BỘ CÔNG THƯƠNG

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP. HCM



LẬP TRÌNH PYTHON

BÀI BÁO CÁO 9

NHÓM 1

Giảng viên : CAO VĂN KIÊN

Sinh viên :

Trần Công Hòa 20017691

Hà Hải Nam 20062631

Vương Minh Mẫn 20097941

TP.HCM – 2022

Đề: Thực hiện 5 ví dụ vẽ các dạng biểu đồ khác nhau sử dụng tập dữ liễu mẫu của seaborn

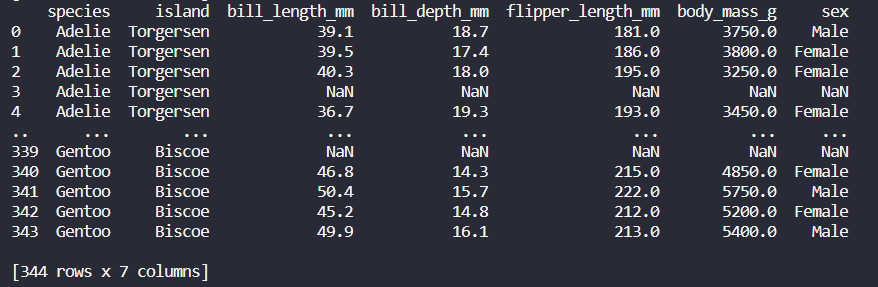
- Phân tích các thành phần trong dữ liệu

- Vẽ đồ thị để thể hiện dữ liệu

- Nhận xét kết quả trên.

Câu 1:

Phân tích dữ liệu:



Dữ liệu gồm 344 dòng và 7 cột

* Cột species thể hiện tên chủng loài của chim cánh cụt
* Cột island thể hiện đảo nơi những loài chim cánh cụt sinh sống
* Cột bill\_length\_mm là thể hiện thông số độ dài của mỏ chim cánh cụt
* Cột bill\_depth\_mm thể hiện thông số độ rộng của mỏ chim cánh cụt
* Cột flipper\_length\_mm thể hiện thông số độ dài cánh của chim cánh cụt
* Cột body\_mass\_g thể hiện thông số khối lượng của cơ thể
* Cột body\_mass\_g thể hiện thông số khối lượng cơ thể
* Cột sex thể hiện giới tính của chim cánh cụt

Vẽ đồ thị để thể hiện dữ liệu

Chart, histogram

Description automatically generated

Nhận xét:

- Ở loài Adelie ở cá thể cái thì có độ dài của cánh trong khoảng giữa 180 đến 200 thì chiếm số lượng lớn, trong khi đó thì cá thể đực có độ dài của cánh dài hơn vì biểu đồ cho nằm trong khoảng từ 190mm đến 200mm có số lượng lớn,

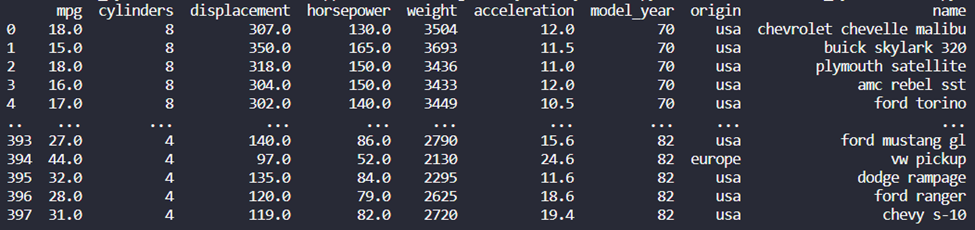
ở chim cái thì có vài cá thể có độ dài cánh dưới 180mm và có vài cá thể có độ dài lớn hơn 200mm, cá thể đực thì đa số đều có độ dài cánh lớn hơn 180mm và vài cá thể có độ dài của cánh lớn hơn 200mm.

- Ở loài Chinstrap ở cá thể này thì còn ít cá thể, cá thể đực chiếm đa số ở độ dài của cánh từ khoảng từ 190mm đến 210mm, cũng có vài cá thể có độ dài của cánh có độ dài của cánh dưới 190mm và trên 210mm, ở cá thể đực và cái thì số lượng cũng tương đương nhau, cá thể cái thì có cánh ngắn hơn cá thể đực đa số độ dài cánh của cá thể cái nằm trong khoảng từ 190mm đến 200mm, cũng có vài cá thể thì độ dài của cánh lớn hơn 200mm và cũng có vài cá thể độ dài cánh dưới190mm.

- Ở loài Gentoo thì độ dài cánh sẽ dài hơn 2 loài còn lại, trong đó thì cá thể đực có độ dài cánh nhiều hơn cá thể cái vì cá thể đực chiếm số lượng lớn ở độ dài của cánh từ 220 đến 240, trong khi đó thì cá thể cái chiếm số đông ở độ dài cánh từ 200 đến 220, cả cá thể đực và cái đều có vài cá thể có độ dài cánh khác với phần chiếm đa số.

Câu 2:

Phân tích dữ liệu:



Dữ liệu gồm 398 dòng và 9 cột

- Cột mpg số dặm đi được trên một gallon

- Cột cylinders là xy lanh mà xe sử dụng

- Cột displacement là thể hiện thông số dung tích xilanh

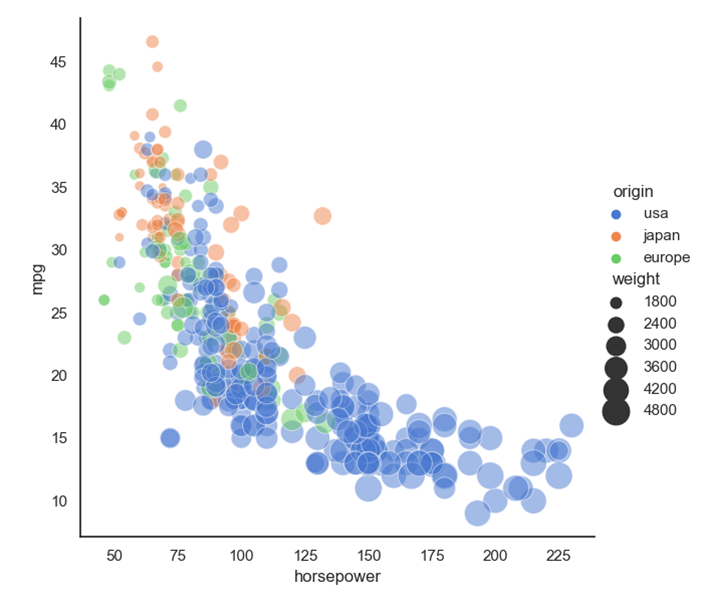
- Cột horsepower là thể hiện thông số mã lực của xe

- Cột acceleration là trọng lượng của xe

- Cột model\_year là mẫu xe của năm

- Cột origin là nguồn gốc của xe

- Cột name là tên xe

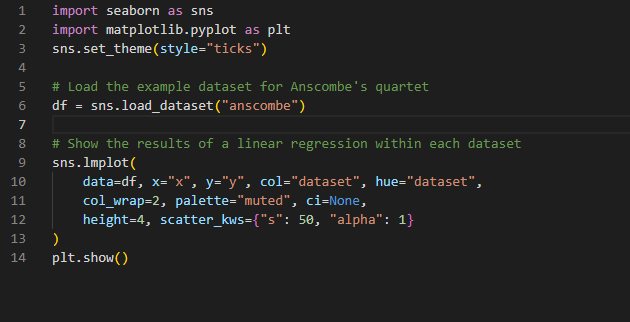


Nhận xét:

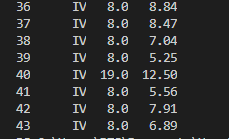
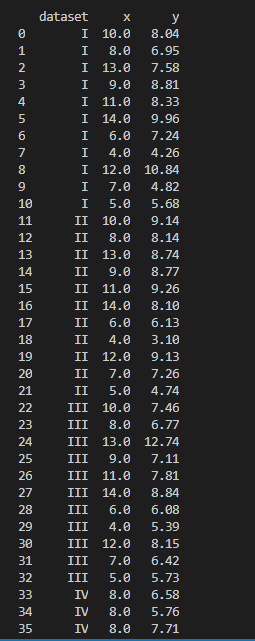
* Các dòng xe có trọng lượng lớn thường rất ít đi thể hiện trên mpg thì không cao, các xe có trọng lượng nhỏ thì số mpg càng cao.
* Trong đó các dòng xe đến từ USA có mã lực lớn hơn so với japan và europe
* Dòng xe Japan thì phân bố ở mpg khá cao.
* Dòng xe europe thì phân bố ở mpg mức trung bình, usa thì phân bố ở mpg ở mức thấp nhất.
* Các dòng xe ở japan và europe thì xe có trọng lượng bé đa số trọng lượng xe có trọng lượng dưới 3600 còn xe của usa thì trọng lượng lớn.

Câu 3:

Code:

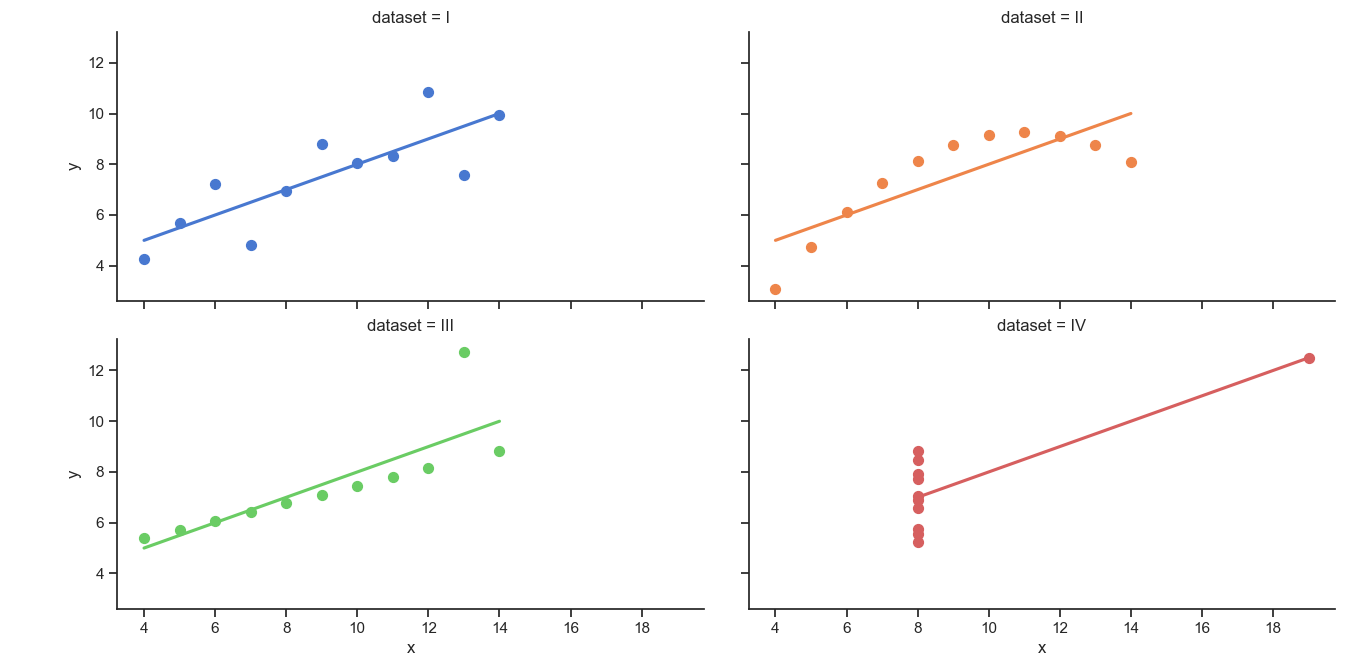


Phân tích dữ liệu:



Ở đây thể hiện dữ liệu do nhà thống kê Anscombe có 44 dòng dữ liệu mỗi dòng có 3 cột thể hiện lần lượt dataset, x, y

Vẽ hình:



Nhận xét:

+ Ở hình dataset 1:

Các dữ liệu phân bố rời rạc không nhất quán với nhau,

các điểm được thể hiện không đồng nhất nằm rải rác.

+ Ở hình dataset 2:

Biểu đồ thứ hai (góc trên bên phải) không phân bố bình thường; tuy mối quan hệ giữa hai ẩn khá rõ ràng, nó không tuyến tính, và [hệ số tương quan Pearson](https://vi.wikipedia.org/w/index.php?title=H%E1%BB%87_s%E1%BB%91_t%C6%B0%C6%A1ng_quan_Pearson&action=edit&redlink=1) không giúp ích được gì ở đây. Một mô hình hồi quy rộng hơn và [hệ số xác định](https://vi.wikipedia.org/w/index.php?title=H%E1%BB%87_s%E1%BB%91_x%C3%A1c_%C4%91%E1%BB%8Bnh&action=edit&redlink=1) tương ứng sẽ phù hợp hơn.

+ Ở hình dataset 3:

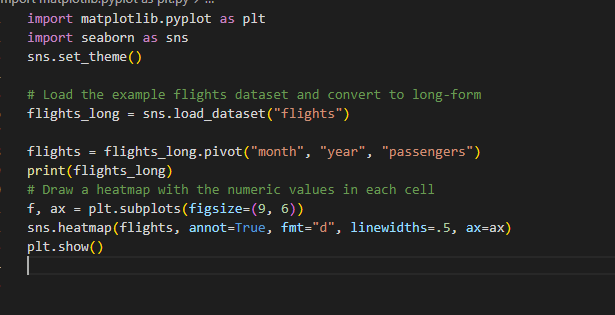
trong biểu đồ thứ ba (góc dưới bên trái), sự phân bố tuy tuyến tính nhưng nên có một [đường thẳng hồi quy](https://vi.wikipedia.org/wiki/H%E1%BB%93i_quy_tuy%E1%BA%BFn_t%C3%ADnh) khác. Mô hình hồi quy có được bị lệch do một [giá trị ngoại lai](https://vi.wikipedia.org/w/index.php?title=Gi%C3%A1_tr%E1%BB%8B_ngo%E1%BA%A1i_lai&action=edit&redlink=1) làm giảm hệ số tương quan từ 1 xuống 0.816.

+ Ở hình dataset 4:

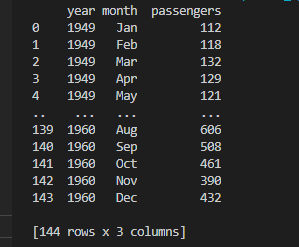
ở cột x tất cả các giá trị đều bằng 8 hiện thị trên đồ thị thì các điểm này nằm trên một cột không giống như những đồ thì khác đồ thị chỉ có một đường hồi quy đi qua hai điểm

Câu 4:

Code:

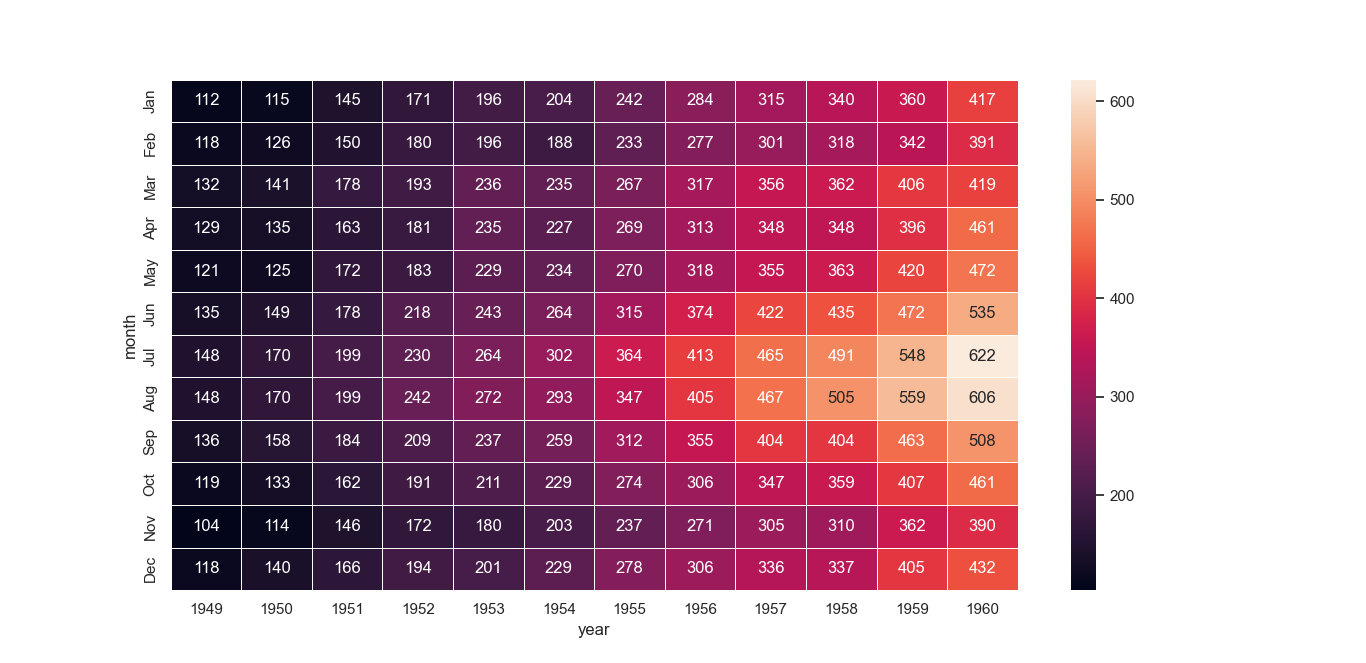


Phân tích dữ liệu đầu vào:



Trong dữ liệu trên ta có được 144 dòng dữ liệu mỗi dòng có ba cột thể hiện lần lượt year, month, passengers

Vẽ hình:



Nhận xét:

Ở năm 1949 trung bình số chuyến bay ở dưới mức 200 biểu diễn bằng màu đen tím và vào tháng 8,7 trong năm có số chuyến bay nhiều nhất trong các tháng còn lại màu sắc của hai tháng này cũng tương tự như vậy có phần nhạt hơn

Cho đến năm 1950 số lượng chuyến bay trong từng tháng đã có sự thay đổi nhưng không đáng kể có tháng chỉ tăng lên 2 - 5 chuyến bay nhưng cao điểm vẫn là tháng 7, 8 với số lượng chuyến bay vượt trội hơn hẳn các tháng còn lại, màu sắc lúc này đã có phần nhạt đi hơn so với năm 1949

Những năm còn lại từ 1951 cho đến 1960 số lượng các chuyến bay đang không ngừng tăng lên và màu sắc của chúng có phần nhạt đi rất nhiều so với năm 1949 cao điểm là vào nằm 1960 trung bình các chuyến bay trong năm này nhiều hơn hẳn so với các năm còn lại màu sắc không còn là đen tím nữa mà đã dần chuyển sang màu cam nhạt và đặc biệt vào tháng 7, 8 đã có số chuyến bay vượt trội hơn hẳn số với các tháng hơn 600 chuyến bay.

Như vậy, với số lượng các chuyến bay đang tăng lên hằng năm như thế biểu hiện cho một nền kinh tế đang trên đà phát triển, số lượng chuyến bay đang tăng cao, nhu cầu người dân đi lại tăng cao, điều kiện kinh tế kéo theo đó cũng tăng trưởng đáng kể hơn.

Câu 5:

* Phân tích các thành phần trong dữ liệu:

Table

Description automatically generated

Total\_bill: tổng hóa đơn

Tip: tiền bo

Sex: giới tính

Smoker: có hút thuốc không

Day: ngày trong tuần

Time: thời gian trong ngày

Size: số lượng

Vẽ đồ thị thể hiện dữ liệu:

Code:

Text

Description automatically generated

Đồ thị:

-Đồ thị dạng violin thể hiện total\_bill theo ngày của người hút và không hút thuốc.

Chart

Description automatically generated

* Nhận xét kết quả trên:

-Tổng hóa đơn trung bình cao nhất của người hút thuốc là vào CN (24.12), thấp nhất là vào thứ 6(16.81)

-Tổng hóa đơn trung bình cao nhất của người không hút thuốc là vào CN (20.51), thấp nhất là vào thứ 5(17.11)

-Tổng hóa đơn trung bình của người có hút thuốc nhiều hơn người không hút thuốc 1,075 lần.

-Tổng hóa đơn của người hút thuốc đồng đều hơn vào thứ 7 và Chủ Nhật.