Best K

Figure 1: Training and Testing Accuracy of kNN

k	Training Accuracy	Testing Accuracy
21	0.81	0.80
29	0.82	0.80
28	0.83	0.80
37	0.83	0.80
36	0.81	0.80
22	0.82	0.80

Table 1: Training and Testing Accuracy for top k values

The best value for k appears to be 21 because the accuracy of training and testing is close, as we can see in Figure 1. Also, from table 1, we can tell that k = 21 gives one of the highest testing accuracies, and the training accuracy is not far off. This shows that the model is not overfitting too much and does a good job with both the train and test data.
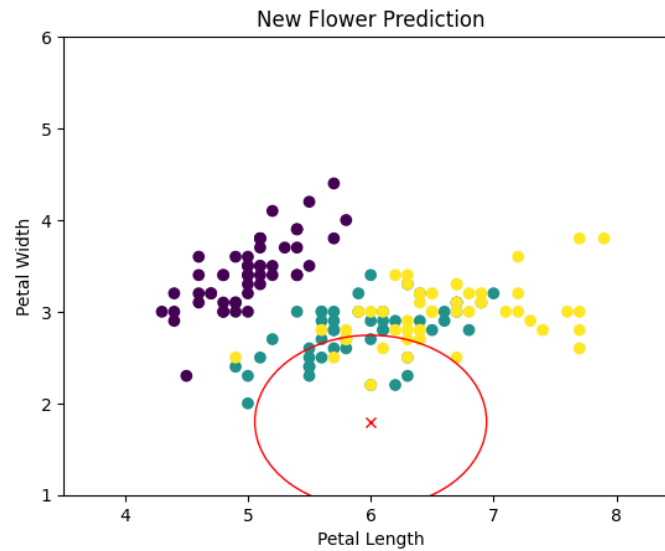
kNN Model

Figure 2: Prediction of New Flower

Figure 2 shows the prediction of a new flower using kNN with k=21. The red x is the new flower and the red circle shows the neighborhood used for prediction.

K-Means Clustering Model

Figure 3 shows the actual groupings of species according to the length and width of the petal. We can see that there are three pretty clear clusters, especially the purple one on the left.

In figure 4, we see the result of the clustering of K-means with k = 3. The model did a good job separating the cluster on the left, but for the other two groups there is some mixing. This could happen because their features are more similar and K-means only look at distance. Overall, K-means did well to find the general structure, but it is not perfect for species with more overlap.
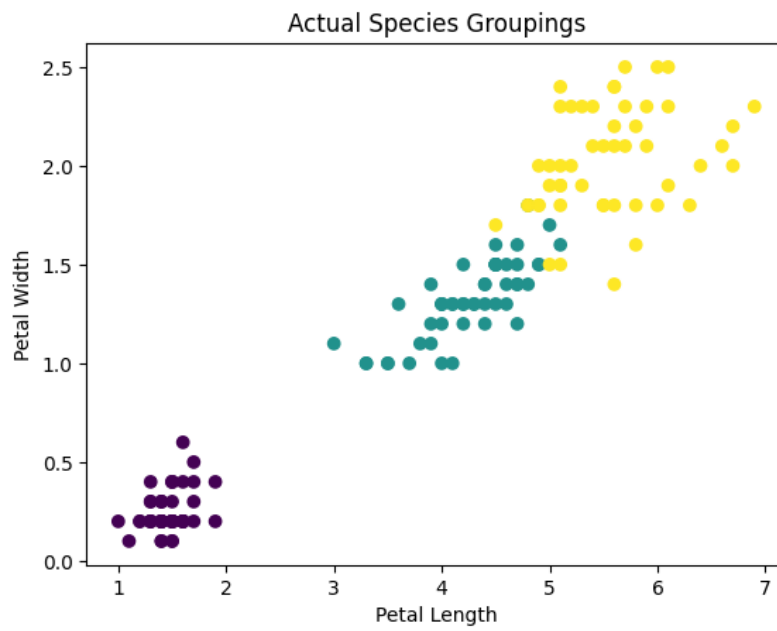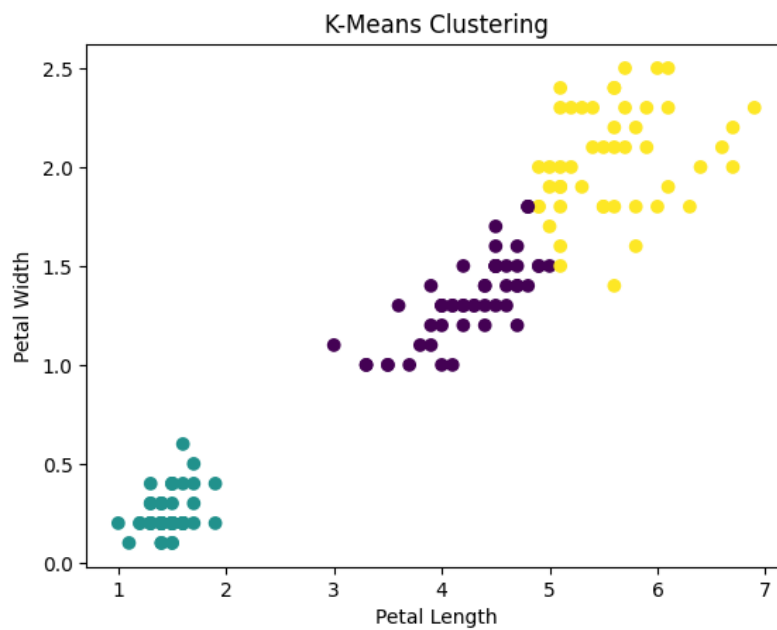
Figure 3: Actual Species Groupings

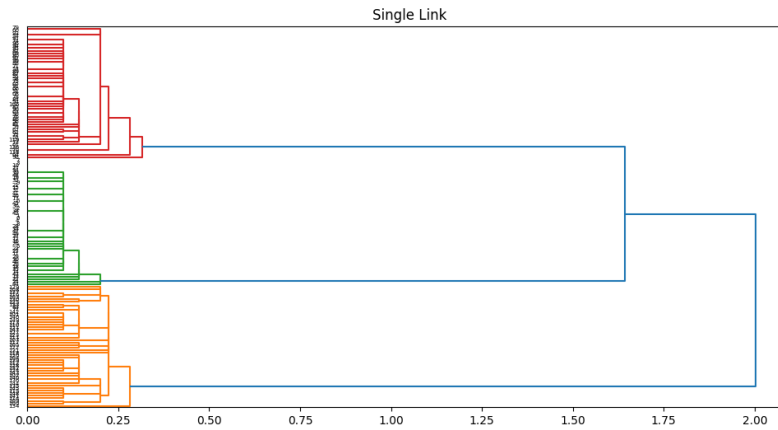Figure 4: K-Means Clustering

Hierarchical Clustering

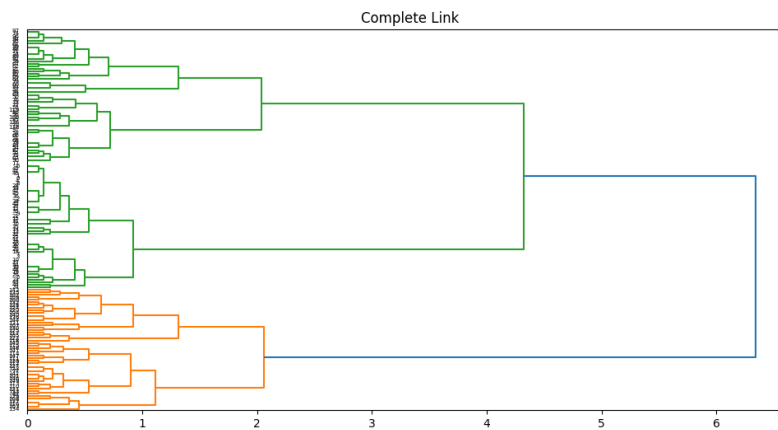Figure 5: Hierarchical Clustering with Single Link

Figure 6: Hierarchical Clustering with Complete Link

In the single link dendrogram (figure 5), we see that clusters start merging early based on the closest points, which creates long chains. This can cause very uneven groupings and is sensitive to outliers.

The complete link method (figure 6) keeps clusters more compact by looking at the farthest points. The structure looks more balanced, and the large clusters are clearer.

The group average link (figure 7) gives a result that is kind of in between like more balanced than single link, but not as tight as complete link.
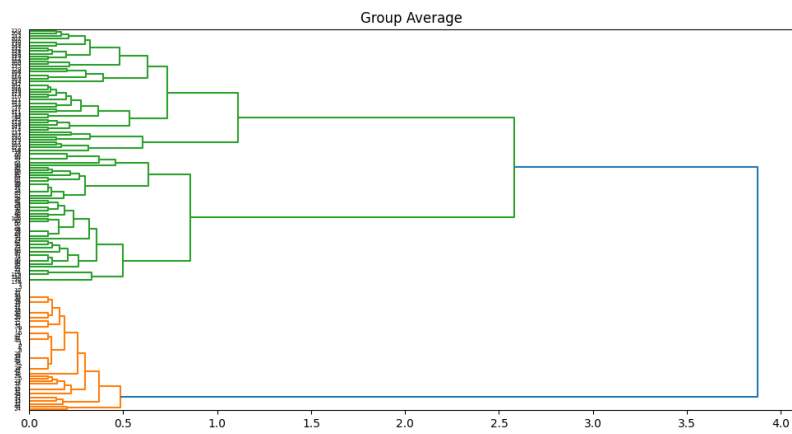
Figure 7: Hierarchical Clustering with Group Average Link

Overall, the three methods give different clustering shapes, and the choice of linkage can really affect how the data is grouped.