# Project 2018 Programming and Scripting

# ADARSHA SACHAN 23/03/2108

Introduction Fisher’s Iris data set

Reference 1 <https://en.wikipedia.org/wiki/Iris_flower_data_set>

Reference 2 <https://www.techopedia.com/definition/32880/iris-flower-data-set>

Fisher's Iris data set introduced by the British statistician and biologist Ronald Fisher in his 1936.

This data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample

The length, width of the sepals and petals, in centimeters, Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.

We have real data 3 species flower.

1 Irish Setosa

2 Irish Versicolor

3 Irish Virginica

In the given attribute data, we have petal length and petal width and Sepal Length and Sepal Width

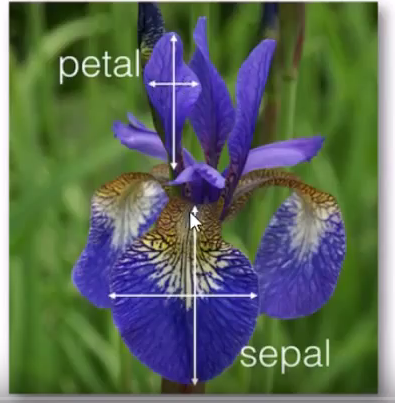
**Question** what are sepal, Petal Length and width

Reference 3 <https://www.youtube.com/watch?v=mE7MWa6N2Vc>

<https://www.ted.com/talks/linus_torvalds_the_mind_behind_linux/transcript?language=en>

Length

Width



Reference Link

**Install pandas**

Reference <https://www.youtube.com/watch?v=khwJY8EeJtU>

In Cmd Type pip install pandas

<https://pandas.pydata.org/pandas-docs/stable/generated/pandas.read_table.html>

**Group series using mapper (function by a series of columns**.

<https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.groupby.html>

http://pandas.pydata.org/pandas-docs/version/0.16.2/generated/pandas.core.groupby.DataFrameGroupBy.plot.html

<https://www.youtube.com/watch?v=wfTABU8VeoY>

**For showing the graph need below library**

Import matplotlib.pyplot as Graph

<https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html>

<https://swcarpentry.github.io/python-novice-gapminder/09-plotting/>

http://pandas.pydata.org/pandas-docs/version/0.16.0/visualization.html

**Create graph using panda dataframe**

https://www.youtube.com/watch?v=F6kmIpWWEdU

<https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.html>

**Research and Descriptive Analysis**

Based on my data analysis I am going to build mathematical way, while providing sepal, Petal length and width it’s will tell me the flower name.

I have written a code various way

**My first Analysis and code**

Project Name **myproject.py**

**While running this project it will ask for Sepal, Petal (length and width) and based on code it will give flower name record as output**

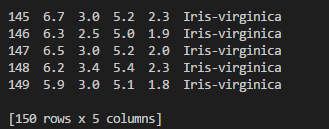
The logic I have written it will work only for this data set.

So my next opinion if I can represent this in Graph so I extend (**myproject.py**) project

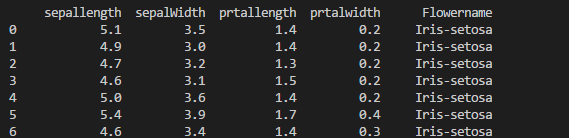
In graph pattern

**Create the graphs**

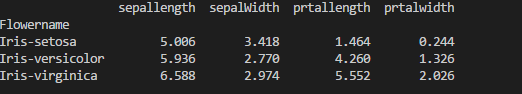
**First I am reading the csv file in Table form so while reading table I am getting below record as per screenshot in that I can see it reads all 150 row and columns**



**But Column name was not there so I added Column after adding it looks likes, so now user can see index id and column name .**

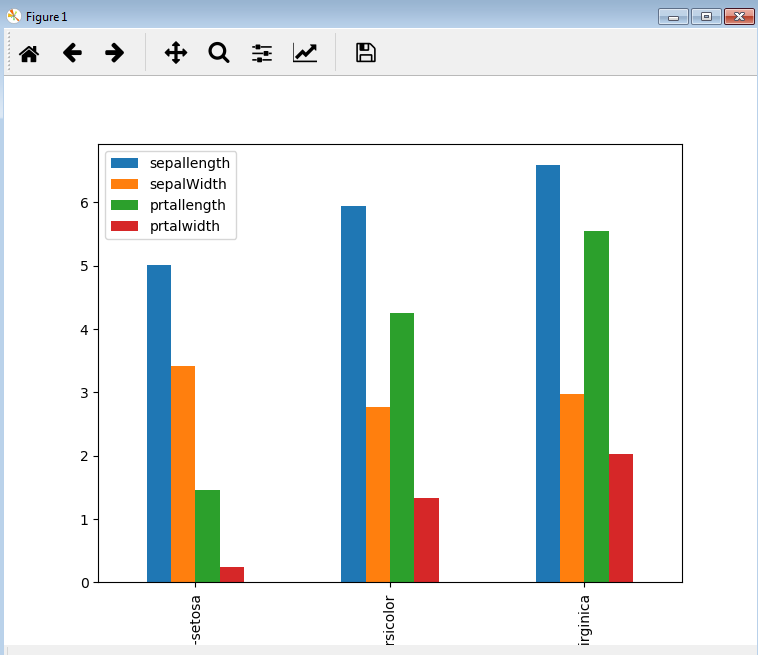


**After reading the index id and column I have created each flower mean value for Sepal length and width and Petal length and width**



Once I got mean value so I can easily represent in graph

Currently I am representing in **bar Graph for each flower looks as per below screen shoot**



When I represent mean value in graph what I see based on each flower I can see X and Y axis my x is flower record and y is sepal length and width, petal length and width min and max value.

Now I want build logic if user provides sepal length and width, petal length and width it gives flower name as output.

I did my analysis in XL sheet and in graph

While analysis the data and graph I can see sepal length is less than 8 Cm and width is less than 5 Cm, Petal Length and width less than 7 and 3 in Cm for all the Data Set.

After analysis I came to below logic.

**Calculation algorithm** data analysis as per mine

**What is perimeter???**

The **perimeter** of a shape is the distance around its edge. Finding the **perimeter** just means adding up the lengths of all its sides.

[**https://homeschool.rebeccareid.com/flower-garden-reviewing-area-perimeter/**](https://homeschool.rebeccareid.com/flower-garden-reviewing-area-perimeter/)

**Petal area = Petal Length \* Petal Width**

**Perimeter = 2(Petal Length +Petal Width + Sepal Length+ Sepal Width)**

**I am writing logic in story in agile way**

**User Story 1**

As a user I want to identify what is the flower name

**I want** sepal, Petal Length and width so I can pass this info to my function.

**User story 2**

As I need a full Irish Data set.

**I want** to pass sepal, Petal Length and width in csv read search function.

Irish data set I can copy from wiki or <http://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data> copy from this link.

User story 3 NA

As I need search logic functions.

**I want to** write a search function when user will pass input, it look in database or csv file and return the result.

If it dint match, Show the description that data not found

User story 4

As I need 4 input value petal, sepal length and width?

**I want to write** a function

While analysis the data I can see sepal length is less than 8 Cm and width is less than 5 Cm, Petal Length and width less than 7 and 3 in Cm for all the Data Set.

**So user can not provide any other input as per above petal, sepal length and width.**

Below user story will be working base on my algorithm data set

User Story 5

As I need how to find **Iris-Setosa flower**

If any time **petal area** is less than 2 it’s Iris-Setosa

If perimeter is less than 26 then 100 % it’s Iris-Setosa

**Condition 1**

If **petal area** is less than 2 and perimeter is less than 25 then it is Iris-Setosa

User Story 6

As I need how to find **ris- virginica and** ris**- versicolor flower**

**Formula for Algorithm**

**Petal area = Petal Length \* Petal Width**

**Perimeter = 2(Petal Length +Petal Width + Sepal Length+ Sepal Width)**

**Condition 2**

If perimeter is less than 33 and petal area less than 9

**Versicolor flower**

**Condition 3**

If perimeter is less than 33 and petal area less than 9 and greater than 7

Than “80% Iris-Virginia and 20% Iris-versicolor”

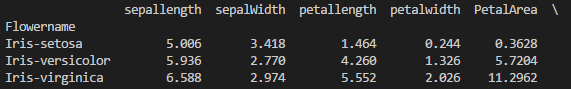
**Else**

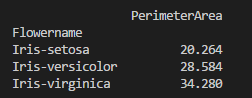
**Condition 4**

**Iris-virginica**

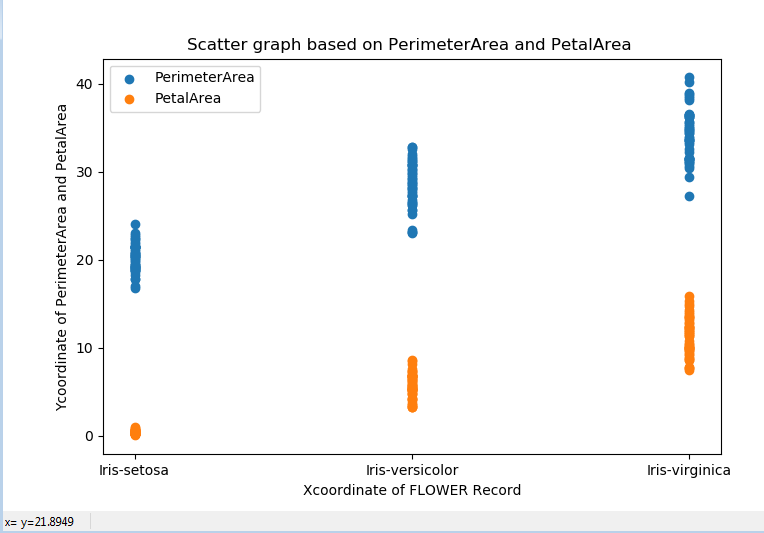
**Conclusion**

Now I am going to use Perimeter and Petal Area in the graph, so I have created two Colum.





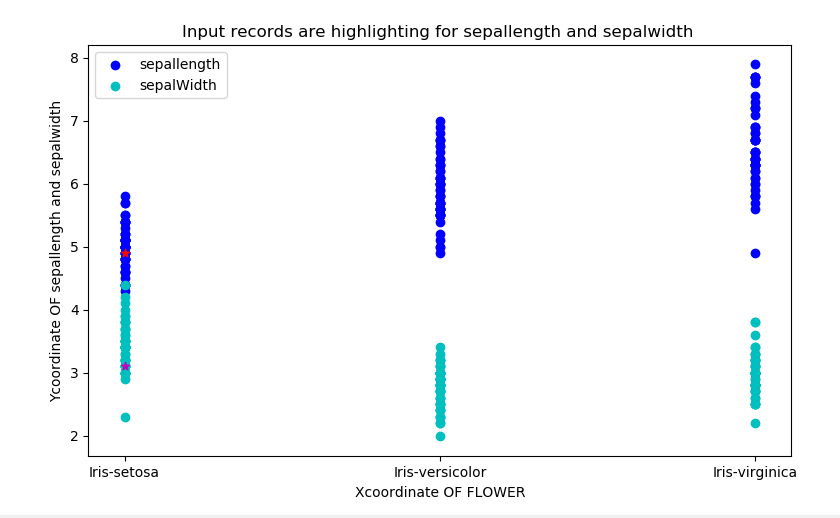
After creating a mean and Graph I came to conclusion. **IN the blow graph user can easily see perimeter and petal area based on graph easily I build my own logic those are Condition 1, 2, 3, 4.**



Now I am representing sepal length and Sepal width, petal length and width, and both sepal and petal length and width in scatter graph.

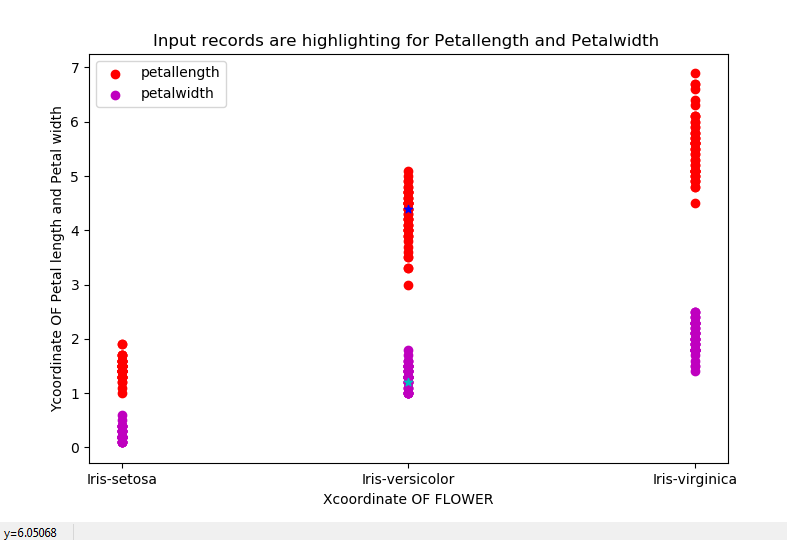
**First graph for Sepal length and Sepal Width vs flower name**

Base on my logic I found flower and input record for sepal length and width are representing in star symbol.



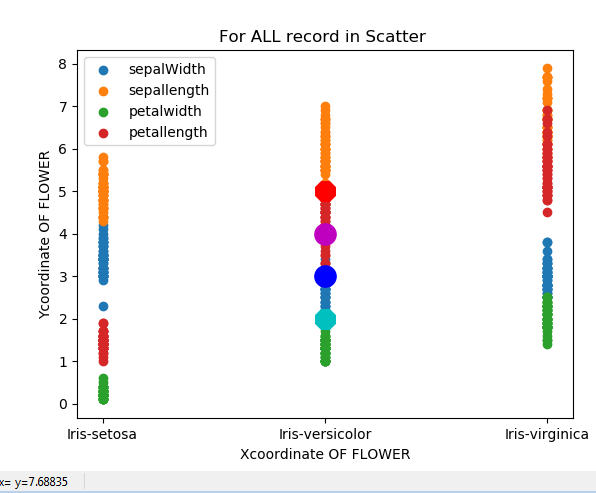
**Second graph for Petal Length and Petal Width vs flower name**

Base on my logic I found flower and it represent in Star Symbol of my input record for petal length and petal width



**Third graph for Sepal and Petal Both**

Base on my logic I found flower and it represent in O Symbol for all coordinate of sepal length and width and petal length and width



Now I have extended this project using sklearn Library it has inbuilt Irish data set

Reference

#http://docs.python-guide.org/en/latest/scenarios/ml/

#http://scikit-learn.org/stable/user\_guide.html

#https://www.python-course.eu/machine\_learning\_with\_scikit.php

#http://scikit-learn.org/stable/tutorial/basic/tutorial.html

#http://www.pythonforbeginners.com/os/pythons-os-module

**Step1**

First read the data set

**Step2**

Read data set x and y where x is array of data and y is flower name

**Step3**

Now using numpy I am picking randomly data

**Step4**

Keeping last 60 entries for testing, rest for training

**Step5**

Classifying using Decision Tree and create the model and model vale will save in desktop with [\\finalized\_model.sav](file:///\\finalized_model.sav) name

**Step6**

Check input record in build model it will predict flower species 0, 1, or 2

Where o indicate Iris-setosa, 1 indicate Iris-versicolor, 3 indicate Iris-virginica