Problem Statement- Liner Regression

A Chinese automobile company Geely Auto aspires to enter the US market by setting up their manufacturing unit there and producing cars locally to give competition to their US and European counterparts.

They have contracted an automobile consulting company since they want to understand the factors on which pricing of a car depends. They want to do this for the cars produced by the American and European countries, since their design is very different from the Chinese cars.

Essentially the company wants to know:

Which variables are highly significant in estimating the price of a car

How well those important variables describe the price of a car

Based on various market surveys, the consulting firm has gathered a large dataset of different types of cars across the world and their attributes.

Goal of this assignment

You are required to model the price of cars with the available independent variables. It will be used by the management to understand how exactly the prices vary with the independent variables. They can accordingly manipulate the design of the cars, the business strategy etc. to meet certain price levels. Further, the model will be a good way for the management to understand the pricing dynamics of a new market.

Data Preparation

There is a variable named CarName whose entries contains the first word as name of "car company" followed by name of "car model". For example, chevrolet impala has 'chevrolet' as the car company name and 'impala' as the car model name. You need to consider only company name as the independent variable for the model building. 