Clustering Approach
We used three clustering techniques to segment customers:

1. **K-Means Clustering**: This algorithm divides data into a fixed number of clusters, where each customer belongs to the closest cluster center. We tried different numbers of clusters (from 2 to 10) to identify the best fit.
2. **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**: DBSCAN groups points that are close to each other, identifying clusters based on density. It also marks outliers as noise.

Features Used for Clustering:

- **Customer Profile Features**:
  - *Region*: Geographical regions where customers live.
  - *Customer Name*: Encoded into numerical values for clustering.
- **Transaction Data**:
  - *Quantity*: The number of items a customer buys.
  - *Total Value*: Total amount spent in a transaction.
  - *Price*: Price of the products purchased.

Clustering Process and Number of Clusters:

1. **Region (K=4 clusters)**:
   - We clustered customers based on the regions they belong to. The optimal number of clusters was 4, representing North America, South America, Asia, and Europe.
   - **Clusters**: Customers from different regions had distinct spending patterns.
2. **Total Value (K=10 clusters)**:
   - Customers were clustered based on the total value of their transactions. This segmentation identified 10 groups, from low spenders to high spenders.
   - **Clusters**: These clusters represent varying spending habits.
3. **Price (K=10 clusters)**:
   - Customers were grouped based on the price of the products they purchased. This segmentation helped identify preferences for cheaper versus more expensive products.
   - **Clusters**: Customers with similar product price preferences were grouped together.
4. **Quantity (K=10 clusters)**:
   - Customers were clustered based on the quantity of products they bought. The goal was to group frequent shoppers versus occasional buyers.
   - **Clusters**: The clusters showed how frequently customers made purchases.
5. **Customer Name (K=10 clusters)**:
   - We clustered based on customer names, though it was a less effective method. By encoding customer names, we tried to find patterns in their shopping behavior.
   - **Clusters**: These clusters gave an idea of customer behaviors, though it was noisy.

Clustering Metrics:

- **DB Index**: This index measures the compactness and separation of clusters. A lower DB index indicates better clustering. For the region and quantity clusters, the DB index showed that K-Means did a good job of separating different groups.

- **Silhouette Score**: This score measures how similar a customer is to their own cluster compared to other clusters. A higher silhouette score indicates better-defined clusters. The total value and quantity clusters had high silhouette scores, indicating good segmentation.
- **Davies-Bouldin Index**: This index helps evaluate the quality of the clusters. Lower values indicate better clustering. For regions and quantity, the DB index was low, meaning the clusters were well-formed.

Visualization:
We visualized the clustering results using Principal Component Analysis (PCA), which reduces the data to two dimensions. This helped us plot and visually inspect how well the clusters were separated. Each customer was assigned a color based on their cluster, making it easy to see the groupings.

Conclusion:
The customer segmentation task was successful in identifying meaningful groups of customers. By using K-Means and DBSCAN clustering, we were able to segment customers based on their region, spending habits, and preferences. This segmentation provides valuable insights for targeting different customer groups with tailored marketing strategies. The clustering results showed that customers could be effectively grouped into segments that will help businesses better understand and serve their diverse customer base.