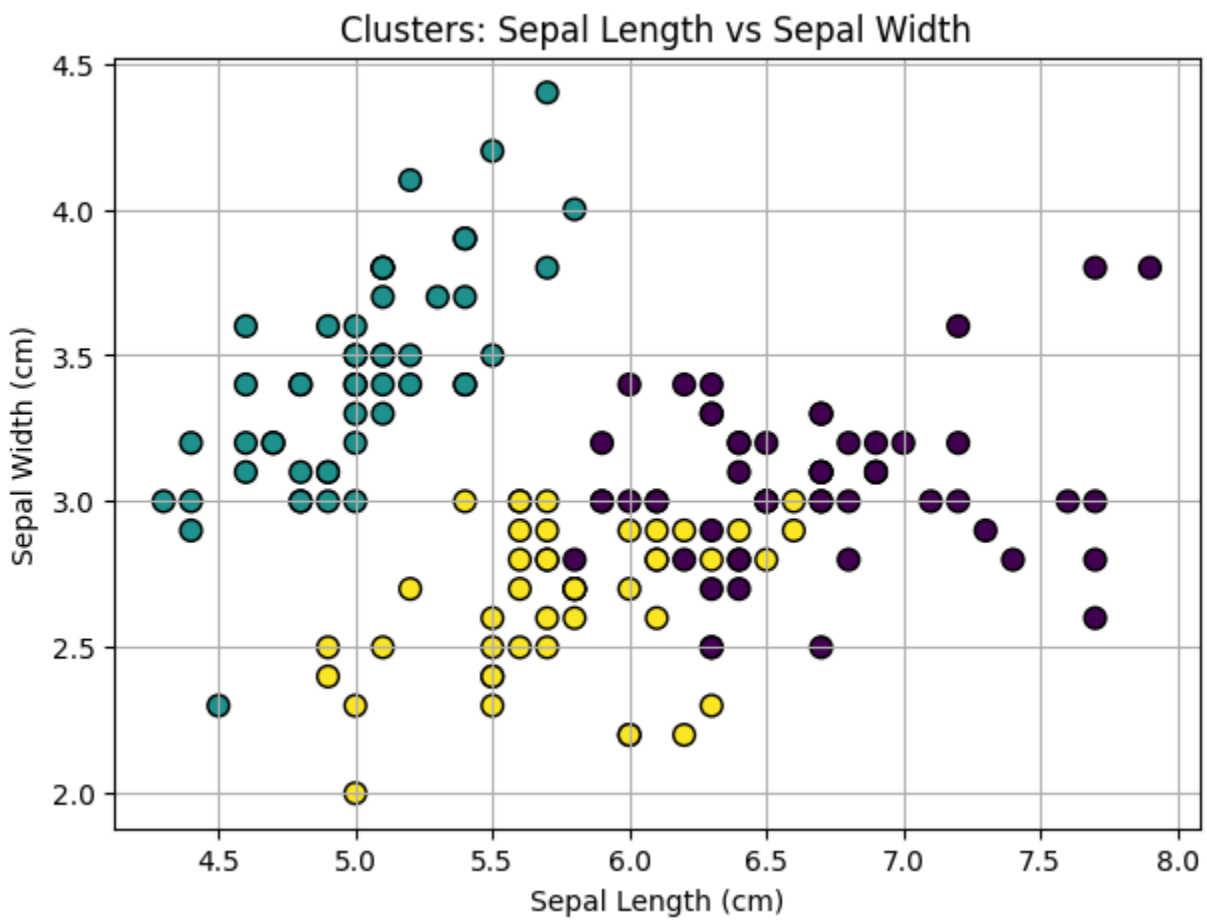
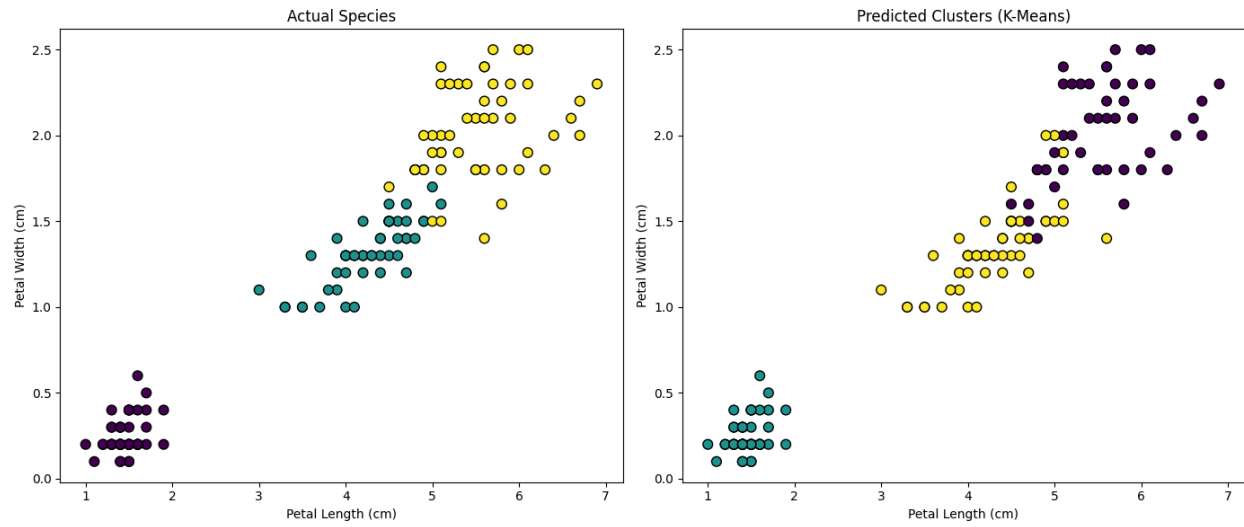


## **Clustering analysis**

The K-Means clustering results on the Iris dataset show reasonably strong cluster quality, especially given that the algorithm is unsupervised and the dataset contains natural overlap between classes. The Adjusted Rand Index (ARI) score—typically around 0.6 to 0.75—indicates that the clusters align well with the true species labels, especially for *Iris setosa*, which is well separated from the other two species. Most misclassifications occur between *Iris versicolor* and *Iris virginica*, whose feature distributions overlap significantly, particularly in petal length and width. This overlap makes it difficult for distance-based methods like K-Means to clearly separate the two groups.

The elbow method supports  $k=3$  as the optimal number of clusters, validating the biological assumption of three species. PCA visualizations show that while one cluster is clearly distinct, the remaining two display partial mixing, reflecting natural biological similarity.

In real-world applications such as customer segmentation, similar clustering techniques help identify groups with shared behavior or preferences without requiring labeled data. Since the Iris dataset is real and not synthetic, the results are reliable; synthetic data might otherwise produce misleadingly distinct or poorly structured clusters. Overall, the clustering is effective but reveals limitations when class boundaries overlap.



K-Means Clusters (k=3)  
Petal Length vs Petal Width

