

## Data Cleaning & Preprocessing Tasks

### 1 Handling Missing Values

Identify missing values in `normalized_losses`, `bore`, `stroke`, `horsepower`, `peak_rpm`, `price`.

Fill numerical missing values using:

**Mean/Median** for `horsepower`, `price`.

**Mode** for categorical columns (`number_of_doors`, `fuel_type`).

Drop records if too many values are missing.

### 2 String & Categorical Data Preprocessing

Convert **categorical data** into lowercase and remove special characters (`make`, `fuel_type`, `aspiration`, etc.).

Replace **"?" or NaN values** with appropriate replacements.

Standardize `number_of_cylinders`:

Convert words ("four", "six", "eight") to integers (4, 6, 8).

Standardize `fuel_system` categories (e.g., **mpfi**, **spdi** → **Multi-Point Fuel Injection**, **Single-Point Fuel Injection**).

### 3 Numeric Feature Processing

Convert **"horsepower" and "peak\_rpm" to integers** (handle missing values first).

Convert `price` to numeric and fill missing values.

Standardize units (e.g., **convert horsepower to kW** if needed).

### 4 Feature Engineering on Numeric Columns

**Create New Features:**

**Power-to-weight ratio** = `horsepower / curb_weight`.

**Engine efficiency** = `horsepower / engine_size`.

**Fuel efficiency** = `city_mpg / highway_mpg`.

Convert `compression_ratio` into **categories**:

Low (<9), Medium (9-11), High (>11).

### **5** *Handling Date or Range Values*

Extract **ranges from normalized\_losses** and create bins (e.g., Low, Medium, High loss categories).

### **6** *Encoding Categorical Features*

**Label Encoding** for fuel\_type, aspiration, drive\_wheels, etc.

**One-Hot Encoding** for body\_style, engine\_type, and fuel\_system.

### **7** *Data Scaling & Normalization*

Scale numerical features (length, width, height, curb\_weight, engine\_size) using **Min-Max Scaling or Standardization**.