DATA ANALYTICS

SQL MINI PROJECT

- Analyzing and manipulating data to gather insight.

- By
- Govindev eyyani pradeesh

INTRODUCTION

- This project involves a thorough examination of a car dataset obtained from CARS24, employing SQL queries to determine valuable insights into the automotive market. By reviewing critical factors such as selling price, fuel type, and transmission, our goal is to identify patterns, preferences, and drivers impacting car sales. Through this analysis, we aim to provide stakeholders with actionable intelligence to guide decision-making processes and deepen their understanding of market dynamics within the automotive sector.

Data set overview

	A	B	C	D	E	F	G	H	I	J	K	L
1	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner	mileage	engine [CC]	max_power	seats
2	Hyundai i20 Asta 1.2	2007	550000	2360457	Petrol	Individual	Manual	Second Owner	18.6	1197	81.83	5
3	Maruti Wagon R LXI Minor	2010	194000	577414	Petrol	Individual	Manual	Second Owner	18.9	1061	67	5
4	Maruti Wagon R VXI BS IV	2011	229999	500000	Petrol	Individual	Manual	Second Owner	18.9	998	67.1	5
5	Maruti Wagon R LXI BS IV	2012	220000	360003	Petrol	Individual	Manual	Second Owner	18.9	998	67.1	5
6	Hyundai Sonata 2.4 GDi MT	2012	550000	330000	Petrol	Individual	Manual	Second Owner	13.44	2359	198.25	5
7	Hyundai Sonata 2.4 GDi MT	2012	500000	330000	Petrol	Individual	Manual	Second Owner	13.44	2359	198.25	5
8	Maruti Ertiga BSIV VXI	2017	700000	227000	Petrol	Individual	Manual	First Owner	17.5	1373	91.1	7

- This is the dataset from this we can get key columns like year of sale ,selling price, mileage etc..
- We can you this to further analyse and get insight

METHODOLOGY

- 1. Data Preparation: The first step involved importing the provided car dataset from CARS24 into a suitable database management system in this case sql

- 2. Database Table Creation:A database table named 'cars24' was created to organize the dataset's attributes effectively.

- 3. SQL Query Execution: SQL queries were formulated to extract insights from the dataset, utilizing various functions, subqueries, and aggregations. Each query was tailored to address specific analytical objectives, such as calculating average selling prices, identifying common fuel types,number of sales per year etc…

- 4.Query Optimization: querys is optimized for simplicity unwanted joins were excluded

- 5.insights:optimal insights we generated to get details about automotive sector from the given data to implement further optimization for this business

SQL QUERY USED

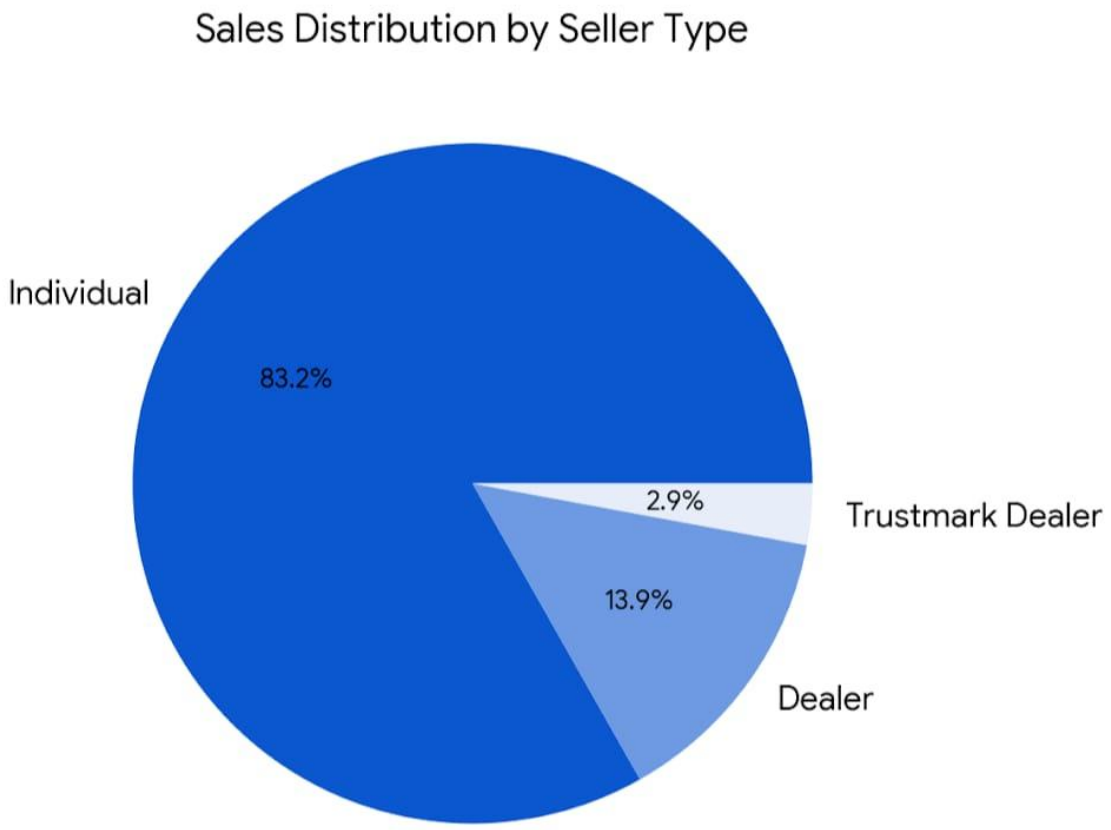
- This project is executed with the help of **functions, subqueries, and aggregations etc..**
- **Functions**: Functions in SQL are used to perform operations on data values.
- **Subqueries:**They can be used in various parts of a SQL statement, such as the SELECT, FROM, WHERE, or HAVING clauses. Subqueries can return a single value, a single row, multiple rows, or even an entire result set.
- **Aggregations**: Aggregations are used to perform operations on groups of rows to return a single result per group. They are commonly used with the GROUP BY clause. Aggregate functions like SUM(), AVG(), COUNT(), MIN(), MAX() are frequently used with aggregations.

INSIGHTS

- Insight 1
- **SELECT ***
- **FROM cars24**
- **WHERE name LIKE '%Honda%';** --and from the number of rows returned we can
- find number of honda cars.
- From this query we can get the details all the cars where they are made by honda
- We can use similar query to find details of every manufactures separately

- Insight 2
- **SELECT ***
- **FROM cars24**
- **WHERE max_power = (SELECT MAX(max_power) FROM cars24);**
- **From this query we can find which car has the most power output**
- **With similar queries we can find which car has the most mileage or least mileage etc..**

- Insight 3
- **SELECT seller_type, COUNT(*) AS num_sales**
- **FROM cars24**
- **GROUP BY seller_type;**
- this query is to find how many different types of sellers and their sales



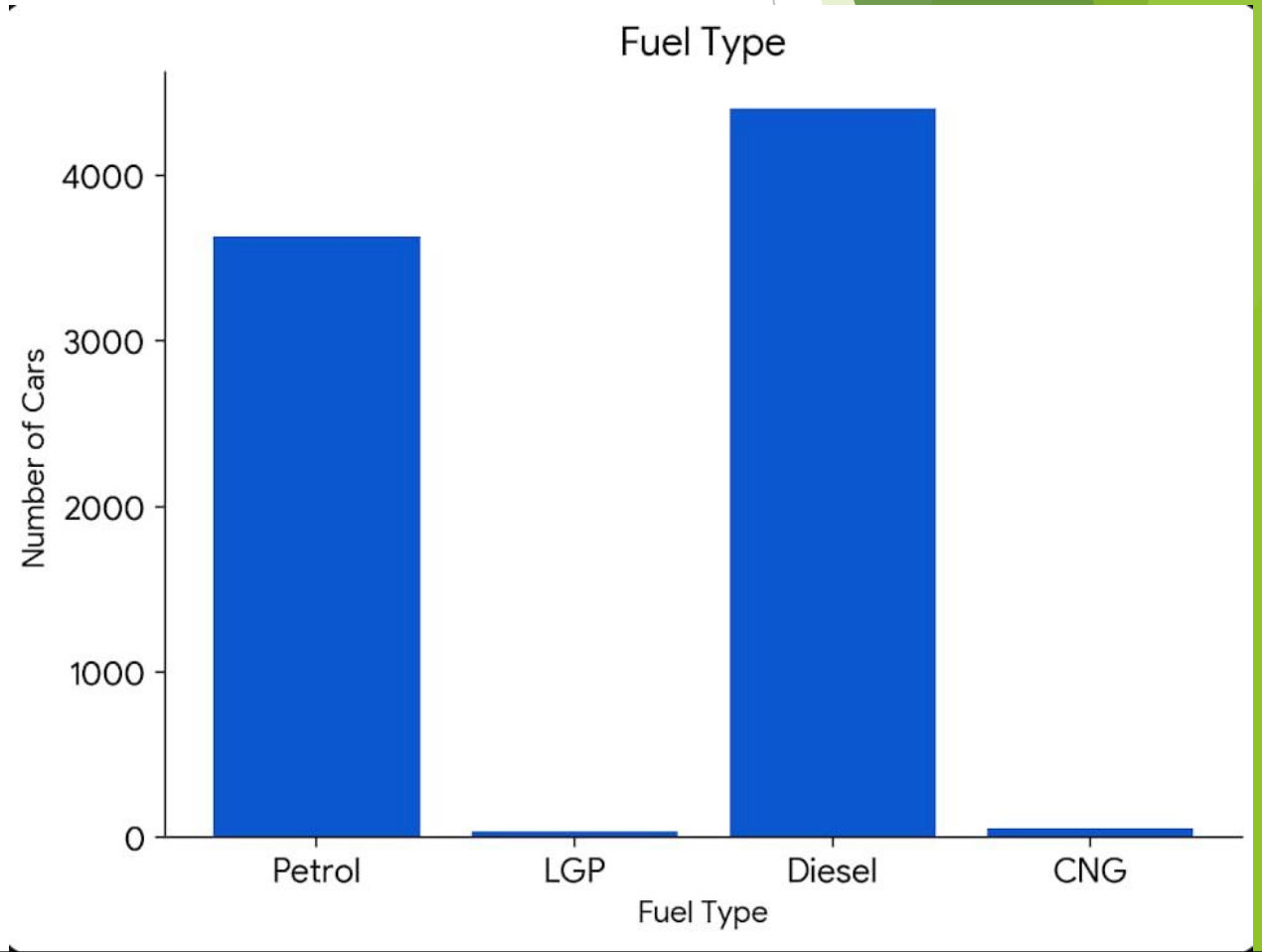
Insight 4

SELECT fuel_type, COUNT(*) AS num_cars

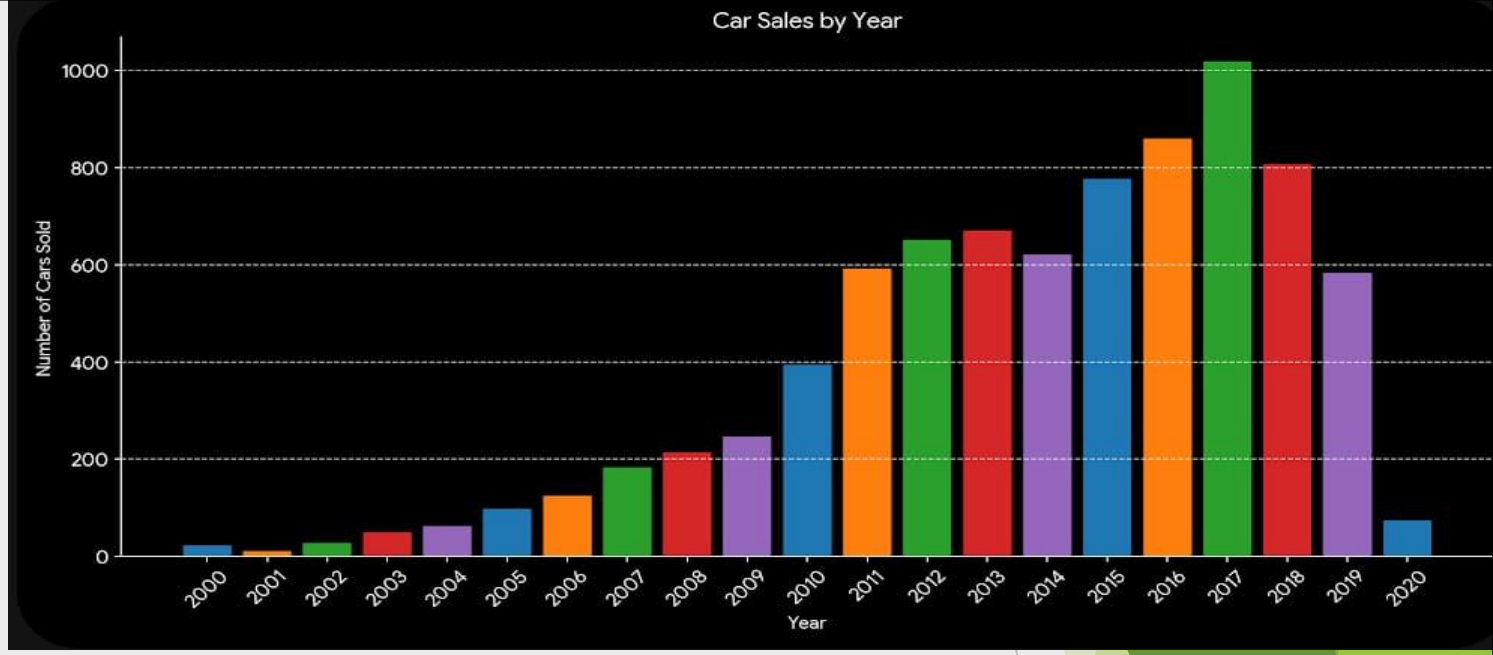
FROM cars24

GROUP BY fuel_type;

this query is to find how many cars are their in each fuel type



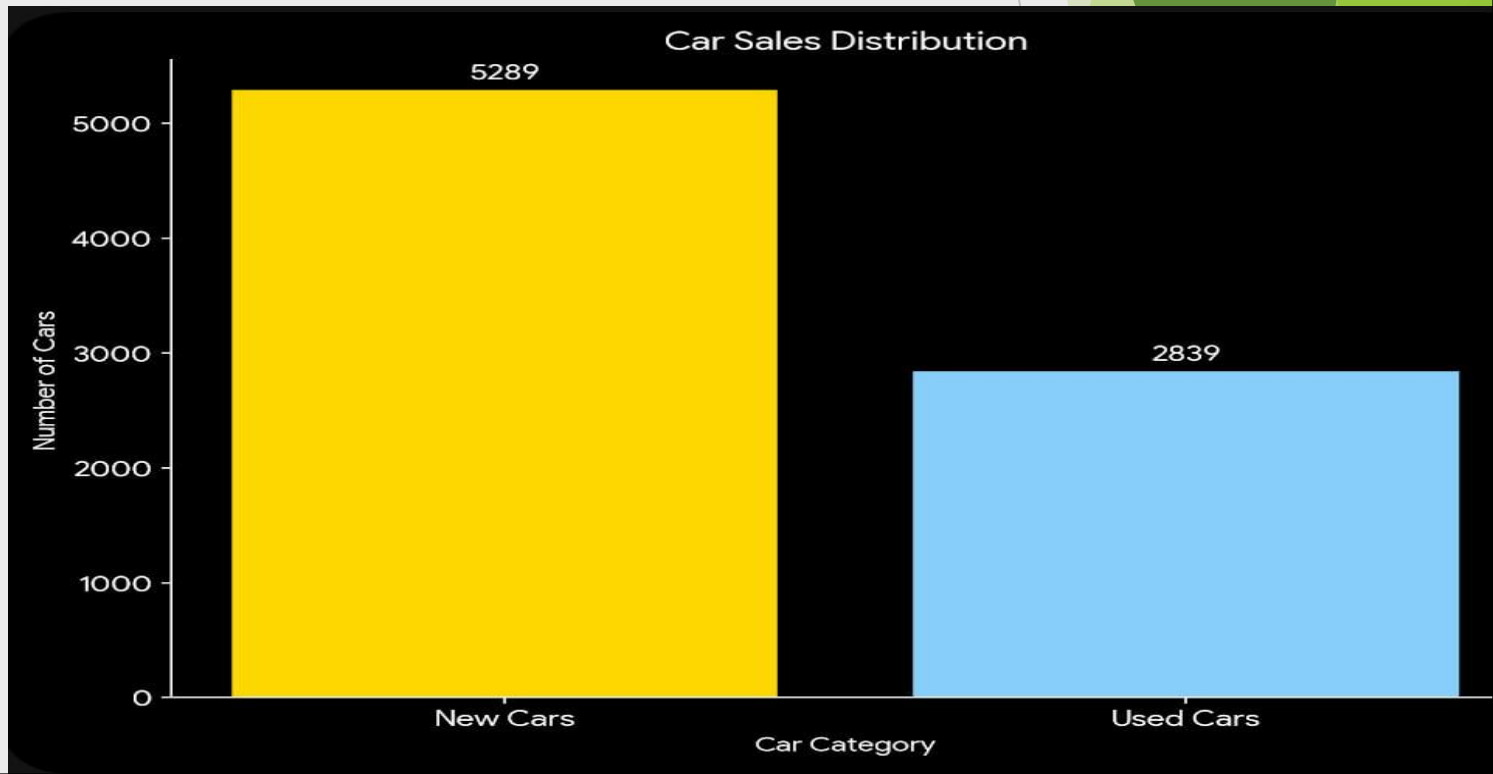
- insight 5
- **SELECT year, COUNT(*) AS num_sales**
- **FROM cars24**
- **GROUP BY year**
- :this query is to find number of sales in each year



Insight 6

```
SELECT     CASE          WHEN owner = 'First Owner' THEN 'New'          ELSE 'Second Hand' END AS carsales_status,     COUNT(*) AS num_cars FROM cars24 GROUP BY carsales_status  ;
```

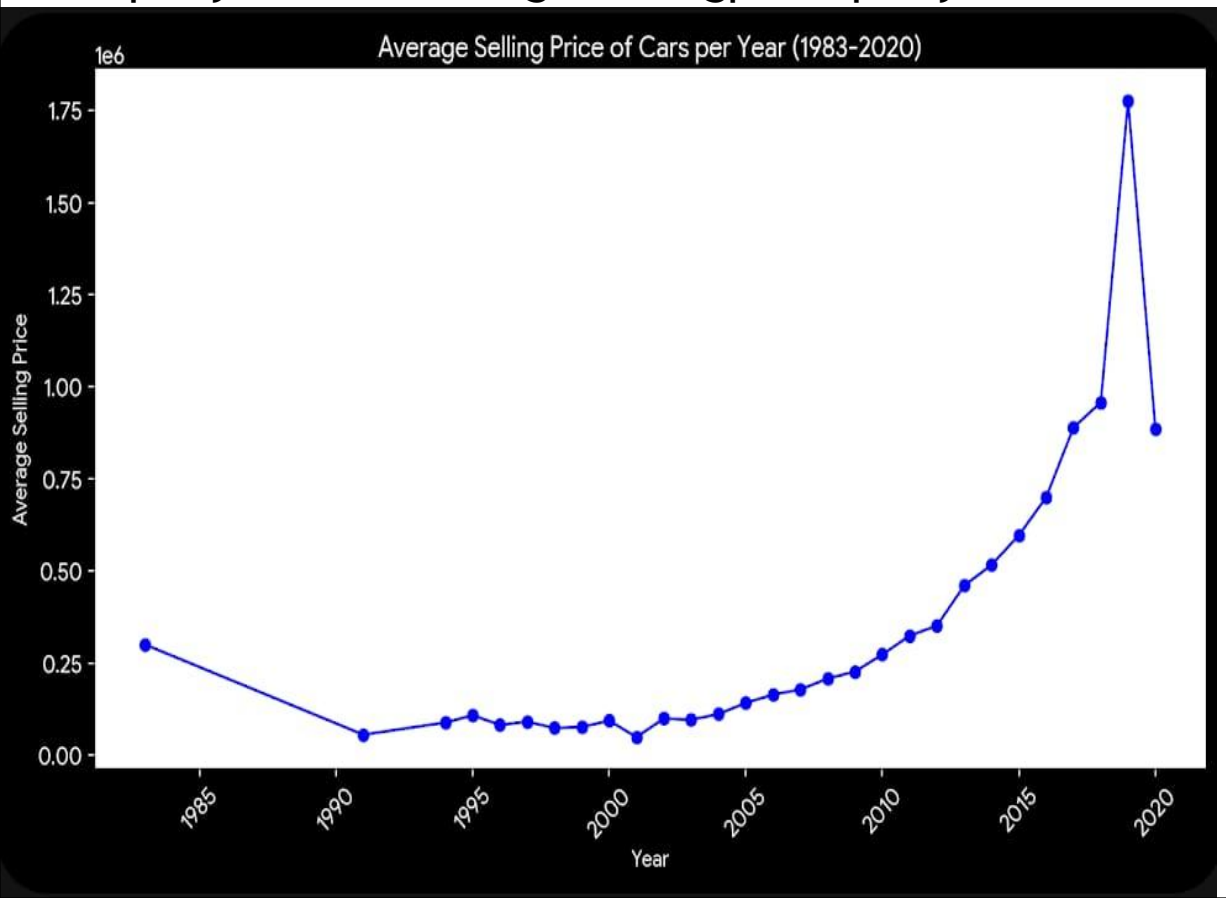
This query is to find how many new cars and how many used cars are sold



Insight 7

```
SELECT year, AVG(selling_price) AS
avg_selling_price
FROM cars24
GROUP BY year
ORDER BY avg_selling_price DESC;
```

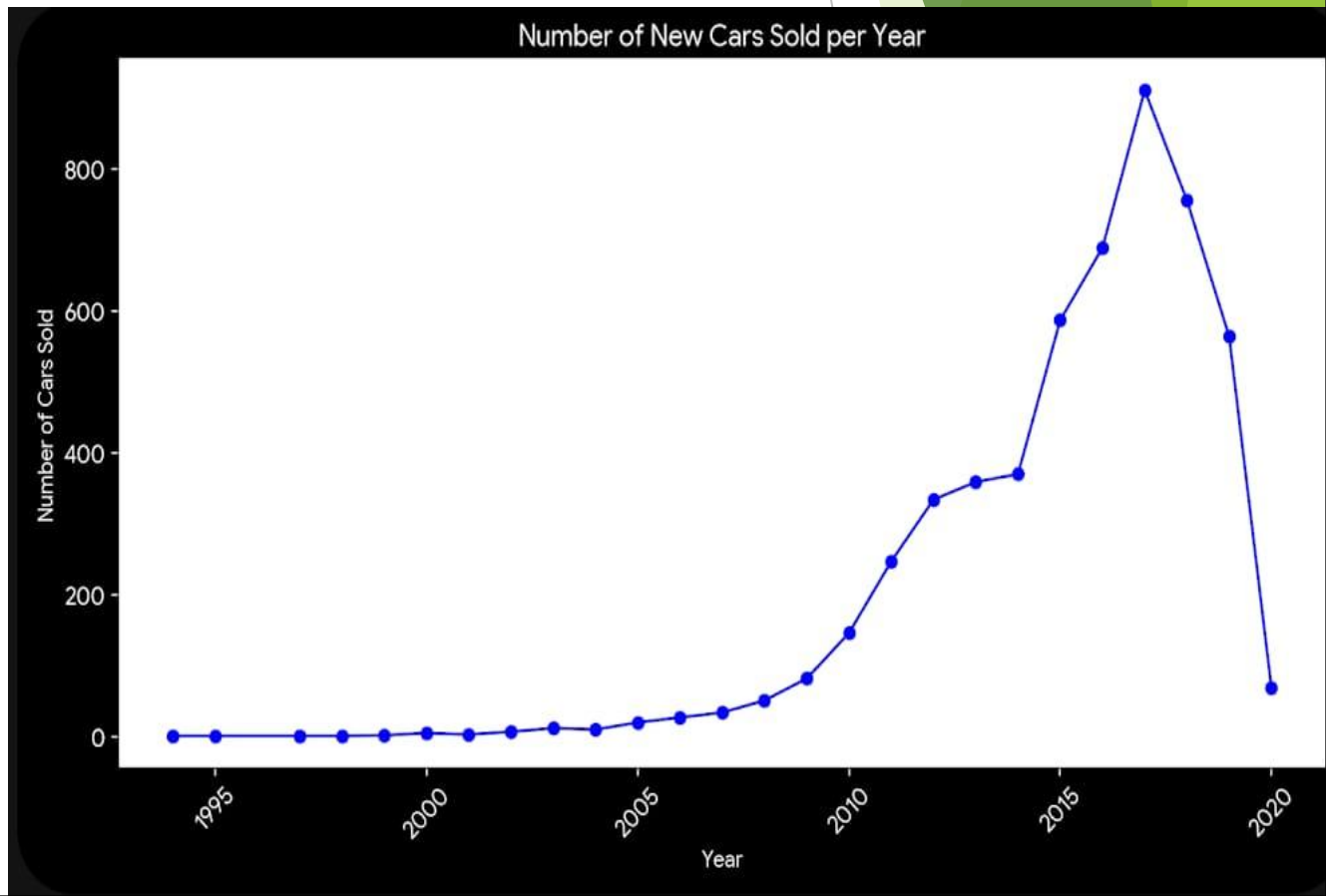
this query to find average sellingprice per year



Insight 8

```
SELECT year, COUNT(*) AS num_newcars
FROM cars24
WHERE owner = 'First Owner'
GROUP BY year
ORDER BY year DESC;
```

this query is to find how many new cars are sold per year



Insight 9

```
SELECT *
FROM cars24
WHERE owner = 'First Owner' AND transmission = 'Automatic';
```

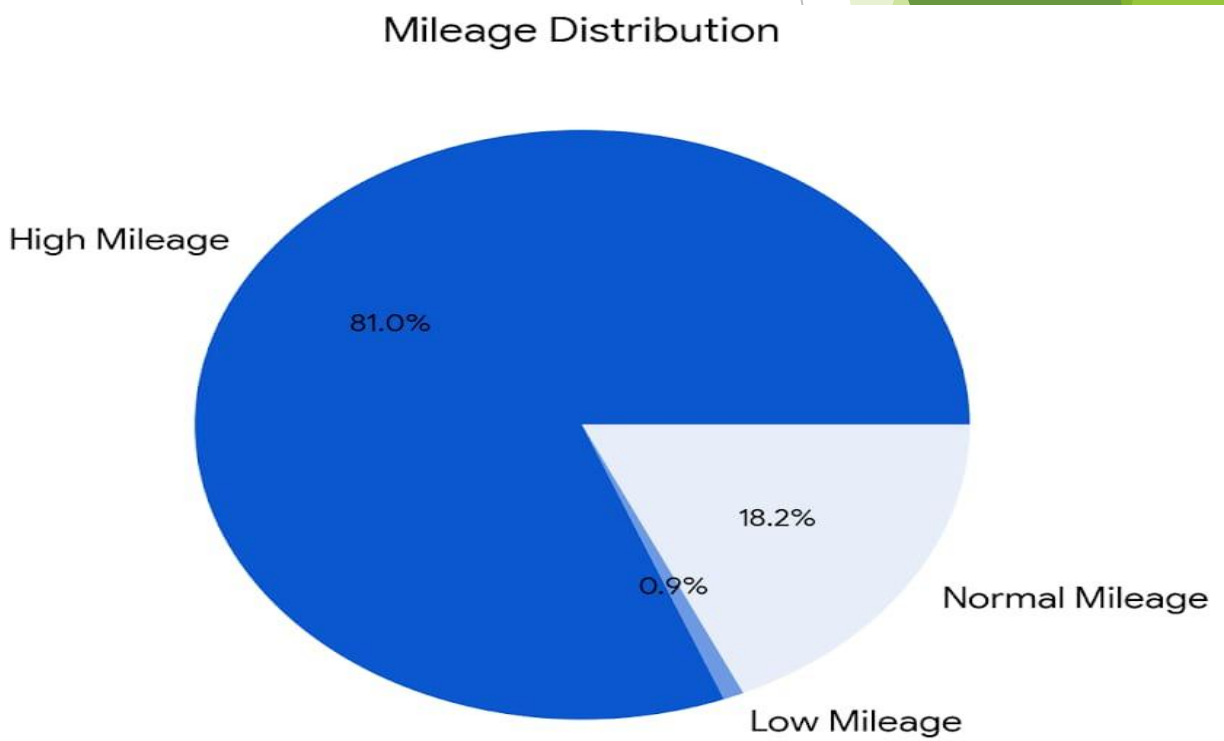
this query is to find the list of automatic new cars

So this is an example for and function if the both conditions are met then only it will retrieve the data.

Insight 10

```
select * from cars24;
SELECT
CASE
WHEN mileage >= 16 THEN 'High Mileage'
WHEN mileage > 11 THEN 'Normal Mileage'
WHEN mileage <= 11 THEN 'Low Mileage'
ELSE 'Unknown' -- Handle NULL values
END AS mileage_category,
COUNT(*) AS num_cars
FROM cars24
GROUP BY mileage_category;
```

this query is to find how many cars are in different mi categories



CONCLUSION

Main findings :

1:in this project we were able to find large portion of the sales are through individuals ,most used fuel type is diesel and petrol .

2:the number of sales peaked at the year 2017 and we can see decline in the sales from that year.

3:since new cars are sold more than used cars we can assume that people prefer new cars.

4:average selling price also peaked at the year 2017 .

5:from the mileage category we know that people prefer high mileage cars which is indicated by large amount of sales of high mileage cars. These are the main findings.

Challenges and lessons learned:

The main challenges we faced is that there where no clear direction to the output we are looking for if there where a clear objective the analyses part will me more accurate.

The lessons learned because there is no specific objective we can you our imagination to get optimum output based on our knowledge and its was a good learning experience.

FUTURE WORK

Potential Future Directions :

Potential future directions include expanding the analysis to include additional variables such as customer demographics, geographical trends, and market competition. Further exploration of the dataset could provide deeper insights into consumer preferences and market dynamics.