**Forecasting the Sale Price of Pre-owned Vehicles**

Team:

Banavath, Sreenu

Bandela, Saheeth

Bobba, Hanumanth

Bokka, Harivardhan Reddy

Chaudhari, Aniruddha

Chenreddy, Adarsh Reddy

Link: [car data (kaggle.com)](https://www.kaggle.com/datasets/athirags/car-data)

Dataset:

The dataset we will utilize is the "automobile data" dataset from Kaggle.com. This dataset comprises details about pre-owned vehicles, such as the manufacturing year, brand, model, sale price, mileage, fuel category, seller type, transmission, and ownership. The dataset consists of 9 columns and 300 rows. Among the columns, three are numerical: manufacturing year, mileage, and sale price. One column is categorical: fuel category. The remaining columns are also categorical but can be transformed into numerical values using one-hot encoding.

Approach: To predict the sale price of a pre-owned vehicle, we will employ various algorithms. We will assess the models' performance using cross-validation and holdout testing. Additionally, we will utilize feature engineering techniques to enhance the models' performance.

Expected Results: Our objective is to develop a model capable of accurately predicting the sale price of a pre-owned vehicle. Furthermore, we anticipate discovering the key factors that significantly impact the sale price of pre-owned vehicles. This knowledge can be utilized by both vehicle buyers and sellers to make more informed decisions.

We firmly believe that this research project has the potential to make a noteworthy contribution to the field of Data Science and the Automotive industry.

We aim to address the following inquiries:

1. What are the key determinants that influence the sale price of a pre-owned vehicle?

2. Can we construct a model that accurately forecasts the sale price of a pre-owned vehicle?

3. How can our model assist both vehicle buyers and sellers in making informed decisions?

Dependent and Independent Variables:

We will consider the sale price as the dependent variable (DV). The other attributes, including manufacturing year, brand, model, mileage, fuel category, seller type, and transmission, will be treated as independent variables (IVs).