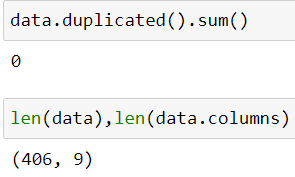
**1. How many cars and how many attributes are in the data set.**

Số lượng xe là số lượng dòng và số lượng thuộc tính là số lượng cột trong tập dữ liệu. Tuy nhiên, để chắc chắn ta sẽ kiểm tra xem có dòng nào bị trùng lặp trong tập dữ liệu hay không.

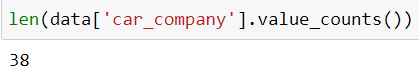


Dựa vào kết quả trên ta thấy, mỗi dòng dữ liệu là phân biệt, có 406 chiếc xe và 9 thuộc tính.

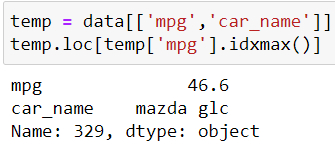
**2. How many distinct car companies are represented in the data set? What is the name of the car with the best MPG? What car company produced the most 8-cylinder cars? What are the names of 3-cylinder cars? Do some internet search that can tell you about the history and popularity of those 3-cylinder cars.**

Ta thêm 1 cột ‘car\_company’ vào tập dữ liệu. Dựa vào cột ‘car\_name’, từ đầu tiên trong tên của các chiếc xe là tên của công ty sản xuất chiếc xe đó.

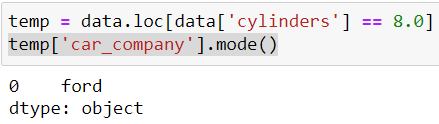
* Có 38 công ty xe phân biệt trong tập dữ liệu.



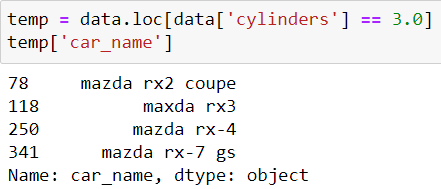
* Xe có MPG tốt nhất là **mazda glc** (46.6mpg)



* Công ty sản xuất xe 8 xi-lanh nhiều nhất là Ford.

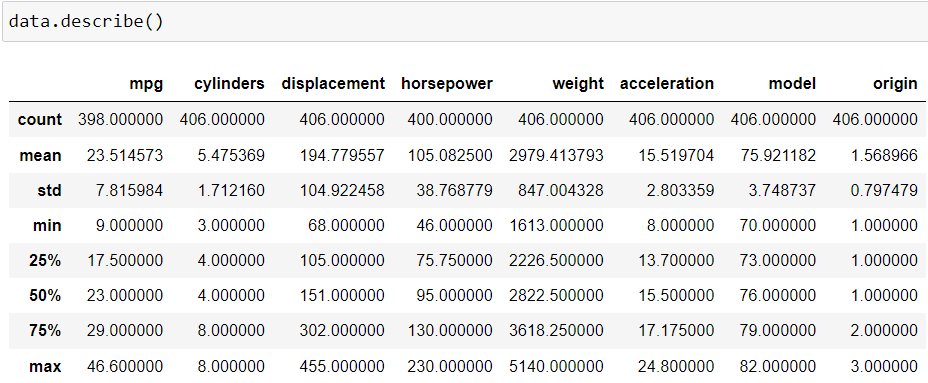


* Các xe có 3 xi-lanh:

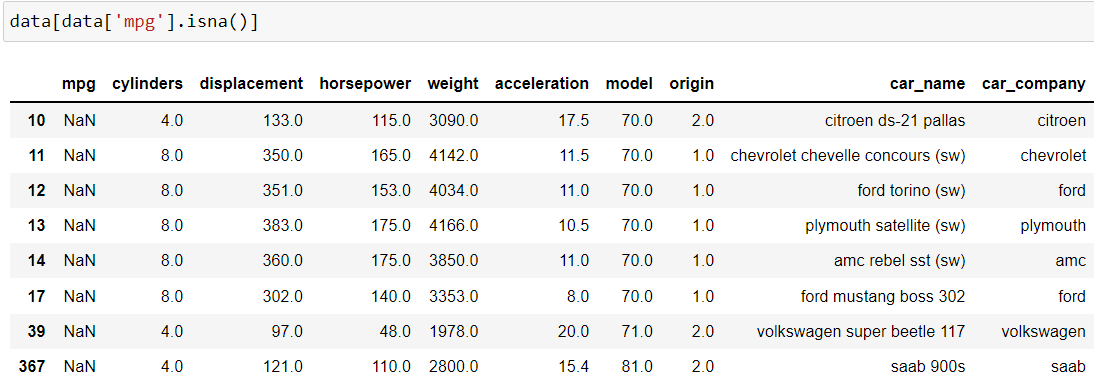


* Động cơ 3 xi-lanh nhỏ gọn, ít tốn nhiên liệu, dễ lắp ráp và dễ bảo trì.

**3. What is the range, mean, and standard deviation of each attribute? Pay attention to potential missing values.**



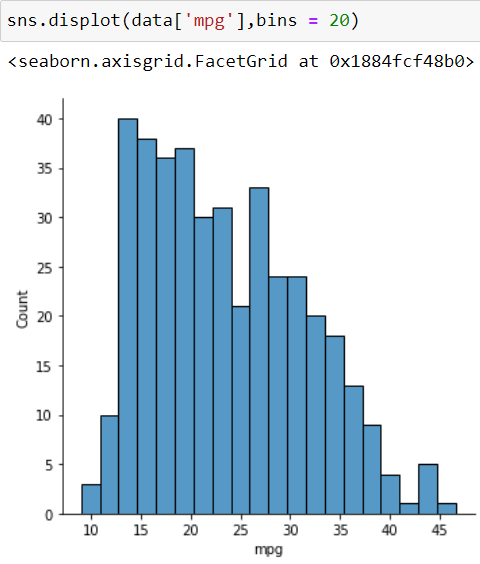
Cột ‘mpg’ có 8 ô bị bỏ trống



**4. Plot histograms for each attribute. Pay attention to the appropriate choice of number of bins. Write 2-3 sentences summarizing some interesting aspects of the data by looking at the histograms.**

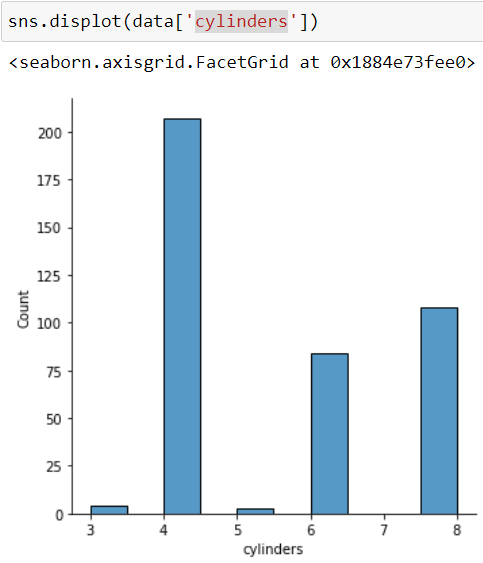
* Mpg

Đa số các xe ô tô có mpg trong khoảng thì 17mpg đến 29mpg. Tuy nhiên đây có thể là số liệu cũ bởi vì theo quy định mới của EU bắt đầu từ năm 2012, xe ô tô có thể thông hành thì phải có mức tiêu thụ nhiên liệu là 52 mpgUS cho xe chạy bằng dầu diesel và 47mpgUS cho xe chạy bằng xăng.



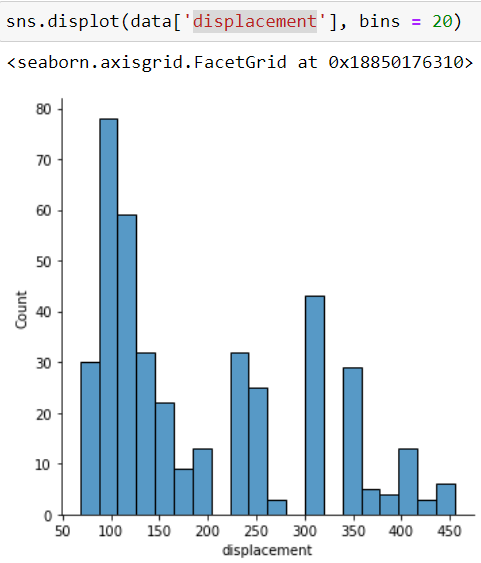
* Cylinders

Ta có thể phân loại ô tô thành 5 loại dựa vào số xi-lanh: 3 xi-lanh, 4 xi-lanh, 5 xi-lanh, 6 xi-lanh, 8 xi-lanh. Dựa vào tập dữ liệu thì số lượng xe có 4 xi-lanh là nhiều nhất.



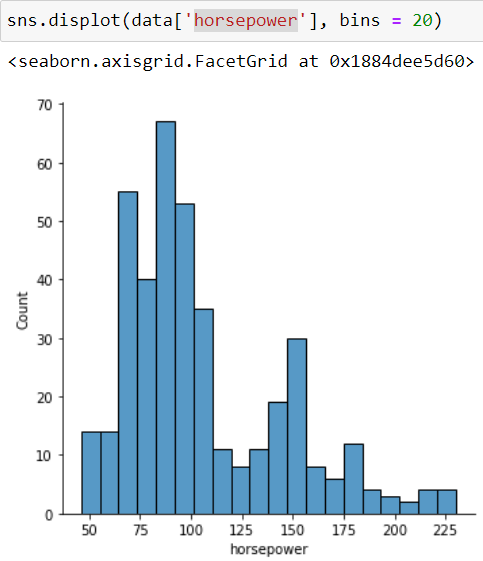
* Displacement

Displacement càng thấp thì xe ít tiêu thụ nhiên liệu hơn. Từ tập dữ liệu, displacement của các xe nằm trong khoảng 105l đến 302l.



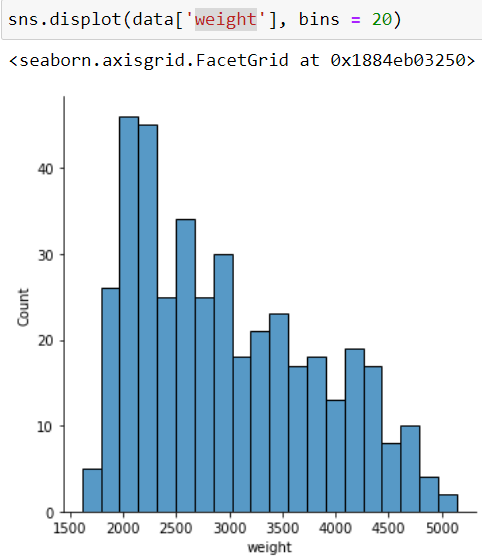
* Horsepower

Mã lực (Horsepower) là thước đo công suất của động cơ xe. Đa số các xe trong tập dữ liệu có từ 46hp đến 130hp.



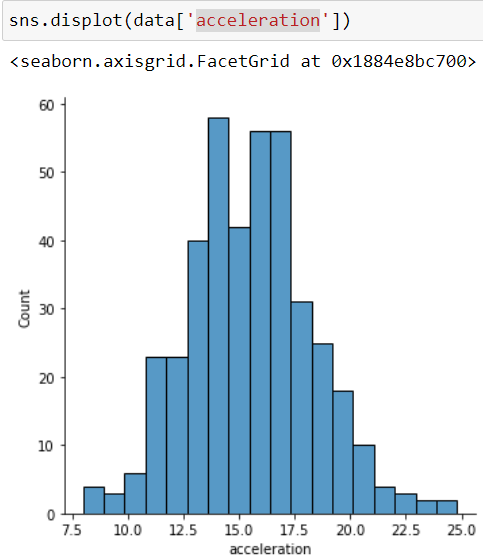
* Weight

Các xe có trọng lượng từ 2226 pound đến 3618 pound. Các công ty sản xuất xe có xu hướng sản xuất các loại xe có trọng lượng nhẹ hơn vì phù hợp cho dân dụng và ít tốn nhiên liệu.



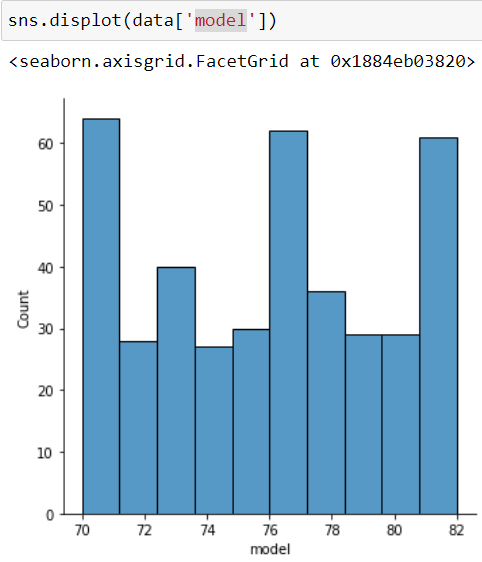
* Acceleration

Tốc độ có thể tăng tốc (Acceleration) là 1 trong các yếu tố được quan tâm hàng đầu của 1 chiếc xe. Các ô tô trong tập dữ liệu có Acceleration trong khoảng từ 8mph - 17.18mph.



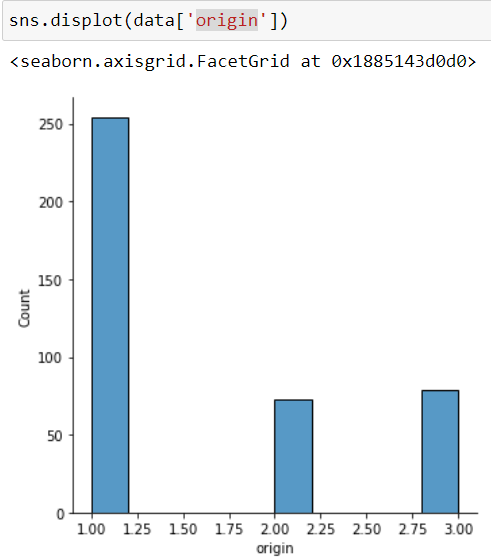
* Model

Model của các xe trong tập dữ liệu chủ yếu là 70, 76, 82.



* Origin

Các xe được chia thành 3 origin là 1,2 và 3. Xe có origin = 1 là nhiều nhất.



**5. Plot a scatterplot of weight vs. MPG attributes. What do you conclude about the relationship between the attributes? What is the correlation coefficient between the 2 attributes?**

* Xe có trọng lượng càng lớn thì sẽ ít tiết kiệm nhiên liệu hơn.
* Hệ số tương quan: -0.83.

