# Project Proposal of Group No. 06

# Analysis of Automotive Features and Pricing

## Objective

We have the following three objectives regarding the *AutoMobile* data set that we have chosen.

* To analyze the relationship between automotive features and pricing.
* Do the exploratory data analysis.
* To identify significant variables that influence the price of automobiles.
* To develop a predictive model to estimate the price of a car based on its features.

## Background

The automotive industry is highly competitive, and pricing plays a crucial role in the success of car manufacturers. In order to draw the business decision, it is essential to be aware about the factors that really affect the car prices. This project seeks to explore the significant variables which are useful for predicting car price. By incorporating them evaluate a proactive car price prediction model by using regression analysis.

## Dataset Description

The current study makes use of a dataset that contains in-depth data on several vehicle features and costs. The dataset includes three different types of entities:

1. The auto's specification in terms of its many features
2. The insurance risk rating given to it
3. Compared to other autos, its normalized use losses

The second rating reflects the extent to which the car is riskier than its asking price suggests. Cars are first given a risk factor symbol according to their cost. Then, depending on how risky it is, this symbol is changed by pushing it up or down the scale. Actuaries refer to this procedure as "symbolling". A score of +3 means the car is unsafe, while a score of -3 means it's probably safe.

The relative average loss payment per insured vehicle year makes up the third component. This number, which indicates the average loss per vehicle per year, is standardized for all vehicles falling within a specific size category (two-door small, station wagons, sports/specialty, etc.).

The dataset includes the following variables:

Variables:

1. symbolling: (***Integer***): The risk rating of the car (-3, to +3).
2. Normalized-losses: (***Numeric***): The normalized losses in insurance claims for the car.
3. make: (***Categorical***): The manufacturer or brand of the car.
4. fuel-type: (***Categorical***): The type of fuel used by the car (gas or diesel).
5. aspiration: (***Categorical***): Whether the car has a standard or turbocharged engine.
6. num-of-doors: (***Categorical***): The number of doors of the car.
7. body-style: (***Categorical***): The body style of the car (e.g., sedan, hatchback).
8. drive-wheels: (***Categorical***): The type of drive wheels (e.g., front-wheel drive, rear-wheel drive).
9. engine-location: (***Categorical***): The location of the car's engine (front or rear).
10. wheel-base: (***Numeric***): The wheelbase of the car (distance between the centers of the front and rear wheels).
11. length: (***Numeric***): The length of the car.
12. width: (***Numeric***): The width of the car.
13. height: (***Numeric***): The height of the car.
14. curb-weight: (***Numeric***): The weight of the car without occupants or baggage.
15. engine-type: (***Categorical***): The type of engine (e.g., dohc, ohc).
16. num-of-cylinders: (***Categorical***): The number of cylinders in the car's engine.
17. engine-size: (***Numerical***): The size of the car's engine in cubic centimeters.
18. fuel-system: (***Categorical***): The type of fuel system used by the car ( mpfi, 2bbl).
19. bore: (***Numerical***): The diameter of each cylinder in the car's engine.
20. stroke: (***Numerical***): The length of the piston stroke in the car's engine.
21. compression-ratio: (***Numerical***): The compression ratio of the car's engine.
22. horsepower: (***Numerical***): The horsepower of the car.
23. peak-rpm: (***Numerical***): The peak rotations per minute of the car's engine.
24. city-mpg: (***Numerical***): The miles per gallon of the car in city driving condition.
25. highway-mpg: (***Numerical***): The miles per gallon of the car on the highway.
26. price: (***Numerical***): This variable represents the price of the car.

## Methodology

* Data Collection: The dataset will be obtained from [source] and imported into R Studio for analysis.
* Data Preprocessing: The dataset will be cleaned by handling missing values, removing duplicates, and correcting any inconsistencies.
* Exploratory Data Analysis: Descriptive statistics, data visualization, and correlation analysis will be performed to understand the distribution and relationships between variables.
* Feature Selection: Statistical techniques, such as correlation analysis and feature importance, will be used to identify the most influential variables.
* Model Development: Various regression techniques, such as linear regression will be employed to build a predictive model for estimating car prices.
* Model Evaluation: The developed model will be evaluated using appropriate performance metrics, such as mean squared error, R-squared, and cross-validation.
* Interpretation and Reporting: The findings and insights from the analysis will be summarized, and a report will be prepared highlighting the key factors affecting car prices.

**Project Timeline:** The project will be conducted over a period of 14th of June - 14th of July. The timeline will be divided into specific phases, including data collection, data preprocessing, exploratory data analysis, feature selection, model development, model evaluation, and reporting. Regular team meetings will be scheduled to track progress, address challenges, and ensure the project stays on schedule.

## Project Team

S/17/342 - A.A.C. Disanayaka: Data collection, preprocessing, and exploratory data analysis.

S/17/458 - R.M.C.M.B Ranathunga: Feature selection and model development.

S/17/070 - W.P.D.M. Jayasinghe: Model evaluation, interpretation, and reporting.

## Desired Outcomes

* Accurate Car Price Prediction
* Identification of Key Factors
* Improved Decision-making
* Cost Reduction
* Competitive Advantage
* Enhanced Customer Experience
* Industry Insights
* Framework for Future Analysis

## References

Kaggle. (n.d.). Automobile dataset. Retrieved June 15, 2023, from <https://www.kaggle.com/datasets/toramky/automobile-dataset?select=Automobile_data.csv>

* <https://www.r-bloggers.com/2015/12/how-to-learn-r-2/>
* <https://r4ds.had.co.nz/introduction.html>

**Group members: (s/17/070, s/17/458, s/17/342)**