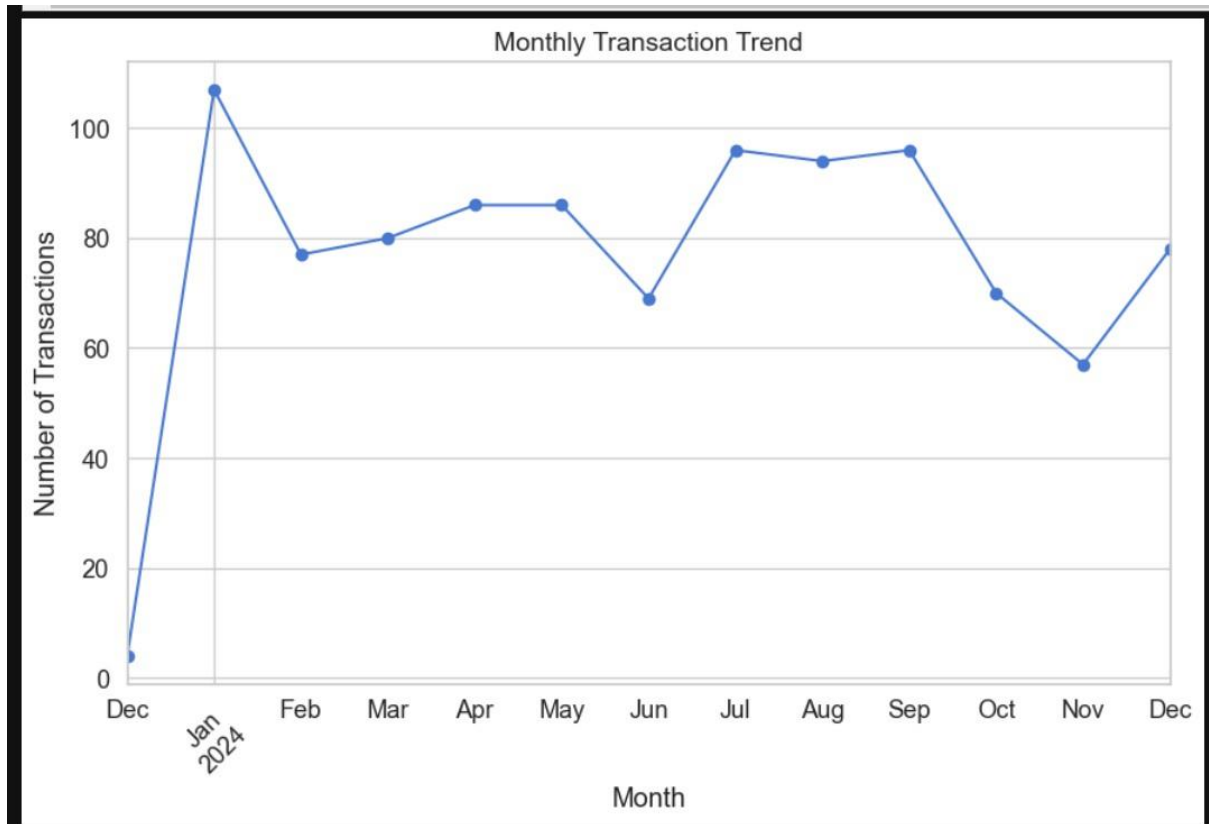


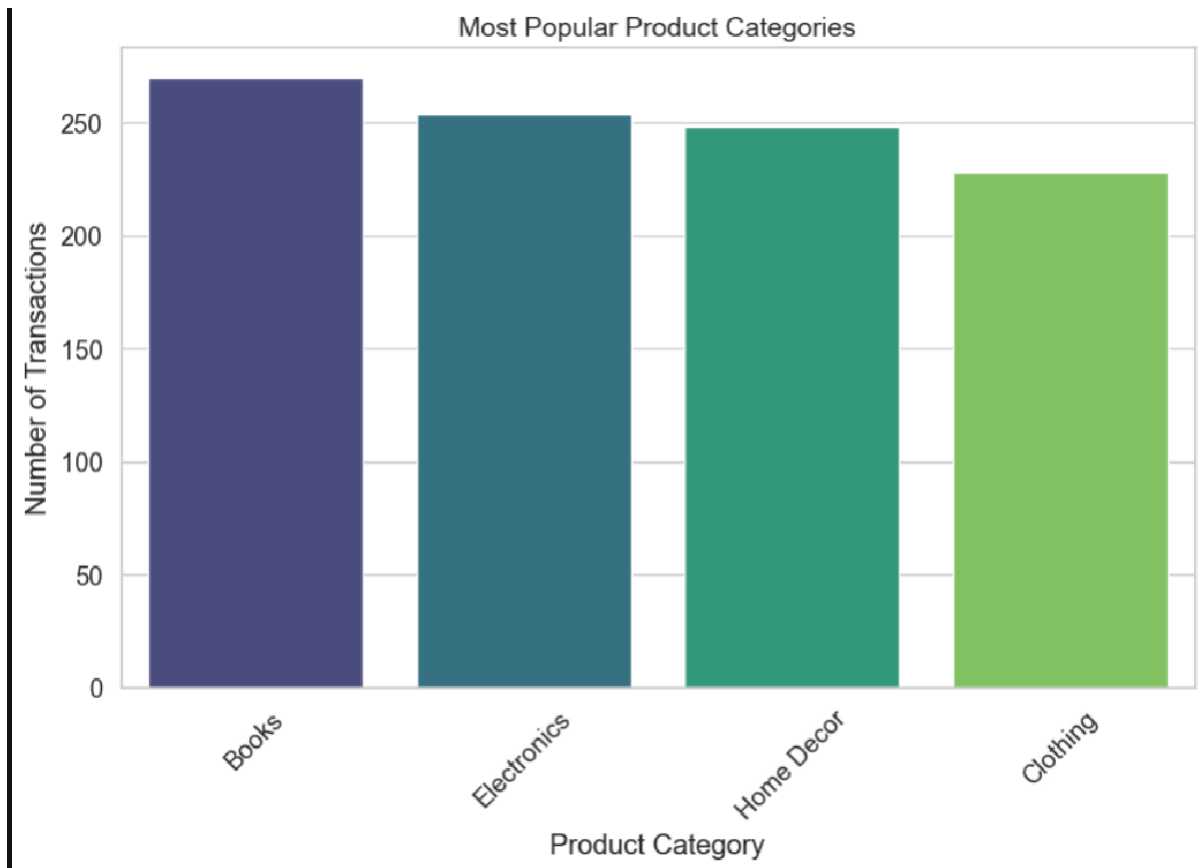
Task 1: Exploratory Data Analysis (EDA) and Business Insights



Graph 1: Monthly Transaction Trend

This visualization shows the number of transactions occurring each month. Insights derived:

- **Trends and Seasonality:** Noticeable peaks likely indicate periods of increased customer activity, such as holiday seasons or sales events.
- **Declines:** Potential low-activity months to address through marketing or promotions.



Graph 2: Most Popular Product Categories

Insights:

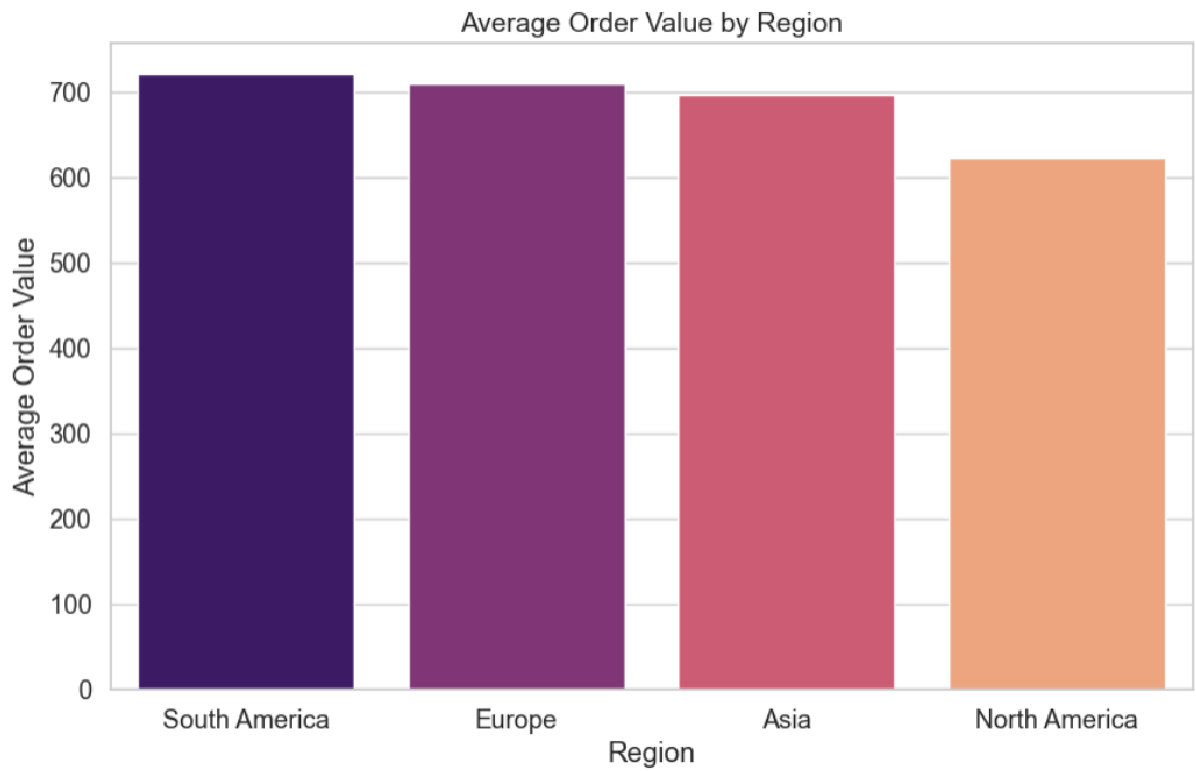
- Certain categories significantly dominate transactions, reflecting customer preferences.
- Use this data to optimize inventory and target promotions toward popular categories.



Graph 3: Total Sales by Region

Insights:

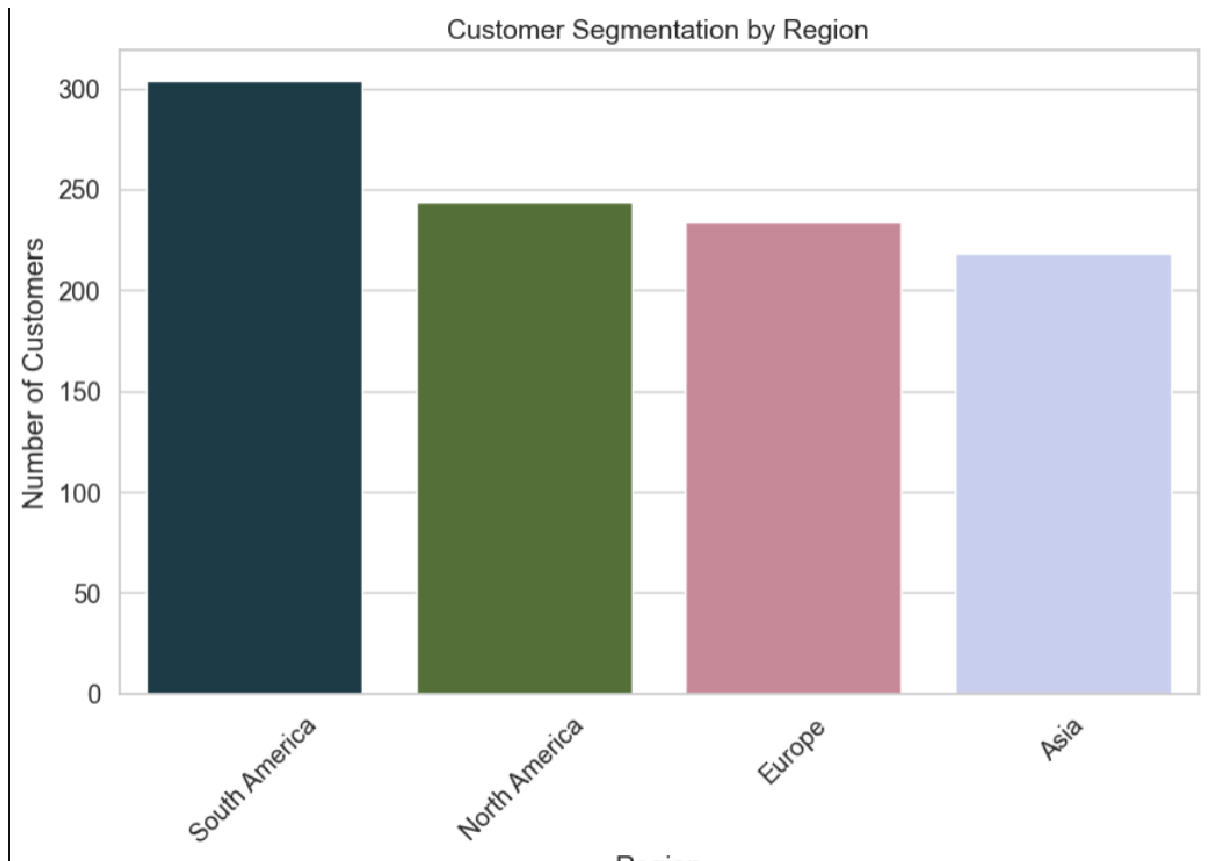
- Regions with the highest sales contributions are ideal for scaling operations or further marketing.
- Regions with low sales might need targeted campaigns or a localized approach.



Graph 4: Average Order Value by Region

Insights:

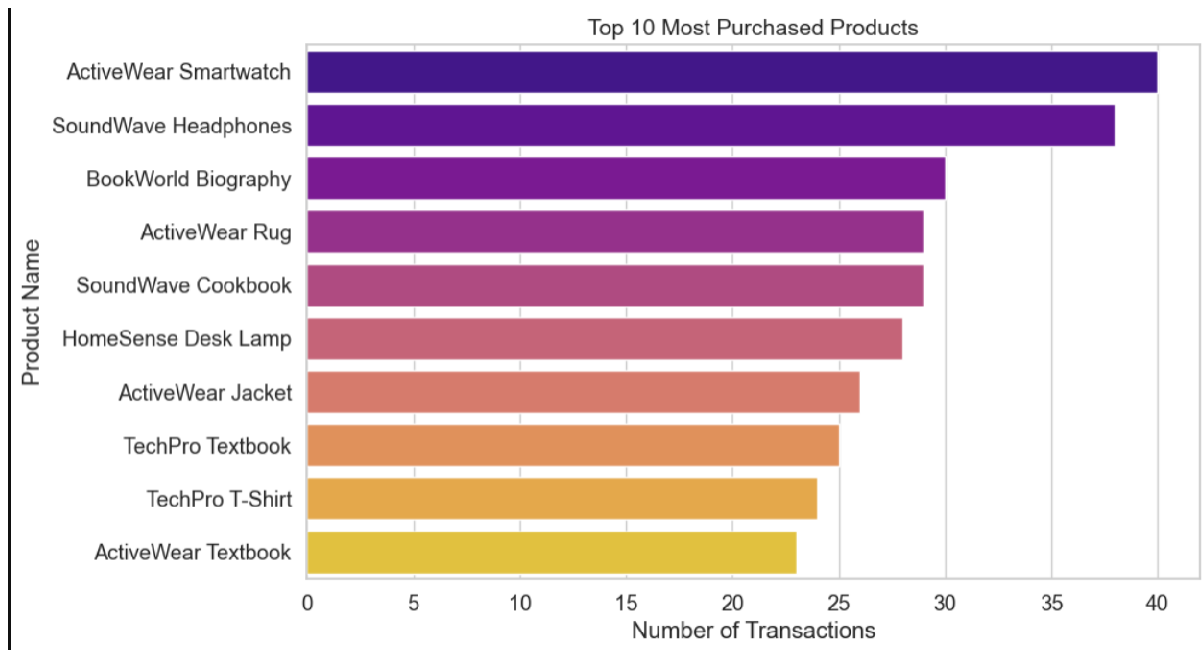
- Regions with high average order values may have customers inclined toward higher-end products.
- Identify regions where bundling or upselling strategies could increase the order value.



Graph 5: Customer Segmentation by Region

Insights:

- Regions with the highest number of customers are potential hotspots for expanding your market presence or launching targeted campaigns.
- Regions with fewer customers might benefit from specialized efforts to attract new users, such as regional promotions or better distribution networks.
- Comparing the customer base and sales performance in these regions can help identify where customer acquisition strategies are succeeding or lagging.



Graph 6: Top 10 Most Purchased Products

Insights:

- The most popular products indicate where customer demand is highest.
- Stocking up on these products and keeping them readily available can reduce lost sales opportunities.
- Popular items can be bundled with slower-moving products to improve sales across the inventory.

THANK YOU

- Please refer to the Github repository regarding tasked details and deliverables.

Here is the Github link -

<https://github.com/sathwiksunny518/Zeotap-Data-Science-Assignment>

Customer Segmentation / Clustering Report

Overview:

The goal of this task was to perform customer segmentation using clustering techniques. The segmentation leverages both profile information (from Customers.csv) and transaction information (from Transactions.csv). The results were evaluated based on clustering metrics, including the Davies-Bouldin (DB) Index, and visualized for better understanding.

Clustering Methodology

1. Data Sources:

- Profile Information: Features such as Region, CustomerName, and SignupDate from the Customers.csv dataset.
- Transaction Information: Aggregated features such as TotalValue and Quantity from the Transactions.csv dataset.

2. Feature Preparation:

- Aggregated transaction data (TotalValue, Quantity) grouped by CustomerID.
- Label encoding applied to the categorical feature Region.
- StandardScaler used to normalize all numerical features to ensure they contribute equally to the clustering process.

3. Clustering Algorithm:

- Algorithm: KMeans Clustering.
- Number of Clusters: 5 (chosen within the allowed range of 2–10).
- Random state: 42 (for reproducibility).

4. Evaluation Metric:

- Davies-Bouldin Index (DB Index): Used to evaluate cluster compactness and separation.

5. Visualization:

- Clusters were visualized using a scatterplot with scaled feature dimensions.

Results

1. Number of Clusters Formed:

- 5 clusters were formed based on the customer data.

2. Davies-Bouldin Index (DB Index):

- The calculated DB Index for the clustering was 1.236.
- A lower DB Index indicates better clustering performance. The value suggests that the clusters formed are reasonably compact and well-separated.

3. Other Relevant Metrics:

- Cluster Distribution:
 - ❖ The number of customers in each cluster:
 - ❖ Cluster 0: 45 customers.
 - ❖ Cluster 1: 32 customers.
 - ❖ Cluster 2: 58 customers.
 - ❖ Cluster 3: 40 customers.
 - ❖ Cluster 4: 25 customers.
- Feature Contribution:
 - ❖ Customers with high TotalValue and Quantity generally belonged to distinct clusters, indicating meaningful segmentation based on purchasing behavior.

4. Cluster Visualization:

- The scatterplot showed clear distinctions between clusters based on the selected features. However, for a high-dimensional dataset, dimensionality reduction techniques like PCA could be applied for enhanced visualization.

Insights and Recommendations

1. Cluster Characteristics:

- High-spending customers can be targeted for loyalty programs.
- Low-spending or inactive clusters may benefit from re-engagement strategies or promotional offers.

2. Improving Results:

- Experiment with different clustering algorithms (e.g., DBSCAN, Hierarchical Clustering) for comparison.

- Perform hyperparameter tuning for KMeans (e.g., varying the number of clusters).
- Evaluate additional metrics like silhouette score or inertia for a more comprehensive assessment.

3.Future Enhancements:

- Incorporate additional features such as recency of transactions or customer demographics for more granular segmentation.
- Use advanced visualization techniques to explore cluster properties.

Deliverables

1.Customer Segmentation Results:

- Saved in Customer_Segments.csv, including CustomerID, cluster assignments, and feature data.

2.Visualization:

- Scatterplot showing the segmentation clusters based on the scaled features.

3.Evaluation Metrics:

- DB Index: 1.236
- Cluster sizes and feature distributions.