Contents

1. [**Project Title:** 2](#_Toc203736318)
2. [**Introduction:** 2](#_Toc203736319)
3. [**Objective:** 2](#_Toc203736320)
4. [**Project Process:** 2](#_Toc203736321)
5. [**Tools and Technologies:** 3](#_Toc203736322)
6. [**Data Collection:** 4](#_Toc203736323)
7. [**Data Cleaning and Preprocessing:** 7](#_Toc203736324)
8. [**Key Matrices Analyzed:** 8](#_Toc203736325)
9. [**Visualization dashboard:** 11](#_Toc203736326)
10. [**Explanation of the dashboard:** 12](#_Toc203736327)
11. [**Findings and Insights:** 12](#_Toc203736328)
12. [**Conclusion:** 14](#_Toc203736329)

# **Project Title:**

A1 Motors Car Sales Analysis Project.

## **Introduction:**

My name is P. Vignesh, and I am a recent B-Tech graduate in Computer Science and Engineering with a strong interest in data analysis and business intelligence.

As part of my learning journey, I worked on a project for a car dealership company called A1 Motors, where I took the role of a Data Analyst.

## **Objective:**

In the car sales industry, data is stored in different files like car details, owner information, sales records, insurance data, and service history.

But when this data is scattered, it's hard to understand overall business performance.

The goal of this project is to combine all these files and analyze them to find useful insights, such as which cars are selling the most, identify top 5 models, and find sales of the year etc.

This project helps in understanding customer behavior, improving sales, and making better business decisions.

# **Project Process:**

**Step 1: Understanding the problem:**

First, I understood the business goal to analyze car sales, service history, ownership, and insurance data to find useful insights for A1 Motors.

**Step 2: Data Loading into SQL server Using ETL process:**

To manage and process the data efficiently, I applied the ETL (Extract, Transform, Load) process:

* Extracted the CSV files from local storage
* Transformed the data (ensured proper formatting and structure)
* Loaded the data into SQL Server Management Studio (SSMS)
* This allowed me to centralize all data and prepare it for structured querying and integration.

**Step 3: Create a Master Table Using SQL Joins:**

Once all datasets were loaded into SSMS, I used SQL joins to merge them into a single master table using keys like Car ID and Owner ID. This helped consolidate all critical business data in one view.

**Step 4: Data Cleaning & Insight Extraction:**

I performed data cleaning operations in SQL, such as:

* Removing duplicates and null values.
* Formatting salary column using round function.
* Filtering out irrelevant or incorrect data. Then, I wrote SQL queries to extract insights like:
* Top-selling cars.
* Cities with the highest sales.
* How much car sales in a month.
* Display sales a year.

**Step 5: Visual Dashboard in Power BI:**

I imported the master table into Power BI to create an interactive dashboard. The dashboard included:

* Display Total Cars.
* Display total fuel types.
* Display total models.
* Display Difference between transmissions.

**Step 6: Automation using Python & Task Scheduler:**

To make the process efficient and repeatable, I automated the ETL (Extract, Transform, Load) pipeline using Python:

* Automated data extraction from source files
* Updated the SQL Server tables
* Refreshed the Power BI dataset  
  I scheduled the automation using Windows Task Scheduler to run at regular intervals, making the process fully automated.

# **Tools and Technologies:**

Throughout the A1 Motors Car Sales Analysis project, I used the following tools and technologies to handle data processing, analysis, visualization, and automation.

**Microsoft Excel:**

* Used for exploring raw CSV files.
* Initial review of data structure and values.

**SQL (SQL Server Management Studio):**

* Stored and managed all CSV data in table format.
* Performed joins, data cleaning, and complex queries.
* Created the master table for final analysis.

**Power BI:**

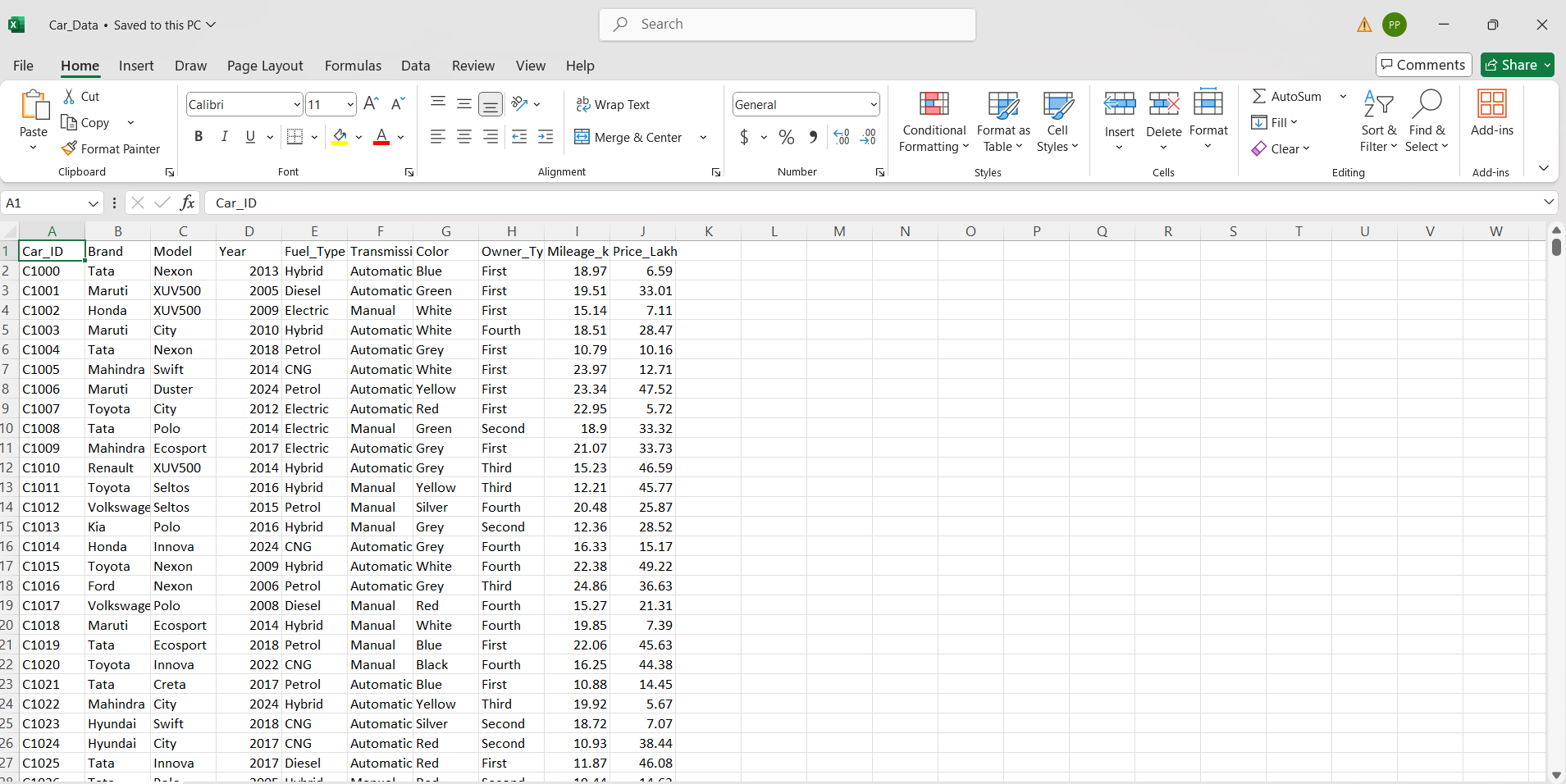
* Connected to the SQL master table.
* Built interactive dashboards and visual reports.
* Used charts and filters to highlight key business insights.

**Python:**

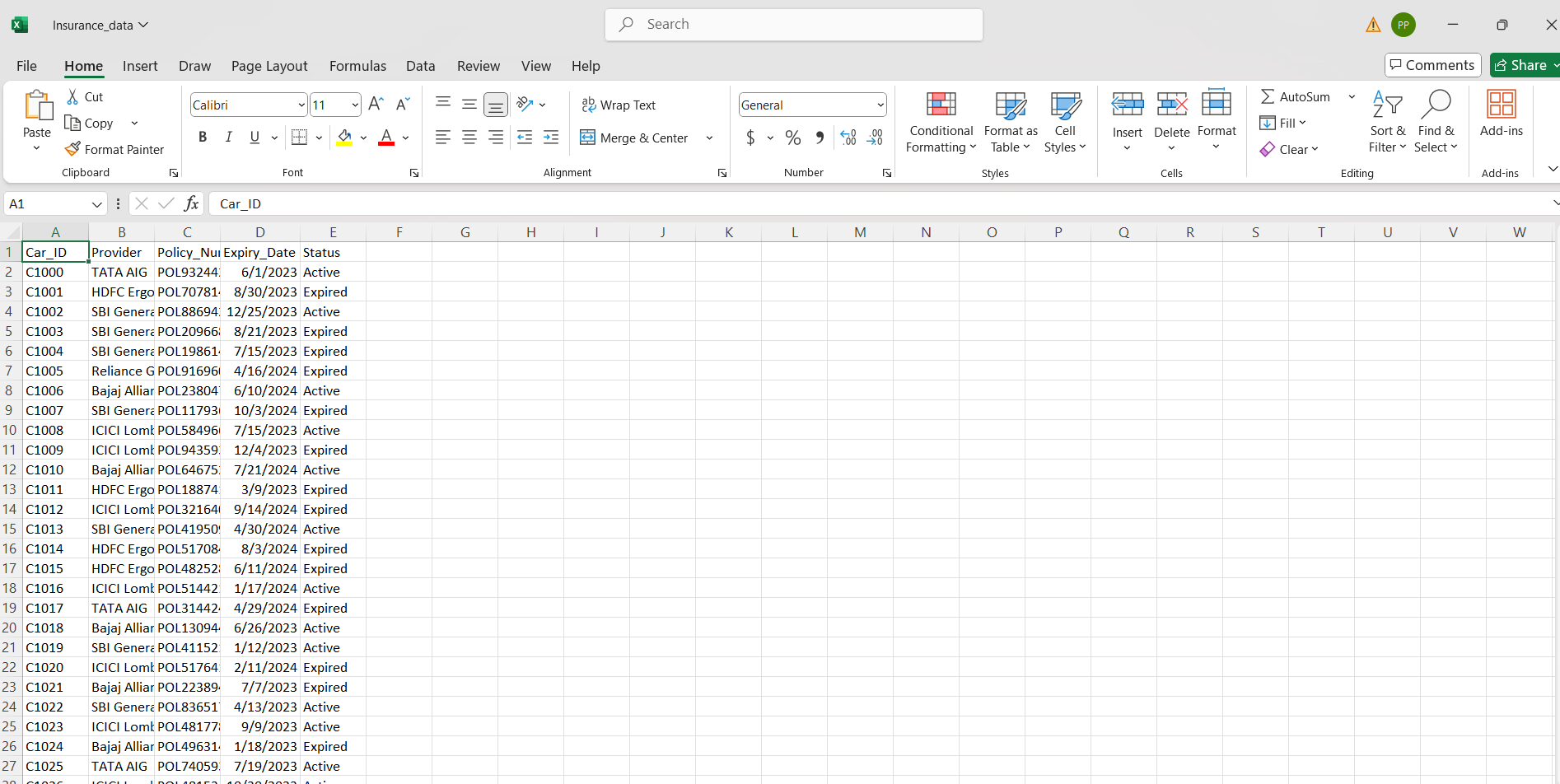
* Automated the ETL pipeline (data extraction, transformation, and loading)
* Refreshed SQL data and Power BI connection
* Scheduled automation using Windows Task Scheduler

# **Data Collection:**

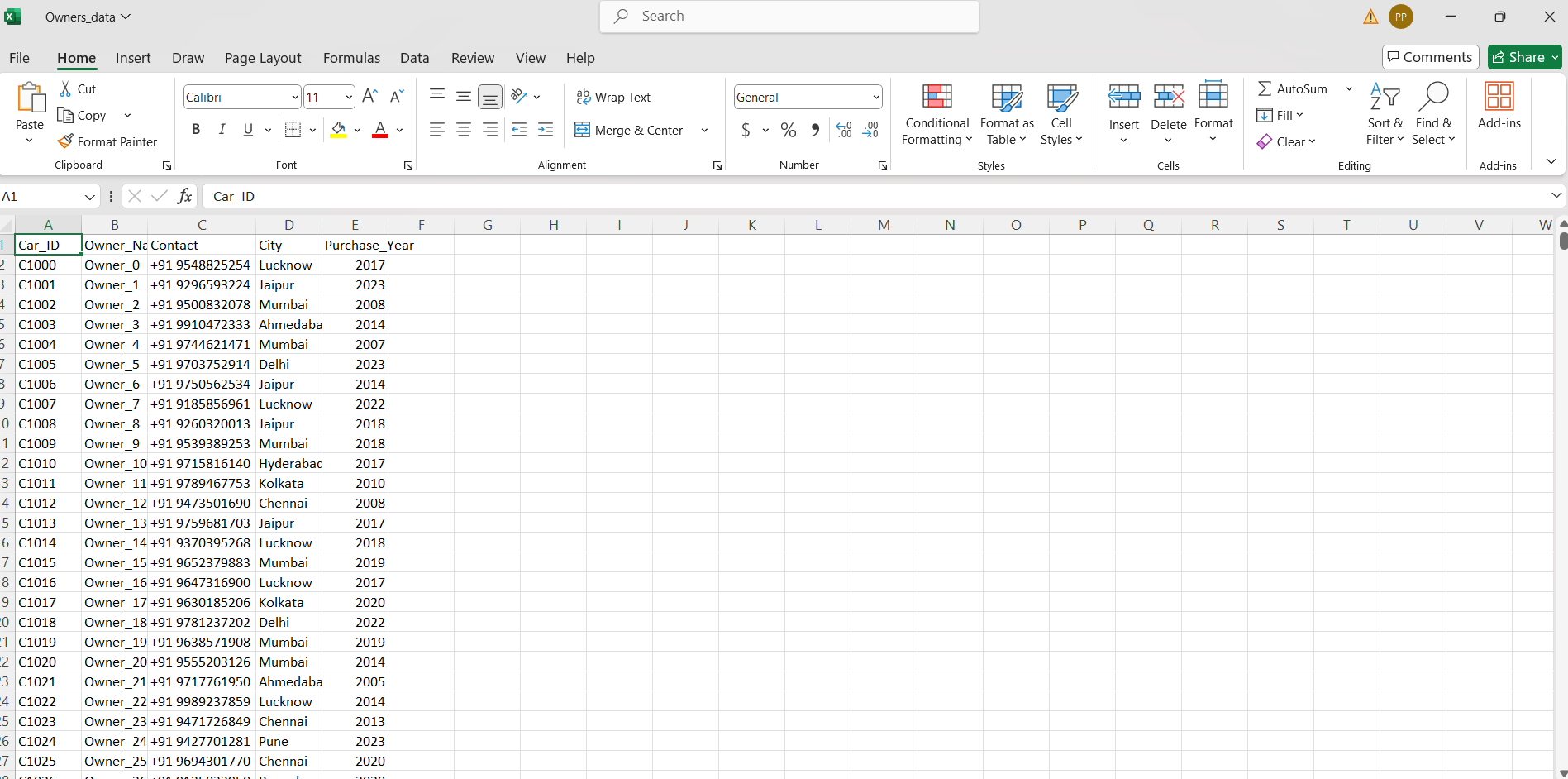
Car Data:



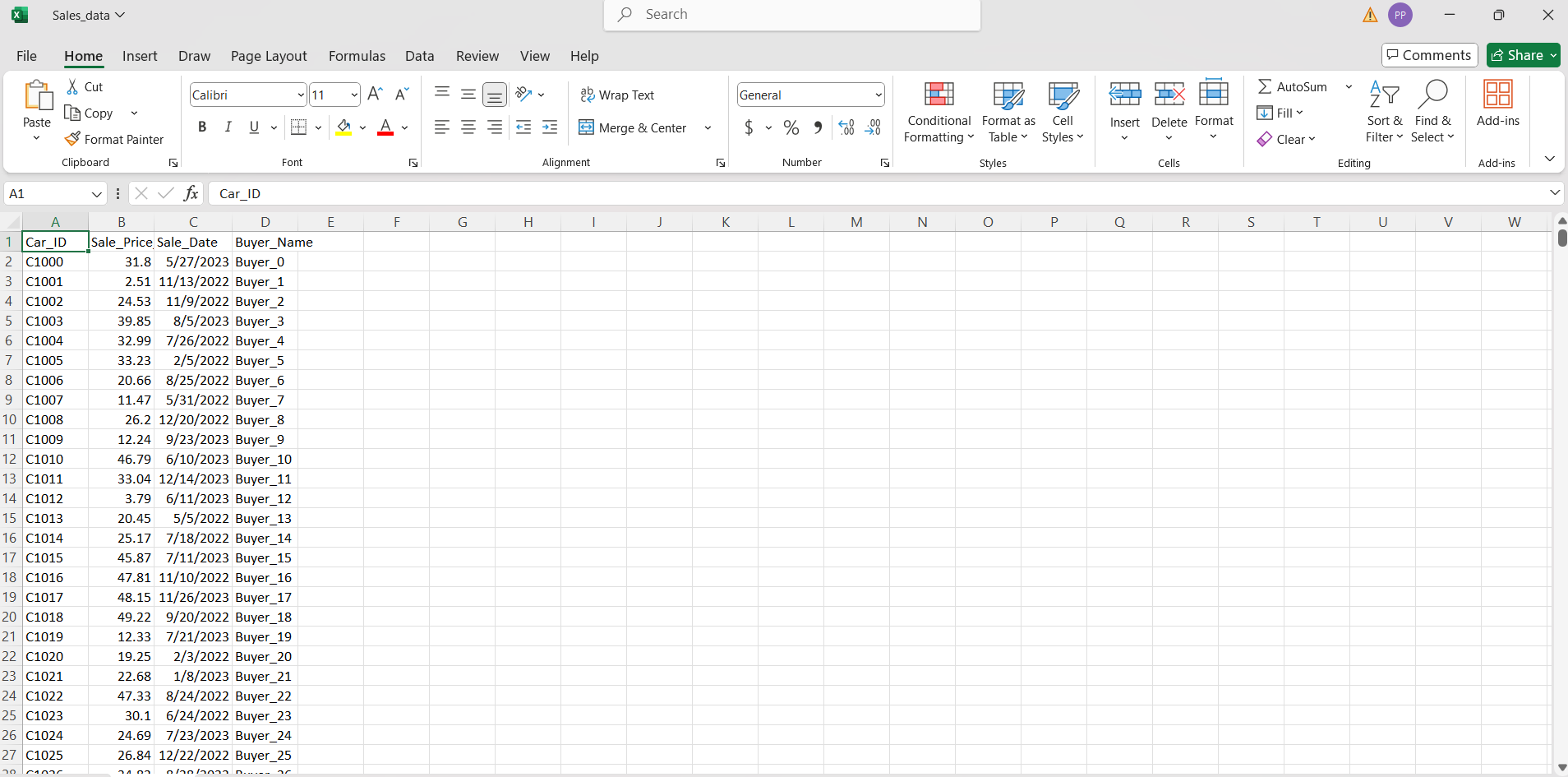
Insurance Data:



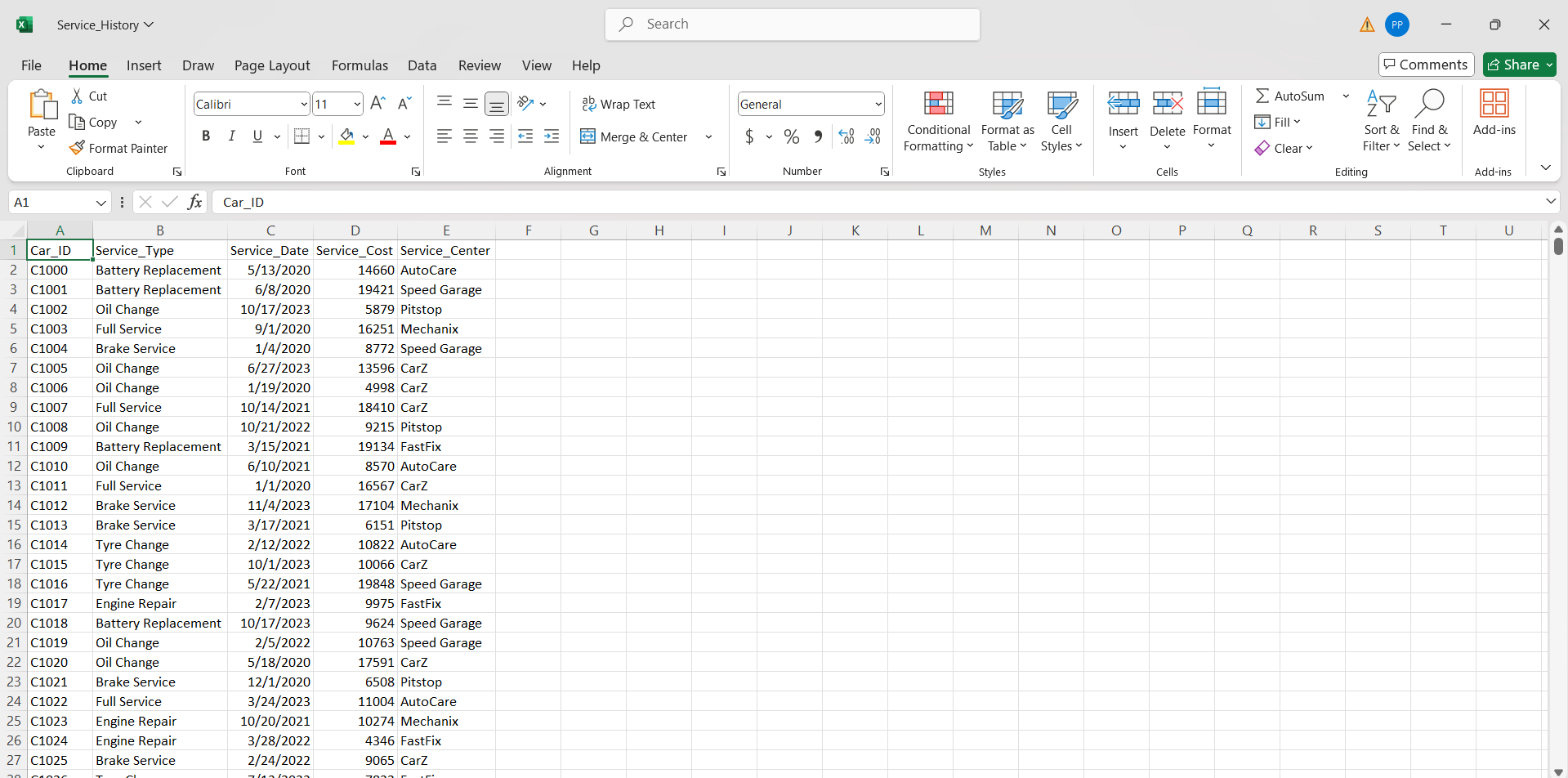
Owners Data



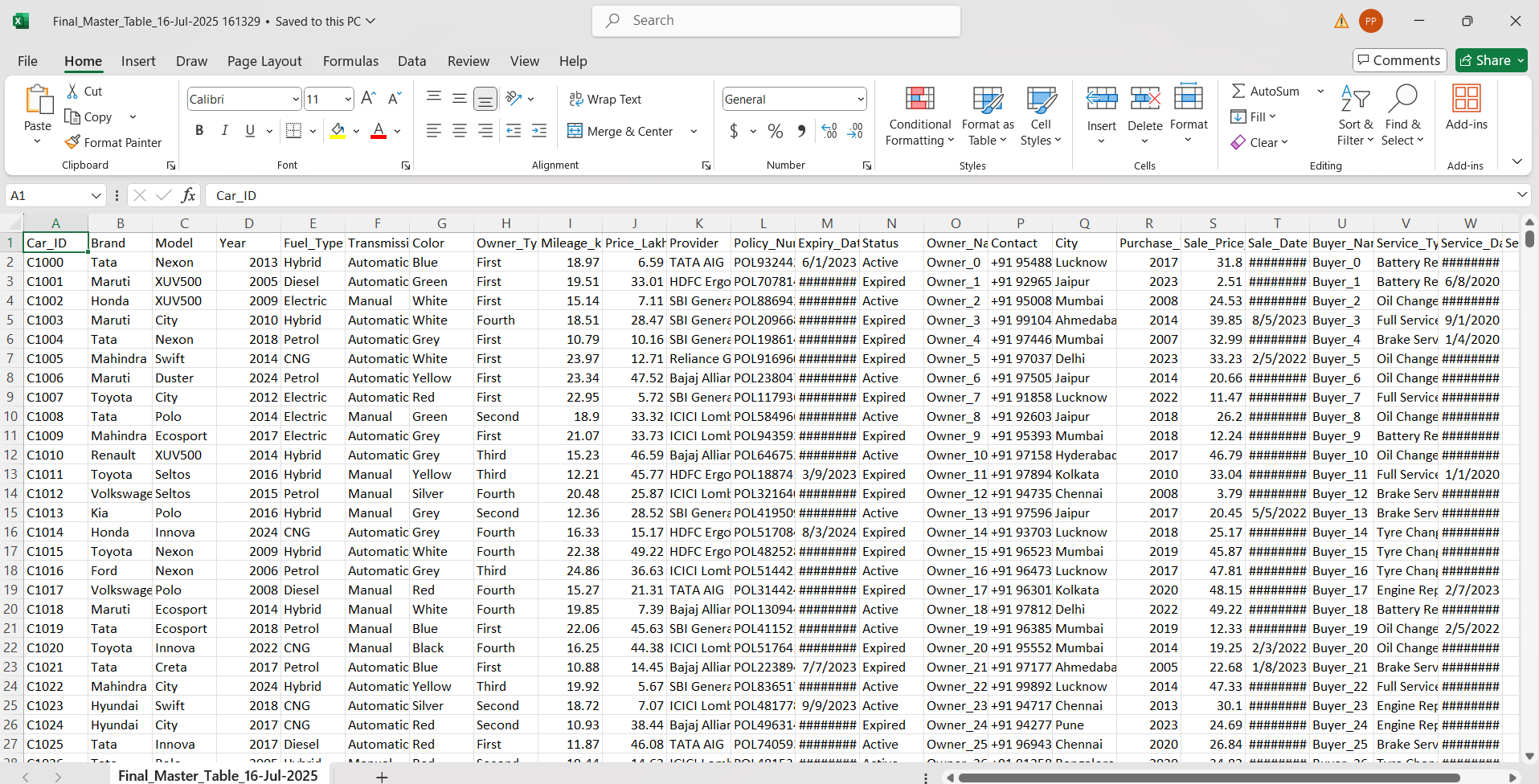
Sales Data:



Service History:



Final Master Table:



# **Data Cleaning and Preprocessing:**

**1.Checking for Duplicate Records:**

* After merging multiple datasets using joins, I identified rows that had repeated values (especially based on Car ID and Owner ID).
* I used SQL queries to detect and remove these duplicates to avoid counting the same entry multiple times.

**2. Handling Missing (Null) Values:**

* I checked columns like Sale\_price, price\_lakh for NULL values.
* In some cases, I removed the rows, while in others, I filled them based on logic or left them as-is if not critical.

**3.Transforming Column Values:**

* I used SQL's ROUND() function to round off prices and sales for better readability.

# **Key Matrices Analyzed:**

**1.Remove distinct values**

with cte as (

select \*,

Row\_Number() Over(partition by Car\_ID, Brand,Model,Year,Fuel\_type,Transmission,color,Owner\_type,Mileage\_kmpl

order by(select null)

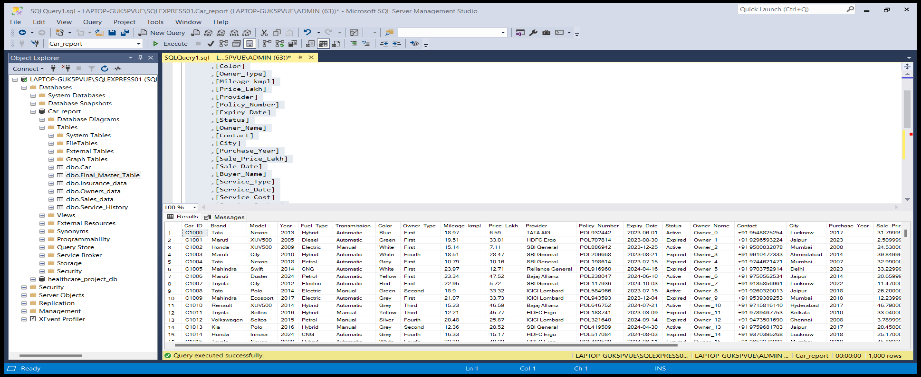
) as rn

from Final\_Master\_Table

)

delete from cte

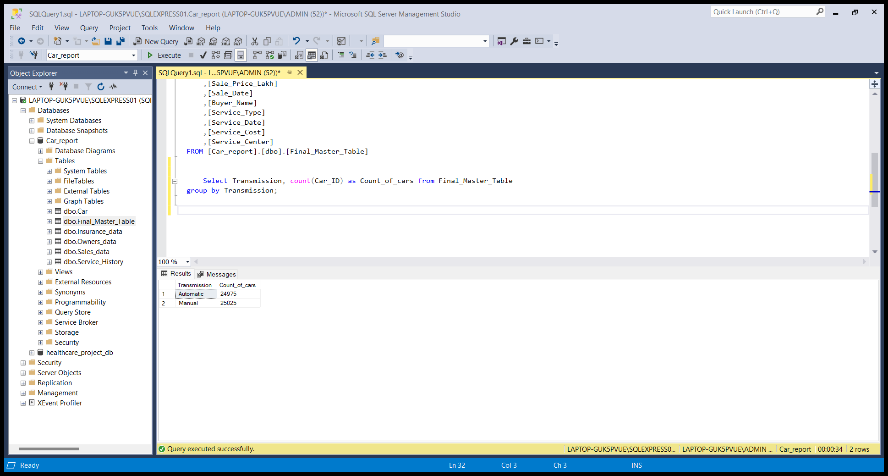
where rn > 1;



**2. Display the Count of cars group by transmission.**

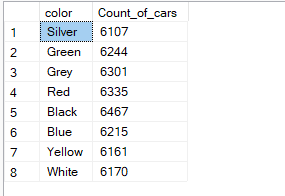
Select Transmission, count(Car\_ID) as Count\_of\_cars from Final\_Master\_Table

group by Transmission;



**3. Display count of cars basis on the colors**

Select color, count(Car\_ID) as Count\_of\_cars from Final\_Master\_Table  
group by color;



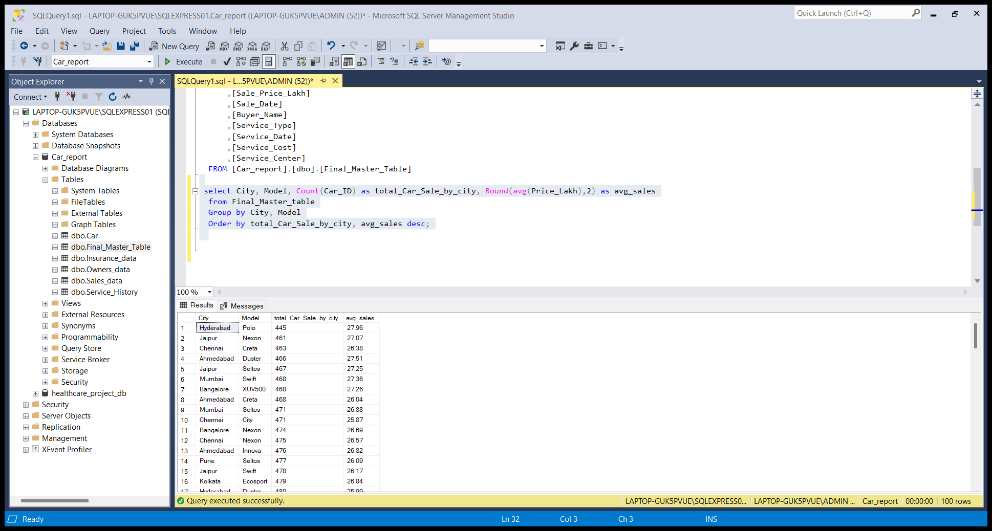
**4. Display average sales of cars group by city, model.**

select City, Model, Count(Car\_ID) as total\_Car\_Sale\_by\_city, Round(avg(Price\_Lakh),2) as avg\_sales

from Final\_Master\_table

Group by City, Model

Order by total\_Car\_Sale\_by\_city, avg\_sales desc;

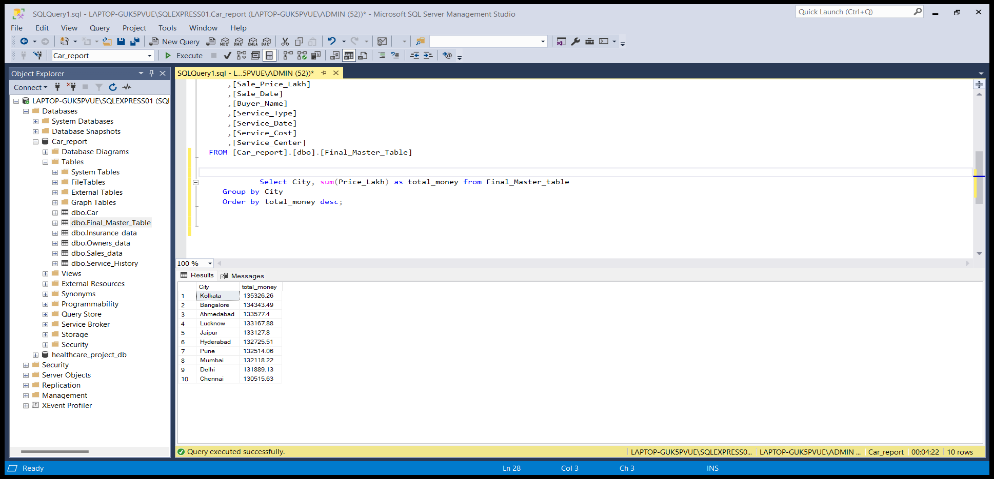


**5. find total sales of a cars group by city**

Select City, sum(Price\_Lakh) as total\_money from Final\_Master\_table

Group by City

Order by total\_money desc;

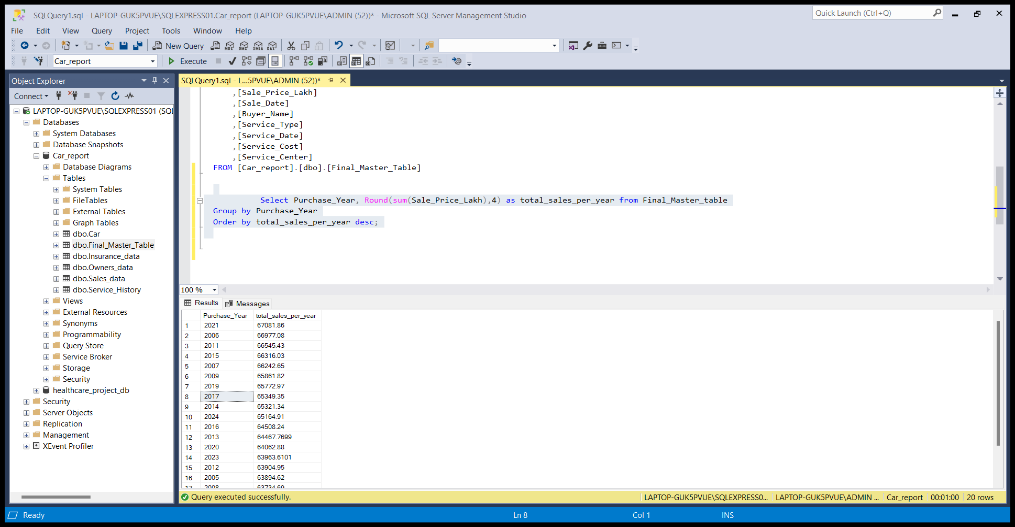


**6. Find total sales group by year.**

Select Purchase\_Year, Round(sum(Sale\_Price\_Lakh),4) as total\_sales\_per\_year from Final\_Master\_table

Group by Purchase\_Year

Order by total\_sales\_per\_year desc;



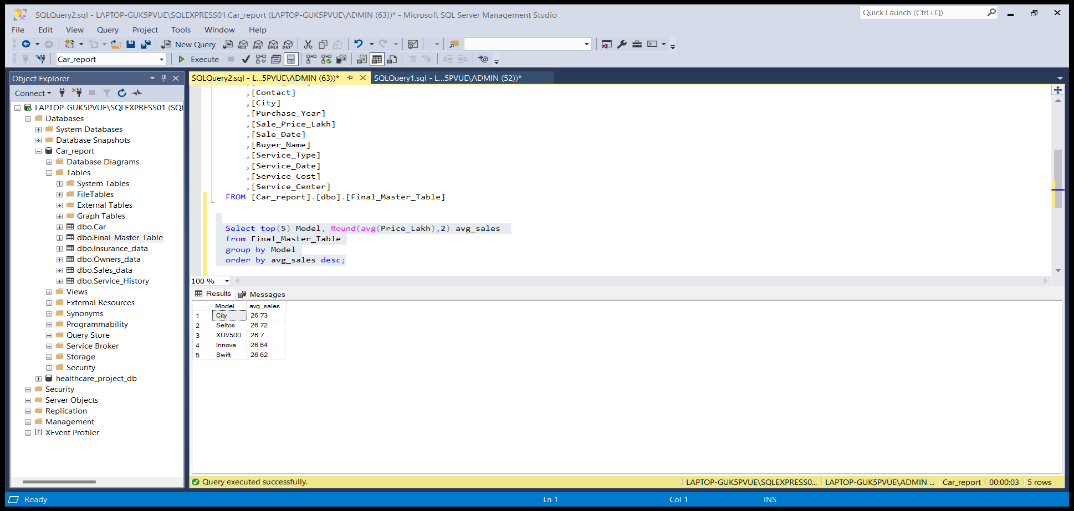
**7. average sales by top 5 model cars**

Select top(5) Model, Round(avg(Price\_Lakh),2) avg\_sales

from Final\_Master\_Table

group by Model

order by avg\_sales desc;

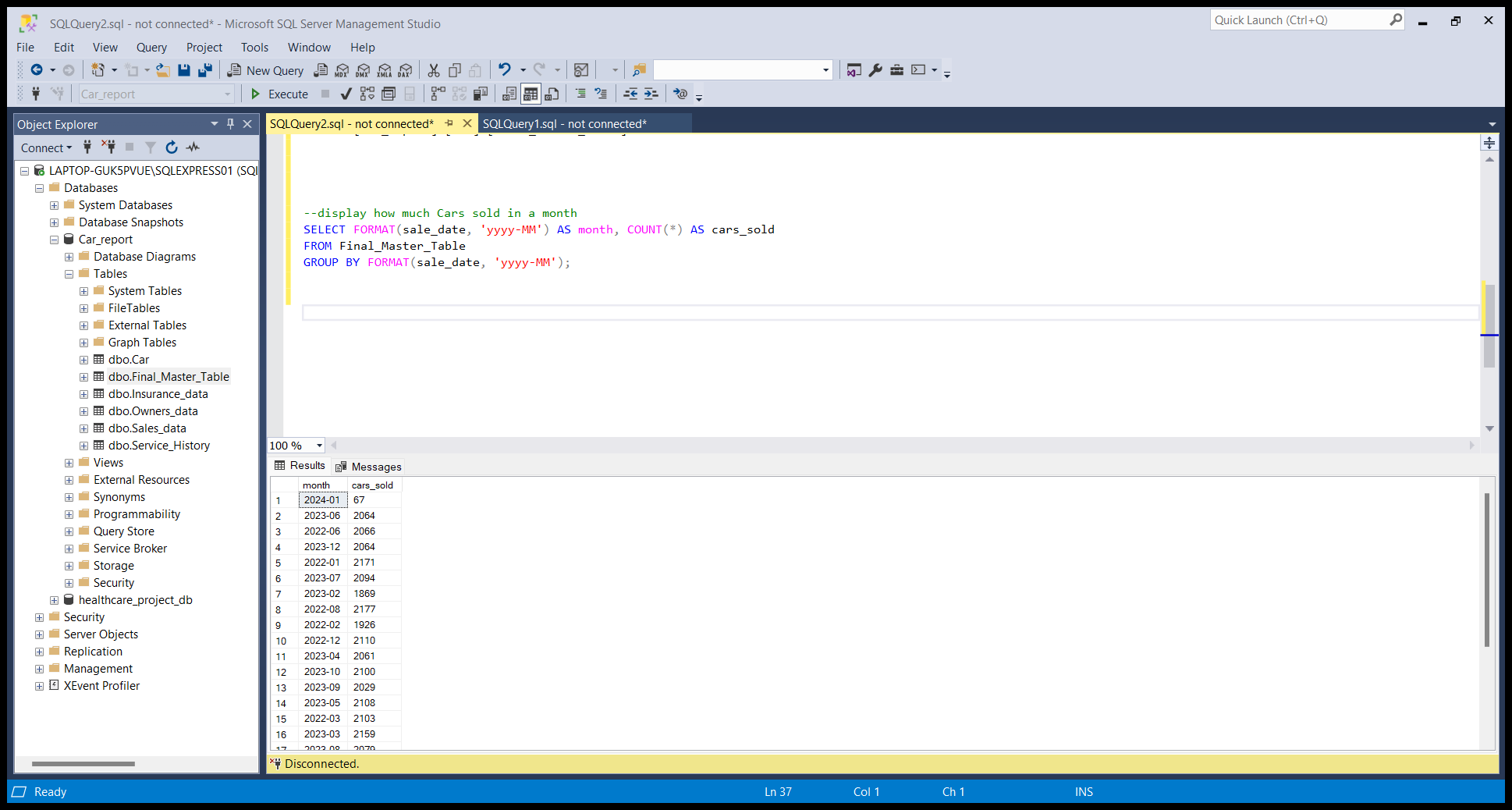


**8. Display how much Cars sold in a month**

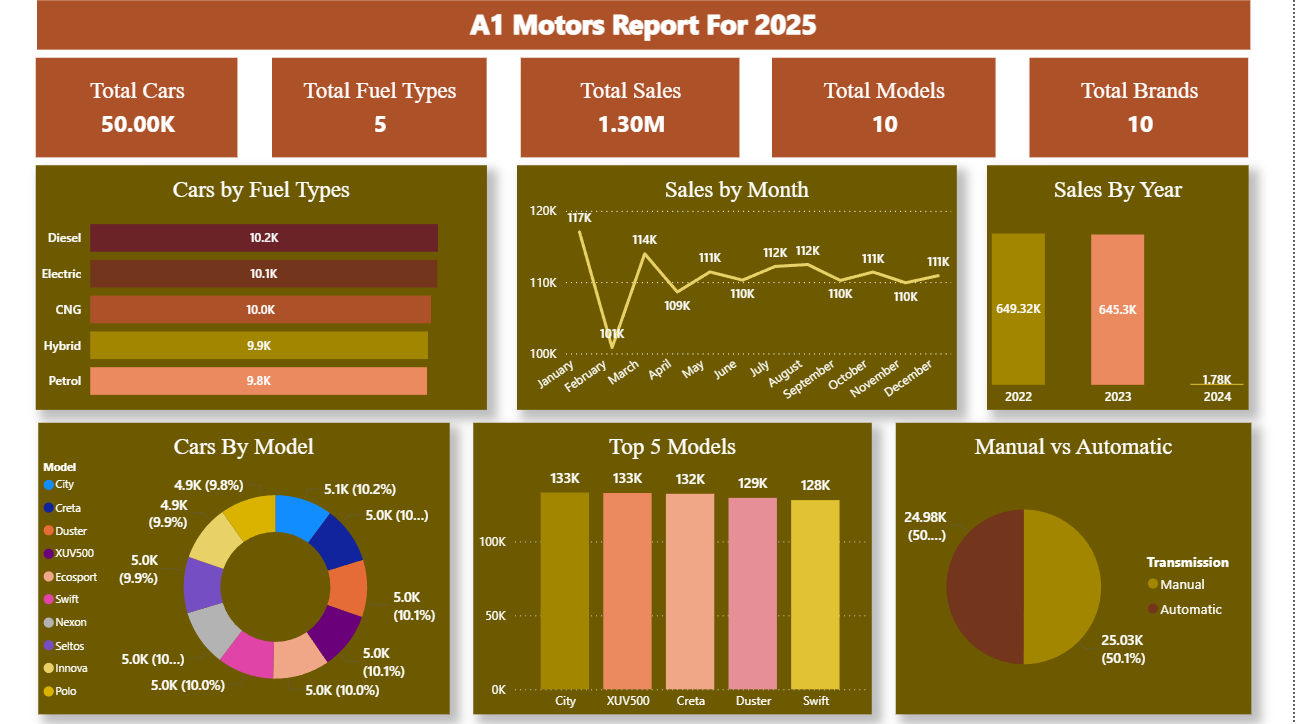
SELECT FORMAT(sale\_date, 'yyyy-MM') AS month, COUNT(\*) AS cars\_sold

FROM Final\_Master\_Table

GROUP BY FORMAT(sale\_date, 'yyyy-MM');



# **Visualization dashboard:**

****

## **Explanation of the dashboard:**

First display the count of cars, fuel types, sales, total models and total brands using title cards.

1. **Cars by Fuel Type (Bar Chart):**

* Visualizes the distribution of cars based on fuel type.
* Diesel (10.2K) and Electric (10.1K) dominate the list, followed closely by CNG, Hybrid, and Petrol.
* Indicates a balanced fuel type inventory, showing the company is adapting to electric trends.

1. **Sales by Month (Line Chart):**

* Shows monthly sales trends for the year.
* January (117K) and March (114K) recorded higher sales.
* February (101K) had the lowest, possibly due to seasonal effects.
* Sales remain stable from June to December (~110K–112K range).

1. **Sales by Year (Bar Chart):**

* 2022 and 2023 contributed the highest sales (649.32K and 645.3K respectively).
* 2024 has a very low count (1.78K), possibly due to data not being fully updated or ongoing.

1. **Top 5 Models (Bar Chart):**

* Shows the five most sold car models:
  + City and XUV500 top the chart with 133K sales each.
  + Followed by Creta (132K), Duster (129K), and Swift (128K).
* Helps in identifying high-performing products.

1. **Manual vs Automatic (Pie Chart):**

* Splits cars by transmission type:
* Manual – 24.98K (50%)
* Automatic – 25.03K (50.1%)
* The split is almost equal, indicating a balanced offering to customers.

# **Findings and Insights:**

1. **Top Models Driving Sales:**

**Insight:** Certain car models significantly outperform others in sales volume.

**Evidence:** City and XUV 500 both are sold in 133k units, followed closely by creta (132k), duster (129k) and swift (128k).

**Recommendation:** Focus marketing campaigns and inventory planning around these top 5 models to maximize revenue and meet demand.

1. **Fuel Type Distribution:**

**Insight:** Sales are evenly distributed across different fuel types.

**Evidence:** Diesel (10.2K), Electric (10.1K), CNG (10K), Hybrid (9.9K), and Petrol (9.8K) vehicles have very similar sales figures.

**Recommendation:** Maintain every types of fuels it will be helpful your customers. As well as increasing your electric vehicle stocks because its going trends on future.

1. **Monthly Sales Performance:**

**Insight:** Monthly sales are consistent, with slight fluctuations.

**Evidence:** Monthly sales between 109k to 117k throughout the year, but in the month of march there will be dip or loss.

**Recommendation:** As we can see march month facing a loss as compared to other months, due to external factors like budget month, working days etc. so analyze march data deeply and concentrate on marketing, promotions, advertisements , discounts etc.

**4. Manual vs Automatic Transmission Preference:**

**Insight:** Customers are almost equally divided between manual and automatic preferences.

**Evidence**: Manual: 24.98K (50%), Automatic: 25.03K (50.1%).

**Recommendation:** Maintain an even inventory of both transmission types. Introduce customer preference surveys to predict future demand.

**5. Sales by Year – Missing/Incomplete Data in 2024:**

**Insight:** There is a huge loss in 2024 as compared to 2022,2023.

**Evidence**: 2022: 649.3K, 2023: 645.3K, 2024: only 1.78K units.

**Recommendation:** Review the 2024 data carefully. If data is incomplete please be active or data is accurate take a action immediately.

**6. Sales by Year – Missing/Incomplete Data in 2024:**

**Insight:** All 10 models contribute relatively evenly to overall car distribution.

**Evidence**: Each model contributes roughly 9.8%–10.2% of total sales.

**Recommendation:** Maintain this product line, it shows business stability. Which model as average sales push towards the top selling categories using strategies.

# **Conclusion:**

The Car Sales Analysis Project for A1 Motors provided a comprehensive view of the company’s sales performance, customer preferences, and product distribution. By integrating and analyzing data from multiple sources—Car\_data, Insurance\_data, Owners\_data, Sales\_data, and Service\_history—the project delivered key business insights.

**The process involved:**

* Building a master dataset using ETL, SQL joins, and data cleaning,
* Extracting valuable insights using SQL queries, and
* Creating an interactive Power BI dashboard to visualize performance trends.

From the analysis:

* Top-selling models like City, XUV500, and Creta were identified,
* Sales across fuel types and transmission modes were well-balanced,
* Monthly sales remained stable with minor dips to improve,
* All models contributed evenly, ensuring a diversified and resilient sales portfolio.

In the final phase, the project was automated using Python scripting and Task Scheduler, enabling real-time updates and scalable reporting.

This end-to-end approach reflects how a Data Analyst can turn raw data into actionable insights and automate workflows for smarter, faster business decisions.