**Car Sales Data Analysis Report**

**Description:**

The dataset consists of **558,837** records, with **22** features related to cars and their sales, including key attributes:

* **year**: Year of manufacture
* **make**: Car manufacturer (Kia, BMW, Volvo)
* **model**: Car model
* **trim**: Specific car version or configuration
* **body**: Car body type (e.g., SUV, sedan)
* **transmission**: Type of transmission (e.g., automatic)
* **vin**: Vehicle Identification Number
* **state**: The U.S. state where the sale took place
* **condition**: Numerical rating of the car’s condition
* **odometer**: Mileage of the car
* **selling price**: The price at which the car was sold
* **mmr**: Manheim Market Report value for the car
* **sale date**: Date of the sale

The dataset allows us to explore various aspects of car sales, such as factors influencing selling prices, market trends, and brand performance.

**Objectives:**

1. What are the top Car Makes regardless Total selling price?
2. Which Car Model contributed the most regarding Total selling price?
3. What are the Top Vehicle seller companies?
4. What are the Percentage of car makes and models which tend to sell above or below market value?
5. Which type of car transmission has been sold?
6. Is there any noticeable seasonal trend in car sales over the months and years?

**Data Cleaning and Analysis Phases:**

**1. Initial Data Exploration**

**-** Data has no Duplicates.

- Data has many null values.

**2. Missing Data Handling**

- Missing data in ‘**Condition**’ column was imputated using Median.

- Missing values ‘**make**’, ‘**model**’ and ‘**trim**’ columns were removed as it indicates that there are no cars, maybe it was a human filling error.

- ‘**Selling Price’** column had 12 null values as well as ‘**sale date**’ column that were removed as they indicate that those cars have not been sold.

- Missing Data in ‘**color**’ and ‘**interior**’ columns was filled with ‘Black’ color (The Dominated color).

- 90 Missing transmission values were filled based on the most frequent **transmission** type for the same **car make.**

**-** 90 Missing odometer values were filled with their mean valuebecause the ‘**odometer**’ data is right skewed.

- Missing data in ‘**body**’ and ‘**trim**’ were filled with ‘Unknown’.

- Missing values in ‘**model**’ column were filled with the most common '**model**' for each '**make**' [BMW, Audi].

**3. Data Formatting:**

**- ‘Sale Year’** was converted from float to int then to datetime format**.**

- ‘**State**’ column was modified to ‘**State Modified**’ to convert from abbreviations to the full name of each state.

**4. Feature Engineering**

Several new features were created to improve analysis:

* ‘**Sale Date**’ was broken down into ‘**Sale Year’, ‘Sale Month’ and ‘Sale Day’** columns for easier seasonal analysis.
* ‘**Condition Rating**’ was generated as the result of converting ‘**Condition**’ column into 5 bins (4 Categories) which are [‘Week’, ‘Good’, ‘Very Good’, ‘Excellent’].
* **‘Market’** column: Categorizing cars into Above Market or Below Market based on their sale price relative to the market value (**Manheim Market Report)**.

**Insights:**

* 1. Top Car Makes are: ‘**Ford**’ and ‘**Chevrolet**’.
  2. Car Models ‘**F-150** ‘ and ‘**Altima**’ were contributed the most regarding Total selling price across 2014 and 2015.
  3. Top Vehicle sellers are :
     + ‘**ford motor credit company llc’**.
     + ‘**Nissan-infiniti It’**.
     + ‘**The hertz corporation’**
  4. ‘**California**’ and ‘**Florida’** were the highest states the cars had been sold in them.
     + ‘**California**’ had high number of cars sold (Above Market) in **2015** while ‘**Florida**’ had the highest number of cars sold (Below Market) in **2014.**
  5. Factors Influencing Selling Price**:**
     + The **cars condition** is the primary driver of the selling price.
     + Cars with better condition ratings (closer to 5) and lower mileage tend to command higher selling prices.
  6. Vehicles with higher condition ratings are more likely to be sold **above market value**.
  7. **Seasonal Trends in Car Sales:**
     + Totalselling prices tend to slow down in **April and July.**
     + Stock high-demand cars in **February and January.**