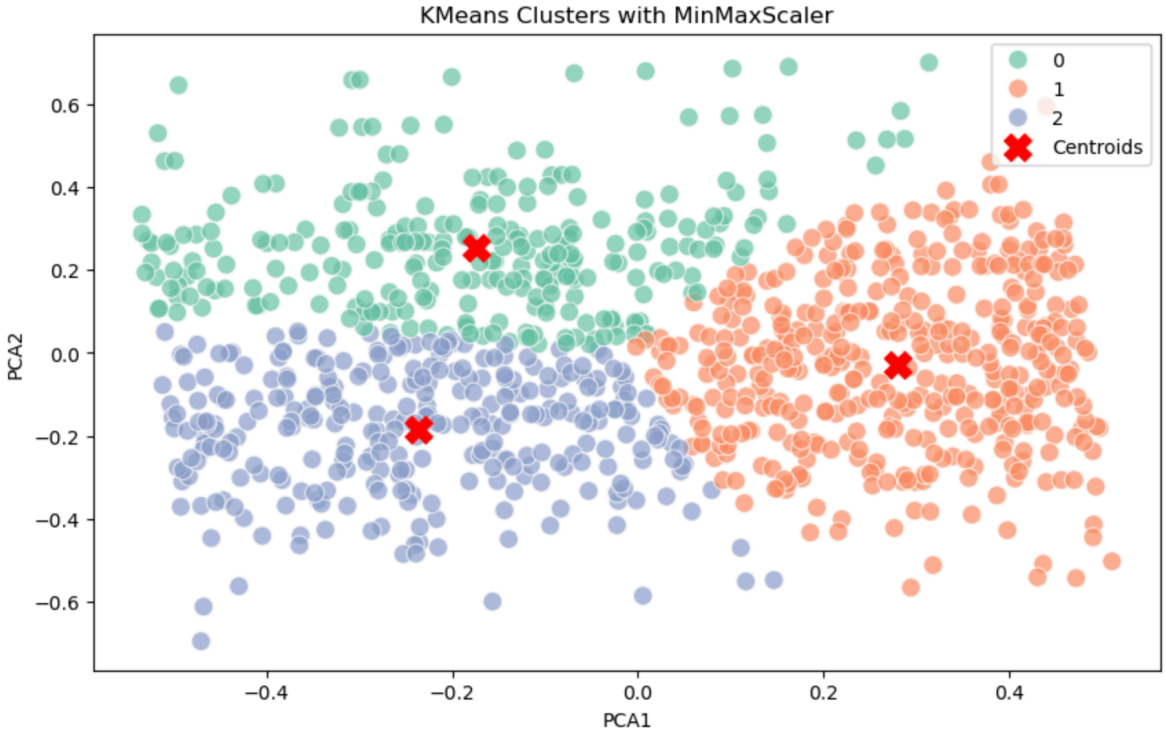
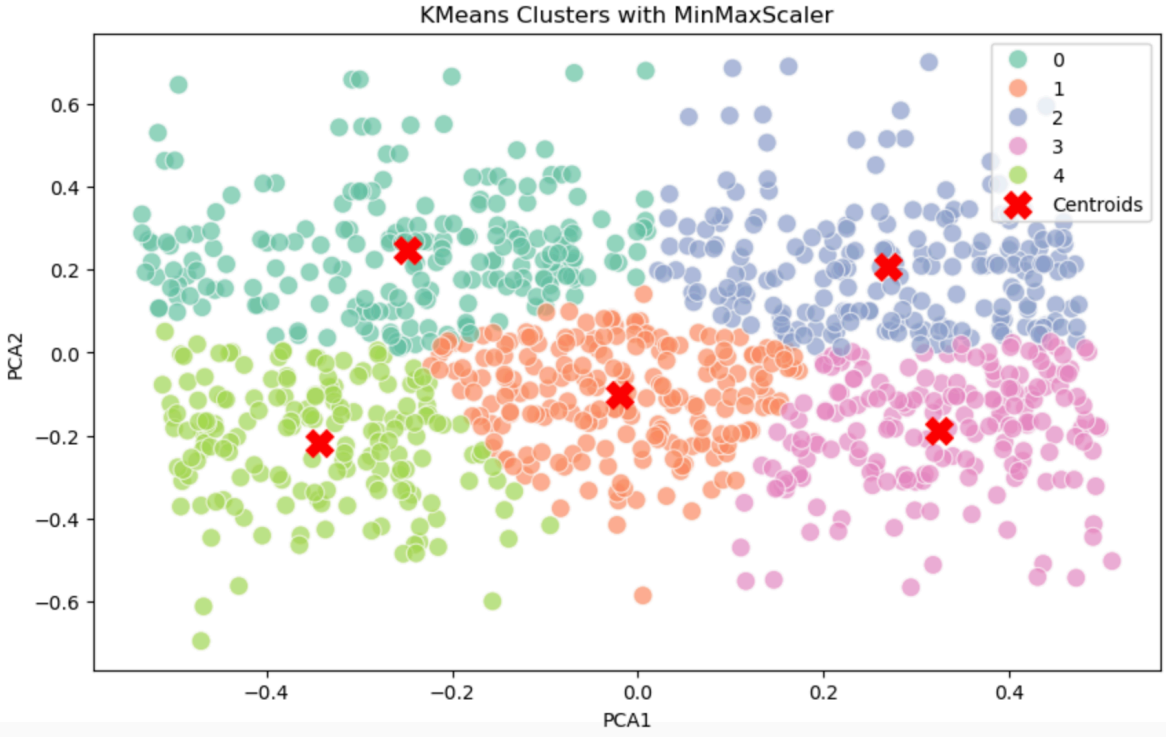
Clustering Results Report

1. Number of Clusters Formed:

- **Number of clusters used:** 3, 5, and other tested values
  - **For 3 clusters:** 3 clusters were formed based on the evaluation from KMeans clustering with MinMax scaling. Below is the Visual representation of 3 clusters:



  - **For 5 clusters:** 5 clusters were formed using KMeans with MinMax scaling.
  - Below is the Visual representation of 5 clusters:



2. DB Index (Davies-Bouldin Index):

- **DB Index for KMeans with 3 clusters:**
  - **Value:** 0.9804
  - **Interpretation:** The Davies-Bouldin Index for 3 clusters is below 1, indicating good cluster separation, but still leaving room for improvement.
- **DB Index for KMeans with 5 clusters:**
  - **Value:** 0.9703
  - **Interpretation:** The Davies-Bouldin Index for 5 clusters is slightly lower than that for 3 clusters, indicating a somewhat improved separation between clusters.
- **DB Index for Other Cluster Counts:**
  - **DB Index above 1.0** (for values other than 3 and 5 clusters) suggests that the clusters are poorly separated and not well-formed. This indicates that increasing the number of clusters beyond a certain point may result in overly fragmented or ill-defined clusters.

3. Silhouette Score:

- **Silhouette Score for KMeans with 3 clusters:**
  - **Value:** 0.3548
  - **Interpretation:** A moderate positive silhouette score suggests that the clusters are somewhat well-separated.
- **Silhouette Score for KMeans with 5 clusters:**
  - **Value:** 0.3112
  - **Interpretation:** The silhouette score for 5 clusters is slightly lower than that for 3 clusters, suggesting that adding more clusters leads to worse cluster cohesion and separation.

4. Cluster Metrics:

To analyze the performance and quality of clustering, additional metrics and evaluations were performed:

- **Cluster Visualization (PCA-based):** Using PCA for dimensionality reduction, clusters were visualized in a 2D space. The results showed the clustering structure for 3 clusters and 5 clusters, with some overlap in 5 clusters, suggesting the difficulty in discerning meaningful separations with more clusters.

5. Conclusion:

- The best results were observed with **3 clusters**, with a **Silhouette Score of 0.3548** and a **DB Index of 0.9804**.
- Increasing the number of clusters to **5** resulted in a slightly lower **Silhouette Score (0.3112)** and a lower **DB Index (0.9703)**, indicating potential overfitting or less meaningful divisions in the data.
- **DB Index greater than 1** in other cluster scenarios suggests that increasing clusters beyond a certain threshold may lead to poorly defined or less meaningful clusters.

6. Recommendations:

Based on the analysis of Silhouette Score and Davies-Bouldin Index, **3 clusters** appear to provide the best balance of cohesion and separation. Therefore, **3 clusters** would be the recommended choice for the current clustering task.