**Project on Used Cars**

**Case Study for analysis**: A used-car-seller giant in India is reviewing its business closely to improve its prediction of “On-spot price estimate”. Which factors decide the price of a used plays an important role in this process. The reseller has data on an existing lot of about 6019 used-cars (File name: Sample data-used cars India) with prices determined already and the other set of 1234 used-cars (test-data) on which it is trying to use analytical methods to fasten its speed of determining the prices for future.

**Your tasks is to:**

(Ensure that you edit and clean the data as per the need of the analysis without losing any important information (datapoints/records) at the same time interpret the operations at every step is very crucial)

1. Perform the basic Exploratory Data Analysis on the sample data.

1. Identify the categorical and continuous variables
2. Make relevant charts, Pivot tables on your own and determine the expected behaviour of variables and their interaction with each other.
3. Make sure to find out the number of cars sold by each brand of the car (Like Maruti, Hyundai, Honda etc…)
4. Give your interpretation for observations made in points (b) and (c)

2. Edit the data to obtain dummy variables for relevant variables, do a correlation analysis and proceed with regression analysis on the “Sample data-used cars India”. Use the output of Regression analysis done to determine the prices for test-data (the set of cars for which the price is yet to be determined):

**Note:** (You don’t need to remove the outliers as having outliers in such analysis is natural as various cars can be in varying physical condition, performance and hence the prices)