Report: Customer Clustering and Segmentation

The script implements customer clustering using the **K-Means algorithm**. It follows a structured approach to preprocess data, engineer features, and visualize clusters. The primary goal is to **group customers** based on **transactional behaviour and demographic attributes** to facilitate better **customer segmentation.**

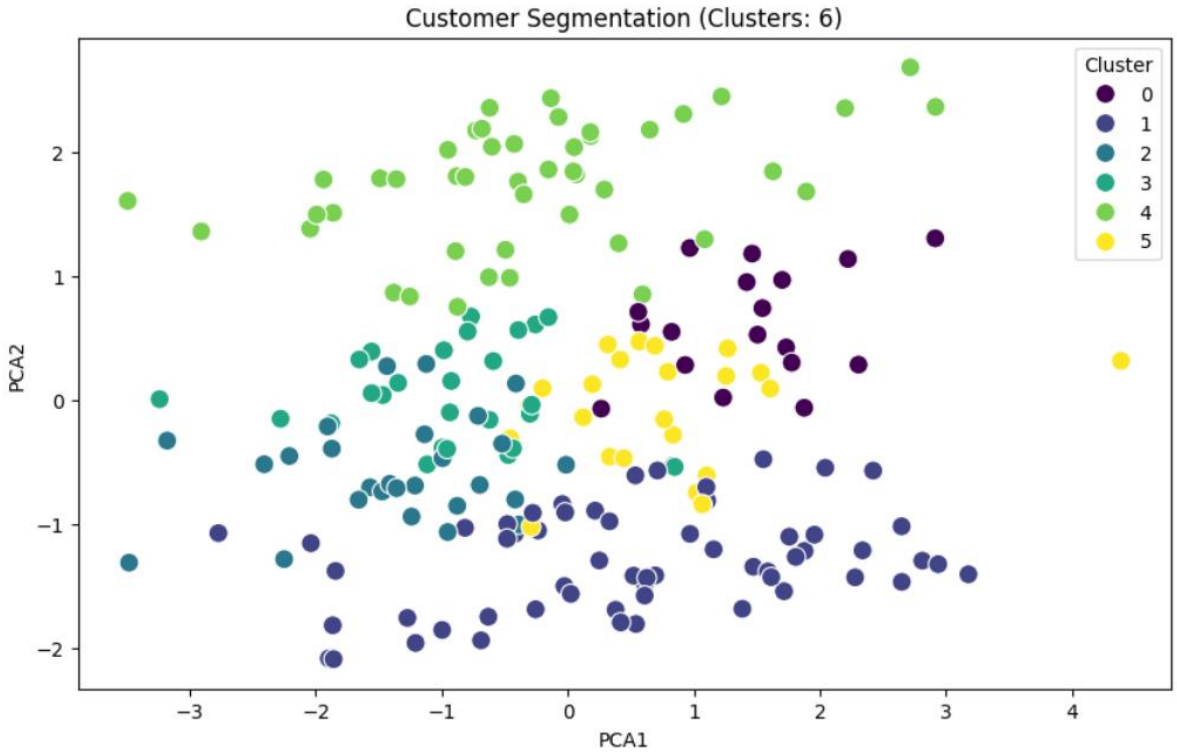
```
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score, silhouette_score
```

Key Observations

1. **Optimal Clusters**:
    - The optimal number of clusters was determined based on the Davies-Bouldin Index. The DBI evaluates the ratio of the sum of within-cluster scatter to the between-cluster separation. A lower DBI indicates better-defined and more compact clusters.
2. **Evaluation Metrics**:
    - **Davies-Bouldin Index**: Measures cluster compactness and separation.
    - **Silhouette Score**: Evaluates the quality of clustering. It ranges from -1 to 1, where values closer to 1 represent clear and well-defined clusters, values near 0 indicate overlapping clusters, and negative values suggest points may be assigned to the wrong cluster.

```
Davies-Bouldin Index: 1.1182346408751787
Silhouette Score: 0.3403143458825825
```

3. **Visualization**:
    - Scatter plot of clusters in reduced dimensions using PCA. The resulting plot provides a clear view of how clusters are distributed and separated in the reduced space. Different colors were used to distinguish clusters, making it easier to identify their boundaries and analyze the spread of data points.



Metrics Summary

- **Davies-Bouldin Index**: *1.1182346408751787*
- **Silhouette Score**: *0.3403143458825825*