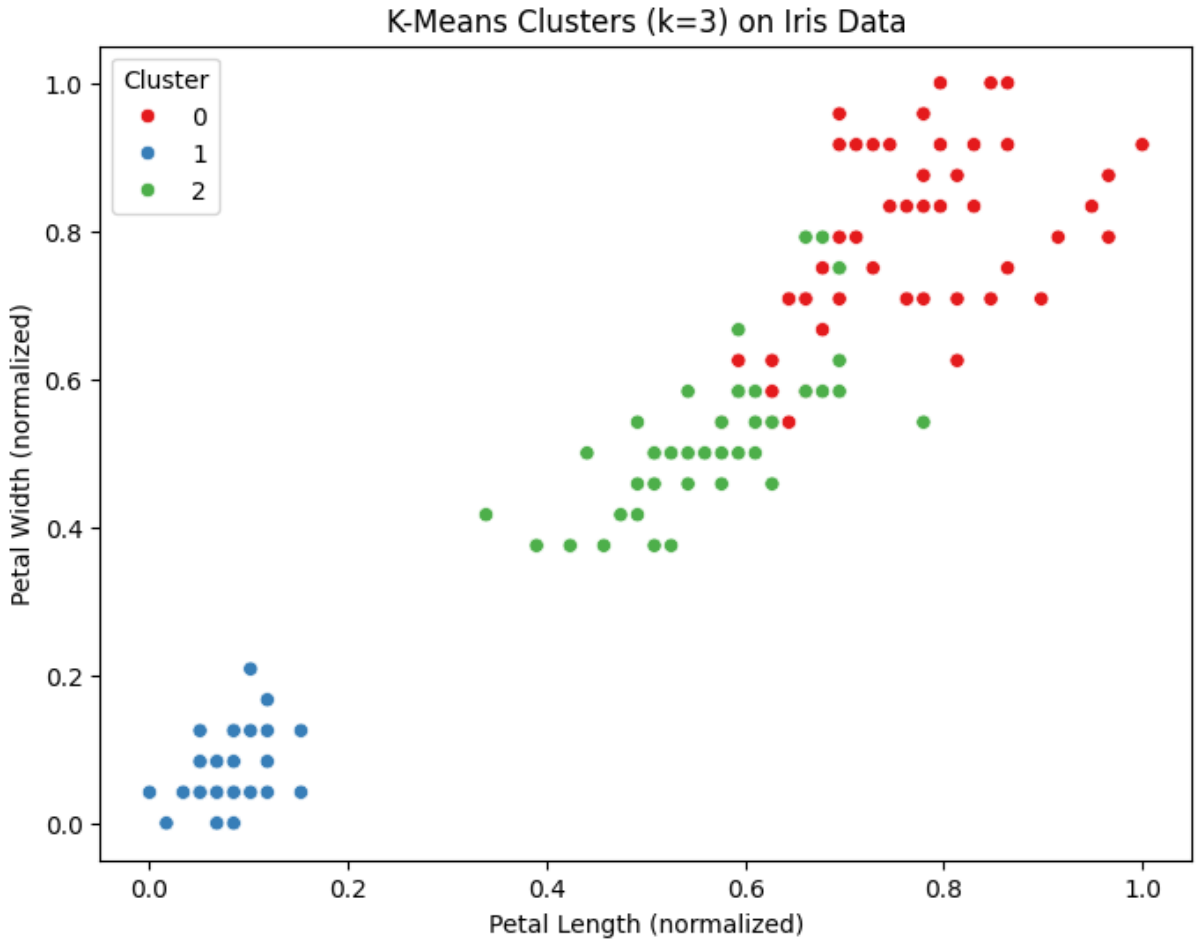
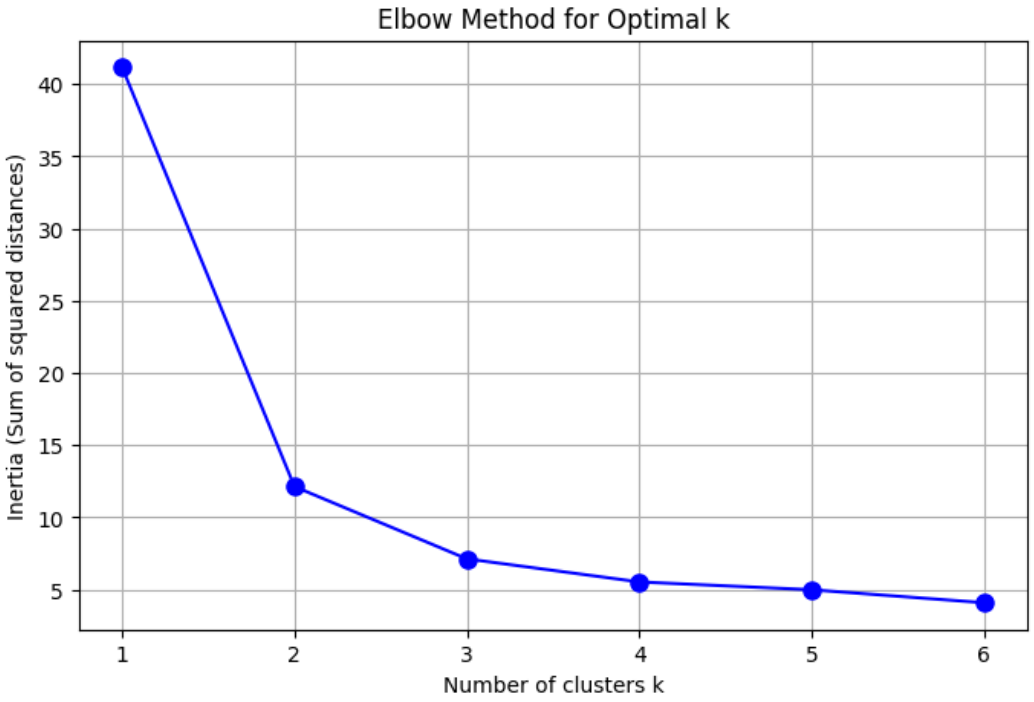
## CLUSTERING ANALYSIS

The K-Means clustering with k=3 demonstrated strong overall cluster quality by effectively grouping the majority of samples according to their true species. One cluster was distinctly separated, reflecting a species with more unique feature characteristics. However, the other two clusters exhibited some overlap, resulting in misclassifications between those species. This overlap highlights natural similarities in their measured features, making perfect separation difficult for an unsupervised algorithm that does not use class labels.




The elbow curve analysis confirmed k=3 as the optimal number of clusters, which aligns with the known number of species in the dataset. Visualization of the clusters using petal length and width showed mostly distinct groupings, but also revealed blurred boundaries where species features are less differentiated.




In real-world applications such as customer segmentation, clustering algorithms help identify distinct groups based on purchasing behavior, demographics or preferences, enabling more personalized marketing and improved resource allocation. Misclassifications can reduce the effectiveness of these efforts by incorrectly grouping individuals, so understanding cluster quality is critical.

If synthetic data were used instead of real data, clustering performance could be affected depending on how accurately the synthetic features mimic actual distributions. Poorly generated synthetic data may introduce unrealistic overlaps or noise, thus reducing cluster accuracy and interpretability.