Used Car Price Prediction

avl

Machine Hack Dataset.

Name: Brand name + Car name.

Location: City.

Year: Year or Edition of the model.

Kilometer Driven: km driven

Fuel type: CNG, Petrol, electric, diesel etc.

Transmission: Manual or Automatic

Owner Type: 1st, 2nd or other.

Mileage: x Km/L or x km/kg

Engine: Engine displacement volume in

Power: Max Power of the engine in bhp.

Seats: No. of seats in the car.

New price: Price of a brand new car from same model.

PART 1

Data Cleaning & EDA.

1. Import the dataset
   - → train
   - → test.
2. Get basic info abt the data.
   - → Shape
   - → dtypes.
   - → Missing values for each feature.
   - → Unique values & value_count of categorical features.
3. Cleaning.
   - Name → to many unique values, thus to simplify the data, extract only the company name & model name.
   - Mileage → Remove units from the data as they are useless for the machine/model.
   - Engine → Same as above.

- Power → Same as above.
- New Price → Too many missing value, Not possible to impute it. So drop the entire col.

(4) Taking care of categorical features

→ Label encode the values. [Not the name feature, will handle it later.]

(5) Imputing missing values.

two options:

a) Simple imputer of sklearn.

b) Write custom fn to impute the values.

(6) Feature Engineering.

- Get car age.
- try divide cars into diff categories.
  - hatchback, sedan, suv etc.
- fuel: clean or not.