CAR DATA ANALYSIS

Introduction

The dataset used is about cars

It includes :

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

Aim

The primary goal of this project is to analyze this car data to uncover insights that can help improve customer satisfaction, and increase revenue.

Objective

1. **Data Collection:** By acquiring a well-structured dataset. This dataset included details such as car information,type of car,transmission and more.

2. **Data Cleaning and Preprocessing :** Performed data cleaning tasks to handle missing values, duplicate records, and outliers, ensuring the dataset's integrity.

3. **SQL Queries :** Designed and executed SQL queries to extract relevant information from the database. This involved a range of SQL operations, including SELECT, JOIN, GROUP BY, and aggregation functions.

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

Analysis

- The transmission in the given dataset were of 4 types “Manual”,“ Manual”, “Automatic”, “ Automatic”. Changed the transmission into 2 types as “Manual” and “Automatic”.
- Trimmed the Name column to get rid of unwanted spaces and ordered it in alphabetical order.
- Added a new column named “Company” by cutting the Name into a substring and adding the substring into that column. Moved the column named “Company” to the first of the table.
- Updated the datatype of mileage from text to int by first trimming it to remove spaces and then cutting it into substring
- Updated the datatype of max power, torque, and engine the same way.
- Categorizing the cars based on the kms driven by them.
- Categorizing mileage into low, avg, high.

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 then 'Less than 10k' 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 then 'Low' 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,name,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,name,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

In this car data there was about 8000 entries at first but after the removal of the duplicates the data is sorted to 6722 entries.

There is 5 fuel types ie, petrol diesel, CNG, LPG, electric and about 3658 number of cars uses diesel and it the highest used fuel. Average selling price of the fuel is lowest for LPG.

Maximum mileage from the whole data is 42kmpl. Mileage 0 vehicles are deleted from the dataset. Mileage between 10 and 20 is categorised into 'Average' and it has the highest number of vehicles.

The company 'Volvo' has the highest with 10000000 selling price of vehicles and 'Opel' has the lowest with 68000.

Manually transmitting vehicles are more than automatically transmitting.

Average selling price of the 'test drive' car is the highest and 'fourth and above' owner is the lowest.

Highest number of cars sold by 'Maruti'.

Five seater cars are available at most and the lowest is 14 seater cars.

Average torque is the highest for the company 'Jaguar' and lowest for 'Daewoo'.

Total selling price is '3553243742'.

The highest driven company is 'Hyundai' with km_driven '2360457'.

The highest seller type is individual and lowest is Trustmark dealer.