

CASE STUDY

Car Data Analysis

This case study is a programming case study wherein you have to build a multiple linear regression model for the prediction of car prices. You will need to submit a Jupyter notebook for the same.

Problem Statement

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** to understand the factors on which the pricing of cars depends. Specifically, they want to understand the factors affecting the pricing of cars in the American market, since those may be very different from the Chinese market. The company wants to know:

- Which variables are significant in predicting the price of a car
- How well those 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 American market.

Business Goal

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 management to understand the pricing dynamics of a new market.

Data Preparation

- There is a variable named **CarName** which is comprised of two parts - the first word is the name of 'car company' and the second is the '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 model building.