USED CARS PRICE ANALYSIS USING SQL

INTRODUCTION

The used car market offers valuable insights into consumer preferences, vehicle performance over time, and pricing trends. This dataset encompasses key details about various used cars, including make and model, mileage, engine specifications, ownership type, and pricing information. By performing a comprehensive Exploratory Data Analysis (EDA), this project aims to identify patterns that influence used car values, understand demand for different car types and features, and provide data-driven insights to support buyers, sellers, and automotive industry stakeholders in making informed decisions.

OBJECTIVE

The objective of this analysis is to derive actionable insights from the dataset to inform and support stakeholders in the used car market, including buyers, dealers, and analysts. By examining a range of variables, such as car specifications, fuel type, transmission, and geographic factors, this analysis aims to identify the primary determinants of used car pricing and demand. Additionally, it will reveal market trends, customer preferences, and pricing strategies, providing a comprehensive understanding of market dynamics and empowering stakeholders to make data-driven decisions in a competitive marketplace.

SCOPE

- Market Demand and Consumer Behaviour: Insights into customer preferences for factors like brand, fuel type, and transmission, helping to understand shifts in consumer trends.
- Pricing Dynamics: Analysis of variables influencing car resale values, such as mileage, condition, location, and brand, which is essential for pricing strategies.
- Regional Trends and Variations: Identifying geographic factors that impact demand,

pricing, and availability, aiding regional dealerships and sellers in optimizing their offerings.

OVERVIEW OF DATASET

The dataset comprises 5,779 rows and 13 columns, each providing detailed information about individual used cars. Below is an overview of the columns in the dataset:

Column	Description
Name	The specific make and model of each car.
Location	The city where the car is listed for sale.
Year	The year the car was manufactured.
Kilometers Driven	Total distance covered by the car in kilometers.
Fuel Type	The type of fuel used by the car (e.g., Petrol, Diesel).
Transmission	The transmission type (e.g., Manual, Automatic).
Owner Type	It indicate whether it’s a first, second, or higher ownership.
Mileage	The car’s mileage in kilometers per liter.
Engine	The engine capacity of the car in cubic centimeters (cc).
Power	The engine power output in horsepower (HP).
Seats	The number of seats in the car.
Price	The listed price of the car in the dataset.
Company	The car's brand.

DATA CLEANING AND PREPARATION

To optimize the dataset for accurate analysis, various data cleaning and preparation steps were implemented to remove inconsistencies, enhance clarity, and organize the data for optimized querying efficiency.

- **Remove Duplicates**

  Ensure that there are no repeated rows in the dataset. This helps maintain accuracy and prevents overcounting in analysis.
- **Handle Missing Data (NULLs)**

  Addressed missing values in columns like Mileage and Power by imputing averages or removing incomplete entries where necessary.
- **Data Type Conversion**

  convert columns to the appropriate data types (e.g., converting strings to integers) to ensure consistency and prevent errors in calculations.
- **Removing Unnecessary Columns:** Dropped any irrelevant columns not essential for primary analysis.
- **Standardizing Units:** Removed unit suffixes (e.g., "kmpl" in mileage and "cc" in engine) for consistency in data types.

DATA ANALYSIS

1. Mean and Standard Deviation of Engine and Power:

```
SELECT AVG(ENGINE) AS AVG_ENGINE, STDDEV(ENGINE) AS STD_ENGINE,
AVG(POWER) AS AVG_POWER, STDDEV(POWER) AS STD_POWER
FROM USED_CARS;
```

AVG_ENGINE	STD_ENGINE	AVG_POWER	STD_POWER
1628.925074	599.1203297	113.691476	53.82991482

2. Distribution by price ranges:

```
SELECT CASE WHEN PRICE < 250000 THEN 'UNDER 250K'
WHEN PRICE BETWEEN 250000 AND 500000 THEN '250K-500K'
WHEN PRICE BETWEEN 500000 AND 1000000 THEN '500K-1M'
ELSE 'OVER 1M'
END AS PRICE_RANGE,
COUNT(*) AS COUNT FROM USED_CARS GROUP BY PRICE_RANGE  ORDER BY COUNT
DESC;
```

PRICE_RANGE	COUNT
250K-500K	1867
500K-1M	1858
Over 1M	1448
Under 250K	606

Most price are in the '250K-500K' range and Least price are in the 'Under 250k' range.

3. Minimum and maximum price:

SELECT MIN(PRICE) AS MIN_PRICE, MAX(PRICE) AS MAX_PRICE FROM USED_CARS;

MIN_PRICE	MAX_PRICE
44000	16000000

4. Kilometers Driven:

SELECT MAX(KILOMETERS_DRIVEN) AS MAX_KILOMETERS_DRIVEN,

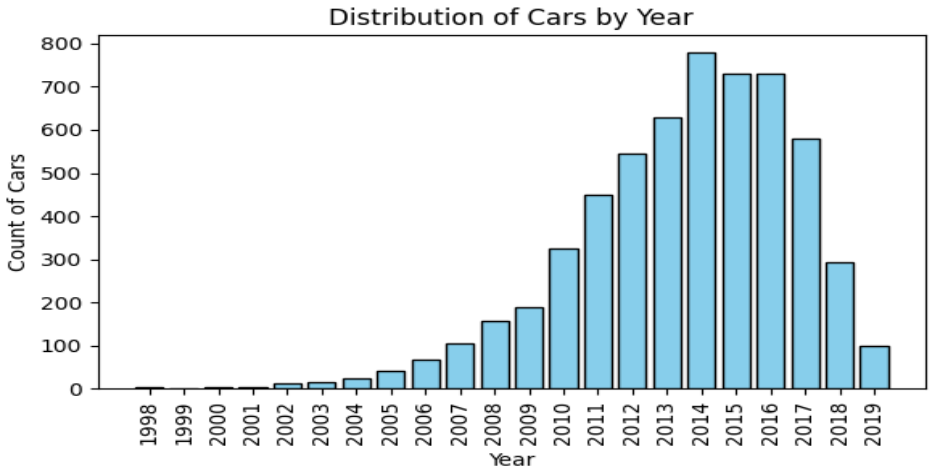
MIN(KILOMETERS_DRIVEN) AS MIN_KILOMETERS_DRIVEN,

AVG(KILOMETERS_DRIVEN) AS AVG_KILOMETERS_DRIVEN FROM USED_CARS;

MAX_KILOMETERS_DRIVEN	MIN_KILOMETERS_DRIVEN	AVG_KILOMETERS_DRIVEN
6500000	171	58359.2525

5. Distribution of Years:

SELECT YEAR, COUNT(*) AS COUNT FROM USED_CARS GROUP BY YEAR ORDER BY YEAR;

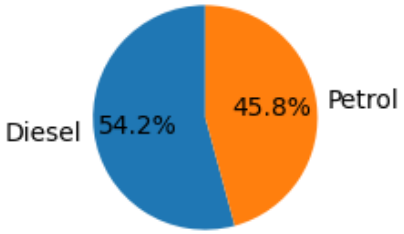


Car count increased steadily until 2014, peaking at 780, then remained high until 2017. After 2017, the count dropped

6. Count the number of cars by fuel type:

SELECT FUEL_TYPE, COUNT(*) AS COUNT_OF_CAR FROM USED_CARS GROUP BY FUEL_TYPE;

FUEL_TYPE	COUNT_OF_CAR
Diesel	3134
Petrol	2645



Diesel has the highest count of cars with 3134, while petrol has 2645.

7. Find the average price of cars based on transmission type:

SELECT TRANSMISSION, AVG(PRICE) AS AVG_PRICE FROM USED_CARS GROUP BY TRANSMISSION;

TRANSMISSION	AVG_PRICE
Manual	543317.716
Automatic	1980857.823

The average price for cars with an automatic transmission is higher than those with a manual transmission.

8. Top 3 most expensive cars based on Price Distribution:

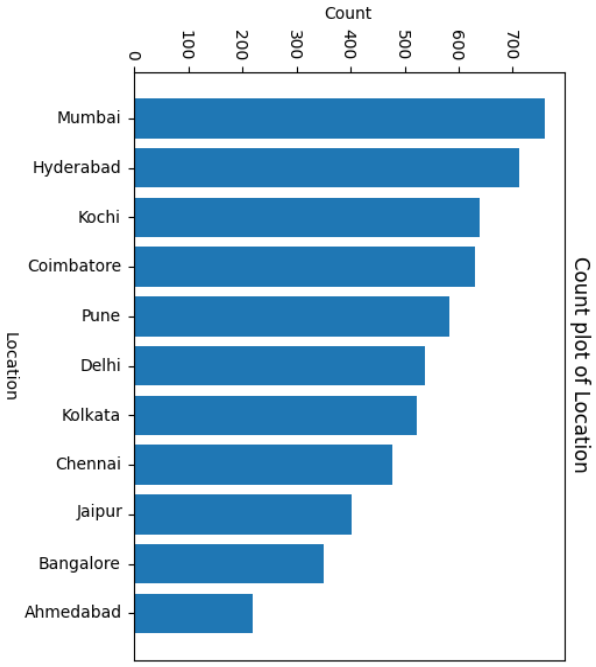
SELECT NAME, PRICE, FROM USED_CARS GROUP BY NAME, PRICE ORDER BY PRICE DESC LIMIT 3;

NAME	PRICE
Land Rover Range Rover 3.0 Diesel LWB Vogue	16000000
Lamborghini Gallardo Coupe	12000000
Jaguar F Type 5.0 V8 S	10000000

9.Find the total number of cars available in each location:

SELECT LOCATION, COUNT(*) AS COUNT FROM USED_CARS GROUP BY LOCATION ORDER BY COUNT DESC;

LOCATION	COUNT
Mumbai	755
Hyderabad	704
Kochi	635
Coimbatore	626
Pune	577
Delhi	533
Kolkata	519
Chennai	473
Jaipur	399
Bangalore	343



10. Average Price vs. Kilometers Driven:

SELECT KILOMETERS_DRIVEN, AVG(PRICE) AS AVG_PRICE FROM USED_CARS GROUP BY KILOMETERS_DRIVEN ORDER BY KILOMETERS_DRIVEN LIMIT 5;

KILOMETERS_DRIVEN	AVG_PRICE
171	360000
600	625000
1000	1018777.778
1001	739000
1011	1350000

11. Price by Fuel Type:

SELECT FUEL_TYPE, AVG(PRICE) AS AVG_PRICE FROM USED_CARS GROUP BY FUEL_TYPE;

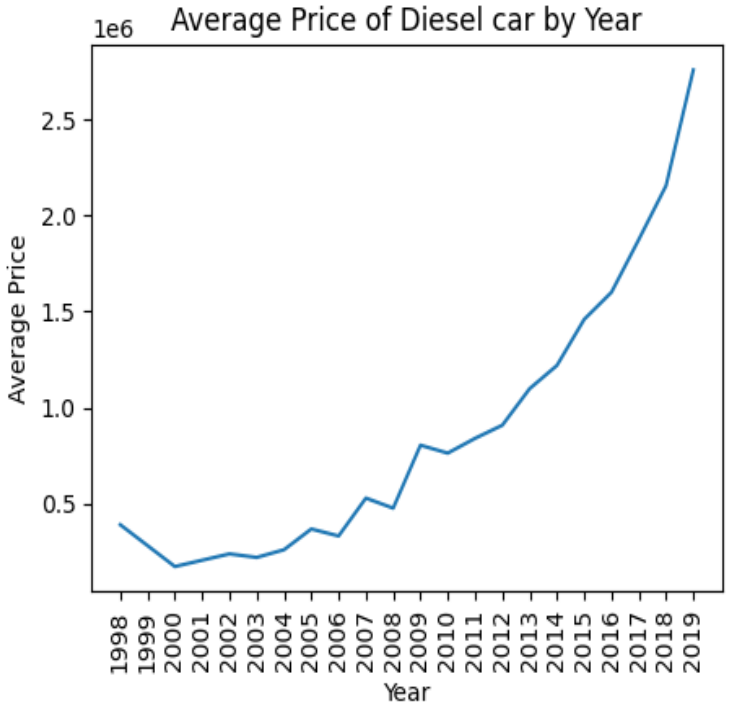
FUEL_TYPE	AVG_PRICE
Diesel	1285831.201
Petrol	577142.9112

This suggests that diesel cars are generally priced higher than petrol cars on average.

12. Price trends for Diesel fuel types over the years:

SELECT YEAR, AVG(PRICE) AS AVG_PRICE FROM USED_CARS WHERE FUEL_TYPE='DIESEL' GROUP BY FUEL_TYPE, YEAR ORDER BY YEAR;

YEAR	AVG_PRICE
2019	2727226.415
2018	2141586.957
2017	1873990.446
2016	1583142.5
2015	1445837.9
2014	1208724.28
2013	1093194.03
2012	887028.49
2011	839571.4286
2010	759901.6393
2009	803911.3924
2008	476385.9649
2007	528633.3333
2006	331578.9474
2005	368000
2004	260000
2003	220000
2002	238333.3333
2000	172500
1998	390000



13.How the price of a car is influenced by the year of manufacture, the number of kilometers driven, and the fuel type.

SELECT NAME, YEAR, KILOMETERS_DRIVEN, FUEL_TYPE, AVG(PRICE) AS AVG_PRICE FROM USED_CARS GROUP BY NAME, YEAR, KI LOMETERS_DRIVEN, FUEL_TYPE ORDER BYEAR, KILOMETERS_DRIVEN LIMIT 9;

NAME	YEAR	KILOMETERS_DRIVEN	FUEL_TYPE	AVG_PRICE
Mercedes-Benz E-Class 250 D W 210	1998	55300	Diesel	390000
Maruti Zen LXI	1998	95150	Petrol	45000
Maruti Zen LX	1998	95150	Petrol	53000
Maruti Zen VX	1999	70000	Petrol	77000
Honda City 1.3 EXI	1999	140000	Petrol	90000
Maruti 800 DX BSII	2000	78000	Petrol	55000
Maruti Zen VXi - BS III	2000	90000	Petrol	70000
Mahindra Bolero ZLX BSIII	2000	124000	Diesel	195000
Tata Sumo Delux	2000	158000	Diesel	150000

14. Location-wise Price Comparison:

SELECT LOCATION, AVG(PRICE) AS AVG_PRICE FROM USED_CARS GROUP BY LOCATION;

LOCATION	AVG_PRICE
Pune	689010.3986
Mumbai	953777.4834
Kolkata	570583.815
Kochi	1129382.677
Jaipur	592278.1955
Hyderabad	999491.4773
Delhi	990956.848
Coimbatore	1499864.217
Chennai	792266.3848
Bangalore	1349612.245
Ahmedabad	840330.2326

This data shows Coimbatore has the highest average price, and Kolkata the lowest, with price variation across the Location.

15. Analyzes how the car price varies based on engine size, power, and transmission type.

SELECT ENGINE, POWER, TRANSMISSION, MAX(PRICE) AS MAX_PRICE FROM USED_CARS GROUP BY ENGINE, POWER, TRANSMISSION ORDER BY ENGINE, POWER LIMIT 8;

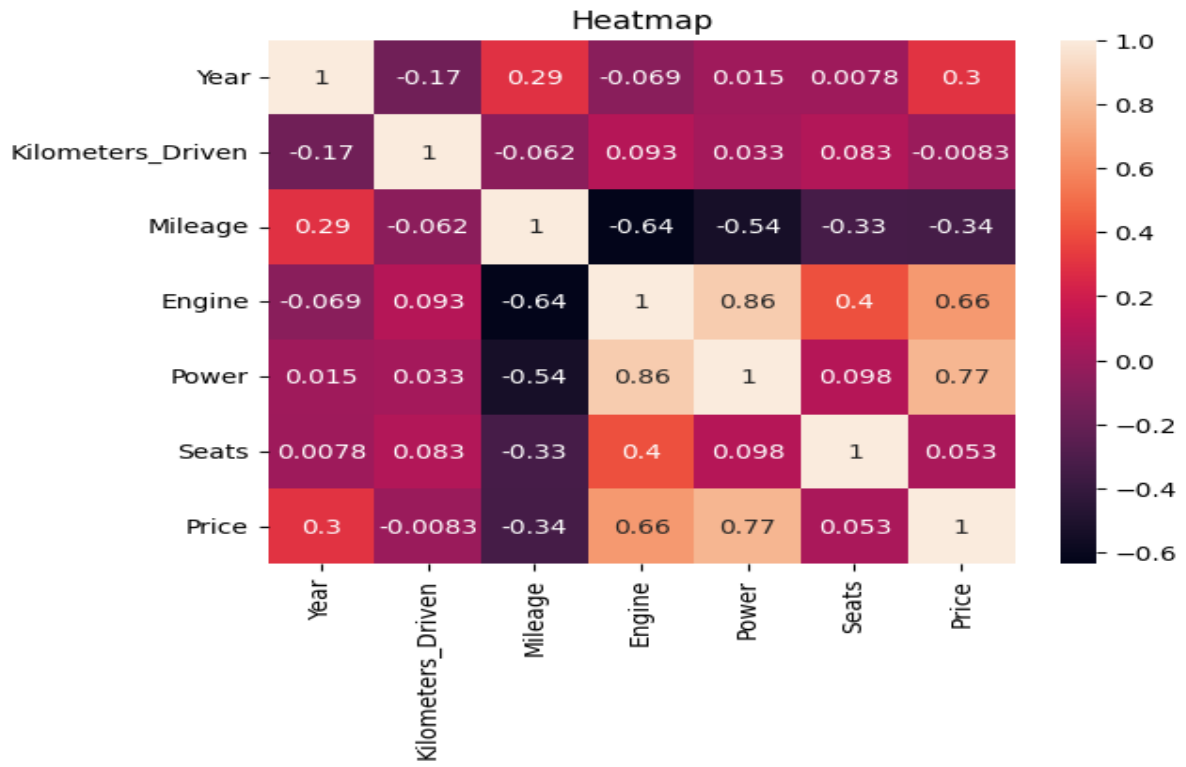
ENGINE	POWER	TRANSMISSION	MAX_PRICE
1047	69	Manual	746000
1061	47	Manual	122000
1061	64	Manual	275000
1061	67	Manual	302000
1086	62.1	Manual	332000
1086	66.7	Manual	230000
1086	68	Manual	70000
1086	68.05	Manual	427000

16. Understand how Fuel_Type, Location, Owner_Type these three categorical variables impact the price.

SELECT FUEL_TYPE, LOCATION, OWNER_TYPE, AVG(PRICE) AS AVG_PRICE FROM USED_CARS GROUP BY FUEL_TYPE, LOCATION, OWNER_TYPE ORDER BY FUEL_TYPE, LOCATION LIMIT 10;

FUEL_TYPE	LOCATION	OWNER_TYPE	AVG_PRICE
Diesel	Ahmedabad	First	1036327
Diesel	Ahmedabad	Second	1159318
Diesel	Bangalore	First	1798310
Diesel	Bangalore	Second	1689897
Diesel	Bangalore	Third	852857.1
Diesel	Chennai	First	1158576
Diesel	Chennai	Fourth & Above	510000
Diesel	Chennai	Second	825000
Diesel	Chennai	Third	467857.1

CORRELATION MATRIX



CONCLUSION

This analysis of the used car market provides valuable insights into the factors that influence the pricing and demand for used cars, allowing for informed decision-making by buyers, sellers, and industry stakeholders. Below are the key findings:

Price Distribution and Trends

- The most common price range for used cars is between 250K and 500K, which accounts for the highest number of cars listed in the dataset.
- The minimum price of a car is 44,000, while the maximum price is 16,000,000 indicating a wide range of cars with different conditions, brands, and features.
- Cars listed in the 'Over 1M' range are fewer in number, but they command significantly higher prices, such as luxury company like Land Rover and Lamborghini.

Factors Influencing Price

- Mileage and Kilometers Driven: Cars with lower kilometers driven tend to have higher prices, as expected. This supports the notion that car condition, as influenced by mileage, is crucial in pricing.
- Fuel Type: Diesel cars, on average, are priced higher than petrol cars, with an average price of 1,285,831 for diesel and 577,142 for petrol.
- Transmission Type: Cars with automatic transmissions are more expensive on average than those with manual transmissions. The average price for automatic cars is 1,980,857, compared to 543,318 for manual cars.
- Engine Size and Power: Larger engines and higher power output tend to correlate with higher car prices. However, other factors like condition and model also play significant roles.

Location-Based Insights

- Regional price differences are significant, with Coimbatore having the highest average price (1,499,864), while Kolkata has the lowest average price (570,584).
- Certain cities, such as Mumbai, Hyderabad, and Kochi, have higher average prices, which could indicate a greater demand for used cars in these locations, likely driven by higher-income areas or better market conditions.

Ownership and Demand

- Cars with first ownership generally command higher prices, especially when compared to second or third-hand cars. First-owner cars tend to be in better condition, contributing to

their higher value.

Year and Condition

- Older cars, especially those made before 2008, are usually much cheaper. For instance, diesel cars from 2008 and earlier can cost as low as 476,385. On the other hand, there is more demand for newer cars, as shown by the larger number of cars from 2017-2019 in the market.

This analysis provides valuable insights into the key factors that influence used car pricing and market demand, offering useful information for both sellers and buyers. By understanding these factors, sellers can better position their vehicles for sale, while buyers can make more informed decisions when navigating the used car market.