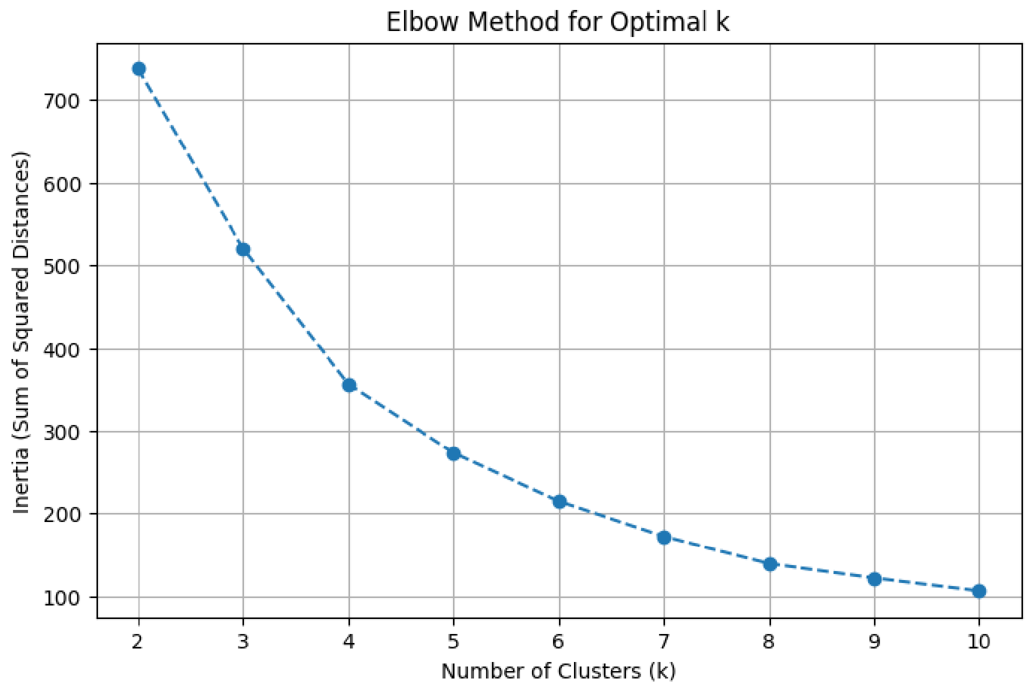
Customer Segmentation Report

Objective

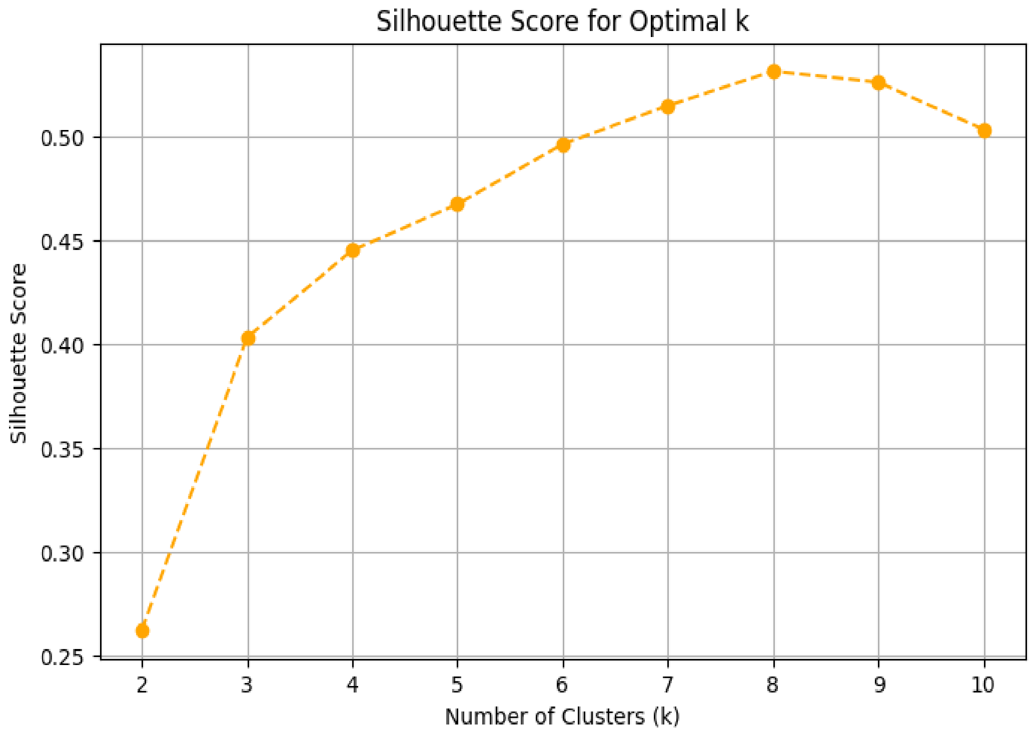
As part of my customer segmentation analysis, I aimed to identify distinct groups of customers based on their transactional and profile data. Using clustering techniques, I segmented customers into meaningful groups to better understand their purchasing behaviours and provide actionable insights for targeted marketing and operational improvements.

Summary of Results

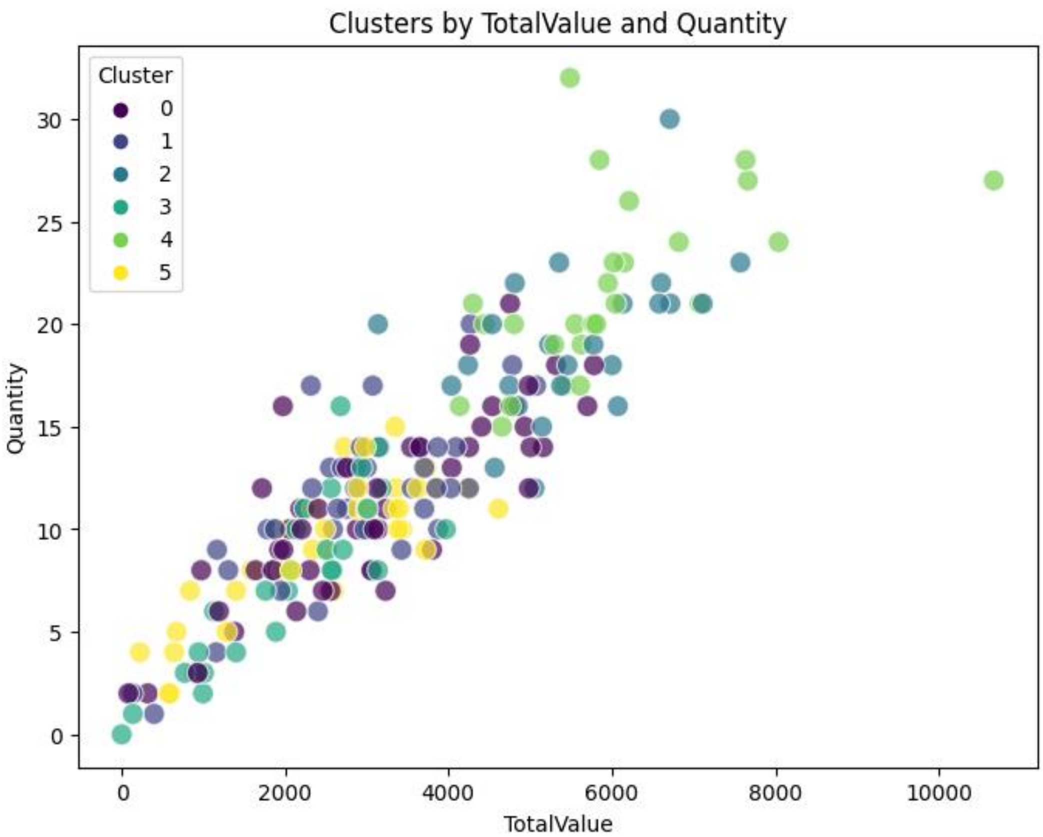
1. **Number of Clusters Formed**:
   After analysing the data using the Elbow Method and Silhouette Score, I determined that the optimal number of clusters was **6**.



2. **Clustering Metrics**:
   - **Davies-Bouldin Index (DBI)**: **0.77**
     A lower DBI indicates better cluster compactness and separation. The DBI of 0.77 suggests that my clusters are moderately well-separated and compact.
   - **Silhouette Score**: **0.50**
     The Silhouette Score of 0.50 shows that the clusters are moderately distinct but exhibit some overlap, indicating a balance between cohesion and separation.



3. **Clustering Algorithm Used**:
I used the K-Means algorithm for clustering after standardizing the data. The results were evaluated based on the above metrics and visualized to understand the segmentation patterns better.



Observations from Cluster Visualization

Distinct Clusters

To visualize the clusters, I plotted the data points using TotalValue (total spending) and Quantity (total units purchased). The scatter plot revealed six distinct clusters (labeled 0 to 5), each represented by different colors. This visualization confirmed that the K-Means algorithm successfully grouped customers based on their purchasing behaviors.

Clustering Pattern

- **TotalValue vs. Quantity**:
  The clusters appear to form along a diagonal pattern, indicating a **positive correlation** between TotalValue and Quantity. As customers purchase more quantities, their total spending also increases proportionally.

- **Cluster Distribution**:
  The clusters are **not uniformly distributed**. Some clusters contain a higher density of data points, representing larger customer groups, while others are more sparsely populated, potentially indicating niche customer segments.

Potential Interpretations

1. **Cluster 0 (Purple)**:
   This cluster contains customers with the highest TotalValue and Quantity values. These likely represent **high-value customers** or **bulk buyers** who contribute significantly to overall revenue.

2. **Cluster 5 (Yellow)**:
   This cluster has the lowest TotalValue and Quantity values. These customers might be **low-value buyers** or **infrequent purchasers** who spend and order in smaller amounts.

3. **Clusters 1 to 4 (Intermediate Clusters)**:
   These clusters fall between the extremes of TotalValue and Quantity. They likely represent **medium-value customers** or customers with varying purchase behaviours, such as consistent small orders or occasional bulk purchases.

Business Implications

1. **Customer Segmentation**:
   These clusters can help segment customers into distinct groups based on their purchasing behaviours. I can use this information to create targeted marketing campaigns or personalized offers tailored to each group. For example:
   - **High-value customers** (Cluster 0) could be targeted with loyalty programs or exclusive deals.
   - **Low-value customers** (Cluster 5) could be incentivized with promotions to encourage higher spending or more frequent purchases.

2. **Inventory Management**:
   Understanding the relationship between TotalValue and Quantity can assist in optimizing inventory levels. Clusters with higher purchasing patterns might indicate

products or categories that need higher stock levels, whereas low-value clusters may signal items at risk of overstocking.

3. **Order Fulfillment**:
   By identifying high-value clusters, I can prioritize order fulfillment strategies. For instance, expedited shipping or special handling could be allocated to customers in Cluster 0, ensuring a premium experience for the most valuable buyers.

Further Analysis

1. **Cluster Characteristics**:
   To enhance my understanding of each cluster, I could dive deeper into their characteristics by analysing additional variables, such as:
   - **Customer demographics**: Age, location, or profession could provide insights into regional or socio-economic purchasing trends.
   - **Product categories**: Identifying which products are popular in each cluster can inform product recommendations and marketing.
   - **Purchase frequency**: Frequent buyers might require different engagement strategies compared to occasional customers.
2. **Clustering Algorithm**:
   While K-Means performed well for this analysis, exploring alternative algorithms like hierarchical clustering or DBSCAN could provide a different perspective. These methods might better handle overlapping clusters or non-linear patterns in the data.

Conclusion

Using K-Means clustering, I identified **6 distinct customer segments**, as confirmed by the scatter plot of TotalValue and Quantity. The clustering evaluation metrics, including the **Davies-Bouldin Index** (0.77) and **Silhouette Score** (0.50), indicate a moderately successful segmentation. This analysis provides actionable insights that can drive targeted marketing, optimize inventory levels, and improve order fulfillment processes.