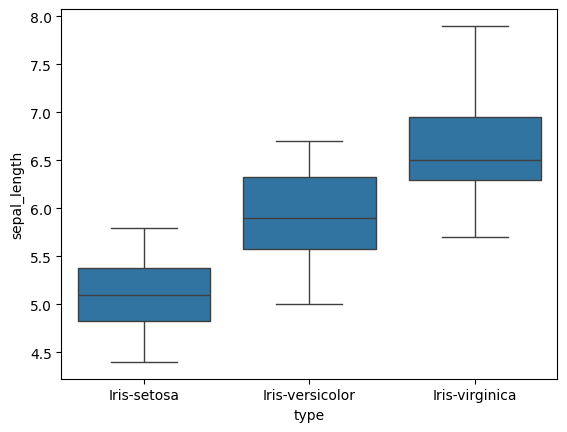
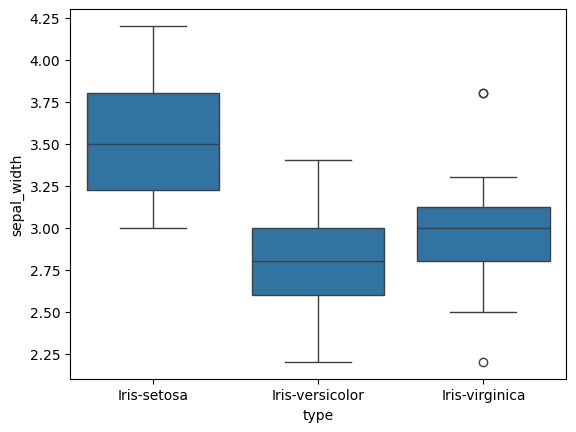
**Understanding the Problem and Data:**

The data set contains three different types of flowers, setosa, versicolor, and virginica. Along With contains the lengths and widths for both the petals and the sepals of each individual flower. The purpose of this analysis is to create a model that can accurately predict the type of flower based on the features of the sepal and petals.

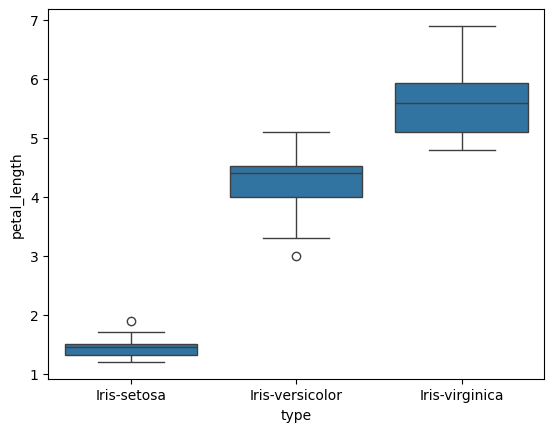
To understand the relationships between the lengths and widths of the different sepals and petals I will create boxplots to visualize each relationship with the type of flower.



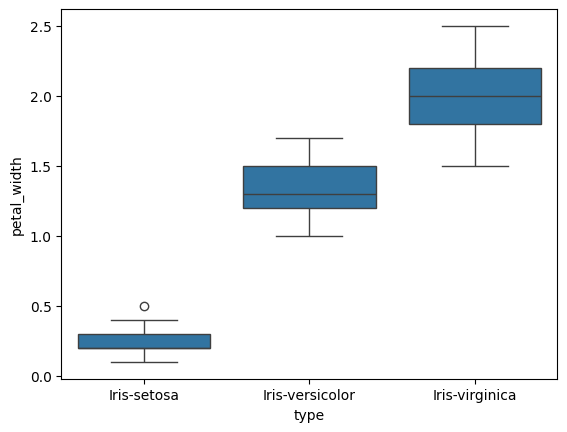
It appears that the setosa has the lowest median sepal length, the virginica has the highest median sepal length, and the veriscolor inbetween.



The setosa appears to have the highest median sepal width, the versicolor has the lowest median sepal width, and the virginica slightly above the versicolor.



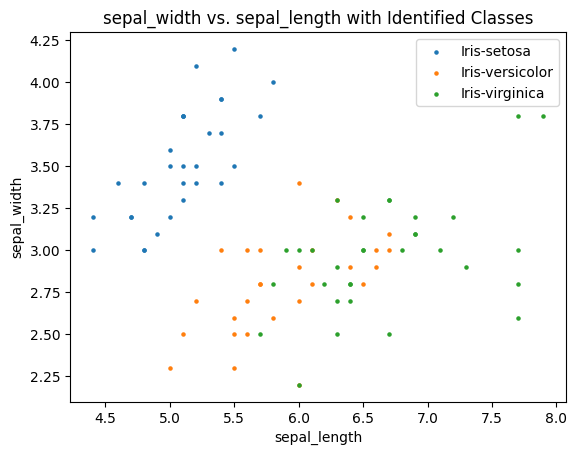
The setosa appears to have a much lower petal median petal length, while the virginica has the largest median petal length with the versicolor not far below.



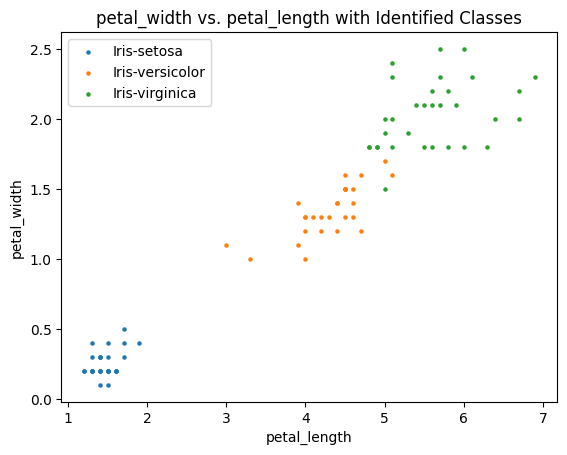
Again, we see the setosa much lower than the versicolor and virginica in terms of petal median width. Along with the virginica having the largest petal width, and the versicolor being close behind.

**Model Specification:**

For our analysis, we are looking to create a model that can predict the correct flower type based on the width and lengths of their petals and sepals. When looking at graphs we are hoping to see separation between the different types of flowers to make it easier to draw a line/plane that can differentiate between the different types.



Looking at the sepal graph, we can see that the setosa flower is clearly separated from the other two flowers. However, veriscolor and verginica are closely related and may make it hard to determine the type.

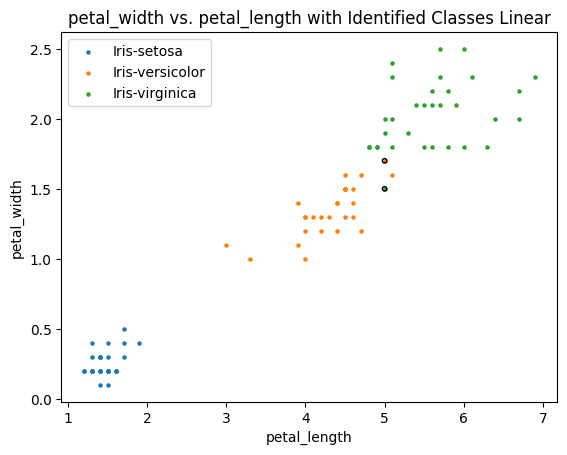


Looking at this graph the setosa is very separated again and while the versicolor and virginica are somewhat close they are more differentiated than the sepal graph.

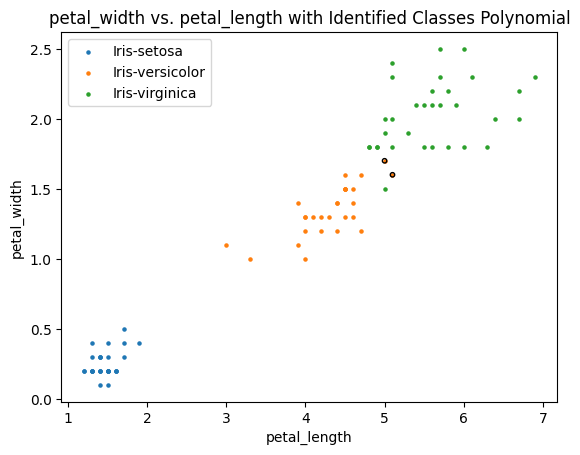
Knowing this I believe that using the petal graph is correct as it is easier to differentiate between the different flower types when using the petal graph.

**Model Estimation 2 Feature:**

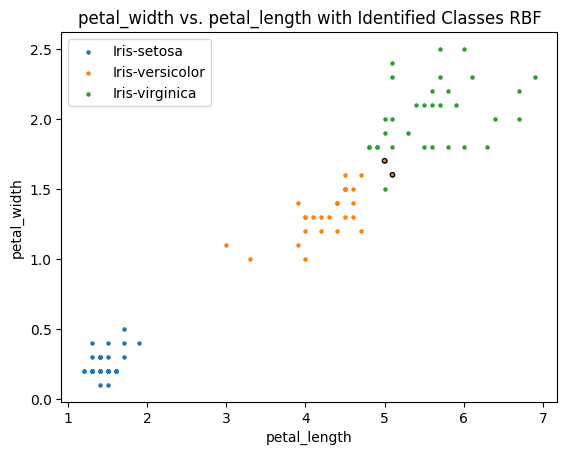
Moving forward with the predictor profile that uses the petal length and width, I tuned the support vector classifier with a linear, polynomial, and rbf kernel. After doing this I created a graph for each showing the misclassified cases.



For the linear kernel, the best paramaters was C=100 which produced this graph.



For the polynomial kernel, the best parameters were C=1 and Degree=3 which produced this graph.



With the RBF kernel, the best parameters were C=1 and Gamma=1

**Model Estimation 4 Feature:**

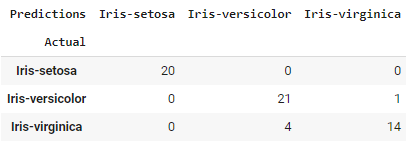
For the model that contained all four features, the best parameter values were as follows:

* Linear: C=10
* Polynomial:C=10, D=2
* RBF: C=1, Gama=1

**Model Assessment 2 Feature:**

To assess the models, we will create a classification matrix that will allow us to calculate the misclassification rate.

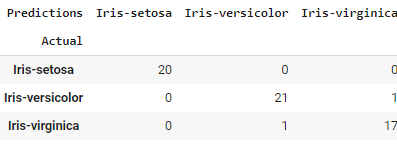
Linear Misclassification Rate: .083



Polynomail Misclassificaiton Rate: .033



RBF Misclassification Rate: .033



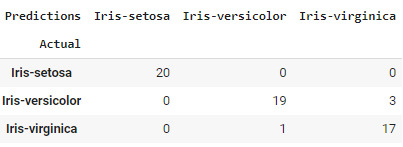
**Model Assessment 4 Feature:**

We will then do the same with the model containing all four features.

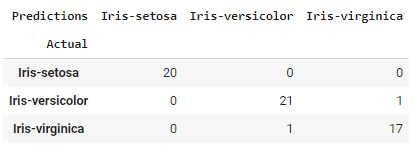
Linear Misclassification Rate: .067



Polynomial Misclassification Rate: .067



RBF Misclassifictation Rate: .05



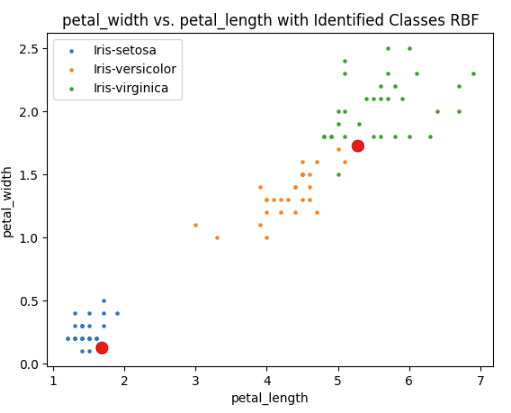
**Model Deployment:**

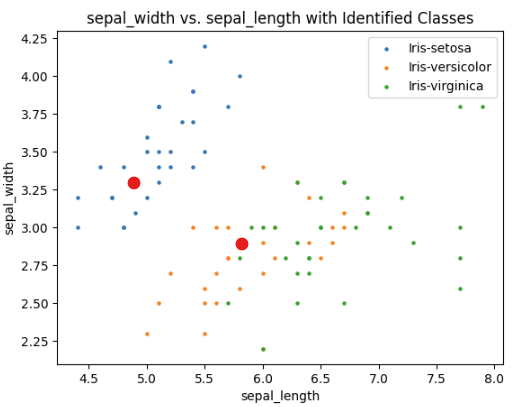
To evaluate the models, we will look at which have the lowest misclassification rate. The best linear model is the four feature model as it has the lower misclassification rate of .067 as opposed to the two-feature model that had a misclassification rate of .083.

The best non-linear models are the two feature models as they both had a misclassification rate of .033 for RBF and polynomial, while the four feature models had a .067 and .05 misclassification rate for polynomial and RBF respectively.

When looking at these points:

∗ Iris 1: sepal length of 4.75, sepal width of 3.25, petal length of 1.75, and petal width of 0.25.  
∗ Iris 2: sepal length of 5.75, sepal width of 2.75, petal length of 5.25, and petal width of 1.75.





We should be more confident with the first iris because it is in the bottom left on the petal graph and the top left on the sepal graph where it is mostly setosa flowers. Whereas iris 2 is in the gray area between versicolor and virginica on both the sepal and petal graphs.