CAR DATA ANALYSIS

Introduction

The dataset analyzed focuses on various car attributes like name, year, price, kilometers driven, fuel type, transmission, and performance metrics. After cleaning and organizing the data, the goal of the project is to uncover insights that can improve customer satisfaction and boost revenue. Key tasks included removing duplicates, handling missing data, and categorizing factors like mileage and kilometers driven. SQL queries were used to explore trends, such as average prices, transmission types, fuel usage, and vehicle categories, offering valuable insights for the automotive industry.

This analysis helps identify key patterns in the car market, such as the most popular fuel types, transmission preferences, and the selling prices of different car brands. By categorizing and organizing the data, we can better understand customer preferences and market trends, which can guide strategies to enhance customer satisfaction and increase sales.

Aim

The analysis also aims to streamline the data by categorizing key features like mileage, kilometers driven, and seating capacity, making it easier to interpret patterns and trends. By identifying which car brands, fuel types, and transmission modes are most popular, along with factors affecting resale value, businesses can tailor their offerings to meet customer demand more effectively. Additionally, insights into market trends and consumer preferences can help optimize inventory management, pricing strategies, and targeted marketing efforts.

Datasets

1.Name

2.year

3.selling\_price

4.km\_driven

5.fuel

6.seller\_type

7.transmission

8.owner

9.mileage

10.engine

11.max\_power

12.torque

13.seats

14.mileage\_category

Objectives

* Data Collection: Gathered a comprehensive dataset with detailed information on cars, including attributes such as car brand, manufacturing year, transmission type, fuel type, mileage, engine specifications, and ownership history.
* Handling Missing Values and Duplicates: Addressed missing data by using imputation techniques and removed duplicate records to ensure data consistency.
* Outlier Detection and Treatment: Identified and treated outliers using statistical methods to maintain the dataset's accuracy and reliability.
* SQL Query Design and Execution: Created and executed SQL queries to extract key insights from the database, utilizing operations like SELECT, JOIN, and GROUP BY.
* Data Aggregation and Analysis: Employed aggregation functions (e.g., SUM, AVG) to analyse data trends and generate summary statistics for informed decision-making.

Data Overview

company varchar(50)

year int

selling\_price int

km\_driven int

fuel text

seller\_type text

transmission text

owner text

mileage text

kmpl int

engine text

CC int

max\_power text

bhp int

torque int

seats int

name text

km\_driven\_category varchar(50)

mileage\_category varchar(50)

current\_date1 date

QUERIES

1. avg selling price of each company -select company,avg(selling\_price) from car\_table group by company;

2. count of each type of transmission vehicles-select transmission,count(transmission) from car\_table group by transmission;

3. count of vehicles with bhp=110.5 -select count(bhp) from car\_table where bhp=100.5;

4. avg selling price of owner- select owner,avg(selling\_price) from car\_table group by owner;

5. count of vehicles of each company-select company,count(company) from car\_table group by company;

6. vehicles with year=2023,owner=Firstowner and CC>1000 - select \*from car\_table where year=2023 and owner='First Owner' and CC>1000;

7. count of vehicles in each category of seats -select seats,count(seats) from car\_table group by seats;

8. km\_driven of cars are sorted to categories

• alter table car\_table add column km\_driven\_category varchar(50);

• update car\_table set km\_driven\_category=case when km\_driven=10000 and km\_driven<=100000 then 'between 10k and 100k' else 'Above 100k' end;

9. mileage of cars is sorted into groups

• alter table car\_table add column mileage\_category varchar(50);

• update car\_table set mileage\_category=case when kmpl=10 and kmpl<=20 then 'Average' else 'High' end;

10. avg selling price of each mileage category -select mileage\_category,avg(selling\_price) from car\_table group by mileage\_category;

11. count of vehicles in each category of seats- select seats,count(seats) from car\_table group by seats;

12. remove the duplicate entries

• create temporary table table1 select company,year,selling\_price,km\_driven,fuel,seller\_type,transmission ,owner,mileage,kmpl,engine,CC,max\_power,bhp,torque,seats,nam e,km\_driven\_category,mileage\_category from car\_table group by company,year,selling\_price,km\_driven,fuel,seller\_type,transmission ,owner,mileage,kmpl,engine,CC,max\_power,bhp,torque,seats,nam e,km\_driven\_category,mileage\_category;

• delete from car\_table; • select count(\*)from table1;

• insert into car\_table select \*from table1;

13. avg torque -select company,avg(torque) from car\_table group by company order by avg(torque);

14.total selling price -select sum(selling\_price) from car\_table;

15.km\_driven by cars -select company,km\_driven from car\_table order by km\_driven desc ;

16.count of each seller type -select seller\_type,count(\*) from car\_table group by seller\_type order by count(\*) desc;

Conclusion

Data Overview: Initially, the dataset contained 8,000 entries, which was reduced to 6,722 after removing duplicates. There are five fuel types: petrol, diesel, CNG, LPG, and electric. Diesel is the most used fuel type, with 3,658 cars, while LPG has the lowest average selling price.

Mileage and Categories: The maximum mileage recorded is 42 kmpl, and vehicles with zero mileage were removed. Vehicles with mileage between 10 and 20 kmpl are categorized as "Average," representing the highest number of vehicles.

Pricing: The highest selling price belongs to Volvo at 10,000,000, while Opel has the lowest at 68,000. Test drive cars have the highest average selling price, and cars owned by "fourth and above" owners have the lowest.

Transmission and Seating: Manual transmission vehicles outnumber automatic ones. Five-seater cars are the most common, while 14-seater cars are the least common.

Company Insights: Maruti has sold the highest number of cars. Jaguar offers the highest average torque, while Daewoo has the lowest. Hyundai cars have driven the highest total distance of 2,360,457 km.

Seller Type and Total Sales: The total selling price across all vehicles is 3,553,243,742. The highest seller type is "individual," while "Trustmark dealer" has the lowest number of sales.

Overall,

The cleaned car dataset, reduced from 8,000 to 6,722 entries after removing duplicates, reveals key insights across various attributes. Diesel is the most popular fuel, while LPG has the lowest average selling price. Vehicles with mileage between 10-20 kmpl are the most common, with zero-mileage cars removed. Volvo has the highest selling price, and Opel the lowest, while manual transmission cars outnumber automatic ones. Five-seater cars are most prevalent, and Maruti leads in car sales, with Hyundai cars having the highest total distance driven. Jaguar provides the highest average torque, and individual sellers dominate the market. These findings highlight trends in fuel, pricing, transmission, and overall vehicle performance.