

Customer Segmentation via KMeans Clustering

*RFM + Interpurchase Time (KMeans / Python)

1. Project Purpose

- *Overview*

This project aimed to perform quantitative customer segmentation on real-world retail transaction data from a UK online store. Beyond standard RFM (Recency, Frequency, Monetary) analysis, I introduced an additional feature: **Interpurchase Time** — capturing customers' average repeat cycle to enable more precise behavioral targeting.

- *Analytical Assumptions*

- Combining RFM with Interpurchase Time can provide a more nuanced segmentation that reflects customers' purchasing rhythms, not just their raw transaction counts.
- The resulting clusters can translate into actionable marketing strategies tailored to different behavioral segments.

2. Data Preparation

- Source: UCI Machine Learning Repository – Online Retail Dataset (541,909 records across ~8 months)
- Key Variables: InvoiceNo, StockCode, Quantity, UnitPrice, InvoiceDate, CustomerID, Country
- Cleaning Process:
 - Removed rows with missing CustomerIDs; ignored missing Description fields (not used in analysis)
 - Removed outliers in Quantity and UnitPrice using IQR filtering to ensure clustering stability.

3. Methodology & Modeling

| RFM + Interpurchase Time Setup

- Recency: Days since last purchase
- Frequency: Total unique transactions per customer
- Monetary: Total purchase value
- Interpurchase Time: $[(T_n - T_1) / (\text{Frequency} - 1)]$

→ **Added this feature because two customers with the same Frequency might show very different purchasing rhythms, and this matters for behavioral segmentation.**

| KMeans Clustering

- Used Elbow Method to determine $k=6$ clusters.
- Evaluated cluster quality using Silhouette Score = 0.6279 (considered a strong separation between clusters).

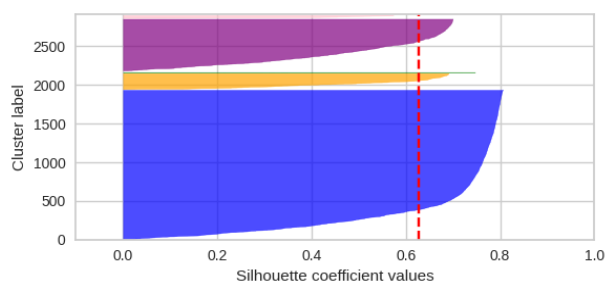
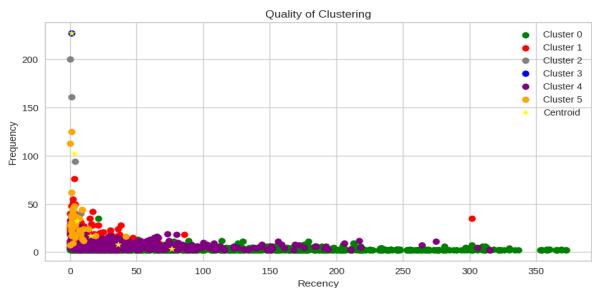
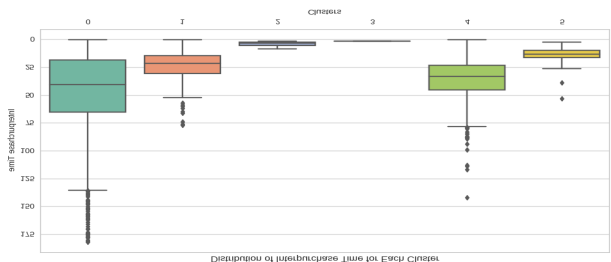
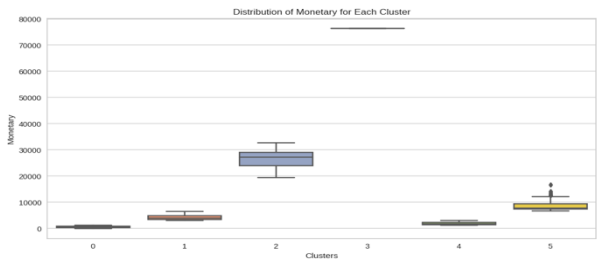
4. Summary

Cluster	Behavior	Strategy
3	Very high frequency, spending, short intervals	Premium targeting, long-term retention
2	High repeat patterns, strong spending power	Membership & referral program enhancement
1	Medium activity level	Custom retention strategies
0, 4	Low frequency, low spending	Re-engagement campaigns, remarketing

→ These insights were derived by interpreting the average RFM + Interpurchase Time metrics per cluster.

5. Visualizations

- Boxplot (Frequency / Cluster): Showed frequency distributions across clusters to identify high-activity groups.
- Scatterplot (Recency vs Frequency): Explored the relationship between recency and frequency to understand structural differences.
- Silhouette Plot: Confirmed the cohesion and separation of clusters.



6. Applicability & Extensions

- Cross-Industry Use:
By capturing mid- to long-term repeat rhythms, Interpurchase Time enables precise prediction of return timing, sharpening remarketing priorities and unlocking cross-industry applications from retail to donor and subscriber retention.