"Driving Data Insights: A Deep Dive into Car Analytics with SQL"

- Analyzing the cars24 table dataset using SQL queries and using SQL queries to uncover insights into the car market.

JOSHIN MATHEWS JOSEPH

CAP-DS-5

2024

Introduction

- **Project Objectives:** To analyze the cars24 dataset using SQL queries to extract insights into car sales trends, consumer preferences, and seller performance in the automotive industry.

- **Dataset**: information about cars, that includes name, year, selling price, kilometers driven, fuel type, seller type, transmission type, owner count, mileage, engine displacement, maximum power output, and number of seats.

- **Significance of Analysis:** This analysis is significant as it provides valuable insights for car manufacturers, dealers, and other stakeholders , making informed decisions regarding product development, pricing strategies, marketing campaigns, and inventory management.

Dataset Overview

- **Highlight key columns**: name, year, selling_price, km_driven, fuel, seller_type, transmission, owner, mileage, engine [CC], max_power, seats
- **Name**:

Useful for identifying different car models, for understanding market preferences and trends.

- **Year**:

Helps in analyzing trends over time, such as changes in car manufacturing or consumer preferences.

- **Selling Price**:

understanding the pricing dynamics of cars in the dataset.

- **Km Driven**:

  Provides insights into the usage and condition of the cars, which can affect their value.

- **Fuel**:

  Helps in understanding the popularity of different fuel types among car buyers.

- **Seller Type**:
  Indicates the source of the cars (individuals, dealers, etc.), which can impact pricing and trustworthiness.

- **Transmission**:
  Allows for analysis of the popularity of manual vs. automatic transmissions.

- **Owner**:
  Indicates the history of ownership, which can affect pricing and perceived value.

- **Mileage**:
  Indicates the fuel efficiency of cars, which is important for buyers concerned about operating costs.

- **Engine [CC]**:
  Provides information about the power and performance of the cars.

- **Max Power**:
  Similar to engine [CC], it helps in understanding the cars' performance characteristics.

- **Seats**:
  Indicates the capacity of the cars, which can be a factor for buyers with specific needs.

	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 [C	max_pow	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
9	Hyundai i20 1.2 Asta	2011	220000	220000	Petrol	Individual	Manual	Fourth & Above Owner	17	1197	80	5
10	Maruti 800 EX	2004	70000	220000	Petrol	Individual	Manual	Second Owner	16.1	796	37	4
11	Honda Civic 1.8 S AT	2007	175000	218463	Petrol	Individual	Automatic	First Owner	12.9	1799	130	5
12	Hyundai Verna XXi ABS (Petrol)	2009	340000	214000	Petrol	Individual	Manual	Second Owner	13.9	1599	103.2	5
13	Renault KWID RXT	2015	210000	210000	Petrol	Individual	Manual	Second Owner	25.17	799	53.3	5
14	Maruti Alto LX	2000	108000	206000	Petrol	Individual	Manual	Fourth & Above Owner	19.7	796	46.3	5
15	Hyundai i10 Magna 1.1L	2010	187000	200400	Petrol	Individual	Manual	Second Owner	19.81	1086	68.05	5
16	Ford Fiesta 1.4 Duratec ZXI	2008	136000	200185	Petrol	Individual	Manual	First Owner	16.6	1388	68	5
17	Maruti Swift Dzire 1.2 Vxi BSIV	2010	210000	200000	Petrol	Individual	Manual	First Owner	17.5	1197	85.8	5
18	Maruti Zen Estilo VXI BSIV	2010	160000	200000	Petrol	Individual	Manual	First Owner	19	998	67.1	5
19	Honda CR-V 2.0L 2WD AT	2006	125000	200000	Petrol	Individual	Automatic	Third Owner	13.1	1997	141.1	5
20	Maruti Wagon R LX	2006	65000	198000	Petrol	Individual	Manual	Second Owner	18.9	998	67.1	5

Methodology

- **Data Collection:** Obtain the cars24 dataset from a reliable source.
- **Data Import:** Import the cleaned dataset into a SQL database for analysis.
- **SQL Query Development:** Develop SQL queries to extract insights from the dataset based on the analysis objectives. Use SELECT, GROUP BY, ORDER BY, and other SQL functions as needed.
- **Data Analysis:** Execute the SQL queries to analyze the dataset and extract relevant information, such as total car counts, sales distribution, average prices, and other metrics.
- **Insights Generation:** Interpret the analysis results to derive meaningful insights about the dataset, such as market trends, consumer preferences, and seller performance.

Insights :

- **Insight number 1**

select * from cars24

This query gives an overview of the full dataset of the cars24 table.

- **Insight number 2**

select count (*) from cars24

This query gives an overview of the total number of cars present in the cars24 table.

- **Insight number 3**

```
SELECT seller_type, COUNT (*) AS total_cars_sold
FROM CARS24
GROUP BY seller_type;
```

This query gives an overview of the distribution of car sales among different types of sellers.

- **Insight number 4**

```
SELECT year, COUNT (*) AS total_cars
FROM CARS24
GROUP BY year
ORDER BY year;
```

This query provides insights into the total number of cars manufactured per year.

- **Insight number 5**

```
SELECT year, fuel, AVG(selling_price) AS avg_price_car
FROM CARS24
GROUP BY year, fuel
ORDER BY year, fuel;
```

This query provides insights into the average selling prices of cars considering both fuel type and seller type.

- **Insight number 6**

```
SELECT year, fuel, COUNT(*) AS total_cars_sold
FROM CARS24
GROUP BY year, fuel
ORDER BY year, fuel;
```

This query gives a breakdown of the total number of cars sold by both fuel type and manufacturing year.

- **Insight number 7**

```
SELECT transmission, seller_type, COUNT(*) AS total_cars_sold
FROM CARS24
GROUP BY transmission, seller_type;
```

This query gives insights into the distribution of car sales considering both transmission type and seller type.

- **Insight number 8**

```
SELECT owner, fuel, COUNT(*) AS cars_sold
FROM CARS24
GROUP BY owner, fuel
ORDER BY owner;
```

This query gives a breakdown of the total number of cars sold by both owner type and fuel type.

- **Insight number 9**

```
SELECT seller_type, MAX (max_power) AS max_power
FROM CARS24
GROUP BY seller_type;
```

This query helps in understanding the maximum power output of cars sold by different types of sellers.

- **Insight number 10**

```
SELECT name, COUNT (*) AS total_sold

FROM CARS24

GROUP BY name

ORDER BY total_sold DESC

LIMIT 5;
```

This query lists the top 5 most common car models sold based on the number of units.

Conclusion:

1. **Total Cars:** The dataset contains information on a total of number of cars.
2. **Seller Type Distribution:** Cars are sold by three types of sellers: individual, dealer, and Trustmark dealer.
3. **Manufacturing Year Analysis:** The dataset includes cars manufactured across multiple years.
4. **Average Selling Price:** The average selling price of cars varies by manufacturing year and fuel type.
5. **Fuel Type Analysis:** The dataset includes cars using different fuel types.
6. **Transmission Type Analysis:** Cars are sold with both manual and automatic transmissions
7. **Owner Type Analysis:** Cars are sold with varying numbers of previous owners.
8. **Maximum Power Output:** The maximum power output of cars varies by seller type
9. **Top Selling Car Models:** The top 5 most common car models sold, based on the number of units,
10. **Fuel and Transmission Type Analysis:** The distribution of car sales by both fuel type and transmission.

Implications of Findings:

1. **Consumer Preferences:** Insights into consumer preferences, such as fuel type and transmission type choices, can help manufacturers tailor their offerings to better meet consumer demand.
2. **Pricing Strategy:** Average selling prices by manufacturing year and fuel type can inform pricing strategies, helping sellers set competitive prices based on market trends.
3. **Seller Performance:** Analysis of car sales by seller type can help evaluate the performance of different types of sellers and guide marketing and sales efforts.
4. **Product Development:** Understanding the maximum power output preferred by buyers can guide product development efforts, helping manufacturers design cars that meet consumer expectations.
5. **Inventory Management:** Knowing the total number of cars manufactured per year can assist in inventory management, ensuring that dealers have the right mix of cars to meet demand.
6. **Marketing and Sales:** Insights into the top-selling car models can inform marketing and sales strategies, helping dealers focus their efforts on promoting the most popular models.

Key Takeaways:

1. **Market Understanding:** A comprehensive analysis of the dataset provides valuable insights into market trends, consumer preferences, and seller performance in the automotive industry.
2. **Sales Distribution:** The distribution of car sales among different seller types highlights the importance of understanding the dynamics of individual sellers, dealers, and Trustmark dealers in the market.
3. **Product Performance:** Analyzing the total number of cars manufactured per year indicates how different car models have performed over the years in terms of production and sales.
4. **Price Analysis:** The average selling prices of cars by manufacturing year and fuel type provide insights into pricing strategies and consumer behavior related to pricing.
5. **Competitive Insights:** Identifying the top-selling car models provides competitive insights that can be used to develop marketing and sales strategies to attract more customers.
6. **Future Planning:** The analysis provides a foundation for future planning, such as product development, inventory management, and marketing strategies, based on the trends and patterns observed in the dataset.

limitations:

1. **Data Quality:** The accuracy and completeness of the dataset may vary, affecting the reliability of the analysis results. Missing or incorrect data could lead to biased conclusions.
2. **Sample Size:** The dataset may not represent the entire market or population, limiting the generalizability of the findings. A larger sample size could provide more representative insights.
3. **Scope:** The analysis is based on the variables included in the dataset. Other factors that could impact car sales or market trends, such as economic conditions or consumer preferences, are not considered.
4. **Data Preprocessing:** The analysis may be influenced by the preprocessing steps applied to the dataset, such as data cleaning and transformation, which could introduce errors or biases.
5. **Time Frame:** The analysis is based on data from a specific time frame, and trends or patterns observed may not apply to other periods.
6. **External Factors:** External factors not captured in the dataset, such as changes in government regulations or industry trends, could impact the analysis results.

Future Work:

- **Data Visualization:** Create interactive data visualizations to present analysis findings more engagingly and understandably.
- **Real-time Data Analysis:** Implement real-time data analysis capabilities to monitor market trends and make timely decisions.
- **Benchmarking:** Compare the performance of cars24 with competitors or industry benchmarks to identify areas for improvement and strategic opportunities.

THANK YOU……..