revenue Potential From Improving Retention Rates



Let's Connect

triantowahyudi@gmail.com

Wahyudi Trianto

wahyuditrianto