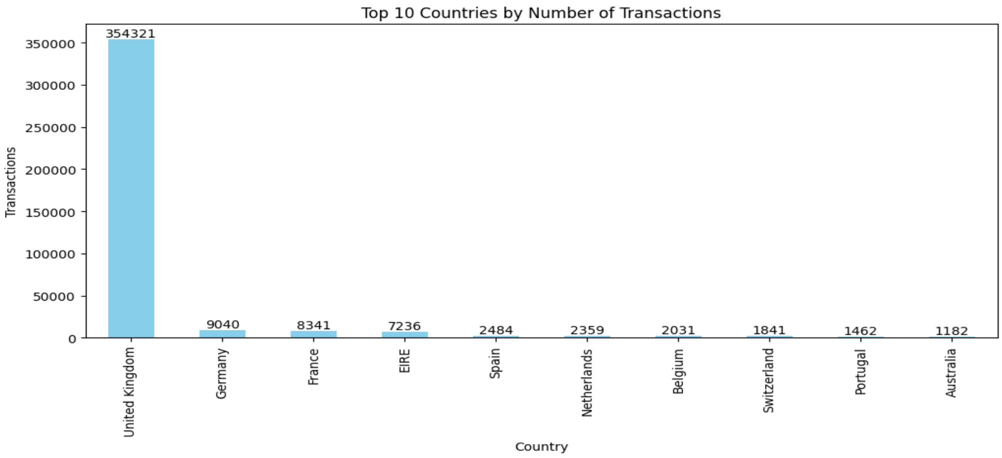
Shopper Spectrum Project Report

Analysis Report

1. Top 10 Countries by Number of Transactions



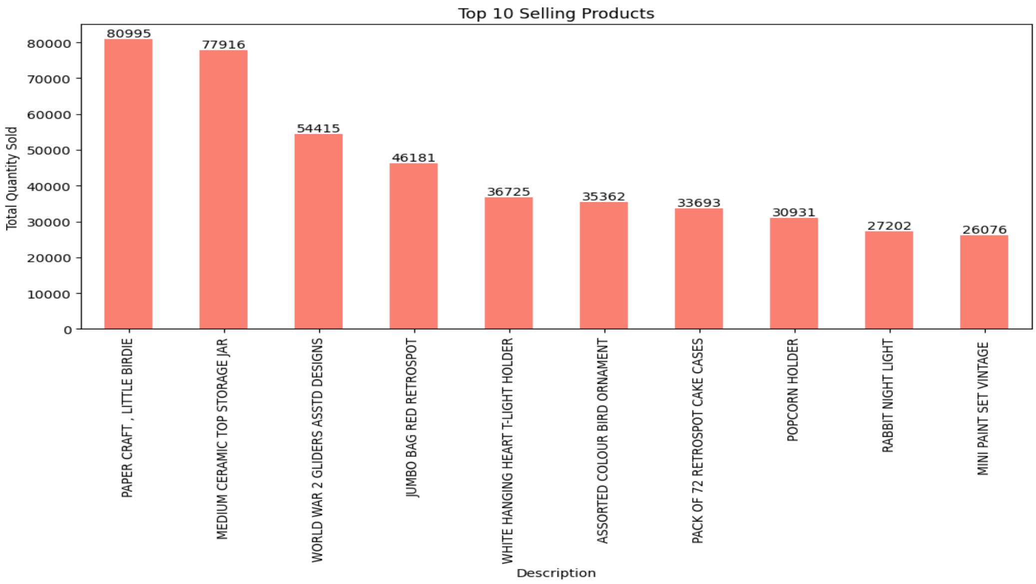
- The graph shows that the United Kingdom clearly dominates transactions with 354,321, making it the main market and largest contributor to business activity. Other countries have much lower transaction counts in comparison, with Germany, France, and EIRE forming a second group with moderate engagement.
- The remaining countries—Spain, Netherlands, Belgium, Switzerland, Portugal, and Australia—each have fewer than 2,500 transactions. This indicates limited market presence or smaller customer bases outside the UK.

Conclusion:
The business is heavily dependent on the UK market. While a few European countries show potential, there is a clear opportunity to strengthen presence and expand transactions outside the UK to reduce dependency on a single country.

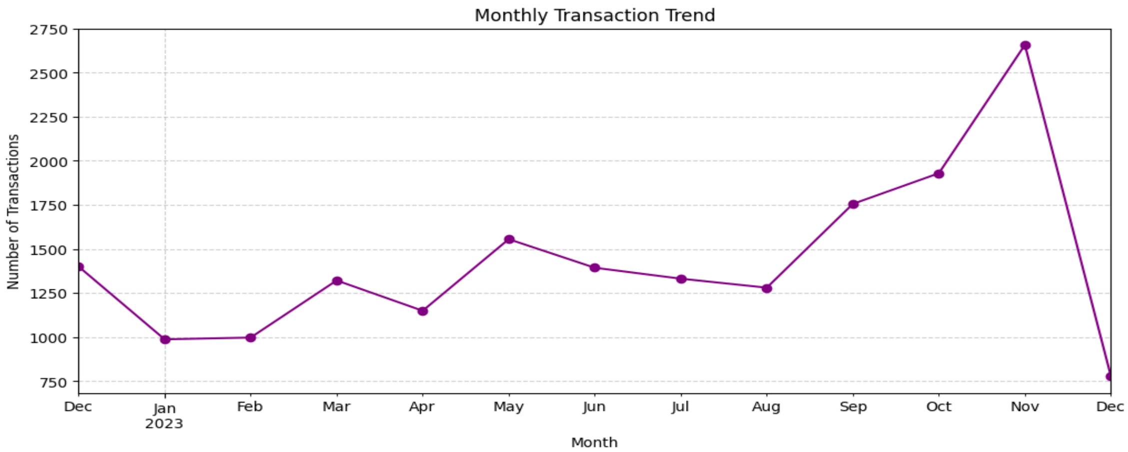
2. Top 10 Selling Products

- The graph shows that Paper Craft, Little Birdie is the top-selling product, followed closely by the Medium Ceramic Top Storage Jar, indicating very high demand for these items.
- Products like World War 2 Gliders Asstd Designs and Jumbo Bag Red Retrospot also perform strongly with good sales volumes. The remaining products show steady but comparatively lower sales. Overall, sales are concentrated among a few popular products, with gradual decline across the rest.

Conclusion:
A few products contribute most of the sales, with *Paper Craft, Little Birdie* leading overall. Sales gradually decrease across the remaining top products, showing that demand is concentrated on a small set of popular items.



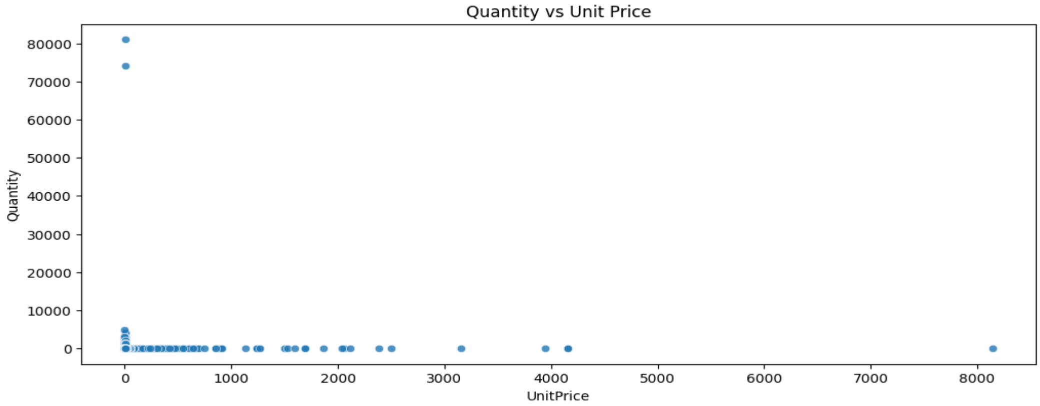
3. Monthly Purchase Trends



- The graph shows monthly transaction fluctuations across the year with no constant pattern in the early months. Transactions remain moderate from January to August with small ups and downs.
- From September onward, there is a clear increase, reaching the highest point in November. December shows a sharp decline in transactions compared to previous months.

Conclusion:
Transactions peak toward the end of the year, especially in November, indicating strong seasonal demand before a significant drop in December.

4. Quantity vs Unit Price (Outlier Detection)

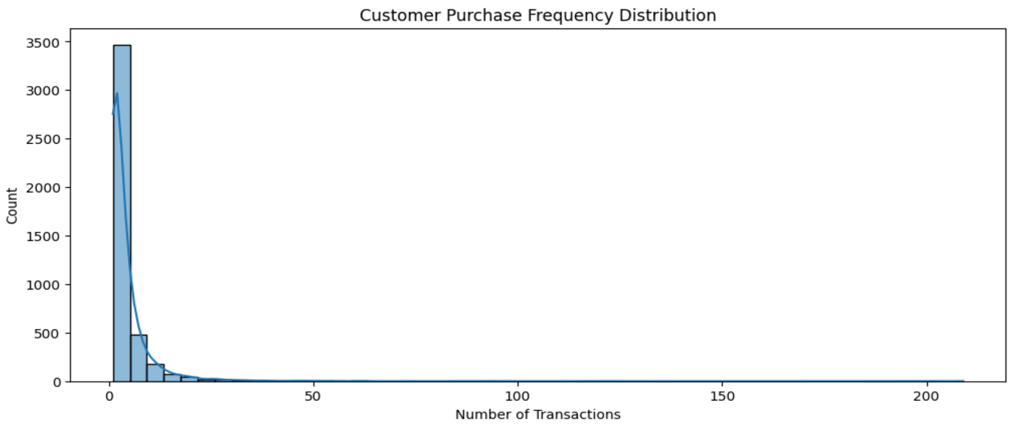


- The scatter plot shows the relationship between unit price and quantity sold. Most transactions are concentrated at low unit prices with higher quantities, indicating that cheaper products are sold more frequently.
- As the unit price increases, the quantity sold generally decreases, showing lower demand for expensive items. A few extreme points indicate very high quantities at very low prices and very high prices with very low quantities.

Conclusion:

There is an inverse relationship between unit price and quantity sold, where lower-priced products drive higher sales volumes while higher-priced products sell in smaller quantities.

5. Current Purchase Frequency Distribution



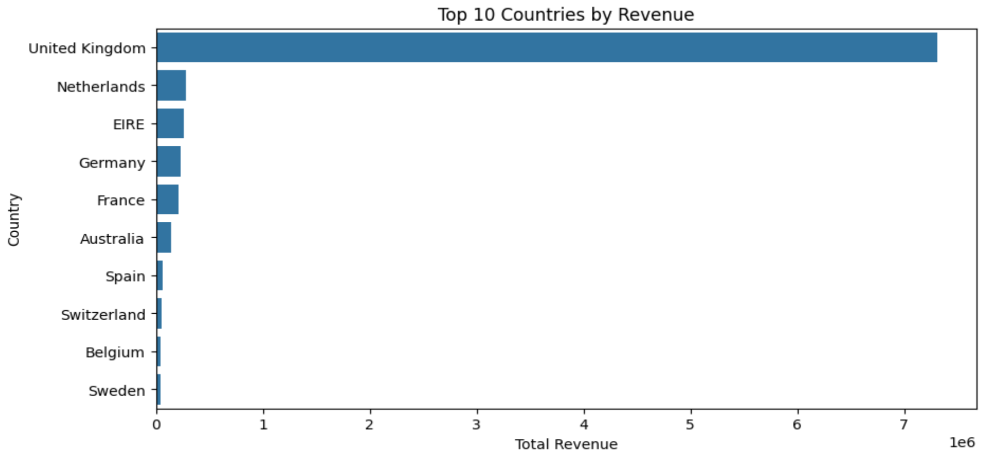
- The histogram shows how often customers make purchases. Most customers have a very low number of transactions, indicating many one-time or occasional buyers.

- As the number of transactions increases, the customer count drops sharply. Only a small group of customers make frequent purchases, forming a long tail on the right side of the distribution.

  Conclusion:

  The customer base is dominated by occasional buyers, while a small number of loyal customers contribute to high purchase frequency.

6. Revenue Contribution by Top 10 Countries

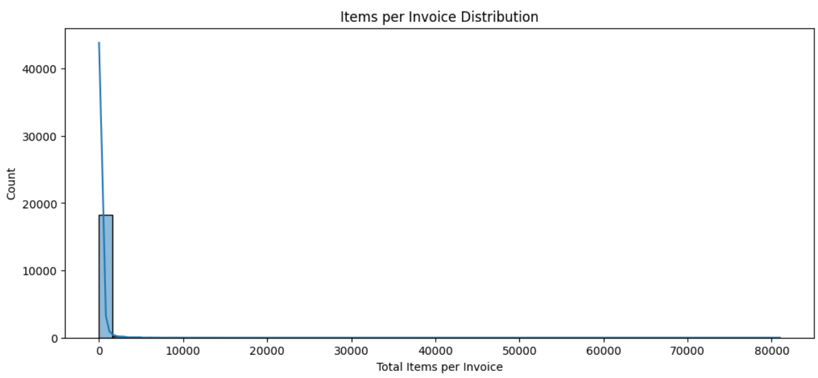


- The chart shows the total revenue generated by the top 10 countries. The United Kingdom contributes the highest revenue by a very large margin, clearly dominating overall sales.
- Other countries such as the Netherlands, EIRE, Germany, and France generate much smaller revenue in comparison. The remaining countries contribute minimal revenue, indicating limited business activity outside the UK.

  Conclusion:

  Revenue is heavily concentrated in the United Kingdom, highlighting strong dependence on a single market and potential opportunities for growth in other countries.

7. Invoice Size Distribution (Items Per Invoice)



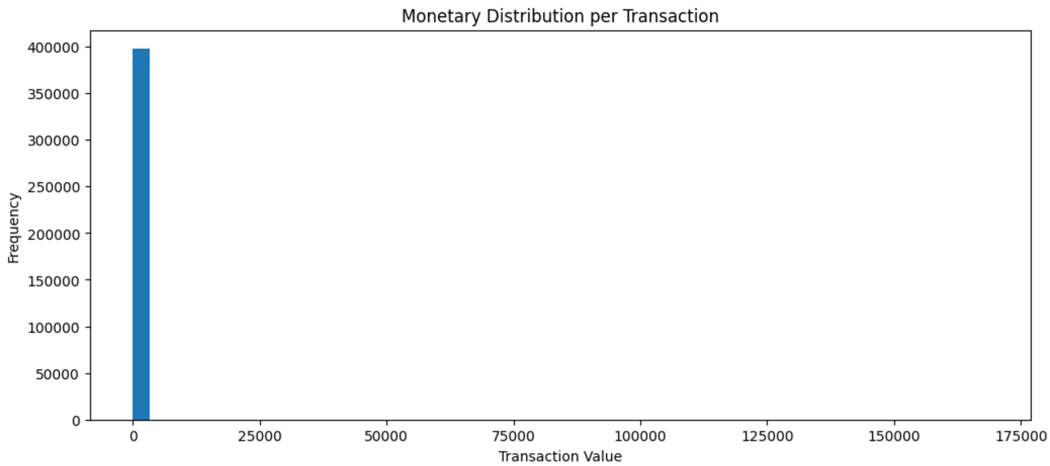
➢ The distribution shows the number of items included in each invoice. Most invoices contain a very small number of items, indicating that customers usually make small purchases.

➢ As the number of items per invoice increases, the frequency drops sharply. A few extreme cases with very large item counts create a long right tail in the distribution.

Conclusion:

Most transactions involve small orders, while a small number of invoices with bulk purchases significantly increase the item count range.

8. Monetary Per Transactions Distribution



➢ The graph shows the distribution of monetary value per transaction. Most transactions have low monetary values, indicating that customers usually make small purchases.

➢ As transaction value increases, the frequency drops sharply. A few very high-value transactions create a long right tail in the distribution.

Conclusion:

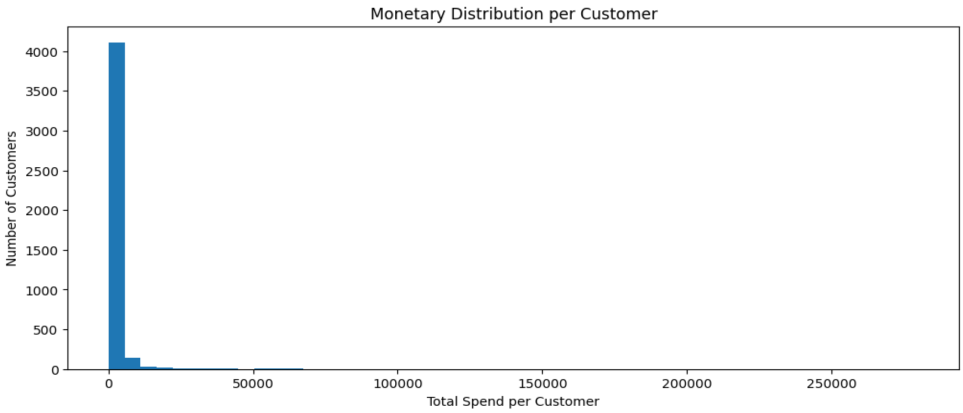
The business is driven mainly by many low-value transactions, with only a small number of high-value purchases contributing to overall revenue.

9. Monetary Per Customer Distribution

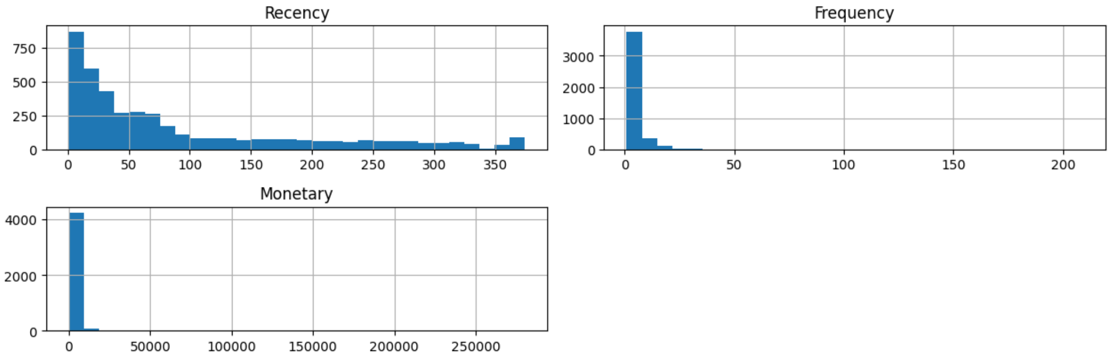
➢ The chart shows the distribution of total spending per customer. Most customers have low overall spending, indicating many small or infrequent buyers.

- As total spend increases, the number of customers drops sharply. A small group of customers contributes very high spending, forming a long right tail.

Conclusion:
Overall revenue is influenced strongly by a small number of high-value customers, while the majority contribute relatively low total spending.



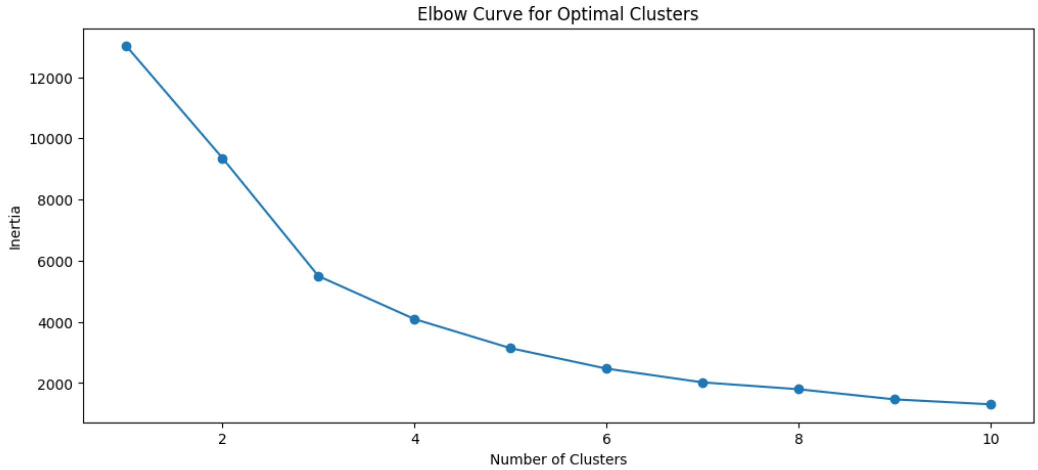
10. Visualization of RFM Distribution



- The RFM charts show the distribution of Recency, Frequency, and Monetary values across customers. Most customers have low recency values, meaning they have purchased recently, while a smaller group has not purchased for a long time.
- The frequency chart shows that most customers make very few purchases, with only a small number buying frequently. The monetary distribution indicates that most customers spend small amounts, while a few high-value customers spend significantly more.

Conclusion:
The customer base mainly consists of recent but low-frequency, low-spending customers, with a small segment of loyal and high-value customers driving a large share of revenue.

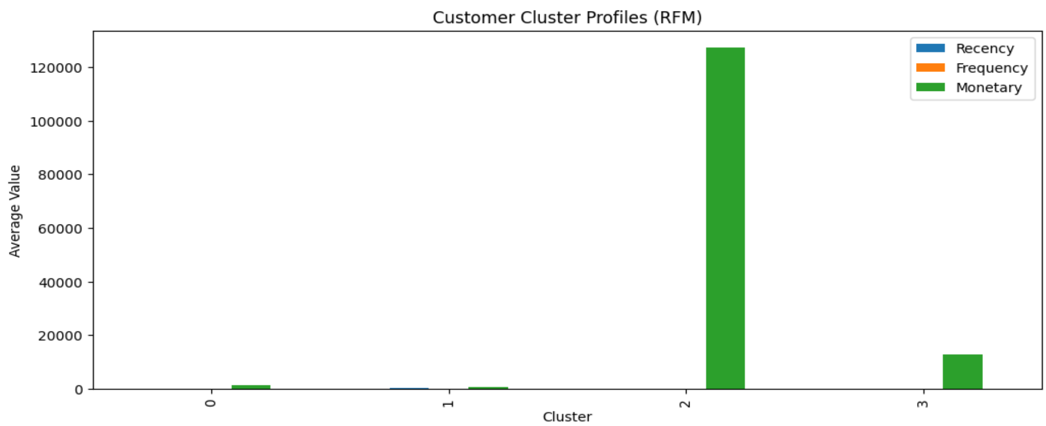
11. Elbow Curve for Optimal Clusters



- The elbow curve shows how inertia decreases as the number of clusters increases. There is a steep drop in inertia from 1 to around 3 clusters, indicating significant improvement in clustering.
- After 3 or 4 clusters, the decrease becomes more gradual, showing diminishing returns. This flattening of the curve suggests an optimal balance between model simplicity and performance.

Conclusion:
The optimal number of clusters is around 3 or 4, as adding more clusters beyond this point provides only minor improvement.

12. Customers Cluster Profiles

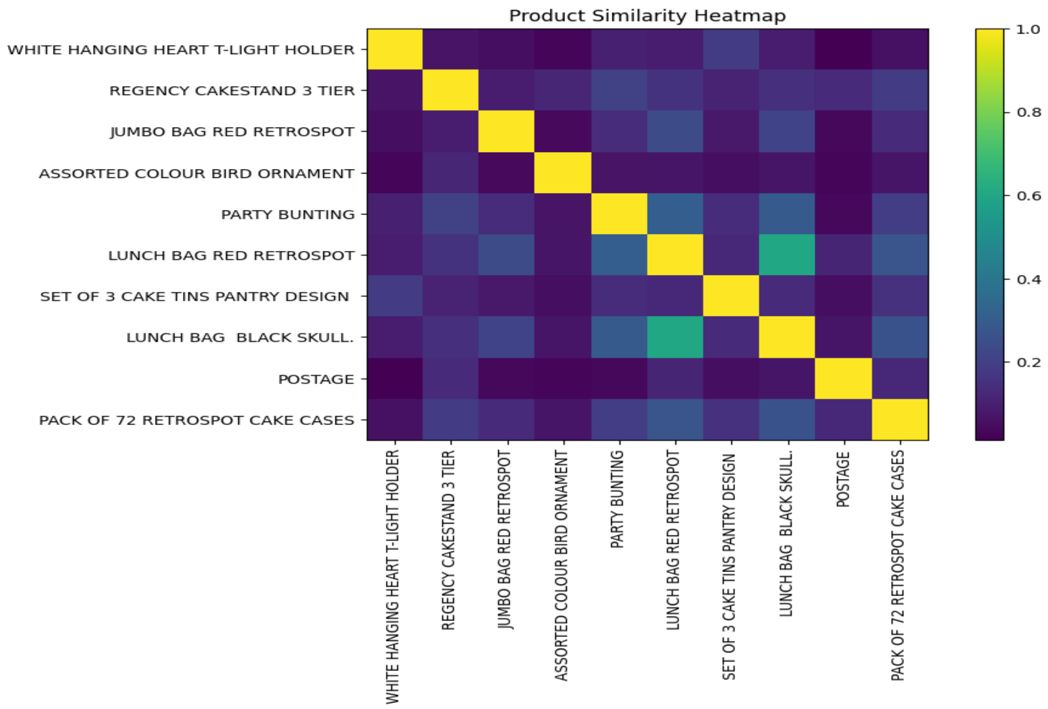


- The chart compares customer clusters based on average Recency, Frequency, and Monetary (RFM) values. Cluster 2 clearly stands out with extremely high monetary value, indicating a small group of very high-value customers.

- Cluster 3 also shows strong monetary contribution but at a lower level. Clusters 0 and 1 have much lower spending and purchase frequency, representing more regular or low-value customers.

Conclusion:
A small number of high-value customers (especially Cluster 2) drive most of the revenue, while the majority of customers belong to lower-value clusters with limited spending.

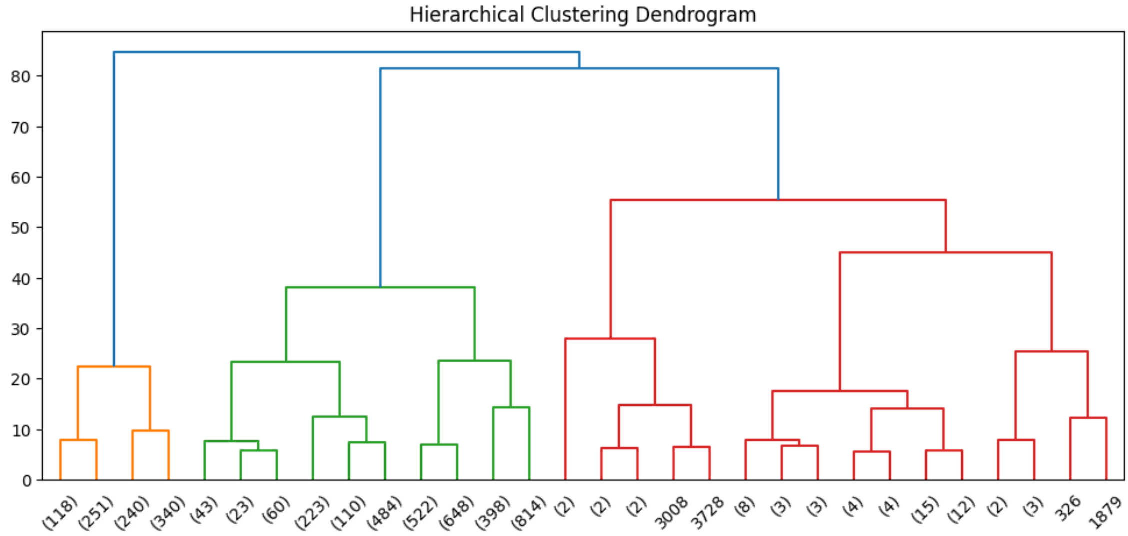
13. Product Similarity Heatmap



- The heatmap shows the similarity between different products based on customer purchasing patterns. Most product pairs have low similarity, indicating they are usually bought independently.
- Some related items, especially similar bag products, show higher similarity, suggesting they are often purchased together. This highlights clear product groupings driven by customer buying behaviour.

Conclusion:
Only a few products have strong similarity and cross-selling potential, while most items are purchased separately, indicating distinct customer preferences.

14. Hierarchical Clustering Dendrogram



- The dendrogram shows how customers are grouped based on similarity in their behaviour. Customers merge into small clusters at lower distances, indicating very similar purchasing patterns.
- As the distance increases, these smaller groups combine into larger clusters, showing broader differences between customer groups. The clear vertical gaps suggest the presence of a few well-defined clusters.

Conclusion:

The dendrogram supports dividing customers into a small number of distinct clusters, as there are clear separation points where clusters naturally form.

➢ <u>Clustering Model Evaluation Report (Silhouette Score)</u>

```
# KMeans Silhouette Score
kmeans_score = silhouette_score(rfm_scaled, rfm["Cluster"])

# DBSCAN Silhouette Score
dbscan_labels = rfm["DBSCAN_Cluster"]

if len(set(dbscan_labels)) > 1:
    dbscan_score = silhouette_score(rfm_scaled, dbscan_labels)
else:
    dbscan_score = "Not Applicable"

print("Silhouette Scores")
print("KMeans :", kmeans_score)
print("DBSCAN :", dbscan_score)
```

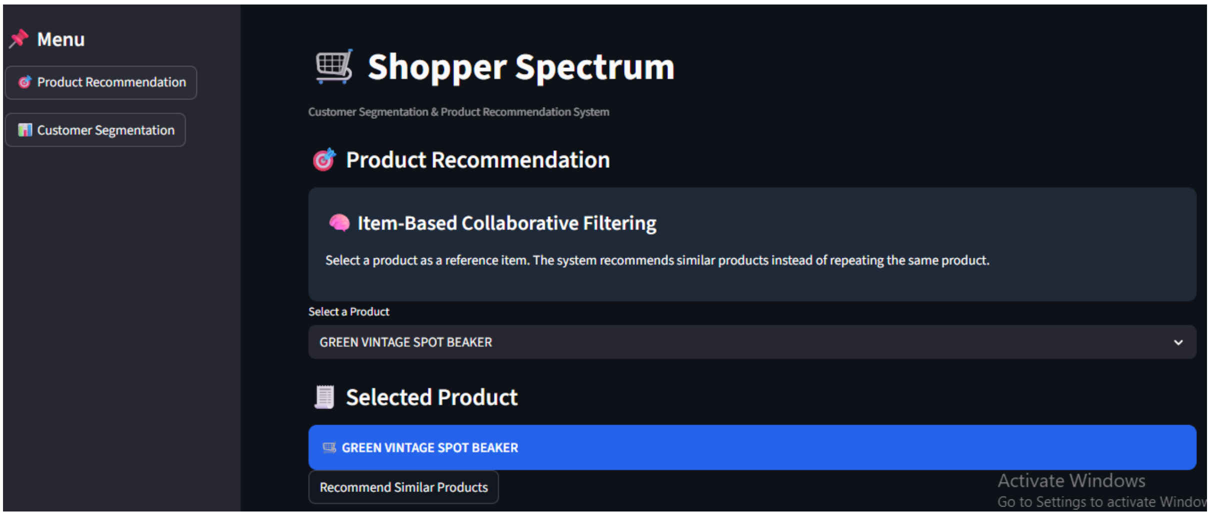
```
Silhouette Scores
KMeans : 0.616212846765192
DBSCAN : 0.8634068574759618
```

This result compares the performance of KMeans and DBSCAN clustering models using the Silhouette Score, which measures how well data points are grouped within their clusters.

- **KMeans** achieved a silhouette score of 0.62, which indicates good clustering with reasonably well-separated and compact clusters. This shows that KMeans has grouped customers in a meaningful and structured way.
- **DBSCAN** achieved a higher silhouette score of 0.86, indicating very strong clustering quality. This means DBSCAN has formed clearer and more distinct clusters, with better separation between different customer groups.

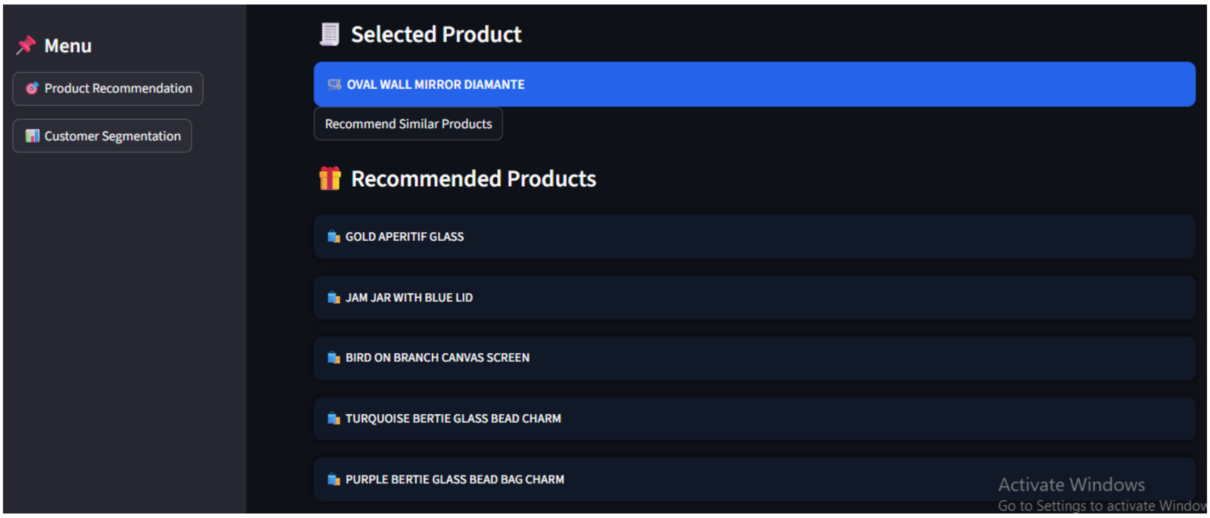
Conclusion:
Both models perform well, but DBSCAN performs better than KMeans for this dataset. DBSCAN captures the natural structure of the data more effectively, while KMeans still provides stable and understandable clusters.

➢ <u>Streamlit UI – Product Recommendation System (Overview)</u>



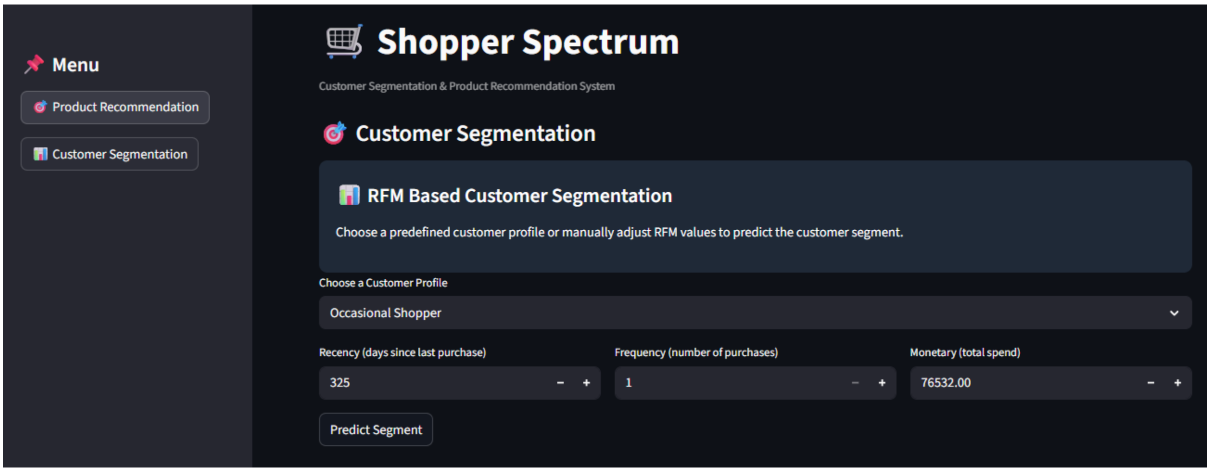
- The interface represents a product recommendation system focused on customer behaviour analysis and personalized product suggestions.
- Users can select a product from a dropdown and receive similar product recommendations using item-based collaborative filtering.
- A clean sidebar menu allows easy navigation between product recommendation and customer segmentation features.
- The UI is simple, interactive, and well-suited for showcasing an end-to-end data science project in a clear and user-friendly way.

➢ <u>Streamlit UI – Product Recommendation Output (Overview)</u>



- ❖ The interface displays the selected product clearly at the top, making it easy for users to understand the reference item.
- ❖ On clicking "Recommend Similar Products", the system generates a list of related products based on customer purchase patterns.
- ❖ Recommended products are shown in a clean, card-style layout, improving readability and user experience.
- ❖ Overall, this screen effectively demonstrates how item-based collaborative filtering provides meaningful and relevant product suggestions in a simple and user-friendly way.

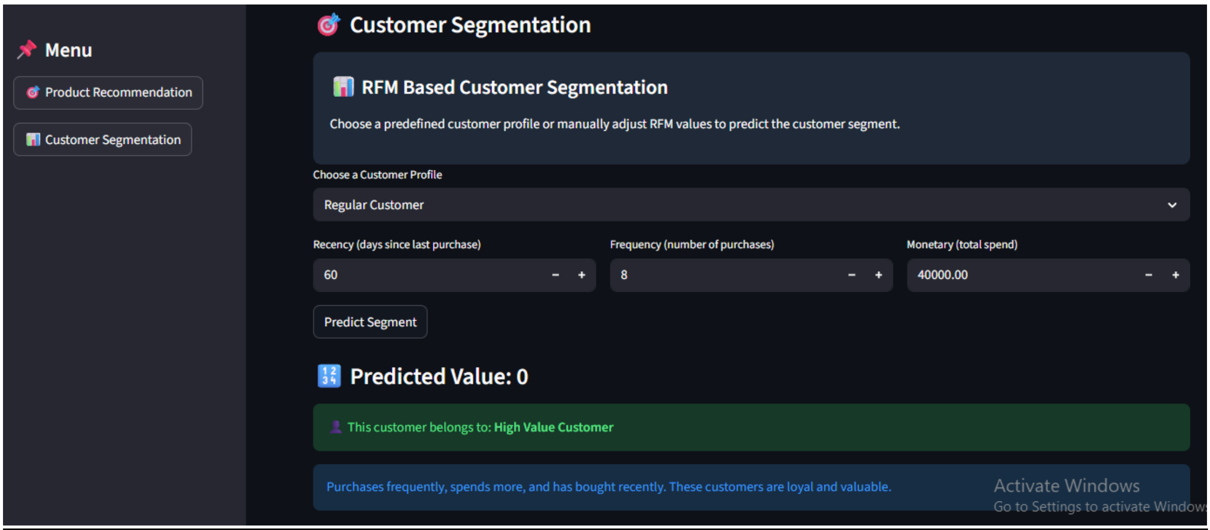
➢ <u>Streamlit UI – Customer Segmentation Module (Overview)</u>



- ❖ This interface represents an RFM-based customer segmentation system used to classify customers based on their purchasing behaviour.
- ❖ Users can select a predefined customer profile or manually adjust Recency, Frequency, and Monetary values to simulate different customer types.

❖ Overall, this module effectively demonstrates how data-driven customer segmentation works in a simple and practical way.

➢ Streamlit UI – Customer Segmentation Result (Overview)



❖ This screen shows the output of the RFM-based customer segmentation model after entering Recency, Frequency, and Monetary values.

❖ Based on the given inputs, the system predicts the customer segment and displays the predicted cluster value clearly.

❖ The result identifies the customer as a High Value Customer, highlighting strong purchase frequency, high spending, and recent activity.

❖ Overall, this output helps users easily understand how RFM values translate into meaningful customer segments.

➢ Conclusions:

❖ This project successfully demonstrates an end-to-end Customer Segmentation and Product Recommendation System using real-world transactional data.

❖ By applying RFM analysis and clustering techniques, the system effectively segments customers into meaningful groups such as high-value, regular, and occasional customers, helping businesses better understand customer behaviour.

❖ The item-based collaborative filtering model provides relevant and practical product recommendations, improving cross-selling opportunities.

❖ The use of Streamlit makes the solution interactive, easy to use, and ideal for presenting complex data science concepts in a simple and visual manner. Overall, the project highlights how data-driven approaches can support better decision-making, personalized marketing, and improved customer engagement.