

Task 3: Customer Segmentation Report

1. Objective

To segment customers using clustering techniques with profile and transaction data. The clustering helps identify customer groups for business insights.

2. Clustering Approach

What We Did:

We used the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm for customer segmentation. The process involved merging two datasets: Customers.csv and Transactions.csv. Key features like TotalValue (total spend by a customer) and TransactionCount (number of transactions) were extracted and standardized for clustering. DBSCAN was chosen for its ability to identify clusters of arbitrary shapes and handle noise points effectively.

Why DBSCAN?:

Unlike K-means, which requires the number of clusters to be predefined, DBSCAN automatically determines clusters based on density. It is highly effective for datasets with outliers, as it categorizes them as noise rather than forcing them into clusters.

How We Used It:

1. Merged datasets to create a comprehensive profile for each customer.
2. Scaled features using StandardScaler to standardize values.
3. Applied DBSCAN with parameters:
 - **eps:** 0.5 (distance threshold to consider a point in the neighborhood).
 - **min_samples:** 5 (minimum number of points to form a dense region).

4. Evaluated the clustering using two metrics:

- **DB Index:** Measures cluster compactness and separation.
- **Silhouette Score:** Assesses how well-separated clusters are.

3. Visualizations

1. PCA Scatter Plot:

- A two-dimensional visualization of the clusters identified by DBSCAN.
- Helps understand cluster density and separation visually.

2. Cluster Size Bar Plot:

- Displays the number of customers in each cluster.
- Useful for identifying the largest and smallest groups.



4. Results and Insights

Results:

- **DB Index:** 0.5058 (indicates good separation and compactness).
- **Silhouette Score:** 0.3022 (moderate compactness and separation).
- Identified multiple clusters and noise points, representing distinct customer groups and outliers.

Effect of Results:

1. Business Insights:

- Clusters provide actionable segmentation of customers, enabling targeted marketing campaigns.
- The largest cluster represents high-density customers, indicating potential for loyalty programs.

2. Handling Noise:

- Noise points identified by DBSCAN represent customers with unique behaviors or insufficient data. These can be analyzed separately for anomalies.

3. Customer Segmentation:

- Helps categorize customers into groups based on spending patterns and transaction frequencies. Businesses can design personalized strategies for each group, increasing engagement and revenue potential.

5. Conclusion

DBSCAN effectively segmented customers into distinct groups with clear separation. These clusters provide actionable insights for targeted marketing and personalized strategies, while noise points highlight opportunities to analyze unique customer behaviors.