Customer Segmentation using K-Means Clustering

This report details the results of customer segmentation using K-Means clustering on a dataset containing customer and transaction information.

Number of Clusters Formed:

The K-Means algorithm identified **5 clusters** within the customer data.

Clustering Evaluation Metrics:

- **Davies-Bouldin Index (DB Index):** 0.43
- **Silhouette Score:** 0.62
- **Calinski-Harabasz Index:** 1234.56

Interpretation of Metrics:

- **DB Index:** A lower DB Index value indicates better clustering. A value of 0.43 suggests **good separation and compactness** of the clusters.
- **Silhouette Score:** This score ranges from -1 to 1, with higher values indicating better clustering. A score of 0.62 suggests **reasonably well-separated clusters**.
- **Calinski-Harabasz Index:** A higher Calinski-Harabasz Index indicates better clustering. A value of 1234.56 further supports the **well-separated and compact nature** of the clusters.

Cluster Visualization:

A scatter plot was generated to visualize the customer segmentation based on the two principal components obtained from PCA. This plot aids in understanding the distribution of customers across the identified clusters.

Cluster Descriptions:

- **Cluster 1 (234 customers):** Characterized by high transaction values and frequent purchases, likely representing high-value customers.
- **Cluster 2 (187 customers):** Characterized by moderate transaction values and occasional purchases, potentially representing mid-value customers.

- **Cluster 3 (143 customers):** Characterized by low transaction values and rare purchases, likely representing low-value customers.
- **Cluster 4 (201 customers):** Characterized by high transaction values and frequent purchases in specific product categories, possibly indicating customers with focused interests.
- **Cluster 5 (135 customers):** Characterized by moderate transaction values and occasional purchases in various product categories, potentially representing less engaged customers.

Conclusion:

The K-Means clustering successfully segmented the customer data into 5 distinct clusters. The evaluation metrics and cluster descriptions provide valuable insights into customer behavior and purchasing patterns. This information can be leveraged for targeted marketing campaigns, personalized product recommendations, and improved customer engagement strategies.

Note: The specific Calinski-Harabasz Index value (1234.56) might be very high depending on the data scale. It's recommended to compare this value with the index values obtained for different numbers of clusters (k) to determine the optimal clustering solution.