**Predicting Flower Species Using Logistic Regression Model**

By Mushfiqur Rahman Khan Majlish

**Methods:**

1. Necessary libraries and datasets are imported. The dataset was downloaded from Kaggle.

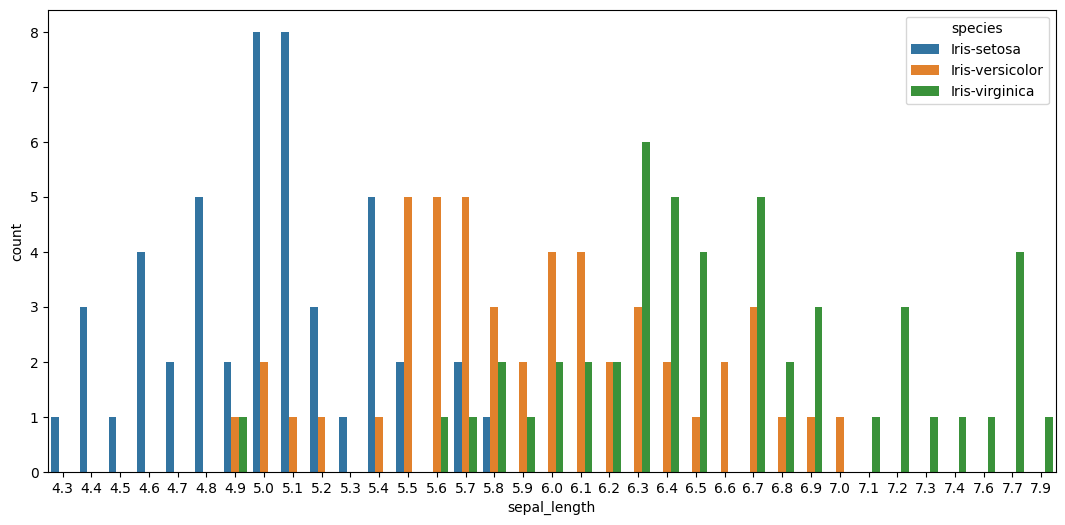
It contains 5 variables and 150 observations. The dataset has no missing values although there are 3 duplicated values.

1. The duplicated values are removed. Now the dataset has 147 observations and 5 variables,
2. Suitable graphical representations are used to notice and compare the pattern of variables.
3. Pearson correlation was calculated and presented in a heatmap.
4. Logistic Regression Model was applied which was used to predict the species.

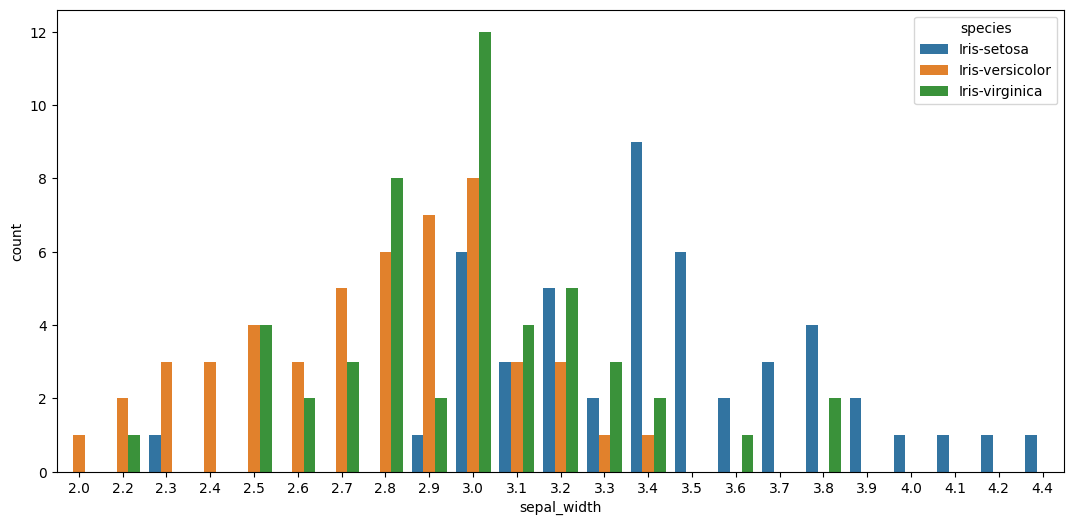
**Variable List:**

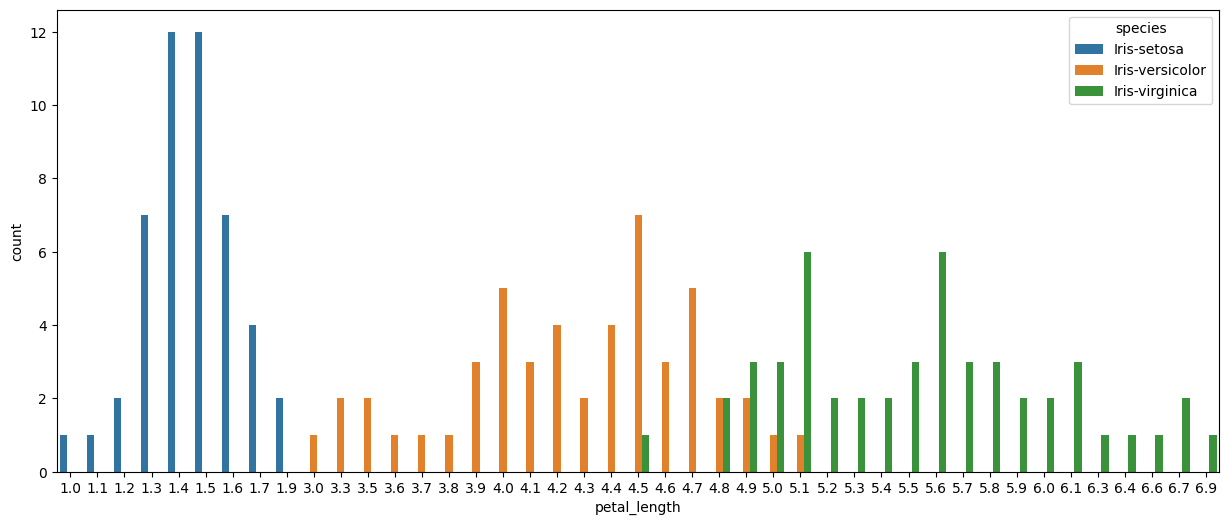
1. Dependent Variable: Species (species of Iris flower)
2. Independent Variable: Sepal length (length of the sepal), sepal width (width of the sepal), petal length (length of the petal), petal width (width of the petal)

**Data Visualization:**

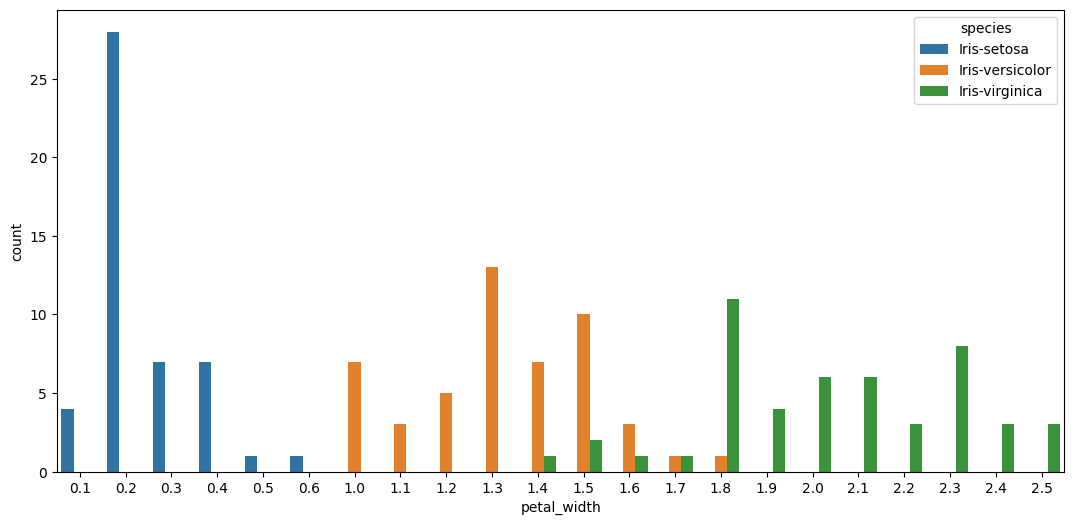


We can see from the graph that sepal length of iris setosa species is comparatively lesser than that of other two species. On the other hand, iris virginica had the largest sepal length among all three species.

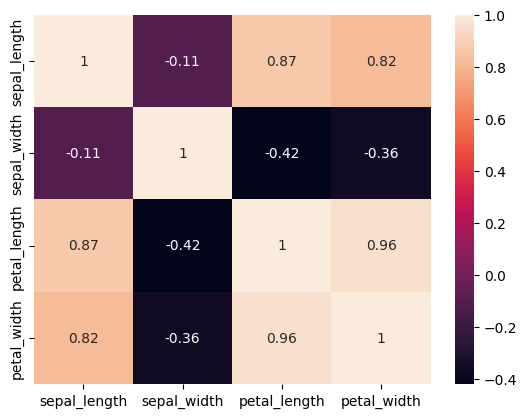
In case of sepal width, iris setosa has the wider sepals of all the three species. On the contrary, iris versicolor has the narrower sepals.



It is evident from the graph that the petals of iris setosa are smaller than that of other species. Iris virginica has the largest petals of the three species.



Petal width follows a similar pattern to petal length. Iris setosa has the narrower petals whereas iris virginica has the wider petals.



The heatmap showcases the correlation of the numerical variables in the dataset. We can see from the heatmap that there are several strong positive relationships as well as some negative weak relations among the variables.

**Result:**

A Logistic Regression Model was employed to predict the species using the numerical variables as independent variables. The model was able to predict the species with 95.5% accuracy score. Hence, it is a pretty good fit.