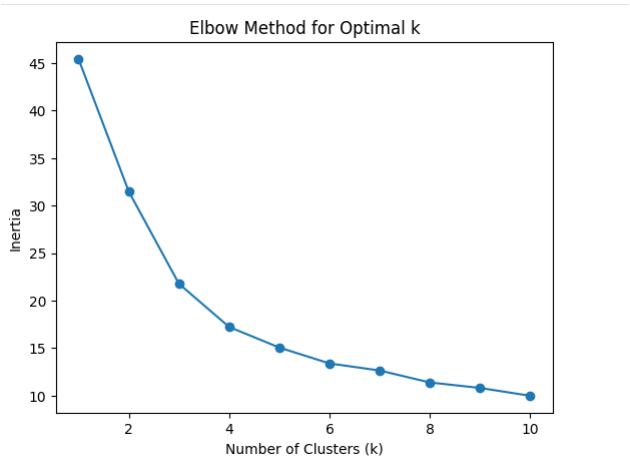
Enhanced Clustering Analysis Report: Customer Behavior Segmentation with t-SNE Visualization

Objective:

The objective of this analysis is to segment customers based on their purchasing behavior using clustering techniques and visualize these clusters in a lower-dimensional space for clearer interpretation. The insights generated from these segments will be valuable for targeted marketing, optimizing customer engagement, and enhancing business strategies.

Methodology:

1. **Data Preprocessing**:
    - **Handling Negative Values in 'Days To First Purchase'**:
        - The 'Days_To_First_Purchase' column may contain negative values if the first purchase occurs before the signup date. These anomalies were corrected by replacing negative values with 0 to maintain data integrity. This ensures that the dataset remains consistent for subsequent analysis.
2. **Feature Scaling**:
    - To enable effective distance-based clustering, **Min-Max Scaling** was applied to numerical features such as **Total Spent**, **Total Quantity**, **Days Between Purchases**, and **Days To First Purchase**. This transformation ensures that all features lie within the same range (0 to 1), making them comparable in magnitude and preventing any feature from dominating the clustering process.
3. **Clustering with K-Means**:
    - **K-Means Clustering** was used to segment the customer data into groups based on purchasing behavior and engagement patterns.
    - To determine the optimal number of clusters, the **Elbow Method** was employed. The "elbow" point, identified at **K=4**, indicates the optimal number of clusters where adding more clusters results in diminishing returns in reducing the **Within-Cluster Sum of Squares (WCSS)**.



4. **Evaluation Metrics**:
    - **Silhouette Score** of **0.3114** indicates moderate cluster separation, suggesting that while there is some distinction between clusters, they are not perfectly separated.
    - **Davies-Bouldin Index** of **1.0024** indicates a reasonable balance between the compactness and separation of the clusters, suggesting a good overall quality of clustering.
    - **K-Means Inertia** of **17.2391** shows moderate compactness within clusters, with room for further improvement in terms of reducing within-cluster variance.
5. **Dimensionality Reduction with t-SNE**:
    - **t-SNE (t-Distributed Stochastic Neighbor Embedding)** was used to reduce the high-dimensional feature space to 2D. This technique is especially effective for visualizing high-dimensional data by preserving the local structure of the clusters and enabling intuitive visualization.
    - The **t-SNE** plot clearly shows the distinct clusters identified by the K-Means algorithm, making it easier to interpret customer behavior visually.

Cluster Analysis and Interpretation:

Based on the clustering results and t-SNE visualization, we identified four distinct customer segments, each exhibiting unique purchasing behaviors:

1. **Cluster 0: High Engagement, Quick Adopters**
    - **Characteristics**:
        - These customers exhibit high total spend and purchase quantities, indicating that they are highly engaged and make frequent purchases.
        - They make their first purchase quickly, suggesting they are early adopters or decisive buyers.
    - **Implications**:
        - Marketing campaigns should focus on loyalty rewards and exclusive offers to further engage these customers and deepen their relationship with the brand.
2. **Cluster 1: Low Engagement, Cautious Decision-Makers**
    - **Characteristics**:
        - Customers in this cluster show lower total spend and fewer transactions.
        - They take a significantly longer time to make their first purchase, reflecting cautious or indecisive behavior.

- **Implications**:
    - Strategies for this group should focus on trust-building activities, such as product trials, extended guarantees, and educational content to accelerate their decision-making process.

3. **Cluster 2: Low Spend, Quick Adopters**
    - **Characteristics**:
        - Customers in this cluster make their first purchase quickly but tend to spend less and purchase fewer items.
        - Despite their low spending, they show a high level of engagement early on.
    - **Implications**:
        - This group may benefit from targeted promotions, cross-selling, or upselling opportunities that encourage them to spend more or increase purchase frequency.
4. **Cluster 3: High Engagement, Deliberate Decision-Makers**
    - **Characteristics**:
        - These customers show the highest levels of total spend and transaction frequency.
        - Despite being highly engaged, they take a longer time to make their first purchase, indicating a more deliberate decision-making process.
    - **Implications**:
        - Tailored content such as detailed product descriptions, reviews, and comparisons can help expedite their decision-making process, while also providing incentives for repeat purchases.

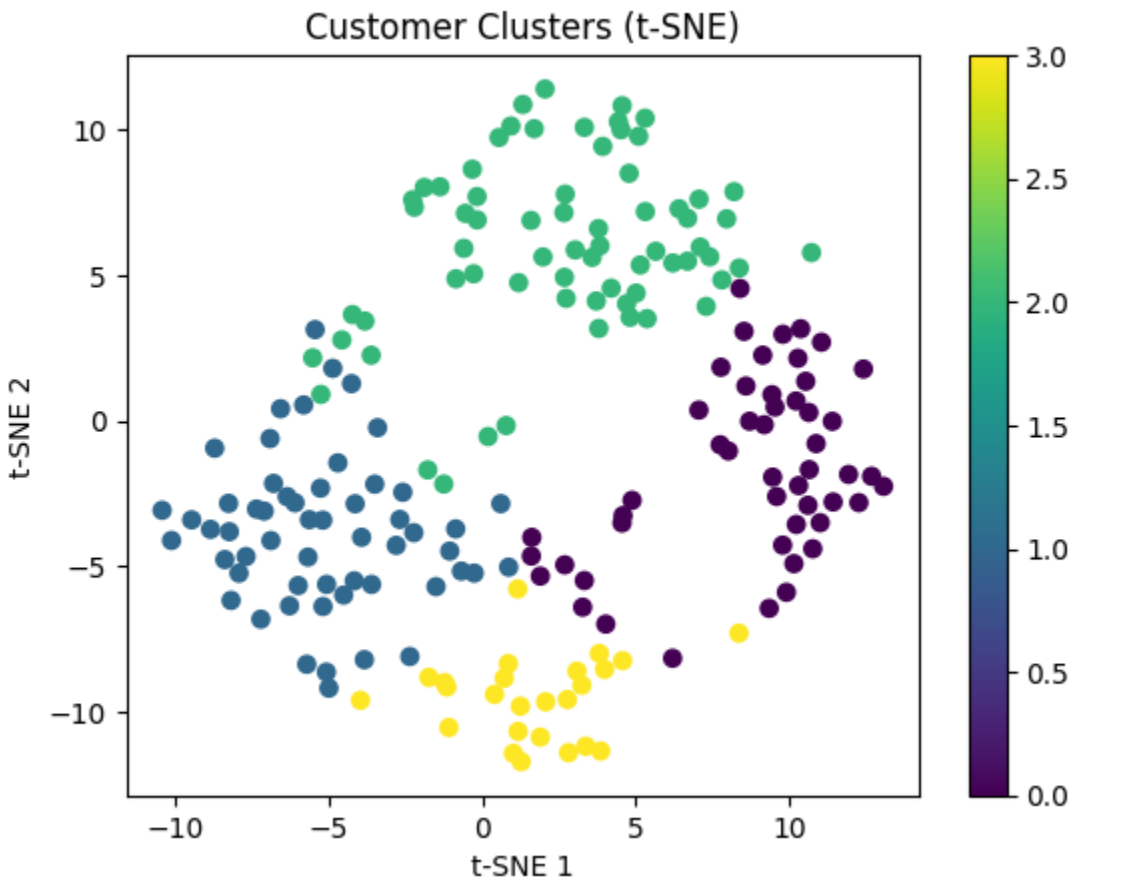
t-SNE Visualization Insights:

The t-SNE visualization provided a clear and intuitive representation of the customer clusters:

- The **distinct grouping** of clusters in the 2D plot highlights the differences in customer behavior across the various segments.
- **Cluster 0** (High Engagement, Quick Adopters) is separated from **Cluster 1** (Low Engagement, Cautious Decision-Makers), with a clear distinction in terms of spending behavior and engagement levels.
- The clusters are visually distinct, making it evident that customer segments exhibit different purchasing behaviors and timelines for decision-making.

t-SNE Visualization Example: *(Figure showing t-SNE plot with color-coded clusters)*

This visual aid serves as a powerful tool for quickly interpreting the structure of customer groups and understanding how different features (such as spending behavior and decision-making time) correlate with customer segmentation.



Feature Correlation Heatmap by Cluster

The image below represents a correlation matrix, visualizing how various features in the dataset are related to each other, specifically considering different customer clusters.

Features:

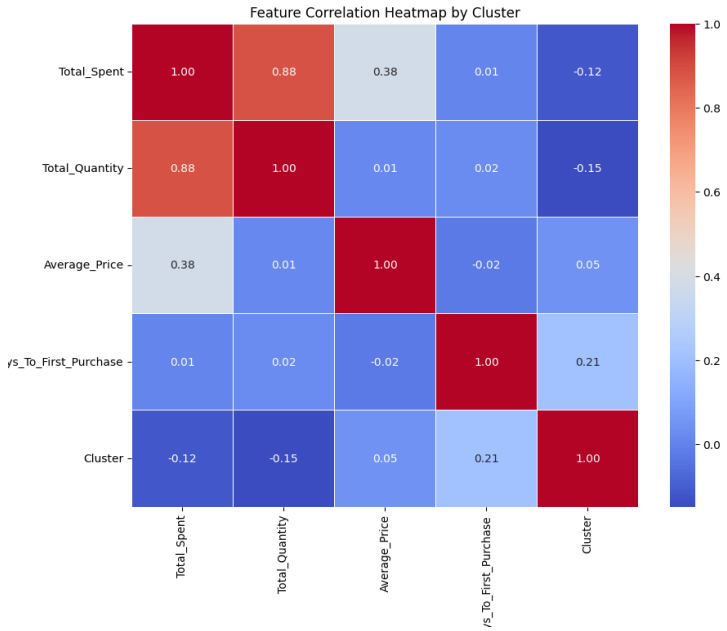
- **Total_Spent**: The total amount of money spent by a customer.
- **Total_Quantity**: The total number of items purchased by a customer.
- **Average_Price**: The average price of items purchased by a customer.
- **Days_To_First_Purchase**: The number of days it took a customer to make their first purchase after becoming a customer.
- **Cluster**: The customer cluster assigned to each individual.

Interpretation:

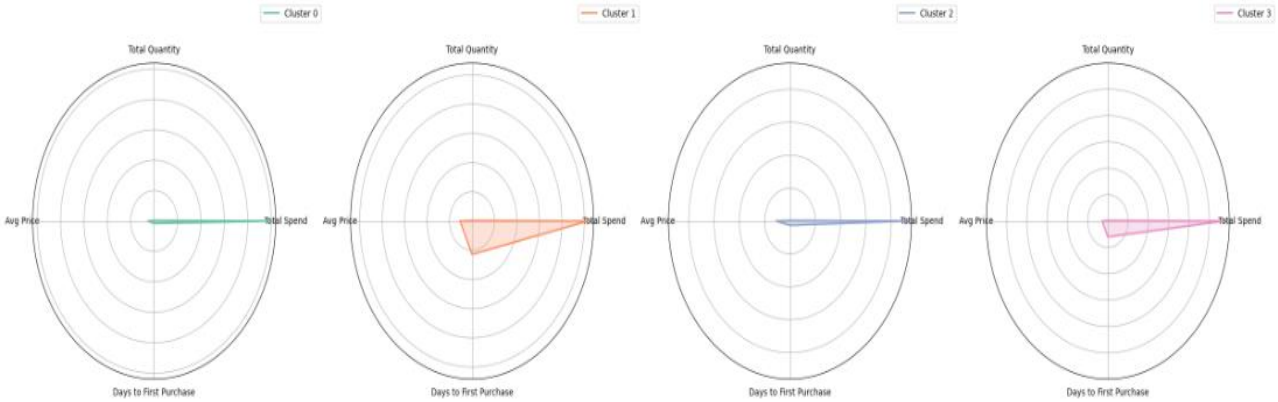
- **Positive Correlations**: Warmer colors indicate positive correlations, meaning an increase in one feature tends to be associated with an increase in the other.
    - **Total_Spent and Total_Quantity**: This strong positive correlation (0.88) shows that customers who spend more money tend to also purchase more items.
- **Negative Correlations**: Cooler colors indicate negative correlations, meaning an increase in one feature tends to be associated with a decrease in the other.
    - **Total_Spent and Cluster**: This slight negative correlation (-0.12) suggests that there might be a difference in spending habits across different customer clusters.
- **Weak Correlations**: Near-zero correlations (white/grey) indicate little to no relationship between the features.
    - **Total_Quantity and Average_Price**: This near-zero correlation (0.01) shows there's a weak relationship between the number of items purchased and the average price of those items.

Key Takeaways:

- **Customer Segmentation**: The correlation with the cluster feature can highlight potential differences in buying behavior among different customer segments. This can inform marketing and product development strategies.
- **Spending Patterns**: The strong positive correlation between Total_Spent and Total_Quantity suggests that customer spending is driven by the number of items purchased.
- **Customer Lifetime Value (CLTV)**: Understanding the relationships between spending, quantity, and price can be important for predicting customer lifetime value.



Customer Segmentation Report



Cluster 0 Analysis

- **High Total Spend**: Customers in this cluster exhibit significantly higher total spend, indicating they are high-value clients.
- **Low Total Quantity**: These customers tend to make fewer purchases, suggesting that while their individual purchases are substantial, they do not shop frequently.
- **Low Days to First Purchase**: Interestingly, these customers may take longer to make their first purchase, which could indicate a more deliberative or careful buying process.

  Insight: Cluster 0 represents high-value customers who are likely making large but infrequent purchases. These customers might engage in periodic, substantial transactions, possibly due to a specific need or purchase behavior pattern.

Cluster 1 Analysis

- **High Total Spend**: Customers in this cluster also exhibit a significant total spend, although this may not be as high as in other clusters.
- **Low Average Price**: Purchases tend to be lower-priced, indicating these customers might prioritize budget-friendly items.
- **Low Total Quantity**: Similar to Cluster 0, customers in Cluster 1 make fewer purchases overall.

  Insight: Cluster 1 is likely made up of customers who buy fewer, higher-priced items. Further analysis on the "days to first purchase" metric is recommended to determine the buying behavior better, as these customers may have a different decision-making timeline compared to other clusters.

Cluster 2 Analysis

- **High Total Spend**: Customers in this cluster spend a significant amount, but their spending behavior is more nuanced compared to other clusters.
- **Average Performance on Other Metrics**: The average price, total quantity, and days to first purchase are all moderate, indicating these customers do not stand out in any one metric.

  Insight: Cluster 2 likely consists of a mix of infrequent but large purchases or customers who buy a diverse range of products. They may not be as loyal or specialized as other clusters, but they contribute moderately to overall sales.

  Cluster 3 Analysis

- **High Total Spend**: This cluster also exhibits a significantly higher total spend, emphasizing their importance as high-value customers.
- **Above-Average Price**: Customers in this cluster tend to purchase more expensive items, which may indicate a preference for premium products.
- **Low Total Quantity**: As with the other clusters, customers make fewer purchases overall.
- **Low Days to First Purchase**: These customers make purchases relatively soon after their first interaction, suggesting a higher level of engagement or a quicker decision-making process.

  Insight: Cluster 3 represents high-value, loyal customers who make fewer but more significant purchases. Their purchasing decisions are faster, indicating a strong intent to buy once they engage with the brand.

Summary of Insights

- **High-Value Customers**: All clusters exhibit high total spend, signaling the importance of these customers to overall revenue.
- **Purchase Behavior**: While most clusters exhibit fewer total purchases, the high-value nature of these transactions suggests that customers are making larger but more selective purchases.
- **Days to First Purchase**: Cluster 3 stands out with a low days to first purchase, indicating quicker decision-making. Further segmentation of the "days to first purchase" could offer valuable insights for marketing strategies.