

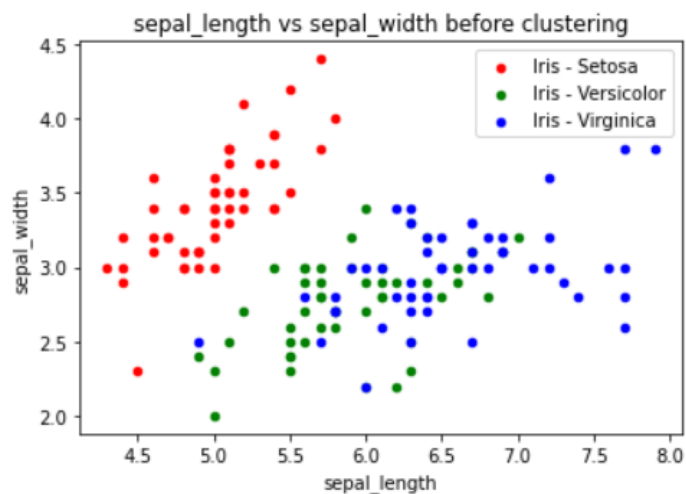
Name: Bhavani Sai Shriya Anumala  
ID: 1001870184

## Project 3 - Report

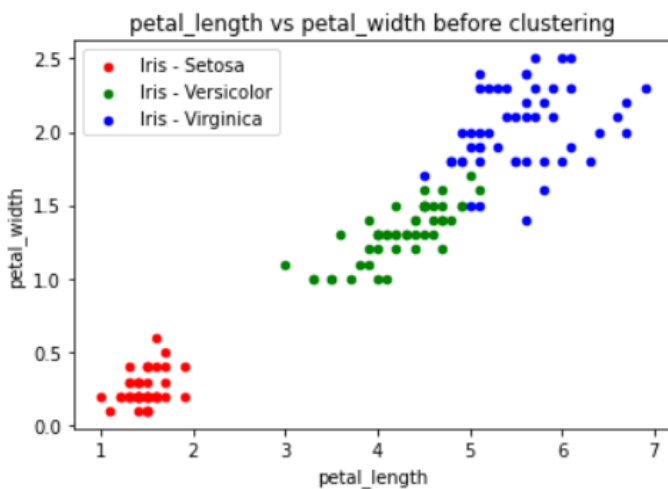
### **K Means Clustering:**

Firstly we loaded all the necessary libraries and the dataset, iris.data and gave column names to it.

Later we visualized the dataset depending on the sepal\_length and sepal\_width.



And we also visualized the dataset based on petal\_length and petal\_width.



Here, we have used the LabelEncoder to convert the strings in the species column into numeric.

Now, we create clusters using the K Means clustering algorithm where the K value is 3 one for each of the class values, and find the cluster values using

`Km.cluster_centers_`

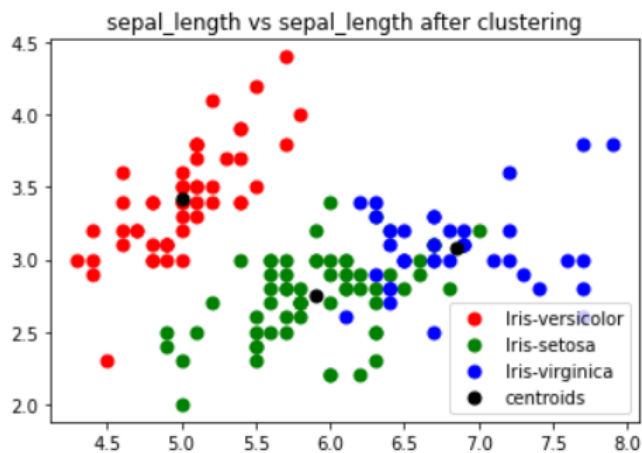
Following are the cluster centers of the iris.data

```
array([[5.9016129 , 2.7483871 , 4.39354839, 1.43387097],  
       [5.006      , 3.418      , 1.464      , 0.244      ],  
       [6.85      , 3.07368421, 5.74210526, 2.07105263]])
```

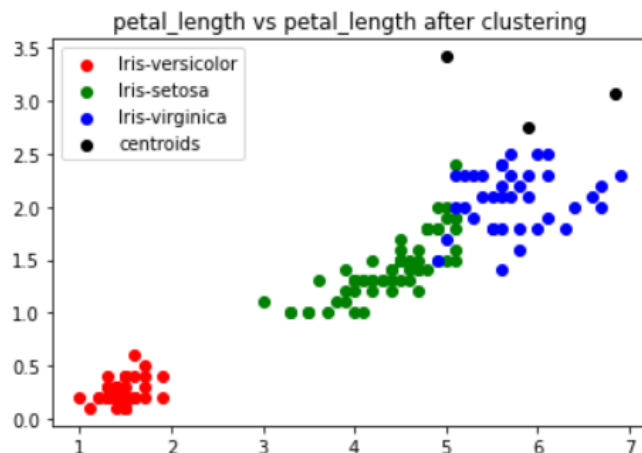
Now we visualize the data after creating the clusters.

Here is the scatter plot for the dataset after clustering with centroids.

This plot is based on the sepal\_length and sepal\_width.



This plot is based on the petal\_length and petal\_width after clustering.



The difference between the plots before and after the clustering is that after

clustering most of the blue dots become green.

## Evaluation:

To calculate the accuracy we have considered the species column and the values predicted after the k Means clustering. The accuracy evaluated is 0.893 or 89.3%

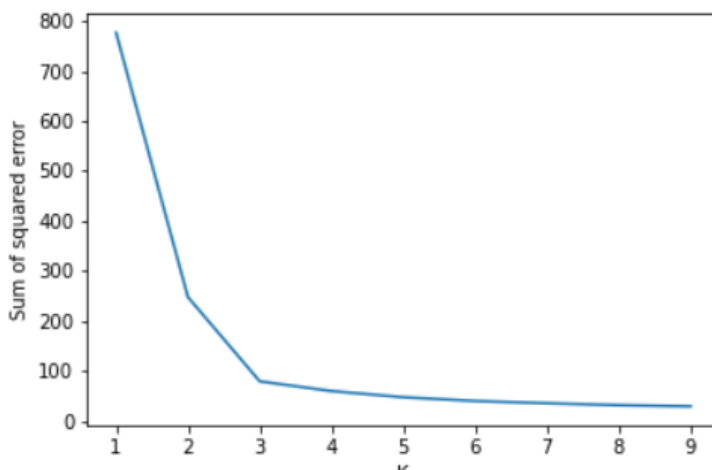
Accuracy\_score is used to calculate the accuracy

```
In [18]: ▶ sm.accuracy_score(pred, s_val)
Out[18]: 0.8933333333333333
```

Confusion matrix is

```
sm.confusion_matrix(pred,s_val)
array([[50,  0,  0],
       [ 0, 48, 14],
       [ 0,  2, 36]], dtype=int64)
```

## Elbow Curve:



Ref:

<https://pythoninai.com/k-means-clustering-python-code/>

