# Exploratory Data Analysis (EDA) on Car Sales Dataset

# Data Overview

- Data Import and Initial Inspection:

- Imported the dataset using `pandas`.

- Initial inspection using `describe()`, `head()`, `info()`, and checking for null values.

- Dataset shape: `(number of rows, number of columns)`

## Data Cleaning

- Handling Missing Values:

- Dropped `transmission`, `interior`, and `color` columns as they were deemed unnecessary for analysis.

- Filled missing numerical values in columns `mmr`, `sellingprice`, `condition`, and `odometer` with their respective means.

- Filled missing categorical values in columns `vin`, `make`, `model`, `trim`, and `body` with 'Unknown'.

- Filled missing dates in `saledate` with the most frequent date.

- Post-Cleaning Check:

- Verified that there are no remaining missing values after cleaning.

# Data Visualization and Insights

1. Distribution of Car Prices:

- A histogram of `sellingprice` shows that most car prices range from 0 to 50,000.

- Insight: The majority of cars in the dataset are within this price range.

2. Car Prices vs Odometer Readings:

- A scatter plot showing the relationship between `odometer` (miles driven) and `sellingprice`.

- Insight: There is an inverse relationship where cars with lower mileage generally have higher selling prices.

3. Condition vs Selling Price:

- A bar plot illustrating the impact of `condition` on `sellingprice`.

- Insight: Cars in better condition fetch higher prices. This visualization helps to understand the price difference between vehicles in different conditions.

4. Top 10 Selling Car Models:

- A bar plot showing the top 10 car models based on the number of sales.

- Insight: Identifies the most popular car models in terms of sales.