



CUSTOMER SEGMENTATION

Insights from Online
Retail Datasets

PRESENTED BY: DANIEL MURUTHI >

Menu

About Us

Add to Cart





AGENDA

[Menu](#)[About Us](#)[Add to Cart](#)

- > Project Overview
 - > Objectives
 - > Data Description
 - > Data Preprocessing
 - > EDA
 - > Feature Engineering
 - > Modelling
 - > Key Conclusions
 - > Recommendations
 - > Future Work
-



PROJECT OVERVIEW

› **Purpose**

Customer Segmentation Using RFM Analysis and KMeans Clustering

- Utilized RFM (Recency, Frequency, Monetary) analysis to quantify customer value.
- Applied KMeans clustering to group customers into distinct segments.

› **Approach**

Identified actionable customer segments to inform business decisions and marketing efforts.

Menu

About Us

Add to Cart





OBJECTIVES

[Menu](#)[About Us](#)[Add to Cart](#)

>Analyze customer purchasing patterns using RFM metrics.

Segment customers into meaningful groups using KMeans clustering.

Evaluate the effectiveness of the clustering model.

Provide strategic recommendations based on customer segments.



DATASET DESCRIPTION

Source : UC Irvine Machine Learning Repository - Online Retail dataset & Online Retail II dataset

Features :

- **InvoiceNo:** Unique identifier for each transaction; a 6-digit number.
- **StockCode:** Unique identifier for each product; a 5-digit number.
- **Description:** Name of the product.
- **Quantity:** Number of units purchased per transaction.
- **InvoiceDate:** Date and time when the transaction occurred.
- **UnitPrice:** Price per unit in sterling (£).
- **CustomerID:** Unique identifier for each customer; a 5-digit number.
- **Country:** Country of residence of the customer.

Total Records :

- **Online Retail - 541,909 Records**
- **Online Retail II - 1,067,371 Records**





DATASET PREPROCESSING

>Data Cleaning

- Removed duplicate entries and missing values.
- Filtered out canceled transactions and negative quantities.

Feature Engineering

- Calculated RFM metrics for each customer.
- Applied log transformation to the Monetary value to reduce skewness.

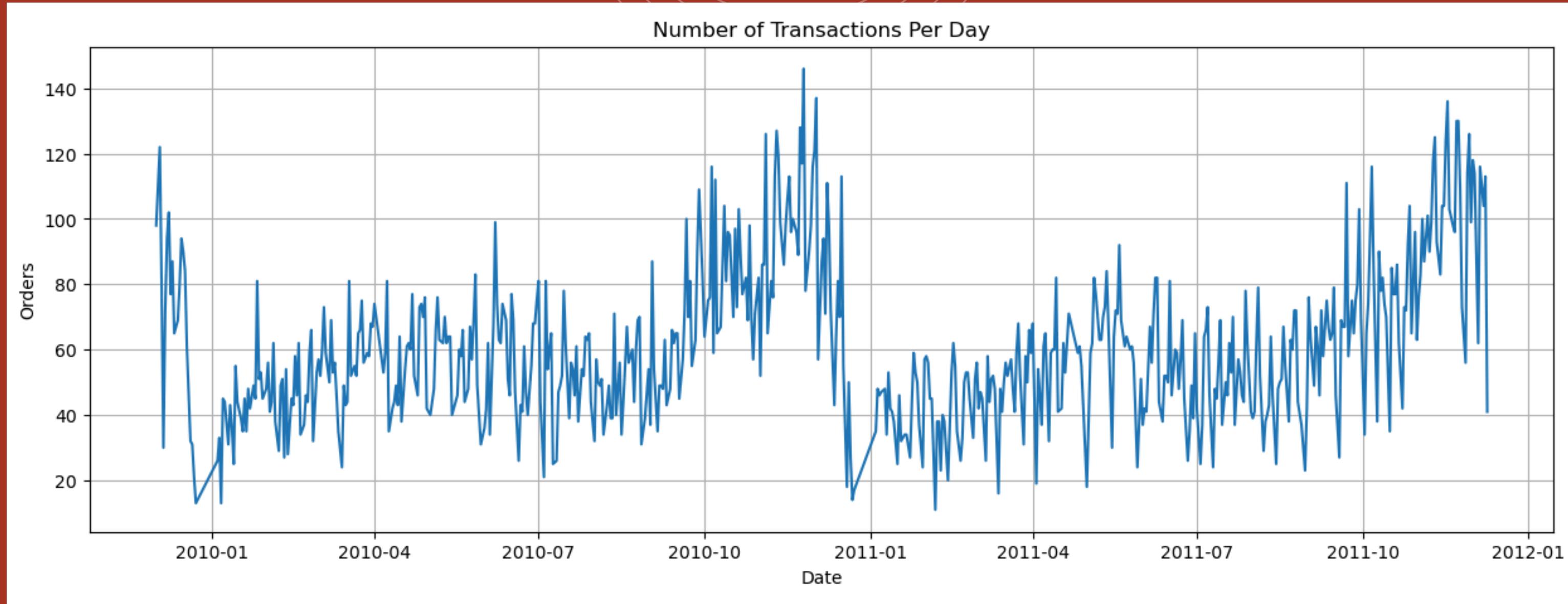
Normalization

- Standardized RFM features using StandardScaler to ensure equal weighting in clustering.



EXPLORATORY DATA ANALYSIS (EDA)

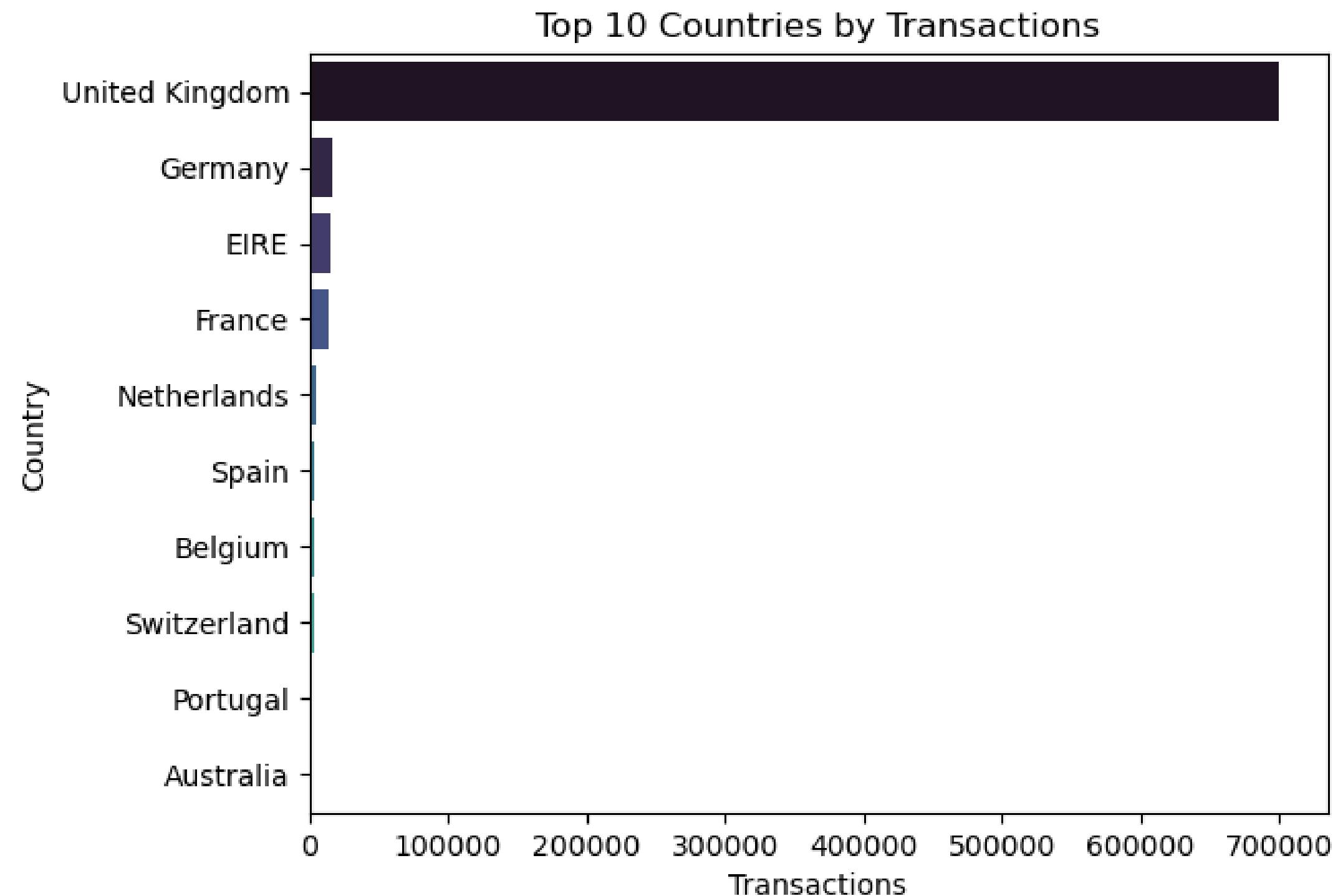
Transaction Patterns



- Seasonal Peaks: Clear Q4 (Oct-Dec) spikes in both 2010 and 2011, reflecting holiday shopping seasons
- Growth Trend: Significant volume increase in late 2011 suggests business expansion
- Daily Volatility: Transactions fluctuate widely (15-145 orders daily), requiring flexible operations

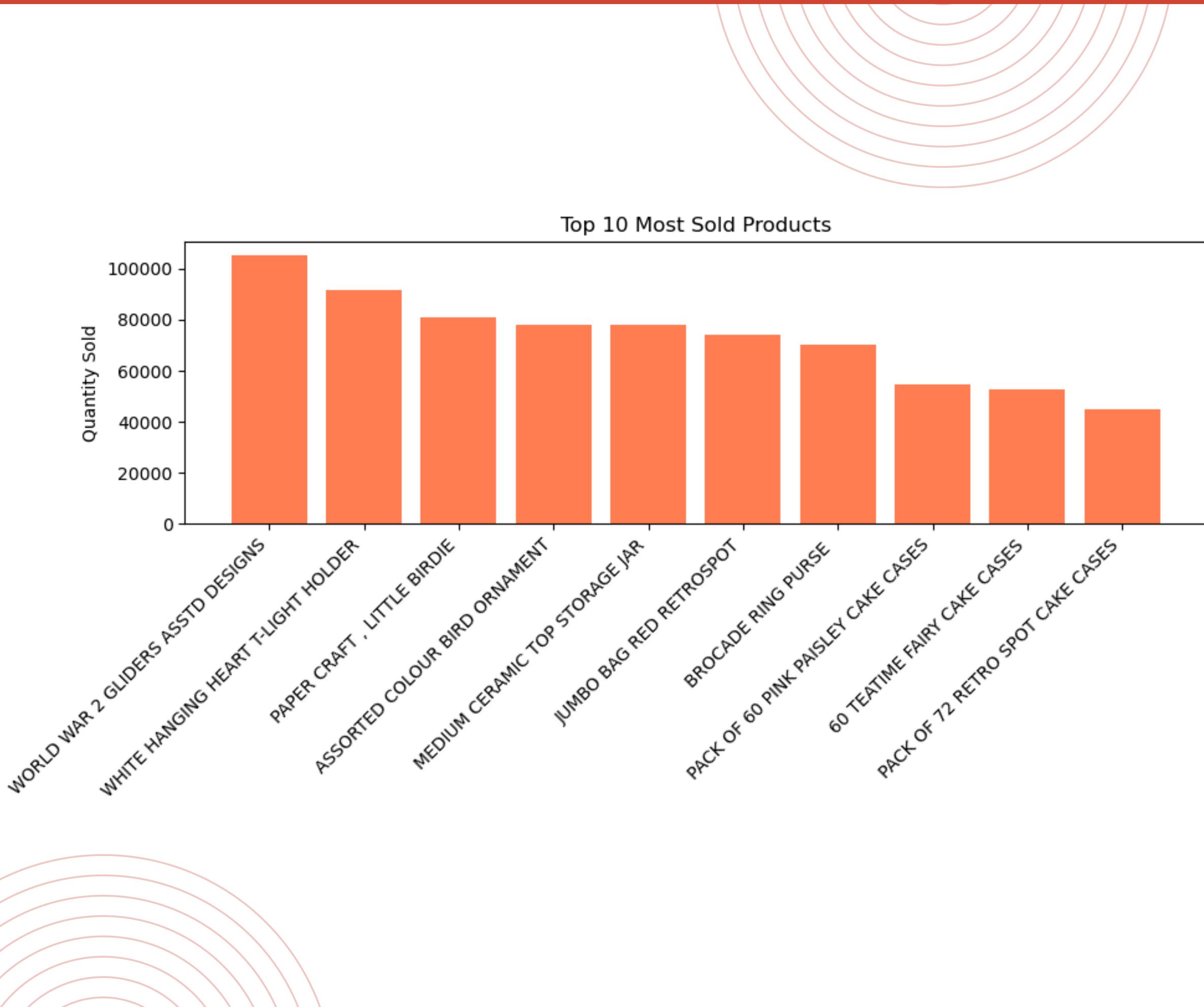
Geographic Distribution

- UK Dominance: UK market represents approximately 25x more transactions than Germany, showing extreme market concentration
- European Footprint: 9 of top 10 countries are European markets (Australia being the only exception)
- Growth Potential: Significant drop-off after UK indicates opportunity for targeted expansion in secondary markets



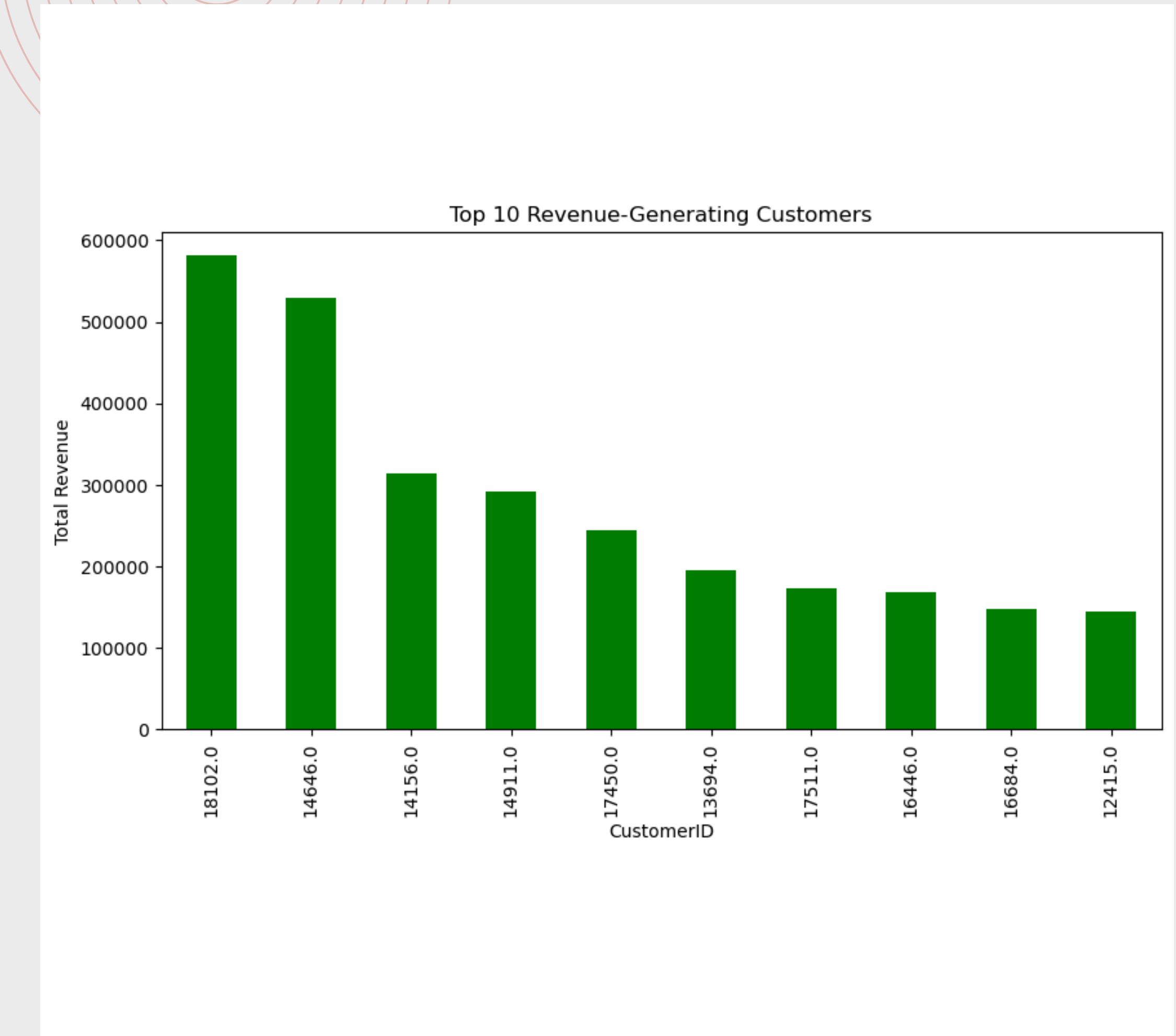
Product Performance

- Top Performer: "WORLD WAR 2 GLIDERS ASSTD DESIGNS" leads sales with approximately 100,000 units
- Category Trends: Decorative items dominate the bestseller list, including hanging ornaments, storage items, and gift products
- Sales Distribution: Gradual decrease in volume across top 10 products, with the 10th bestseller selling ~45% of the top item



Customer Value Analysis

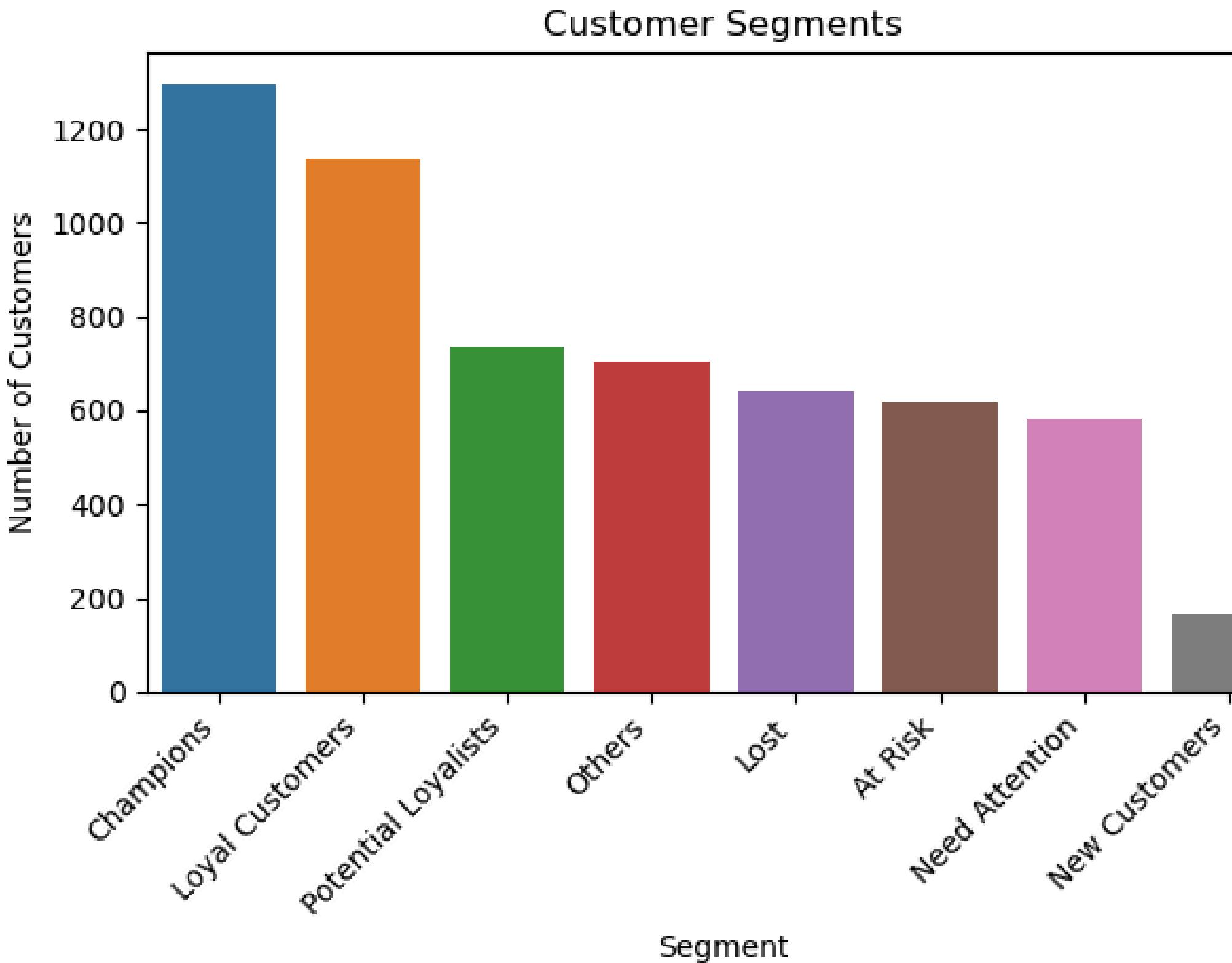
- Revenue Concentration: Top two customers (18102 and 14646) generate significantly higher revenue than others, showing a Pareto-like distribution
- Top Customer Value: Customer 18102 generates nearly 600,000 in revenue, approximately 4x more than the 10th highest customer
- Value Gap: Substantial drop (approximately 40%) between the 2nd and 3rd highest-value customers, highlighting key account dependencies



FEATURE ENGINEERING



RFM Analysis - Customer Segmentation



RFM Segmentation Approach:

- Recency: Days since last purchase
- Frequency: Number of purchases
- Monetary: Total spending

Customer Segments Distribution:

- 8 distinct segments identified
- Champions (1,297) and Loyal Customers (1,138) form our largest segments
- Significant opportunity groups: Potential Loyalists (734) and At Risk (616)



Revenue Contribution by Segment

[Menu](#)[About Us](#)[Add to Cart](#)

Key Revenue Insights:

- **Champions: 68.26% of revenue (highest value segment)**
- **Loyal Customers: 14.77% of revenue (consistent buyers)**
- **At Risk: 8.67% of revenue (need immediate re-engagement)**
- **Top 3 segments contribute >91% of total revenue**



Strategic Focus:

- **Retention strategies for Champions and Loyal Customers**
- **Re-engagement campaigns for At Risk segment**
- **Nurturing programs for Potential Loyalists**
- **New Customer conversion pathway for growth**



MODELLING

a deep understanding of
consumer purchasing
behavior



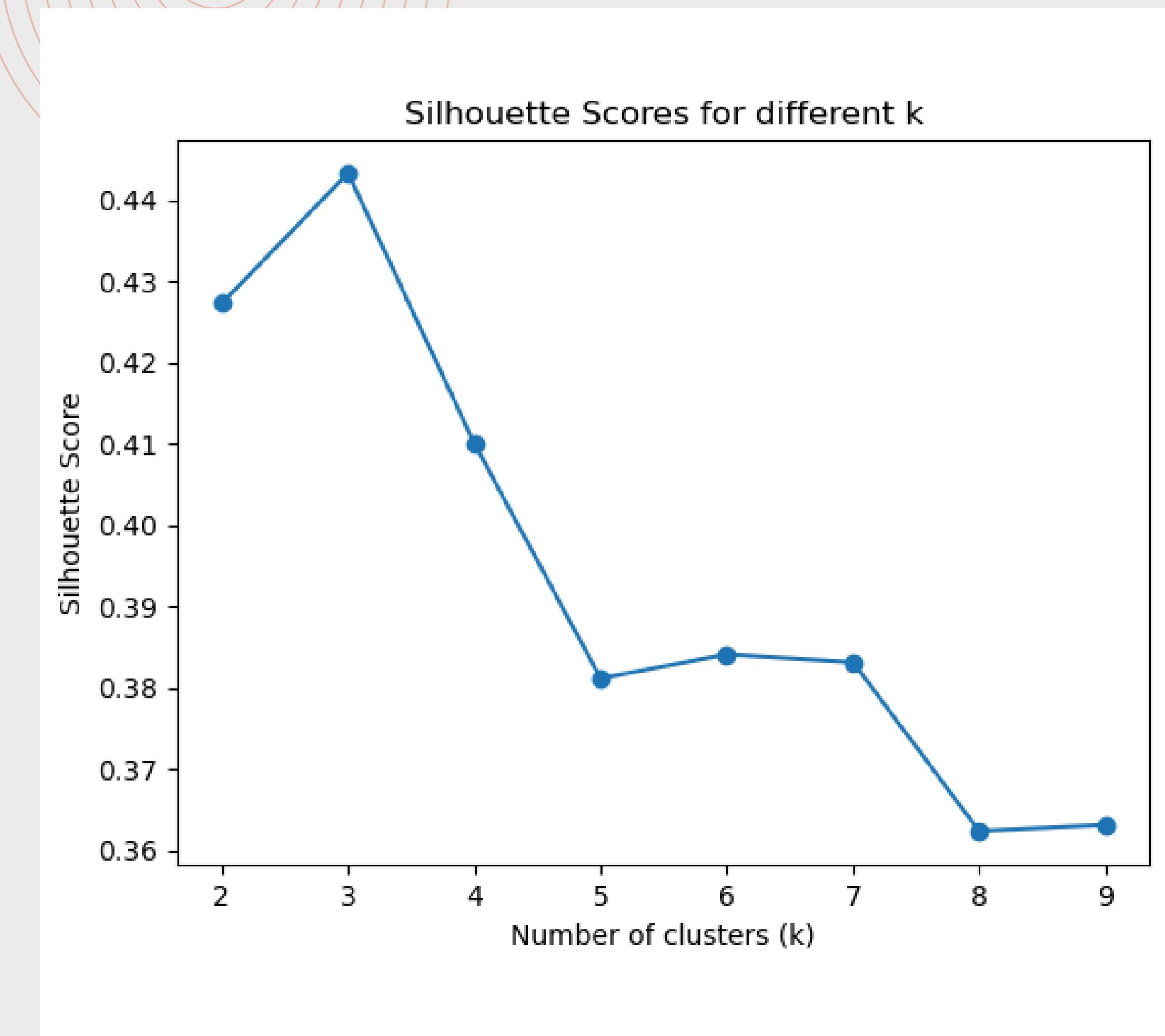
K-means Clustering

Clustering Approach:

- Applied K-means clustering to RFM features (Recency, Frequency, Monetary)
- Standardized features to ensure equal weighting
- Determined optimal cluster count using silhouette analysis

Finding the Optimal K:

- Tested cluster counts from k=2 to k=9
- Highest silhouette score at k=3 (0.4433)
- Score indicates moderately well-defined clusters



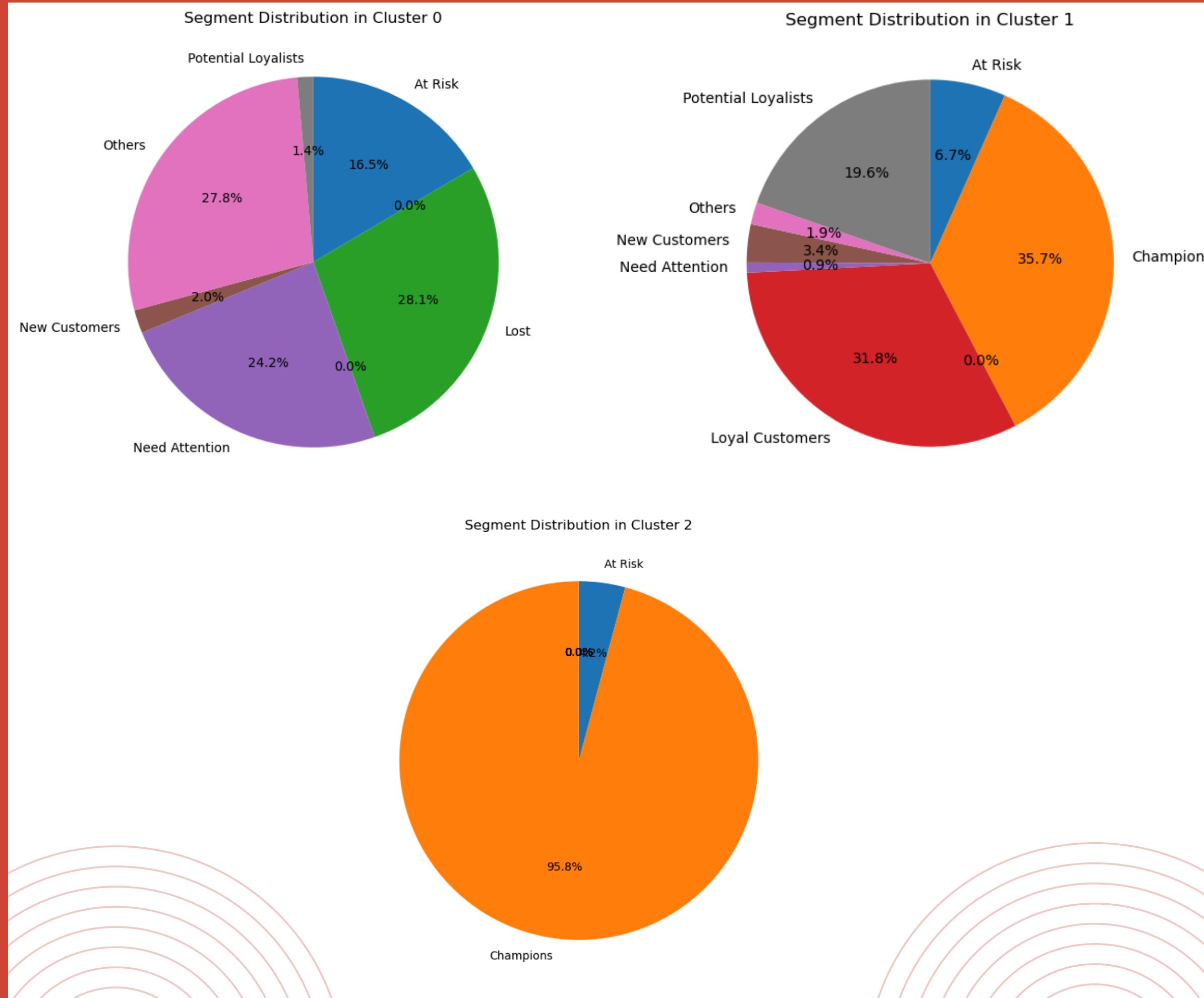
Cluster Analysis & Segment Composition

Cluster Quality Metrics:

- Silhouette Score: 0.4433 (moderate to good separation)
- Davies-Bouldin Index: 0.7397 (well-separated clusters)

Cluster Segment Composition:

- Cluster 0 (Hibernating/At Risk): Primarily Lost (28.1%), Others (27.8%), Need Attention (24.2%), and At Risk (16.5%) customers
- Cluster 1 (Engaged & Growing): Mix of Champions (35.7%), Loyal Customers (31.9%), and Potential Loyalists (19.7%)
- Cluster 2 (VIP / Top Champions): Almost exclusively Champions (95.8%)





Key Conclusions

[Menu](#)[About Us](#)[Add to Cart](#)

High-Value Customer Insights:

- **Champions and Loyal Customers (Clusters 1 & 2) drive majority of revenue**
- **Strong engagement indicators: high frequency and monetary values**
- **VIP segment (Cluster 2) consists almost exclusively of Champions (95.8%)**

At-Risk Customer Patterns:

- **Cluster 0 represents lower engagement (low recency and frequency)**
- **Contains majority of Lost (28.1%), Need Attention (24.2%), and At Risk (16.5%) segments**
- **Significant opportunity for targeted re-engagement**

Segmentation Value:

- **K-Means clustering enhanced traditional RFM segmentation**
- **Created actionable, behavior-based customer groups**
- **Revealed nuanced patterns for strategic decision-making**



Strategic Recommendations



For Champions & VIPs
(Clusters 1 & 2):

- Implement exclusive loyalty programs and VIP benefits
- Create personalized premium offers based on purchase history
- Develop ambassador programs to leverage their high engagement



For At-Risk Customers
(Cluster 0):

- Launch targeted re-engagement campaigns with compelling incentives
- Implement win-back strategies with special limited-time offers
- Analyze churn patterns to address underlying satisfaction issues



For Growth Segments:

- Design specialized onboarding journeys for New Customers
- Create pathway programs to convert Potential Loyalists to Loyal Customers
- Implement monitoring systems to identify customers needing attention

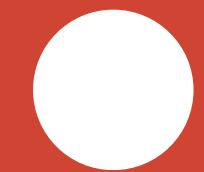
Menu

About Us

Add to Cart



FUTURE STEPS



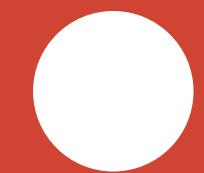
Enhance Data Integration:

- Incorporate demographic and behavioral data to enrich customer profiles.
- Utilize data from multiple channels to gain a holistic view of customer interactions.



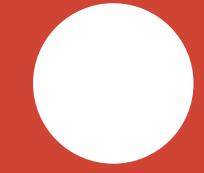
Explore Advanced Clustering Techniques:

- Experiment with algorithms like DBSCAN or Hierarchical Clustering to validate and potentially improve segmentation results.



Develop a Real-Time Dashboard:

- Create an interactive dashboard to monitor customer segments and key metrics in real-time.
- Enable dynamic updates to track the effectiveness of marketing strategies over time.



Implement A/B Testing:

- Test different marketing approaches across segments to determine the most effective strategies.
- Use insights to refine campaigns and improve customer engagement.



Regularly Update Segmentation Models:

- Schedule periodic reviews of segmentation models to account for changing customer behaviors and market trends.
- Ensure that marketing strategies remain aligned with current customer needs.



Menu

About Us

Add to Cart

thank you



Project Code



adinomuruthi1@gmail.com