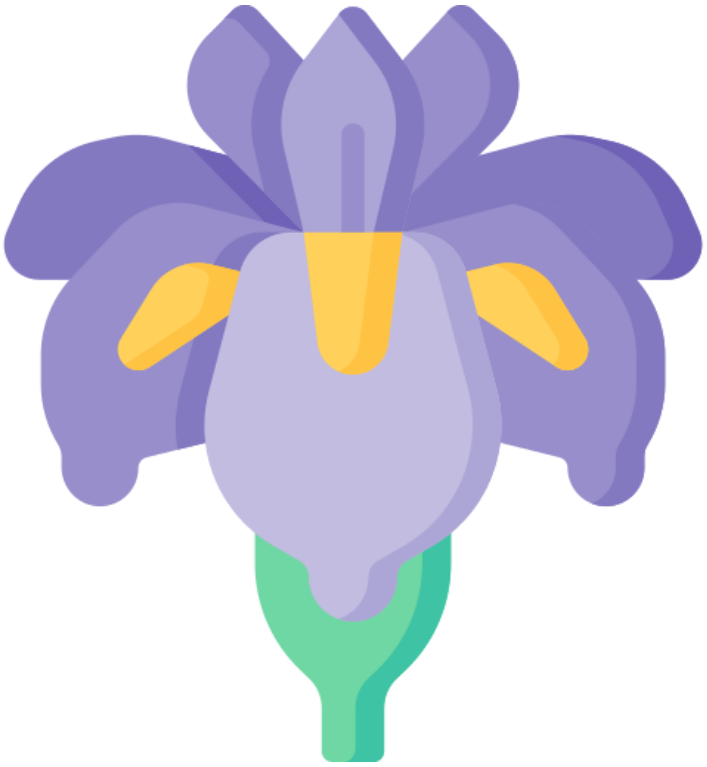
Beautifying the Flowers Data Analysis



Irises are usually purple flowers that bloom between May to June. They attract butterflies and hummingbirds and are lovely additions to gardens.

For this project, you will be given a dataset on iris called *Iris.csv*. You will use data manipulation and visualization, to apply Linear Regression on this data set.

Requirements

- Import pandas, matplotlib, seaborn, sklearn libraries
- Read csv file named as iris.csv
- Show first five records
- Show a dataframe which has sepal width greater than 4
- Show a dataframe which has petal width greater than 1
- Retrieve records which have petal width more than 2
- Figure out the relationship between sepal length and petal length and draw a scatter plot between them and show the relationship between them
- Now apply species as hue in the same scatter plot for better visibility and understanding

Applying Linear Regression

1. Model 1:
    a. Create an object named as y which is storing the dataframe of a dependent variable names as 'sepallengthcm'
    b. Create an object named as x which is storing the dataframe of an independent variable names as 'sepalwidthcm'
    c. Divide the variables into x_train,x_test,y_train,y_test variables using train_test_split method carrying parameters named as x,y and test size should be 30%
    d. Show first five records of all four variables / objects
    e. Create an object named as lr and assign memory from linearregression() method.
    f. Fit both training set into fit method
    g. Predict x_test from predict method and store the result into y_pred object
    h. Show first five records from actual and predicted objects
    i. Try to find out mean_squared_error in prediction using method after passing parameter as y_test and y_pred ,mind the result
2. Model 2:
    a. Create an object named as y and store dataframe of sepallengthcm dependent variable
    b. Store 'sepalwidthcm', 'petallengthcm', 'petalwidthcm' dataframe in x as an independent variables
    c. Do train_test_split like you did in model 1 this time test_size is again 30%
    d. Fit both train set into fit method of linearregression
    e. Predict x_test and store result into y_pred using predict method
    f. Find out mean_squared_error of actual and predicted test set
    g. Describe which model is better and why?

*** Add comments in each section, after drawing any chart you need to describe it in your comment section.

Bonus: If you want to go with some more ML algorithms then please go ahead and present some great outputs with the help of different charts. Can apply Logical Regression as well if you want.

Due Date

You have *1 week* to complete this project, after which you will present your project.