Clustering Analysis Report

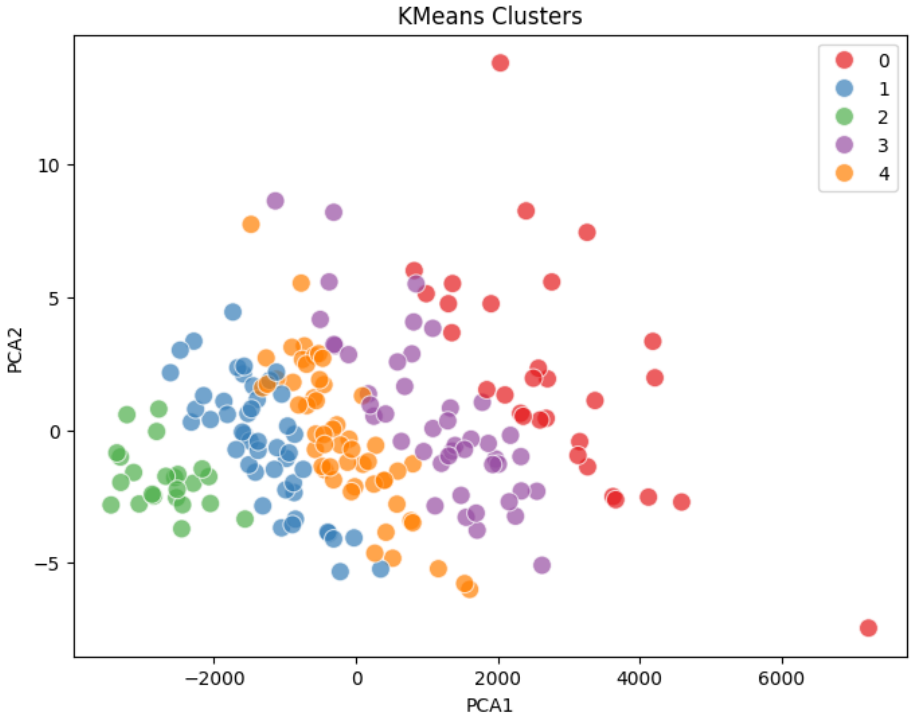
Overview

Customer segmentation was performed using four clustering algorithms: **KMeans**, **DBSCAN**, **Agglomerative Clustering**, and **Gaussian Mixture Model (GMM)**. The aim was to segment customers based on profile and transaction data, generating clusters that could inform marketing and business strategies. The results of each algorithm are summarized below.

Clustering Results

KMeans Clustering

- **DB Index**: 0.895
- **Silhouette Score**: 0.356
- **Cluster Distribution**:
    - Cluster 4: 54 customers
    - Cluster 1: 48 customers
    - Cluster 3: 45 customers
    - Cluster 0: 30 customers
    - Cluster 2: 23 customers

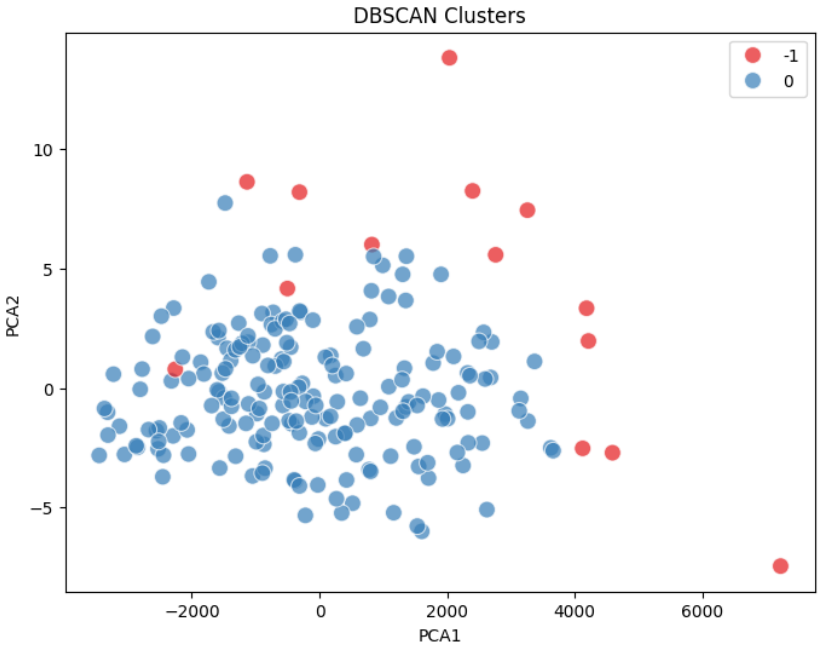


KMeans performed best overall with a **DB Index** suggesting well-separated clusters and a moderate **Silhouette Score** indicating moderate cohesion and separation.

DBSCAN Clustering

- **DB Index**: 1.062
- **Silhouette Score**: 0.392

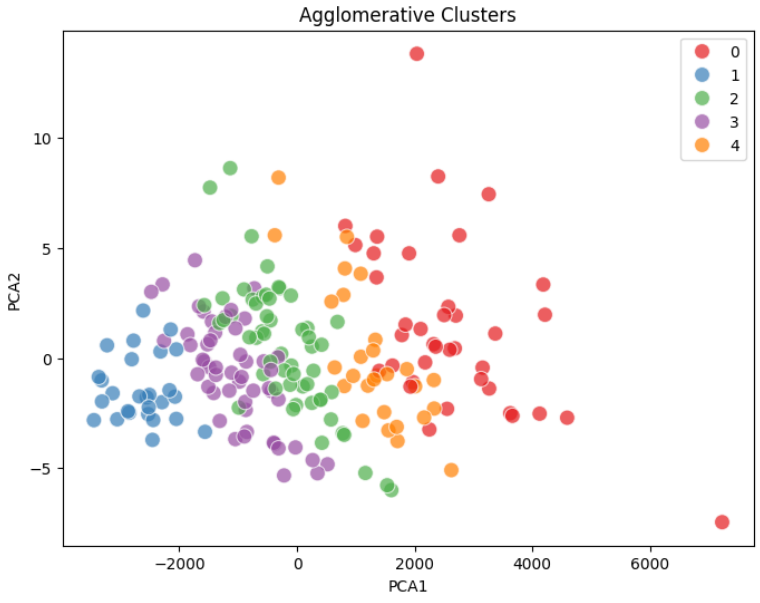
DBSCAN showed strong cluster separation but also sensitivity to noise, reflected in the higher **DB Index**.



Agglomerative Clustering

- **DB Index**: 0.934
- **Silhouette Score**: 0.331

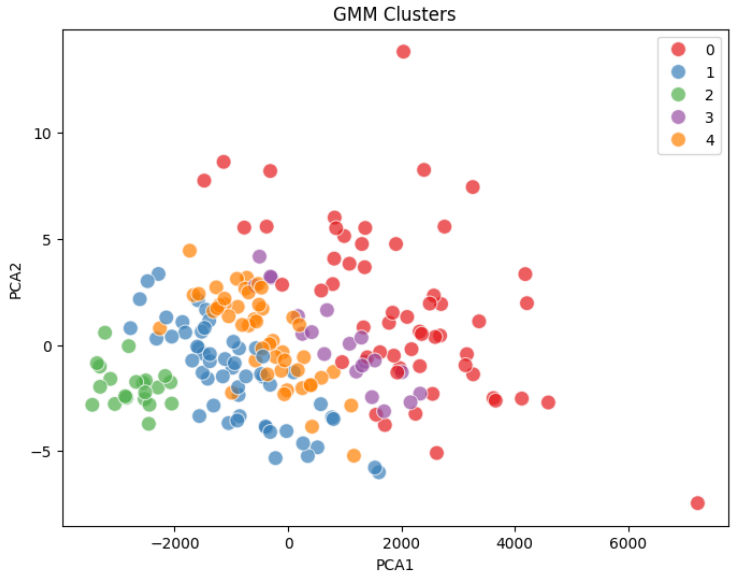
Agglomerative clustering showed moderate performance with a reasonable **DB Index** and **Silhouette Score**, indicating somewhat effective cluster separation.



GMM Clustering

- **DB Index**: 1.211
- **Silhouette Score**: 0.224

GMM showed poor performance with a high **DB Index** and low **Silhouette Score**, indicating weak cohesion and separation.



Best Algorithm: KMeans

KMeans was chosen as the best algorithm based on its balanced performance, with a **DB Index** of 0.895 and **Silhouette Score** of 0.356. This suggests good cluster separation and moderate cohesion, making it the most suitable for this task.

```
PCA1

Best Algorithm: KMeans
DB Index: 0.8950365287786601, Silhouette Score: 0.355876777278508
Cluster distribution:
Cluster
4    54
1    48
3    45
0    30
2    23
Name: count, dtype: int64

In [ ]:
```

Key Insights

- **Cluster 4**: Largest cluster, moderate transaction frequency, and mid-to-high spending. This group should be targeted with **loyalty programs**.
- **Cluster 2**: Small but high-value customers. Target with **premium offers** due to their high spending.

- **Cluster 3**: High spenders but low transaction frequency. Upselling strategies can encourage more purchases.
- **Cluster 0**: Passive customers with low activity. **Re-engagement strategies** such as discounts could help.
- **Cluster 1**: Medium transaction frequency and lower spending. **Cross-selling** campaigns could be effective.

Business Recommendations

- **Loyalty Programs**: Focus on **Cluster 4** to retain high-engagement customers.
- **Premium Offers**: Target **Cluster 2** with exclusive offers for high spenders.
- **Re-engagement**: Use personalized campaigns to re-engage **Cluster 0**.
- **Upselling**: Target **Cluster 3** with upselling opportunities to boost frequency.

Conclusion

The segmentation provides valuable insights for targeted marketing, customer retention, and revenue generation. By tailoring strategies to the identified clusters, the business can improve engagement and drive growth.