**DataScience Homework I**

**Name:陳奕安 ID:0510846**

**Data set :**

This dataset consists of information of car From 1985 Ward's Automotive Yearbook. And it consists detailed information about cars, including physical characteristic of cars and some evaluation from relevant industries. The reason why I chose this dataset is due to its thoroughness and completeness, it provide many information that is very technical and very insightful. Hopefully, we can gain some insight from this dataset.

**Features:**

This data set consists of three types of entities:

(a) the specification of an auto in terms of various characteristics(**“length”,“weight”,“fuel-type”,“engine-location” etc**)

(b) its assigned insurance risk rating **("symboling”)**

(c) its normalized losses in use as compared to other cars **(”normalized-losses”)**.

The second rating corresponds to the degree to which the auto is more risky than its price indicates. Cars are initially assigned a risk factor symbol associated with its price. Then, if it is more risky (or less), this symbol is adjusted by moving it up (or down) the scale. Actuarians call this process **"symboling".** A value of +3 indicates that the auto is risky, -3 that it is probably pretty safe. The third factor is the relative average loss payment per insured vehicle year. This value is normalized for all autos within a particular size classification (two-door small, station wagons, sports/speciality, etc...), and represents the average loss per car per year.

**Goal :**

Because this dataset contains a number of manufacturer-oriented features and customer-oriented features. So it provide a lot technical information about the car(ex,aspiration,engine-location,engine-type). My primary goal is to find

(1) The correlation between horse-power, peak-rpm, city-mpg, highway-mpg and its physical characteristics

(2) How the cars performance affect the final price

(3) With the same performance the effect brought by the car brands.