In this lab, we implemented the K-Means clustering algorithm, which assigns data points to clusters based on their proximity to centroids. One challenge with K-Means is determining the optimal number of clusters, as it does not inherently provide this information like DBScan does. To address this, we evaluated the clustering performance using silhouette scores.

Using the Olivetti face dataset, we examined how different cluster counts affected the quality of the groupings. Our results showed that after five clusters, the variation within clusters became highly volatile. We observed inconsistencies in cluster density and size, indicating that higher values of clusters led to suboptimal grouping. By focusing on the range of three to five clusters, we identified a more balanced distribution, though some distinctions remained.

Comparing silhouette scores across different cluster counts, we found that the difference between four and five clusters was minimal, whereas three clusters showed a significant contrast. While an argument could be made for using two clusters, the imbalance in cluster density made three the better choice.

In conclusion, based on our analysis, three clusters provided the most optimal separation for our dataset. This experiment reinforced the importance of context in clustering—choosing clusters is not just about numerical optimization but also about interpreting the results in a meaningful way.