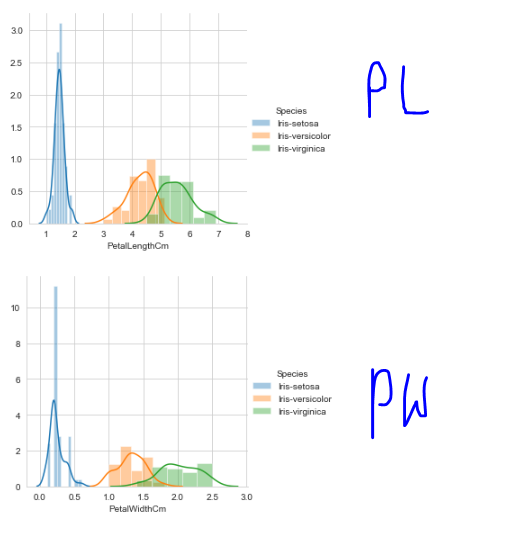
Univariate: ‘Uni’ means one and ‘Variate’ means variable, ie univariate analysis means one variable analysis.

**How can I use PDF while creating models:**

"How can I use PDF while creating models ?"- Probability Density Function (PDF) is used for univariate analysis. Univariate analysis as the name suggests is a single variable analysis. So, PDF is very helpful when we have to dig deeper into a particular feature.   
A histogram is a plot of the data you have collected. It can give you an idea about how the probability distribution of your measurement looks, but it cannot give you an accurate figure for the probability for how often your measurement falls some distance away from the mean, especially as you go further away from the mean.  
  
For example, in our classroom we have 100 students. Since each student will have a unique height the probability for any value will be 0 which is not a very useful number for us as the height can take infinitesimal number of possibilities. In such cases we will use PDF where we will give an interval for which the probability will be computed. Here, if we want to find the probability of student with height of 103 to 107 we will get a probability of say 0.48 or 48%. Now this information is useful for the recipient if he would like to arrive at certain conclusion regarding his theory to check the height of students in that range.  
  
Hence, For continuous random variables, the probability that X takes on any particular value x is 0. i.e. finding P(X = x) for a continuous random variable X is not going to work. Instead, we’ll need to find the probability that X falls in some interval (a, b), that is, we’ll need to find P(a < X < b). We’ll do that using a probability density function (PDF).

Below are the density plot for all features(Sepal length, Sepal Width, Petal length, Petal Width) and on the basis of these plots we’ll try to pick feature that can be used to classify them.

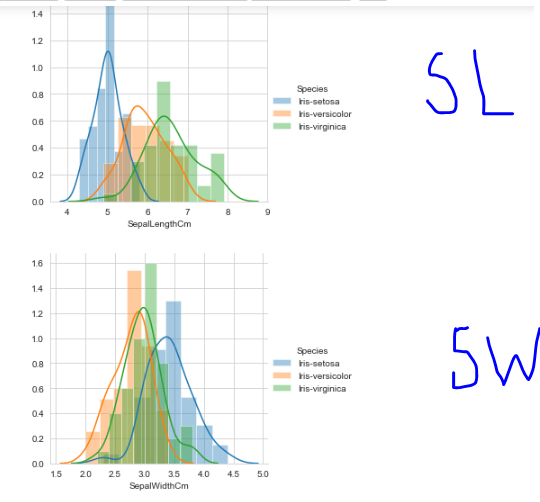


Here one thing to note is: **The more the distance between density plots, the more they are good, because they clearly separate each other.**

Here for **Petal length**, we can see setossa is clearly separated(as there is more distance between setosa and versicolor) from versicolor and virginica.

But in **Petal Width**, we can see setossa and versicolor are meeting at point, still they can separate these two because they are not overlapping,

So in this scenario if we have to make choice based on PL or PW, then we pick PL.



Here for **Sepal length** and **Sepal width** we can’t see any separation between plots, as there are too much overlapping, so if we have to make choice between all 4 we pick as **Petal length.**

Now on basis of above analysis, we can calssify them as:

If PL <= 2 then it is setossa