

CARS INFO DATASET

INTRODUCTION

This dataset provides a comprehensive overview of various car models, offering valuable insights into their characteristics and market trends.

The car dataset offers a wealth of information on various car models, covering everything from specifications and performance to exterior features and safety ratings. This comprehensive collection of data can be used for market analysis, product development, predictive modelling, and data visualization.

AIM

By analysing the dataset, you can gain valuable insights into consumer preferences, pricing trends, and the competitive landscape of the automotive industry

The aim of a car dataset is to provide a comprehensive and structured collection of information about various car models. This data can be used for a variety of purposes, including:

- **Market Analysis:** Understanding consumer preferences, identifying trends, and assessing the competitive landscape.
- **Product Development:** Informing design decisions, feature prioritization, and target market segmentation.
- **Predictive Modeling:** Forecasting sales, predicting resale values.

OBJECTIVE

1. Data Collection:

By acquiring a well-structured dataset. This dataset included details such as the company, model, year, fuel type, transmission, the selling price and other specified information about the cars.

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, GROUP BY, ORDER BY and aggregation functions.

4. Exploratory Data Analysis (EDA):

Used SQL to perform a wide range of analyses to uncover hidden patterns and trends. These patterns can provide valuable insights into customer spending behaviour, seasonal trends, and most bought car model or the company with the most number of cars.

DATA OVERVIEW

Column	Type
Company	varchar
Name	text
Selling price	Int
Km driven	int
Fuel	text
Seller type	text
Transmission	text
Mileage	int
Engine	int
Seats	int
Mileage Category	int
Current date	date

DATA 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.

QUERIES

1. To find the total number of bookings

```
SELECT COUNT(*) FROM cars;
```

2. Average selling price of each company

```
SELECT company,avg(selling_price) FROM car_table GROUP BY  
company;
```

3. Count of automatic transmission vehicles

```
SELECT count(transmission) FROM car_table WHERE  
transmission='Automatic';
```

4. count of vehicles with max_power=110.5

```
SELECT count(max_power) FROM cars WHERE  
max_power=100.5;
```

5. Average selling price of owner

```
SELECT owner,avg(selling_price) FROM car_table GROUP BY  
owner;
```

6. Count of vehicles of each company

```
SELECT company,count(company) FROM car_table GROUP BY  
company;
```

7. To find which mileage category selling price is the highest

```
SELECT mileage_category,max(selling_price) FROM cars  
GROUP BY mileage_category ;
```

8. Company with the max number of vehicles

```
SELECT company,count(company) FROM cars GROUP BY  
company ORDER BY count(company);
```

9. Cars with automatic transmission and petrol

```
SELECT * FROM cars WHERE transmission = 'automatic' and fuel  
= 'petrol';
```

10. Count of cars with year 2023, owner 'first owner', and
engine > 1000

```
SELECT * FROM cars WHERE year = '2023' and owner = 'first  
owner' and engine > 1000 ORDER BY company;
```

11. To find which ownership is common

```
SELECT owner,count(*) FROM cars GROUP BY owner ORDER BY  
count(*) desc limit 1 ;
```

12. Number of vehicles with year=2023,owner=Firstowner and
CC>1000

```
SELECT *from cars where year=2023 and owner='First  
Owner' and CC>1000;
```

13. mileage of cars is sorted into groups
alter table car_table add column mileage_category varchar(50);
update cars set mileage_category=case when kmpl<10 then
'Low' when kmpl>=10 and kmpl<=20 then 'Average' else 'High'
end;
14. To determine the average mileage of automatic versus
manual cars
SELECT transmission,avg(mileage)FROM car_table GROUP BY
transmission ORDER BY avg(mileage) desc;
15. remove the duplicate entries
create temporary table table1 select
company,year,selling_price,km_driven,fuel,seller_type,transmiss
ion,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,transmiss
ion,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;

CONCLUSION

The dataset on information of the cars aims to transform raw data into actionable insights

1. Total Number of Cars :

The total number of cars recorded in the database is 6730. This figure provides a baseline for understanding the volume of data and subsequent analyses.

2. Total number of cars from each company :

Maruti	2093
Skoda	70
BMW	46
MG	4
Tata	634
Hyundai	1211
Renault	211
Mahindra	711
Nissan	74
Datsun	57
Ford	353
Honda	361
Kia	3
Toyota	324
Volkswagen	173
Audi	33
Isuzu	4

Jeep	22
Land	3
Lexus	1
Mercedes-Benz	46
Volvo	9
Force	4
Chevrolet	216
Jaguar	8
Fiat	39
Mitsubishi	11
Ashok	1
Ambassador	4
Daewoo	3
Opel	1

3. Cars based on the type of fuel:

There is 5 fuel types ie, petrol diesel, CNG, LPG, electric

This data helps to understand the percentage of cars with different fuel types, enabling them to effectively convey this information.

Petrol	2987
Diesel	3656

Electric	1
CNG	51
LPG	35

4. Average selling price of cars with different ownership

Test Drive Car – 44,03,800

First Owner – 6,23,523

Second Owner – 3,95,166

Test Drive Car – 44,03,800

Third Owner – 2,86,885

Fourth & Above Owner – 2,32,503

This data helps understand that car prices generally decrease as the number of previous owners increases.

5. Most common seller

Individual sellers are the most common sellers in the site. This data helps the app management in many ways

6. The average mileage of automatic versus manual cars

Manual – 19

Automatic – 15

This data helps customers choose between an automatic or manual car based on their average mileage