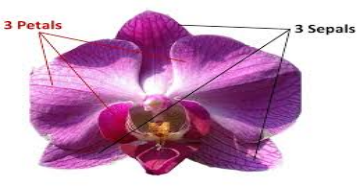
**IRIS Dataset**

The Iris Dataset contains four features (length and width of sepals and petals) of 50samples of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). ... The dataset is often used in data mining, classification and clustering examples and to test algorithms.



A flat coloured, ring-shaped membrane behind the cornea of the eye, with an adjustable circular opening (pupil) in the centre

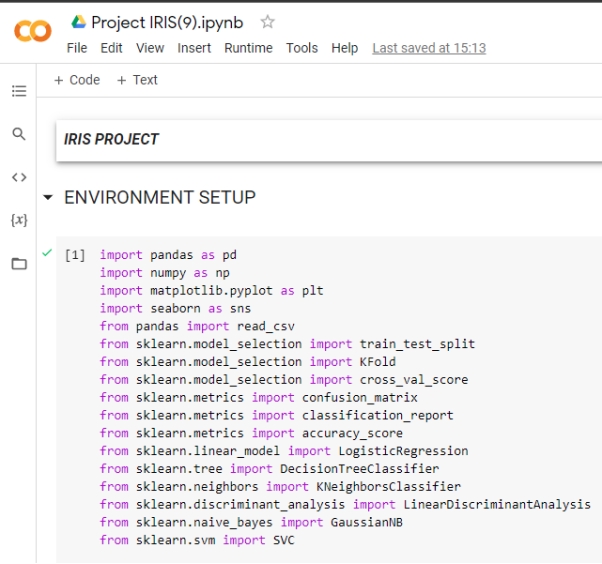
A plant with showy flowers, typically of purple or yellow, and sword-shaped leaves. Irises are native to both Eurasia and North America and widely cultivated as ornamentals

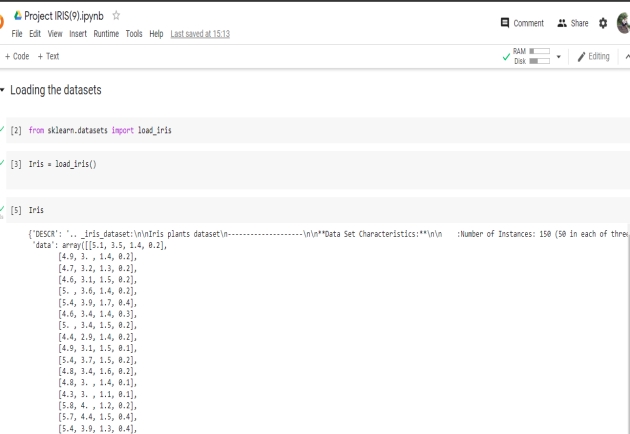
#### **Supervised learning on the iris dataset**

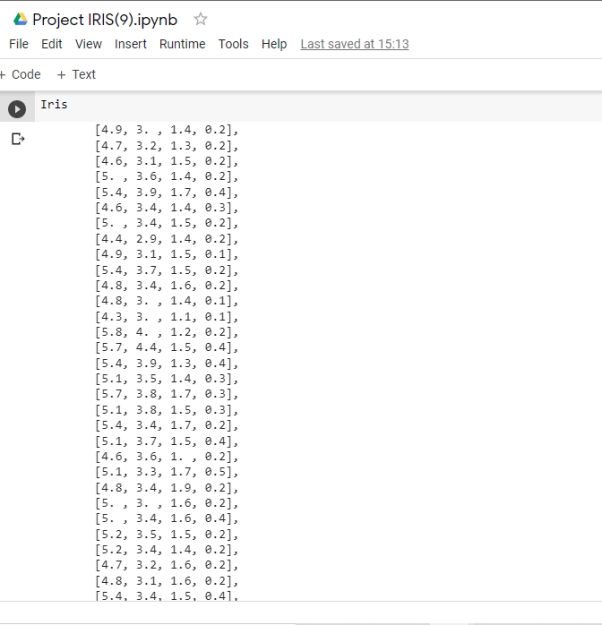
1. Framed as a supervised learning problem

2. Predict the species of an iris using the measurements

3. Famous dataset for machine learning because prediction is easy





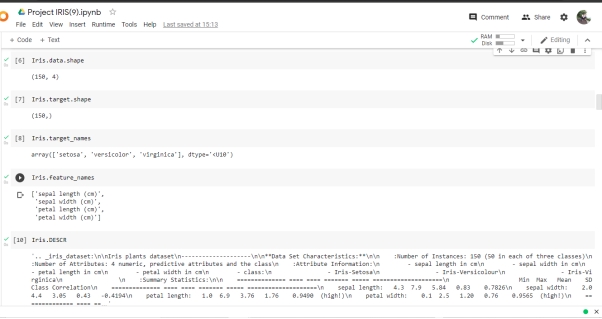




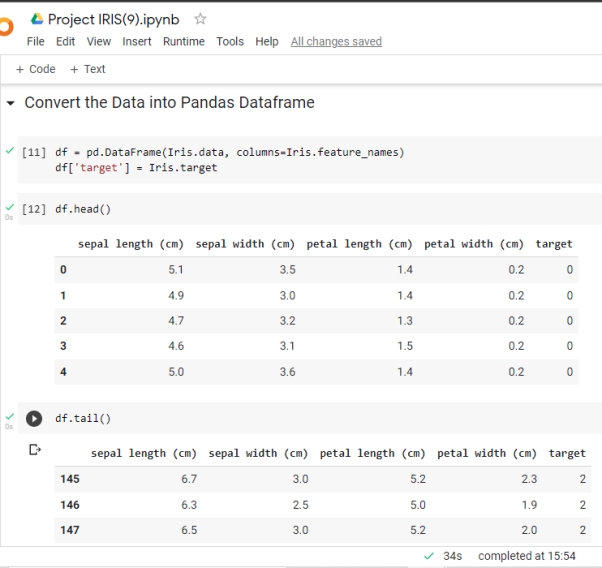
#### **Machine learning terminology**

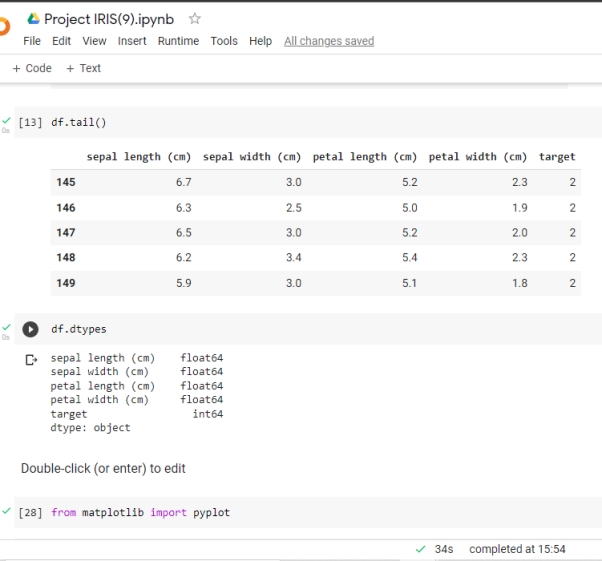
* Each row is an observation (also known as: sample, example, instance, record)
* Each column is a feature (also known as: predictor, attribute, independent variable, input, regressor, covariate)

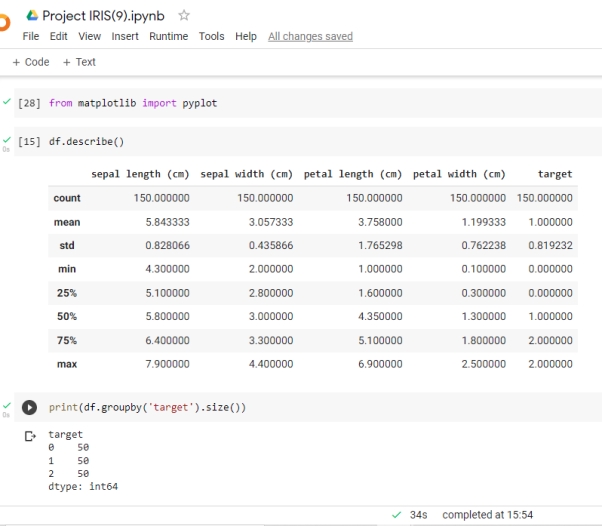
#### **Exploring the Iris dataset**



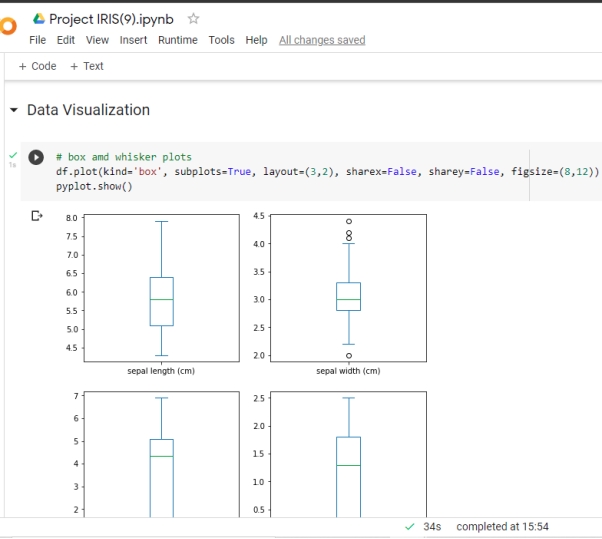
Creating a table like look for our data, with the help of Pandas library.

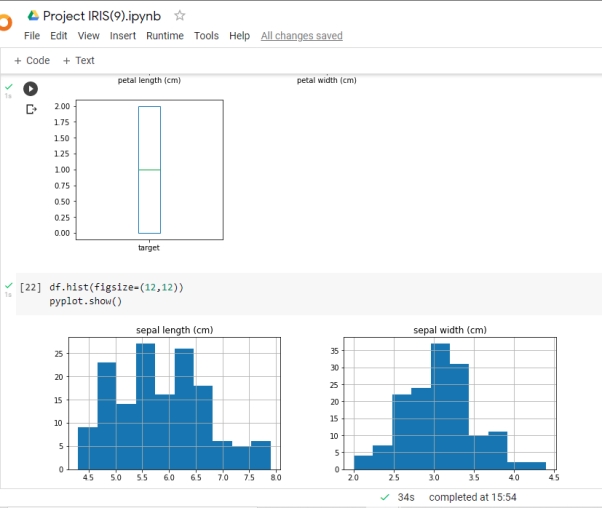


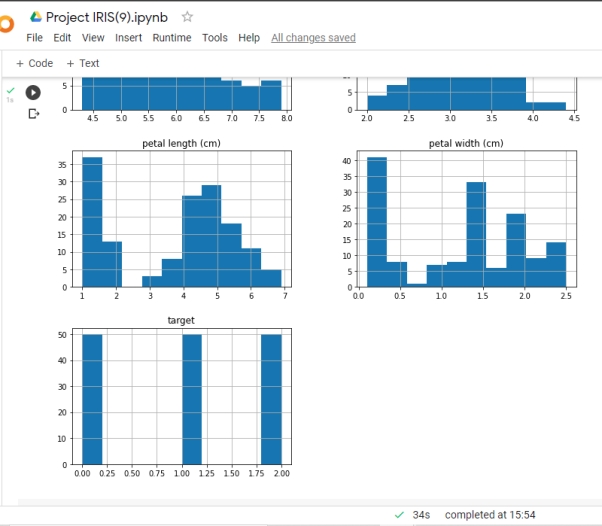


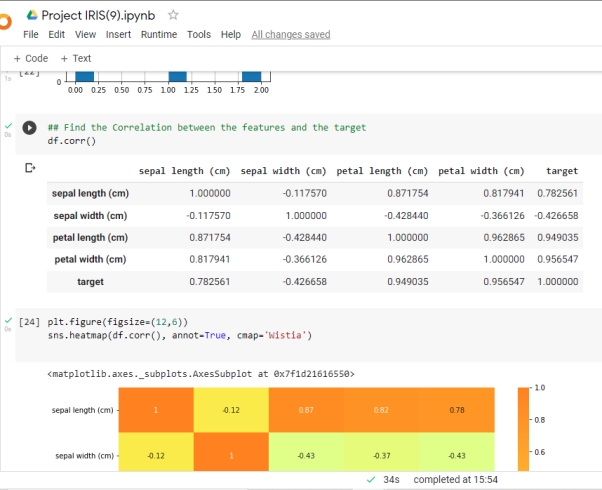


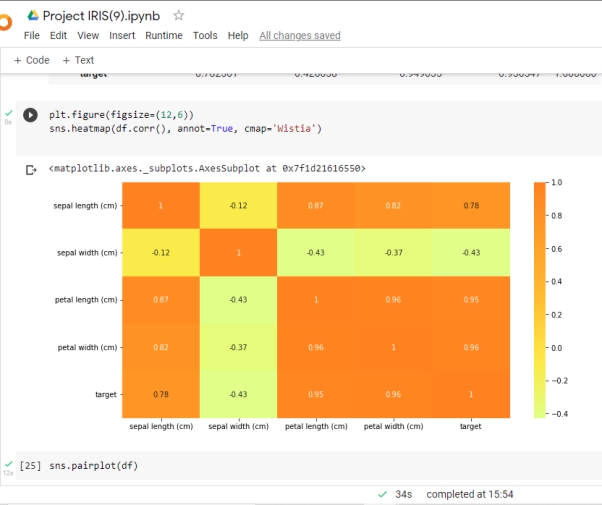
### **Data Visualization**

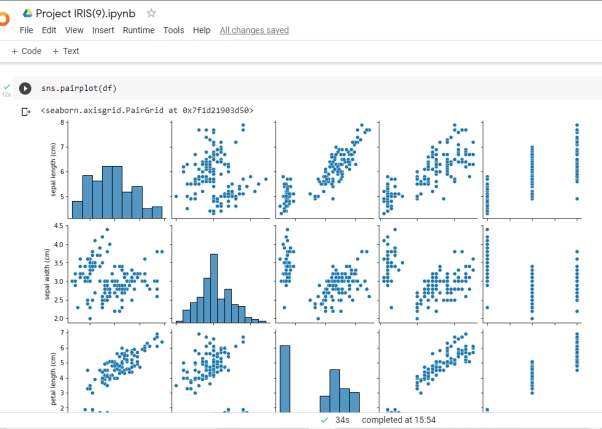


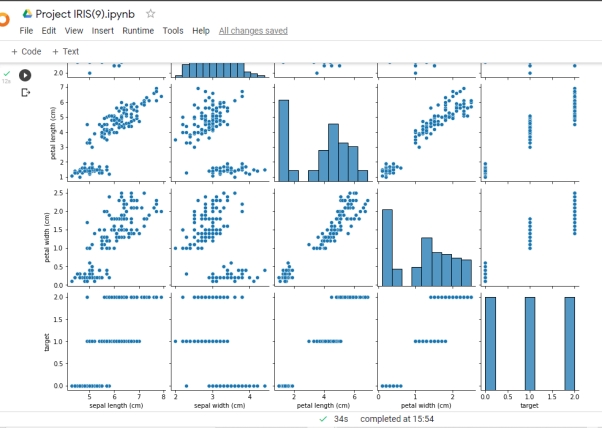




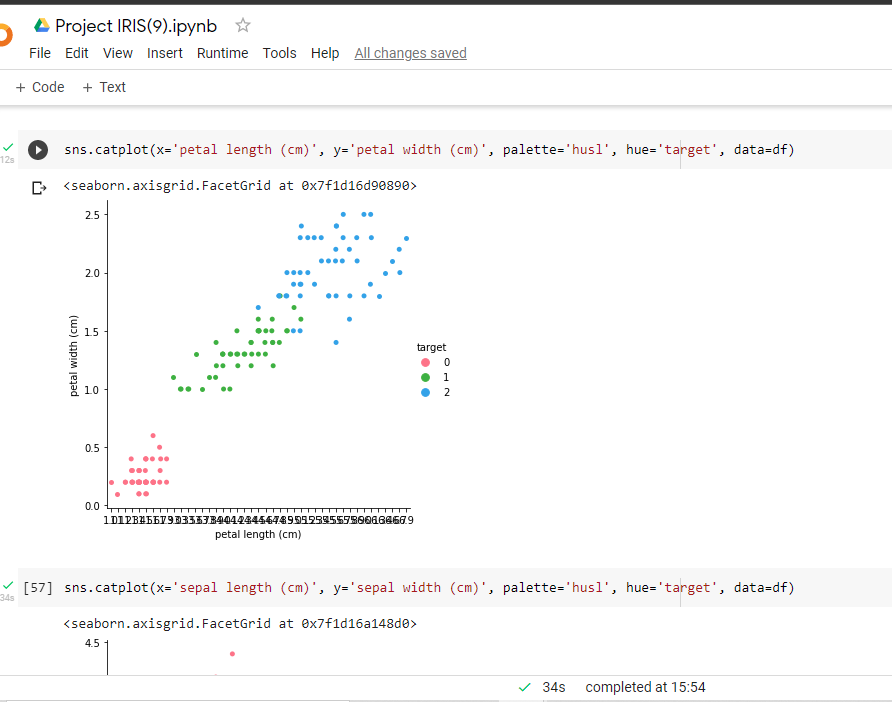


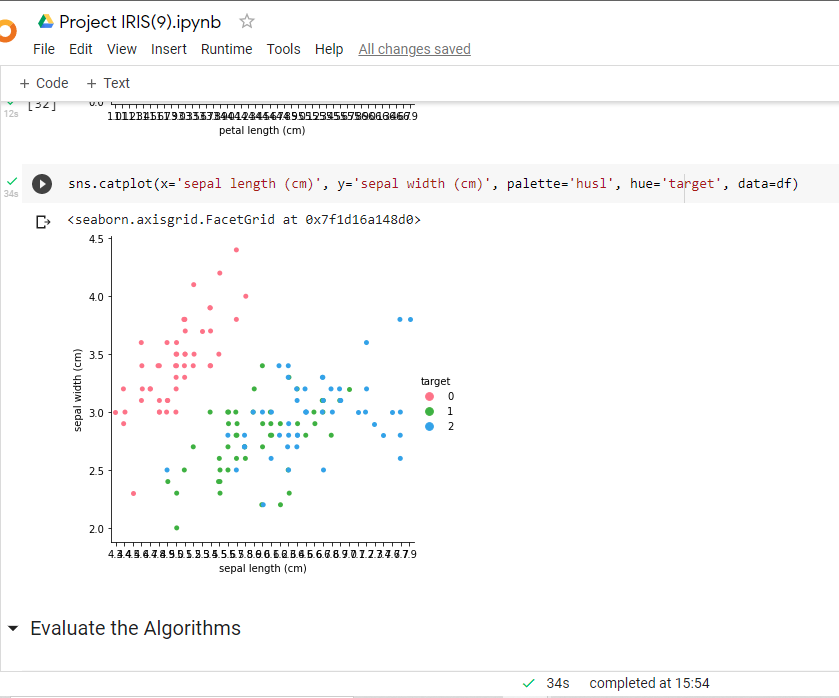






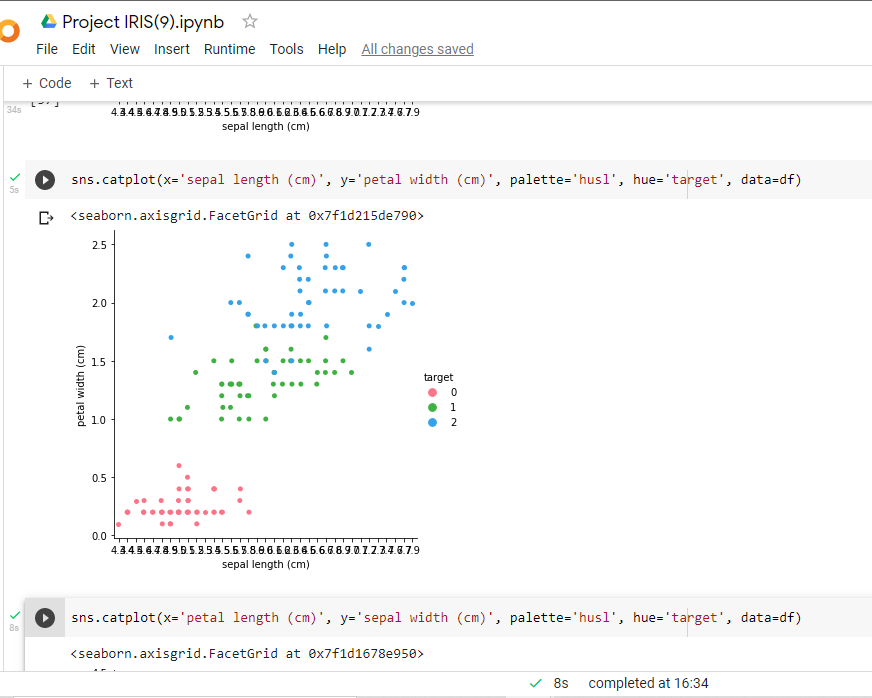
Next step : Visualization of the features. We will plot the combinations of given features in form of scatter to derive the relationship and correlation between the features.





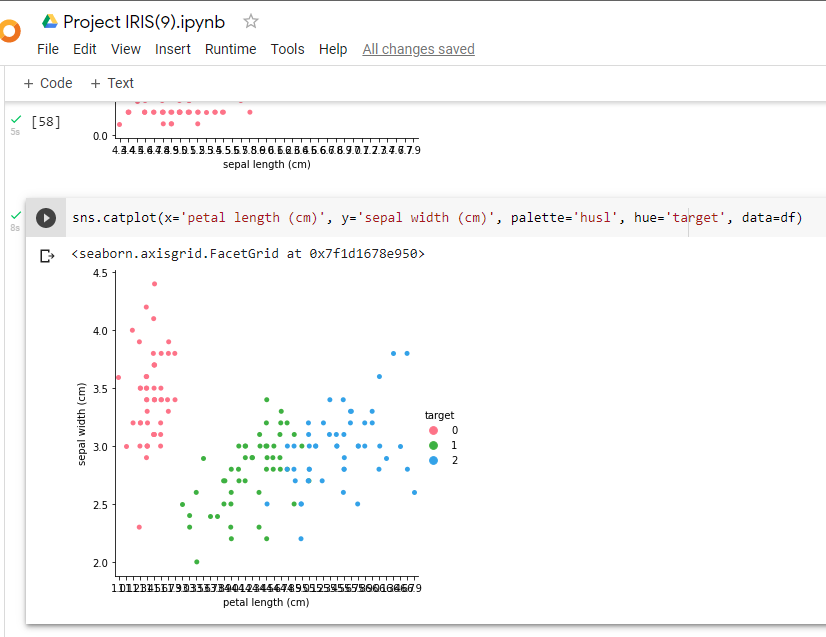
#### **Observation**

we observe that the targets are easily differentiated in the above scatter plot.  
Petal length and petal width are suitable attributes they have  
the ability to predict the output accurately. We also confirm this by  
numerical estimation by finding out the correaltion coefficient r  
r is very close to 1 which infers that correlation is very strong.



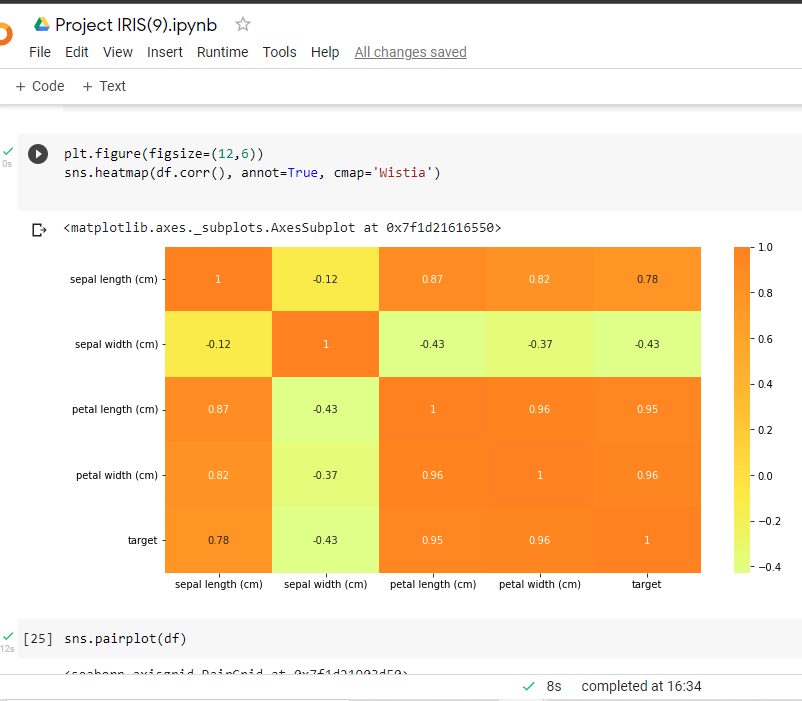
#### **Observation**

here we can distinguish among various targets(species of IRIS flower)  
but we observe that there exists certain values that can cause confusion  
to identify target 1 and 2 as they readily intermix. The values spread over  
a range similar between the two.



#### **Analysis**

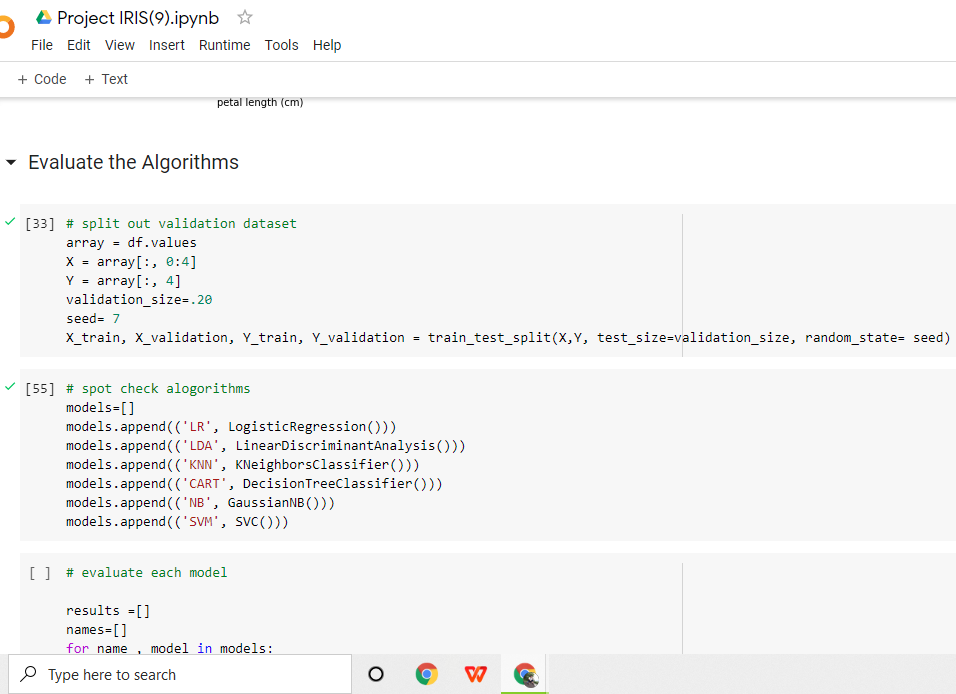
Observe that among the four plotted scatter plots the second one with feature names as petal width and petal length gives a better picture of the relationship with the Species . This plot also shows a strong relationship. We can understand that petal length and petal width can help to predict the target or Species better.

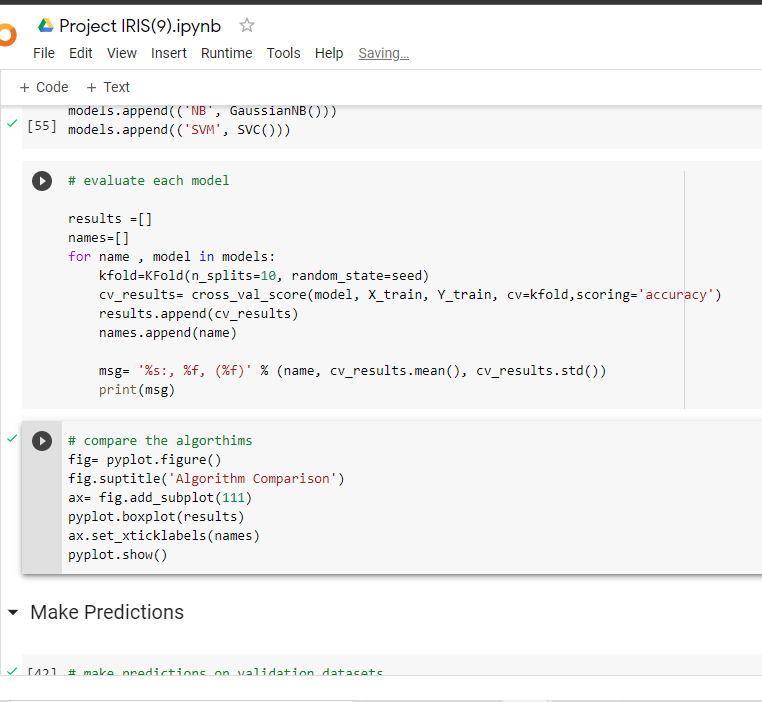


We can also find the correlation between features and Species with the help of **heat map.** Here we notice that there is a **positive correlation** between sepal length, petal length, petal width with the species but sepal width has a **negative correlation** with the species.

The highest correlation can be observed with respect to petal length and petal width. This can also be seen in the above scatter plots.

### **Evaluate the Algorithms**





### **Make Predictions**

