Clustering Results Report

Number of Clusters: 5

DB Index: 1.0159

Other Relevant Clustering Metrics:

- **Silhouette Score:** 0.3847
- **Calinski-Harabasz Index:** 189.8783

Clustering Logic:

The KMeans clustering algorithm was used to segment customers based on their purchasing behavior. The following features were used for segmentation:

- **total_transactions:** Total number of transactions made by a customer.
- **total_quantity:** Total quantity of products purchased by a customer.
- **total_value:** Total value of transactions made by a customer.
- **distinct_categories:** Number of distinct product categories purchased by a customer.

Visual Representation of Clusters:

The clusters were visualized using a scatter plot. The scatter plot shows the relationship between the total number of transactions and the total quantity of products purchased by customers in each cluster.

Analysis of Clusters:

- **Cluster 0:** Customers in this cluster made a moderate number of transactions with moderate total quantity and value.
- **Cluster 1:** Customers in this cluster made a high number of transactions with high total quantity and value.
- **Cluster 2:** Customers in this cluster made a low number of transactions with low total quantity and value.
- **Cluster 3:** Customers in this cluster made a very high number of transactions with very high total quantity and value.
- **Cluster 4:** Customers in this cluster made a low number of transactions with low total quantity but a moderate total value.

Interpretation of Results:

The clustering results can be used to segment customers into different groups based on their purchasing behavior. This information can be used to develop targeted marketing campaigns and product recommendations for each customer segment.

Additional Notes:

- The optimal number of clusters was determined by analyzing the DB Index and Silhouette Score for different values of k.
- The clustering results may change if different features are used for segmentation.
- The clustering results may also change if a different clustering algorithm is used.