## **Project 19 Used Hyundai car price Regression**

[Dataset Link](https://drive.google.com/file/d/14AUeEyL8gxmc7Vf8s5BpNOEeW3bcaiR2/view?usp=sharing)

## **About data:**

This dataset contains used cars sale prices in the UK

Variables

* Model
* Year
* Price
* Transmission
* Mileage
* fuelType
* tax
* mpg
* engineSize

**Procedure**

1. Import Data
2. Check dataset size
3. Find and treat missing values (If any)
4. Check column types and describe which columns are numerical or categorical
5. Perform Univariate analysis
   1. Calculate mean, median, std dev, and quartiles of numerical data
   2. Plot histogram for a few categorical variables
   3. Check the distribution of numerical variables and comment on it
6. Perform Bivariate analysis
7. Plot pair plots
8. Perform a Chi-square analysis to check whether there is a relationship between

* transmission and fuelType
* fuelType and model

1. Calculate Pearson correlation, and plot their heatmap
2. Drop any unnecessary columns
3. One hot encode categorical variables (if any)
4. Split into train and test set
5. Scale the variables
6. Train multiple models like Linear regression, Decision Tree, Random Forest, SVR, etc.
7. Check their performance, and comment on which is the best model

**Compulsory**

1. Use grid search CV to tune the hyperparameter of the best model
2. Train a polynomial regression model with degrees 2, and 3 and compare its performance with other models