**Phạm Thành Hậu - 5951071024**

Chúng ta sử dụng dataset về loai hoa Iris

Đây là ví dụ về 5 hàng đầu của tâp dư liệu của chúng ta

Table

Description automatically generated

Text

Description automatically generated

Chúng ta có 150 hàng và 6 cột cho dư liêu trên , trong đó có 1 cột có thuôc tính phân loại và các cột khác đều có giá trị .

Table

Description automatically generated

Bảng trên cho ta số thiêu liệu thống kê của tập dữ liệu như: giá trị trung bình , độ lệch chuẩn , giá trị nhỏ nhất. , lớn nhất của từng thuộc tính.

Table

Description automatically generatedGraphical user interface, text

Description automatically generated

Chúng ta thấy chỉ có 3 loại hoa duy nhât và môi loại có số lượng dữ liệu đầu vào là bằng nhau

**Mối quan hệ giữa các biến**

Chúng ta sẽ thấy mối quan hệ giữa chiều dài đài hoa và chiều rộng đài hoa cũng như giữa chiều dài cánh hoa và chiều rộng cánh hoa.

Chart, scatter chart

Description automatically generated

Dựa và hình trên chúng ta có thể thấy được rằng:

* Loài Setosa có chiều dài lá đài nhỏ hơn nhưng chiều rộng lá đài lớn hơn.
* Loài Versicolor nằm giữa hai loài còn lại về chiều dài và chiều rộng của lá đài
* Loài Virginica có chiều dài lá đài lớn hơn nhưng chiều rộng lá đài nhỏ hơn.

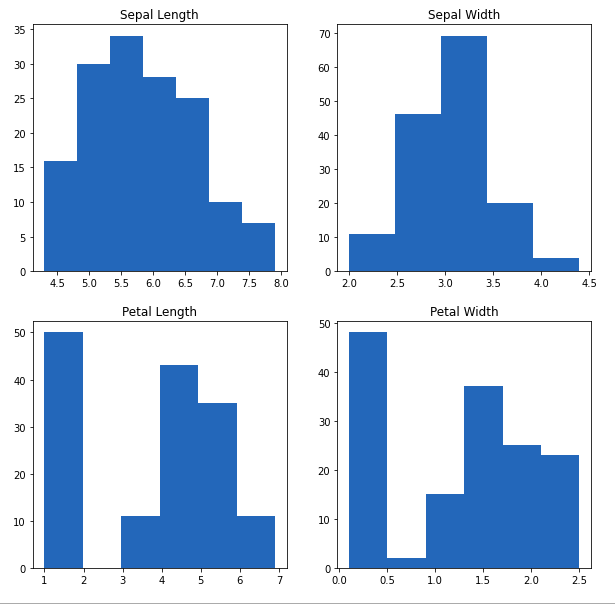
Chart, scatter chart

Description automatically generated

Dựa vào hình trên ta có thể thấy:

* Loài Setosa có chiều dài và chiều rộng cánh hoa nhỏ hơn
* Loài Versicolor nằm giữa hai loài còn lại về chiều dài và chiều rộng cánh hoa
* Loài Virginica có chiều dài và chiều rộng cánh hoa lớn nhất.

**Histogram:**



Hình trên cho chúng ta thấy:

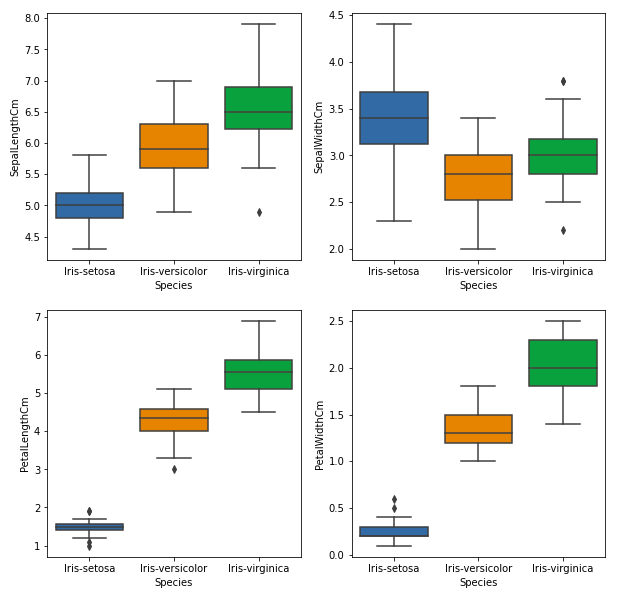
* Tần suất cao nhất của chiều dài lá đài là từ 30 đến 35, trong đó từ 5,5 đến 6
* Tần số cao nhất của chiều rộng lá đài là khoảng 70, trong đó từ 3,0 đến 3,5
* Tần suất cao nhất của chiều dài cánh hoa là khoảng 50, trong đó từ 1 đến 2
* Tần số cao nhất của chiều rộng cánh hoa là từ 40 đến 50, trong đó từ 0,0 đến 0,5

**Boxplot:**

Chart, box and whisker chart

Description automatically generated

Hình trên chúng ta sự phân bố dư liệu của các thuộc tính ( giá trị thấp nhất , cao nhất , các giá trị ngoại lệ)



Sự phân bố dữ liệu của các thuộc tính theo từng loại hoa ( min , max , và outline)