Automobile Price Prediction

The Problem

Our target business is a French car manufacturer **Peugeot Automobiles**. Peugeot aims to enter the Indian Market by setting up a manufacturing unit in India and are planning to launch a new car model in the near future in competition with their European and American 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 Indian market, since those may be very different from the French market.

The Company wants to figure out:

- Which variables significantly affect the price of a vehicle in the Indian Market
- How well do these variables explain the price

Based on a large scale market survey, the company has collected a database of over 1200 cars across the Indian market and now it is upto us to find the solution to their problem.

The Opportunity (Benefit to the customer)

Pricing is one of the most important aspects for the success of a product, especially in a price competitive market like India. If the company misses the mark on price and over-priced the model, it is very likely that its sales will suffer. On the other hand, if the prices are quoted to low, the company can miss out on potential profit. Hence it is a must that the right note is hit with the pricing. This is where we come in and help the company achieve its targets with the new car model.

The Solution

For the pricing model, we are going to use the database available for more than 1200 car models in the Indian market. We will attempt to determine the price by using a linear regression model. We will try different kinds of linear regression models to ascertain the best fit and test it on real-world data.

The Dataset

The dataset chosen is the Indian Cars Dataset from Kaggle (https://www.kaggle.com/medhekarabhinav5/indian-cars-dataset)

The dataset has reliable data though some amount of wrangling will be required due to missing data.

It contains a variety of features such as Model price, Engine related features, Body and comfort related features etc (Over 140). This makes it the optimal dataset for making a pricing model and also use to build a recommendation engine.

The Typical Output

We would be predicting a price bracket in which to place the company's new car. We will also be comparing the accuracy scores of various machine learning algorithms in predicting the price.