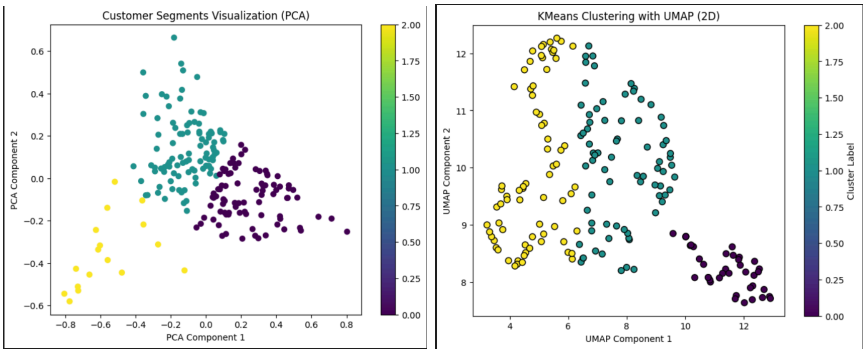
Shreyas Peyyeti Clustering Results

Overview of Clustering Algorithms

I evaluated several clustering algorithms: **KMeans**, **DBSCAN**, **Spectral Clustering**, **Agglomerative Clustering**, and **Gaussian Mixture Models (GMM)**. We assessed their performance using the **Silhouette Score** and **Davies-Bouldin Index (DBI)**.

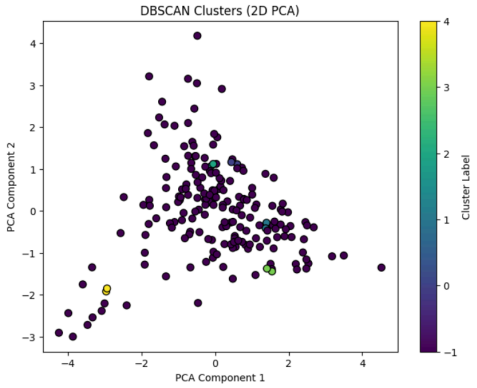
KMeans

- **Clustering**: Great clusters with no overlap.
- **Silhouette Score**:
  - **PCA**: 0.303
  - **UMAP**: 0.442
    KMeans with UMAP performed better, showing better-defined clusters.
- **DBI**:
  - **PCA**: 1.117
  - **UMAP**: 0.829
    KMeans with UMAP produced more compact clusters.



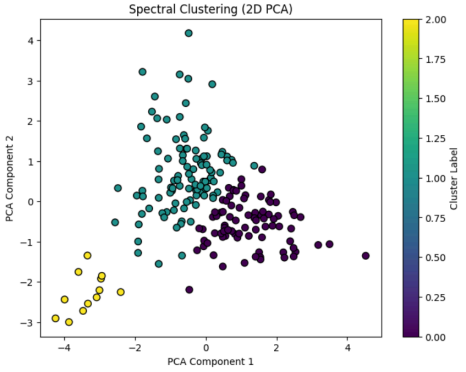
DBSCAN

- **Clustering**: Poor plots with overlapping clusters, but best clustering metrics.
- **Silhouette Score**: 0.887 (very high, indicating well-separated clusters)
- **DBI**: 0.238 (very low, indicating compact clusters)



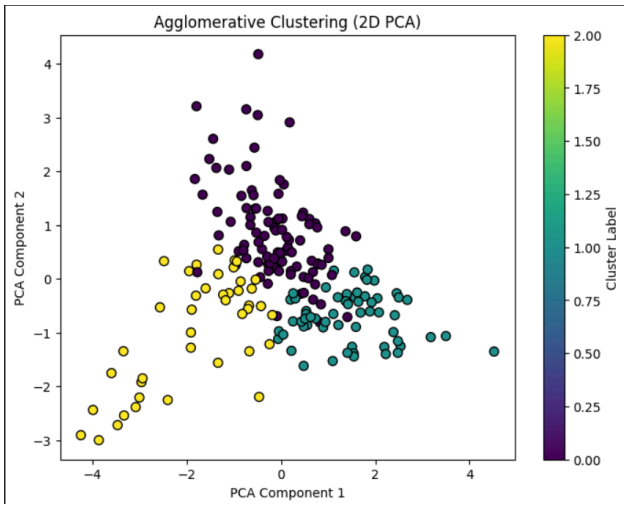
Spectral Clustering

- **Clustering**: Good visual results with no overlap, but performance metrics are moderate.
- **Silhouette Score**:
    - **PCA**: 0.424
    - **UMAP**: 0.364
- **DBI**:
    - **PCA**: 0.735
    - **UMAP**: 0.692



Agglomerative Clustering

- **Clustering**: Decent clusters with slight overlap.
- **Silhouette Score**: 0.227 (indicating moderate clustering quality)
- **DBI**: 1.392 (higher, indicating less compact clusters)



Conclusion

- **DBSCAN** achieved the best clustering quality with the highest **Silhouette Score** (0.887) and the lowest **DBI** (0.238), but the plots were less interpretable.
- **KMeans with UMAP** produced clearer visualizations with a **Silhouette Score** of 0.442 and a **DBI** of 0.829, making it a solid choice for interpretable clusters.
- **Spectral Clustering** showed decent visual results but was less effective than **DBSCAN** and **KMeans with UMAP** in terms of clustering quality.
- **Agglomerative Clustering** and **GMM** produced the weakest results.

Algorithm	Silhouette Score	Davies-Bouldin Index
KMeans (PCA)	0.303	1.117
KMeans (UMAP)	0.442	0.829
DBSCAN	0.887	0.238
Spectral (PCA)	0.424	0.735
Spectral (UMAP)	0.364	0.692
Agglomerative	0.227	1.392