# PROJECT NEBULA

# Report ON CARDEKHO Used Cars Price Analysis





**Submitted by**

**Padala venkata sai**

**ACKNOWLEDGMENT**

I want to extend my sincere gratitude to the entire team at “Analytics Space” for granting me the opportunity to work on the “Used Car Price Analysis” project. I would also like to express my heartfelt thanks to my academic team, 'Data Trained,' for providing me with this enriching experience. Working on this project not only allowed me to enhance my analytical skills but also exposed me to a plethora of new insights. The invaluable suggestions and guidance from both teams played a pivotal role in the successful completion of this project, and I am truly grateful for the learning and growth it has brought to me.

**TABLE OF CONTENTS:**

**Chapter 1**

## **Introduction**

* 1. Project Objective
* Overview of Project Nebula
* Importance of Analyzing Used Car Data
* Importance of web scraping and data preprocessing and data visualization
  1. Business Objectives
* Business Problem Framing

1.3 Conceptual Background of the Domain Problem

1.4 Review of Literature

1.5 Motivation for the Problem Undertaken

**Chapter 2**

**Methodology and Analytical Framming problem**

2.1 Mathematical/ Analytical Modelling of the

2.2 Tools and Technologies

* 1. How to do Web Scraping with Selenium

2.4 Data Sources and their formats

2.5 How to do Data Preprocessing

* Filling null values
* Seperating data from merged columns
* Droping unwanted columns
* Renaming as the data headings as per needed
* Reordering the Formate of data as per needed
* Excell work

**Chapter 3**

**Data Analysis and Visualization**

3.1 Descriptive Statistics and Correlation Analysis

3.2 Observation of visualization graphs

3.3 Result and challanges

**Chapter 4**

**Conclusion**

* 1. In Scope and out of Scope
  2. Conclusion & Reference

Chapter 1. Introduction

# Project Objective

The objective of the "Used Cars Price Analysis" project is to conduct an in-depth analysis of the pricing variations among different car in the used car market. The project aims to identify the best cars with high quality available at competitive prices. This involves collecting data through web scraping from CarDekho, performing data cleaning and manipulation, conducting comprehensive visualizations and analyses on various factors like price, EMIs, fuel efficiency, and the distribution of cars from each manufacturer. The ultimate goal is to draw conclusions and insights that can assist in making informed decisions about purchasing used cars.

* **Overview of Project Nebula**

Discover the main objectives and scope of Project Nebula for web scraping and Data cleaning , Data Modelling , Data preprocessing In Python and data visualization in Tableau.

* **Importance of Analyzing Used Car Data**

Analyzing used car data is crucial for various reasons, and it can provide valuable insights for both consumers and businesses.

* **Importance of web scraping and data preprocessing and data visualization**

Explore why web scraping and data preprocessing and data visualization are critical for obtaining and understanding valuable data.

# 1.2 Business Objectives

You as a Data scientist are required to apply some data science techniques for the price of cars with the available independent variables. That should help the management to understand how exactly the prices vary with the independent variables. They can accordingly manipulate the design of the cars, the business strategy etc. to meet certain price levels.

* **Business Problem Framing**

The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But, due to the increased prices of new cars and the financial incapability of the customers to buy them, Used Car sales are on a global increase. With the covid 19 impact in the market, we have seen lot of changes in the car market. Now some cars are in demand hence making them costly and some are not in demand hence cheaper.

There is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. Existing System includes a process where a seller decides a price randomly and buyer has no idea about the car and its value in the present-day scenario. In fact, seller also has no idea about the car’s existing value or the price he should be selling the car at. To overcome this problem there is need to developed a model which will be highly effective to predict the actual price a car rather than the price range of a car

# 1.3 Conceptual Background of the Domain Problem

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle’s price on the market. The focus of this project is getting analysis of used car prices and the price of a used car based on its features, in order to make informed purchases.

Accurate car price Analysis involves expert knowledge, because price usually depends on many distinctive features and factors. Typically, most significant ones are brand and model, age. The fuel type used in the car as well as fuel consumption per mile highly affect price of a car due to a frequent change in the price of a fuel. Different features like exterior colour, door number, type of transmission, dimensions, safety, air condition, interior, whether it has navigation or not will also influence the car price. In this project, we applied different methods and techniques in order to achieve higher precision of the used car price analysis.

Here web scrapping and data analysis are used because they provide us with continuous value as an output and not a categorized value because of which it will be possible to analysis the actual price a car rather than the price range of a car.

The data associated with the investigation was very large because there are thousands of used cars and each car’s data comprises of values of many features. Both data gathering and analysis are complex. Used cars data scrape from [www.cardekho.com](http://www.cardekho.com/) which is a well-known online platform for reselling used and new cars in India. Features like car’s model, car name, Transmission type, fuel type and emi’s and price were include

Here we are trying to help the client works with small traders, who sell used cars to understand the price of the used cars by the used car price analysis. These would help the client/sellers to understand the used car market and accordingly they would be able to know the how many car have been selling the used car in the market.

# 1.4 Review of Literature

* Literature review covers relevant literature with the aim of gaining insight into the factors that are important to analysis the price of used cars in the market. In this study, we discuss various applications and methods which inspired us to build analysis about the price of used
* Cars in different locations. We did a background survey regarding the basic ideas our project and used those ideas for the collection of data information by doing web scraping from [www.cardekho.com](http://www.cardekho.com/) website which is a web platform where seller can sell their used car.

* This project is more about data exploration, feature engineering and pre- processing that can be done on this data. Since we scrape huge amount of data that includes more car related features, we can do better data exploration and also used to make the beautiful analysis.
* The goal of this project is to build an application which can analysis the used car prices with the help of other features. In the long term, this would allow people to better explain and reviewing their purchase with each other in this increasing digital world.

# 1.5 Motivation for the Problem Undertaken

* The project was the first provided to me by Analytics Space as a part of the internship programme. The exposure to real world data and the opportunity to deploy my skillset in solving a real time problem has been the primary motivation.

Data needed for this project is require to scrap from internet and work over it. Deciding whether a used car is worth the posted price when you see listings online can be difficult. The focus of this project on analysis the price of a used car based on its features, in order to make informed purchases. This study may help online web services that tells a used car’s market selling used car price value.

Chapter 2 Methodology and Analytical Framming

**2.1 Mathematical/ Analytical Modeling of the Problem**

* Selenium may encounter problem with websites that heavily rely on dynamic content loading, requiring you to implement mechanisms to wait for elements to load.
* website implement anti-scraping measures, such as CAPTCHAs or IP blocking. Strategies for handling these problem need to be incorporated into your web scraping process.
* Dealing with missing data is a common . Decide whether to impute missing values, remove incomplete records, or use other strategies based on the nature of the data
* Addressing inconsistencies, outliers, and errors in the scraped data is crucial for accurate analysis. This may involve standardizing formats, handling outliers, and correcting data entry errors.
* More some problems I have face and I overcome this problem and I have done webscraping and data preprocessing and visualization on tableau let us look it now how we done this project

**2.2 Tools and Technologies**

While taking up the project we should be familiar with the hardware and software required for the successful completion of the project. Here we need the following hardware and software.

**Hardware required**: -

1. Processor — core i5 and above
2. RAM — 8 GB or above
3. SSD — 250GB or above

**Software/s required**: -

* 1. (Python)Anaconda Navigator –Jupyter Notebook

**2.3 How to do Web Scraping with Selenium**

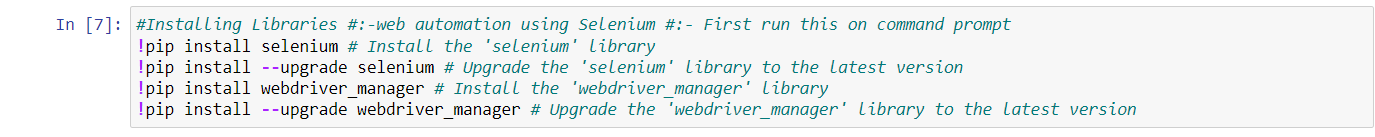
**FIRST STEP:-**

# Dowload Msedge driver through :-#1.Open this link :--https://developer.microsoft.com/en-us/microsoft-edge/tools/webdriver/

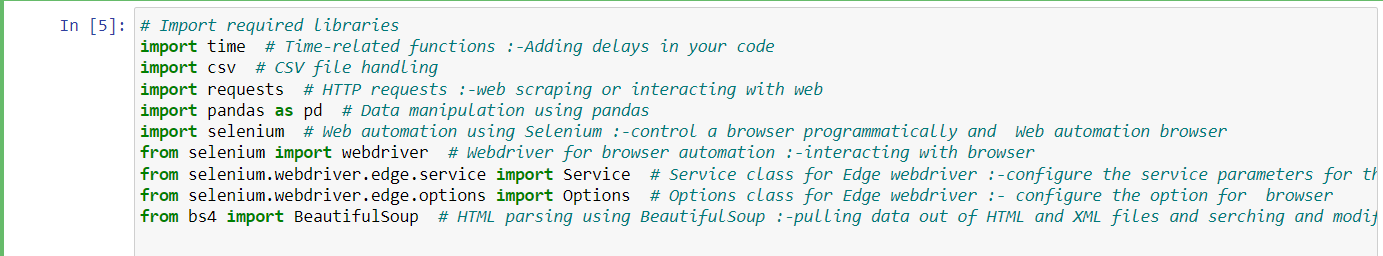
#2.Dowload edgedriver as per version --Version 120.0.2210.61 (Official build) (64-bit)

#3.After dowloding should extract excactly where your python main is located

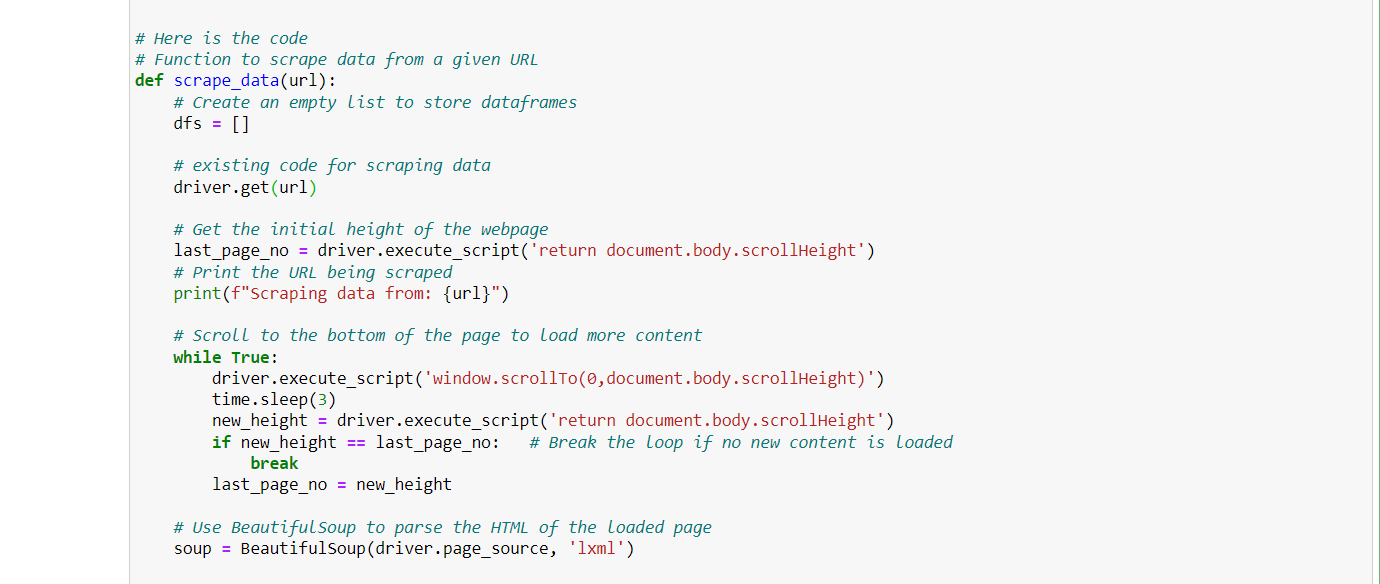
**SECOUND STEP:- # INSTALL THE LIBRARIES**

****

**THIRD STEP:- #IMPORT THE REQUIRED LIBRARIES**

****

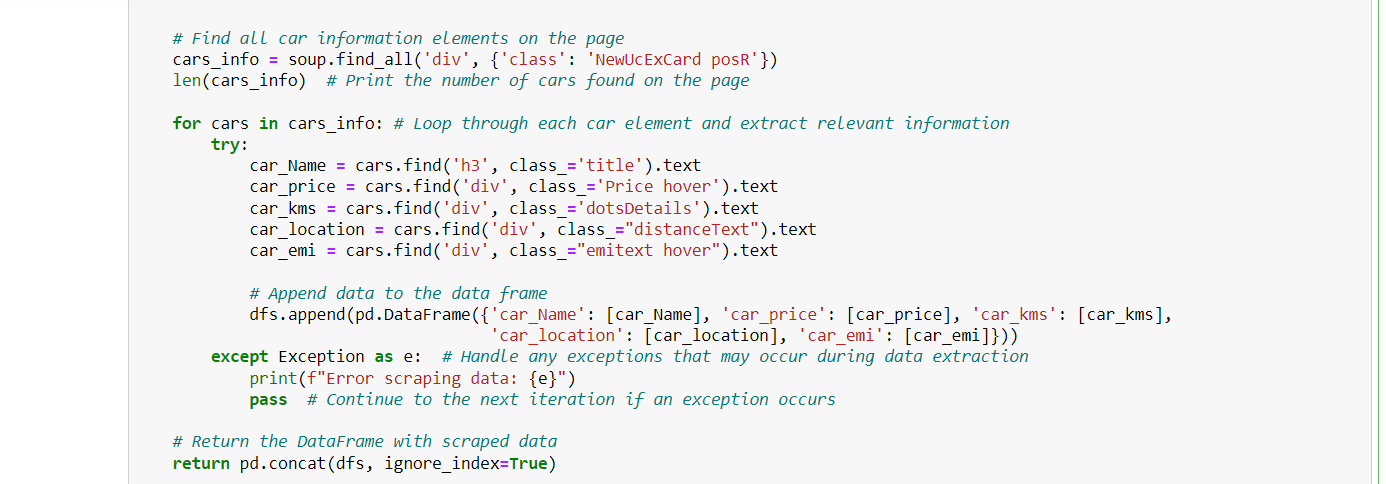
* **Time:** This library provides various time-related functions. In the context of your code, it mentions adding delays in your code, which is often used to control the pace of execution.
* **Csv:** This library is for handling CSV (Comma-Separated Values) files. It's commonly used for reading from and writing to CSV files, which are a popular format for tabular data.
* **Requests:** The requests library is used for making HTTP requests. In your code, it mentions HTTP requests for web scraping or interacting with the web.
* **Pandas:** This library is widely used for data manipulation and analysis. It provides data structures like DataFrames that are helpful for working with structured data.
* **Selenium:** Selenium is a powerful tool for web automation. It allows you to programmatically control a web browser, and it's often used for tasks like web scraping, testing, and browser automation.
* **Webdriver:** Specifically, the Edge webdriver is being imported from the selenium library. This is used to control the Edge browser for web automation.
* **Service and Options:** These are classes related to configuring the service parameters and options for the Edge webdriver, respectively.
* **BeautifulSoup:** This library is used for HTML parsing. It helps in pulling data out of HTML and XML files, making it easier to extract and manipulate information from web pages.

**FOURTH STEP :-WRITE THE MAIN CODE**

**STEP-1:-** This defines a function named **scrape\_data** that takes a URL as an argument. AND Initializes an empty list **dfs** to store DataFrames. It suggests that the code will collect data and store it in DataFrames.

**STEP-2:-Web Scraping Setup**:driver.get(url) Uses Selenium's **get** method to open the specified URL in the web browser controlled by the WebDriver (**driver**).

**STEP-3:-Scrolling to Load More Content.** Gets the initial height of the webpage.Scrolls to the bottom of the page, waits for 3 seconds (**time.sleep(3)**), and checks if new content has loaded. If no new content is loaded, the loop breaks.

**STEP-4:-** Uses **BeautifulSoup** to parse the HTML of the loaded page. driver.page\_source contains the HTML source of the current page. 

**STEP-5:-Finding Car Information Elements**:cars\_info = soup.find\_all('div', {'class': 'NewUcExCard posR'}) Uses BeautifulSoup's **find\_all** method to locate all **<div>** elements with the specified class (**'NewUcExCard posR'**). This is likely to be the class associated with the containers of individual car information on the page. And Prints the number of car information elements found on the page.next then

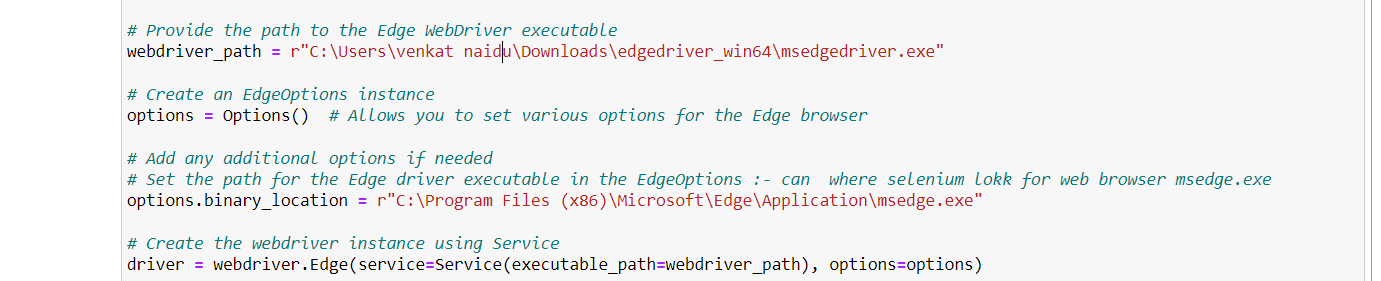
**Step-6:-Looping Through Each Car Element**:for cars in cars\_info: Iterates through each car information element in the list.

**Step -7:-Extracting Relevant Information:**For each car element, it uses the find method to locate specific elements (e.g., name, price, kilometers, location, EMI) within the HTML structure and extracts the text content.

**Step-8:-Appending Data to Data Frame:**Appends the extracted data for each car to the list of DataFrames (dfs). Each iteration creates a new DataFrame with the extracted information for a single car and appends it to the list.

**Step-9:-Handling Exceptions:**except Exception as e: **print(f"Error scraping data: {e}")** pass Handles exceptions that may occur during data extraction. If an exception occurs, it prints an error message and continues to the next iteration.

**Step-10:-Concatenating DataFrames**:**return pd.concat(dfs, ignore\_index=True)** Returns a single DataFrame by concatenating all the DataFrames in the list (**dfs**). The **ignore\_index=True** parameter resets the index of the resulting DataFrame.



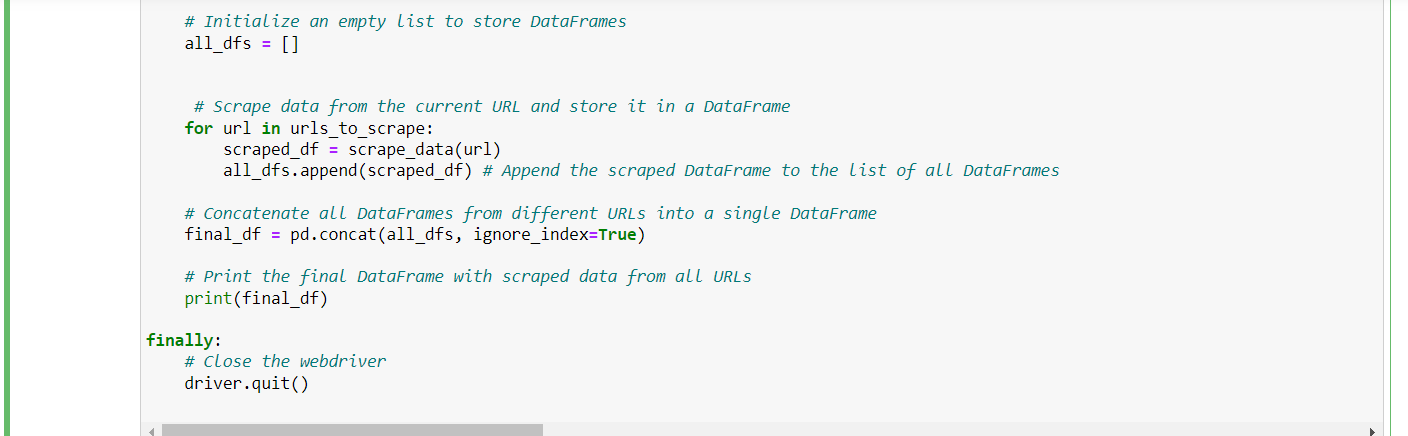
**Step-11:-**Specifies the file path to the Edge WebDriver executable. This is the location where Selenium will look for the WebDriver to control the Edge browser.

* options = Options() Creates an instance of **EdgeOptions**, which allows you to set various options for the Edge browser. This is where you can configure additional settings for the browser.
* options.binary\_location = r"C:\Program Files (x86)\Microsoft\Edge\Application\msedge.exe"
* Optionally sets the binary location for the Edge browser. This is the path to the actual msedge.exe file. If not set, Selenium will try to locate the browser automatically.
* driver = webdriver.Edge(service=Service(executable\_path=webdriver\_path), options=options)

Creates the WebDriver instance for the Edge browser. It uses the webdriver.Edge class, specifying the Service (which includes the path to the WebDriver executable) and any additional options set earlier.

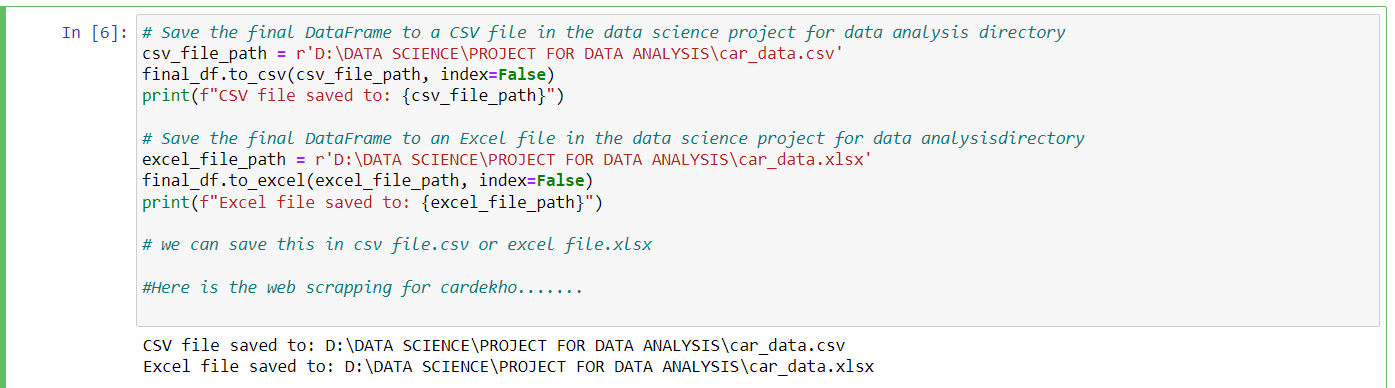
****

**Step-12:-** Enter the different location urls Used cars various filters such as make, price range, and more.

****

**Step-13:-** Initializes an empty list named **all\_dfs** to store DataFrames.

* Iterates through each URL in the **urls\_to\_scrape** list and uses the **scrape\_data** function to scrape data from the respective webpage. The scraped DataFrame for each URL is then appended to the **all\_dfs** list.
* Concatenates all DataFrames in the **all\_dfs** list into a single DataFrame named **final\_df.** The ignore\_index=True parameter resets the index of the resulting DataFrame.
* In a **finally block,** it ensures that the WebDriver (driver) is properly closed using the **quit()** method. This is important to release system resources and close the browser windows.



**Step-14:-**Finally save the data frame to csv or excel file

# 

# Output:-

# 2.4 Data Sources and their formats

The data was collected from the cardekho.com website in excel format. The data was scrapped using selenium. After scrapping ing required features the dataset is saved as an excel file.

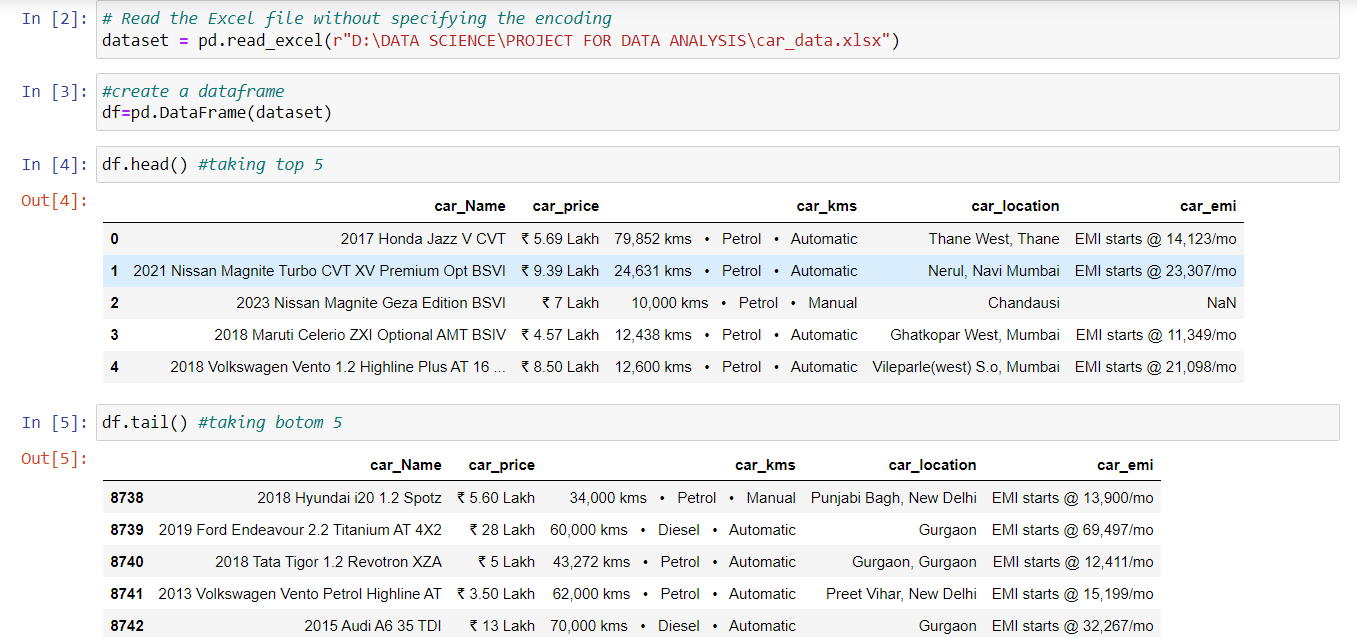
Also, my dataset was having 8742 rows and 6 columns including the target. I n this particular dataset, I have object type of data which has been changed as per our analysis of the dataset. The information about features is as follows.

**Features Information:**

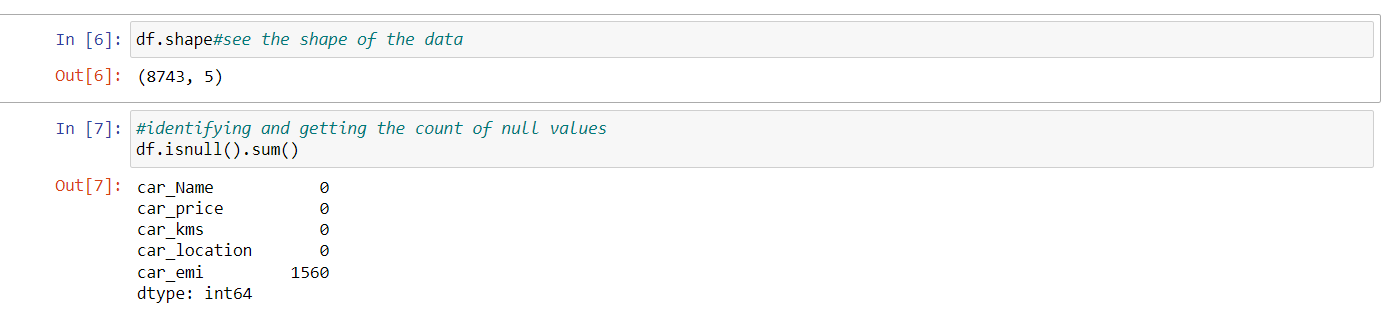
* + - Car\_Name: Name of the cars and their models
    - Car\_Fuel: Type of fuel used for car engine
    - Car\_kms: Car running in km till the date
    - Car\_year:-Age of the car
    - Car\_Transmission-type of transmission automatic or manual
    - Car\_Emi:- no of car has emi and no emis options

# 2.5 How to do Data Preprocessing

**STEP-1:-** Importing libraries

****

**STEP-2:-**Reading the excel file for data prepreocessing Creating the data frame and storing that dataset and df.head() and df.tail to check top 5 and bottom and

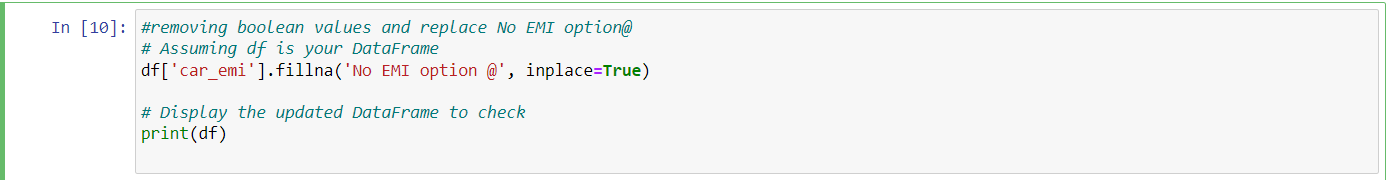
****

**STEP-3:-** Checking shape and null values are present here shape is (8742rows and 5columns) and there 1560 null values are present in emis

* **Filling null values**

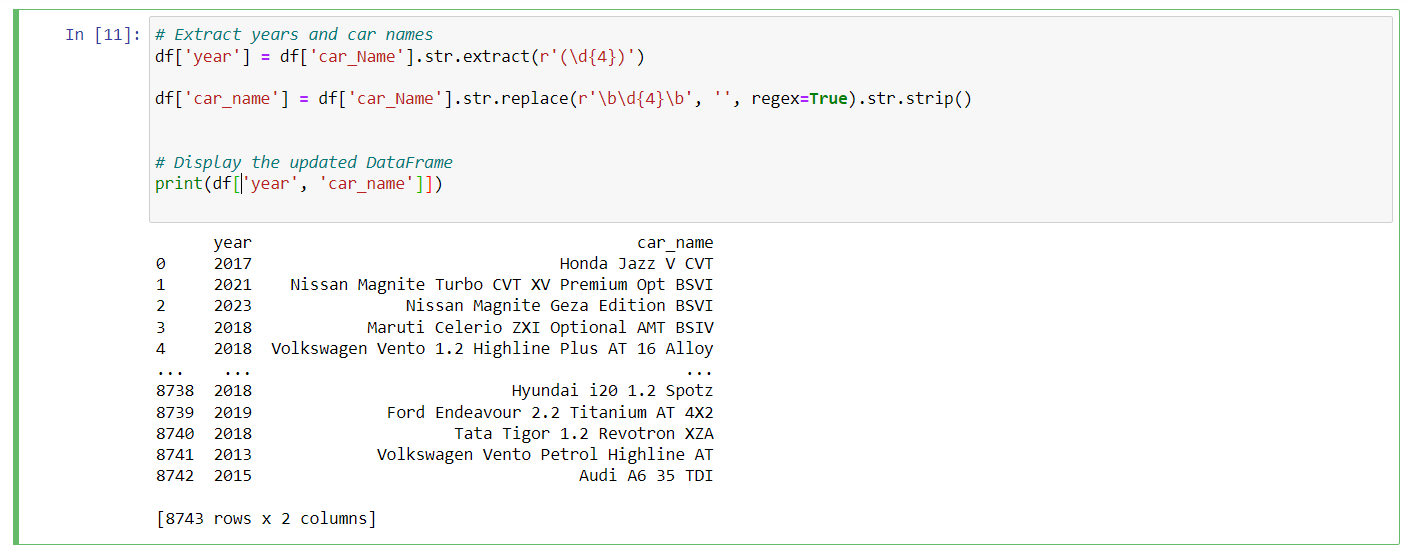


**STEP-4:-**Filling the Null vlues with no emi option @ but it showing Boolean Values again convert this to string



**STEP -5:-**removing bolean values and replacing the no emi option in strings for that we are writing inplace =true and we sucessfuly replaced it

* **Seperating data from merged columns**



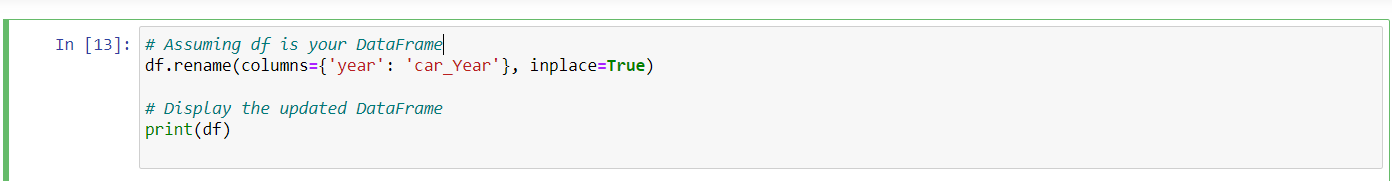
**STEP-6 :-**Firt line extracting year Here, you are creating a new column 'year' in your DataFrame (**df**) by extracting four consecutive digits from the 'car\_Name' column using a regular expression. This assumes that the year information in the 'car\_Name' follows the format of four digits.

* Second line **df['car\_name'] = df['car\_Name'].str.replace(r'\b\d{4}\b', '', regex=True).str.strip() This** line removes any four-digit numbers (presumably the year) from the 'car\_Name' column using **str.replace** with a regular expression. The **\b** represents word boundaries, ensuring that only whole four-digit numbers are replaced. After that, **str.strip()** is used to remove any leading or trailing whitespaces.
* **Droping unwanted columns**



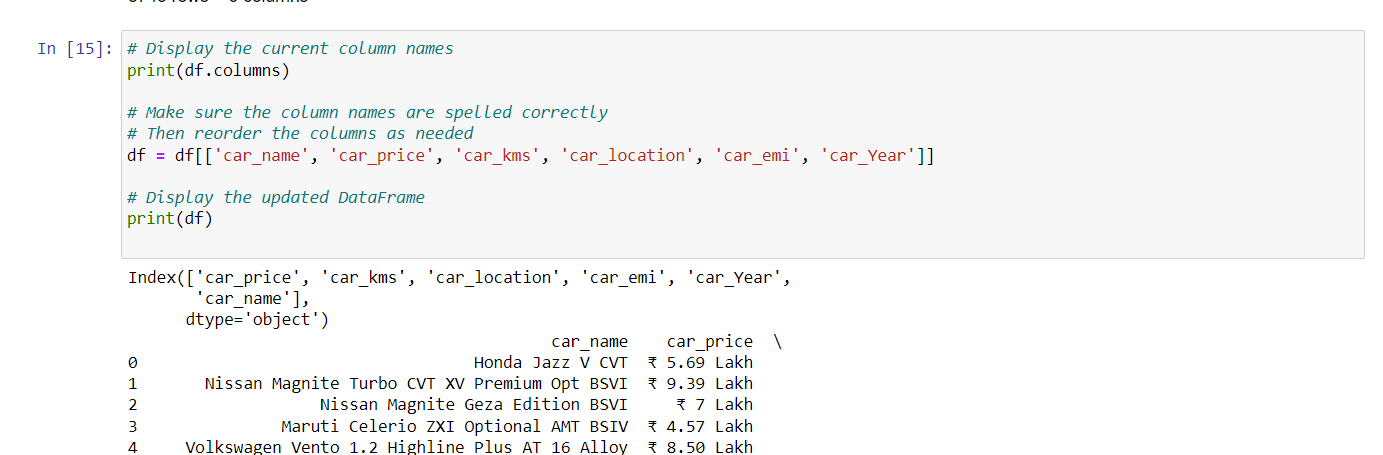
**STEP-7:-** deleting the car name because of it get double time while separating year

* **Renaming as the data headings as per needed**

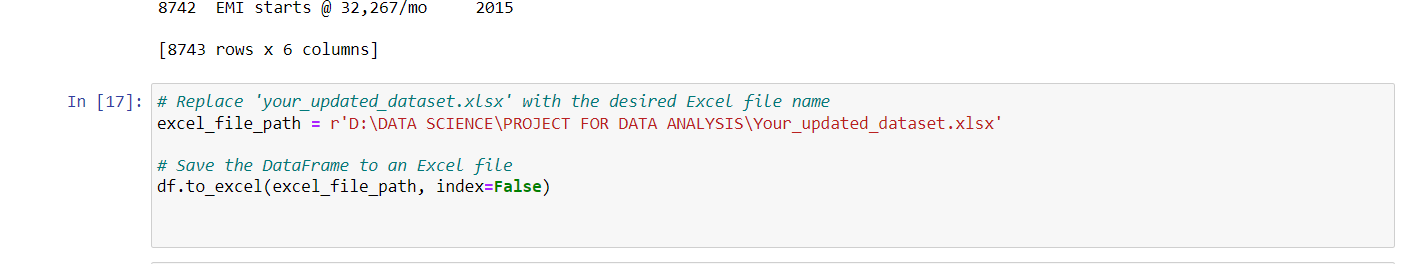


**STEP-8:-** Rename the heading name as per I needed

* **Reordering the Formate of data as per needed**

****

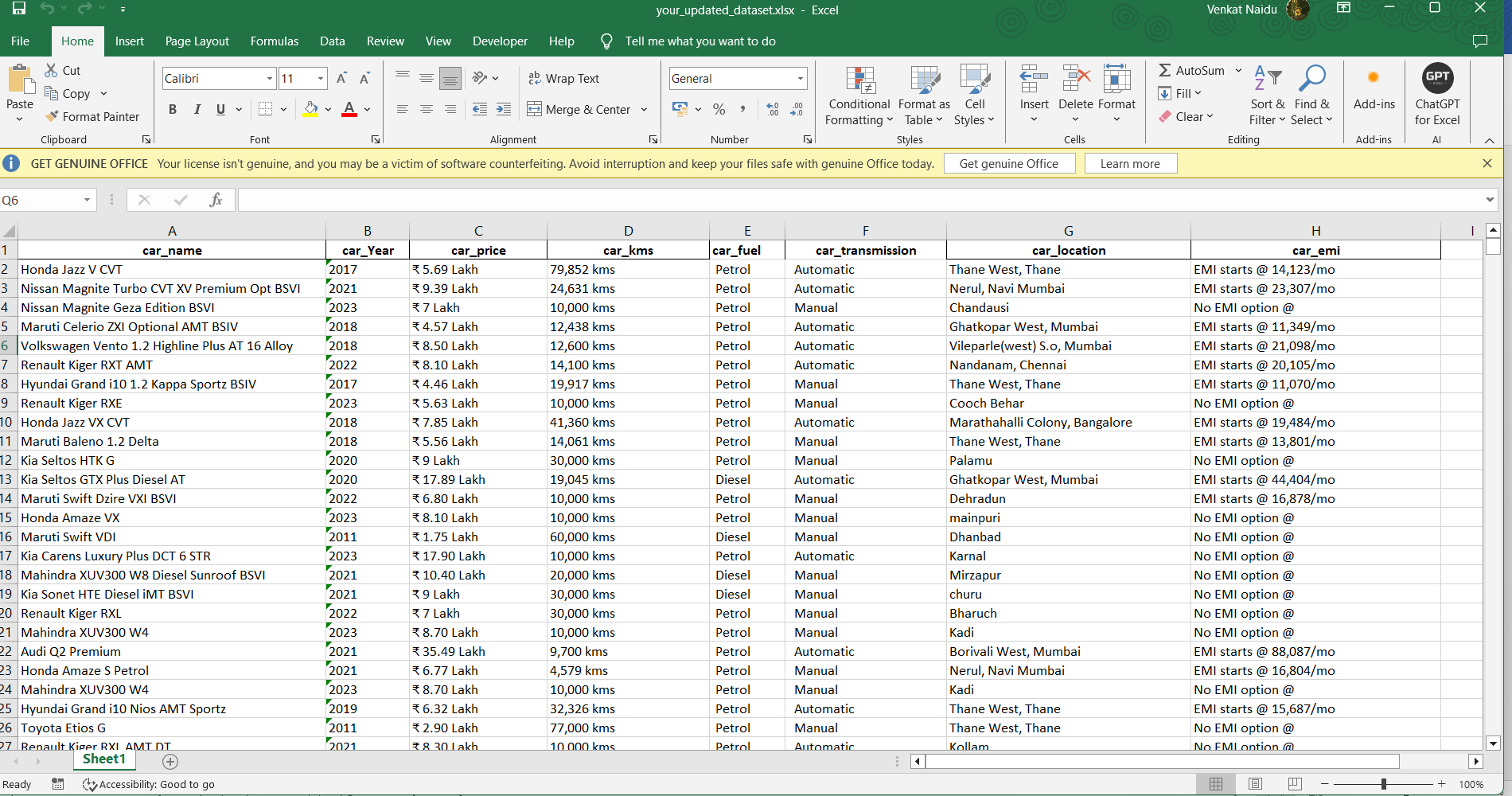
**STEP-9:-** Rordering the names s per need to understand the dat and to look clear view



**STEP-10:-** again saving updata file in to excel to do visualization on tableau

Before doing visualization some work I have done In excell because I don’t want to make work so complex and the excell make work easly to divide the rows the below page show how I have done the work on excell

# Excell WORK

****

I have done some excel work above fig is modified once

Lets discuss what I had done though following steps:-

**STEP -1 :-** Again I had checked the null by ctrl +h if in case while saving the fill any data is missing or not there is no missing data and no null values .

**STEP-2:-** In carkms column first it had got the data including carkms and fuel and transmission here excell makes easy to divide the column.

**STEP-3:-** While dividing the column first I had replaced petrol with delimetries like “petrol”then same I have done to diesel,lpg,cng then after I went to data and text to column and I had modified the data as per needed.

This are the few steps what I had done in excel.

Chapter 3 Data Analysis and Visualization

**3.1 Descriptive Statistics and Correlation Analysis**

**Descriptive Statistics** Descriptive statistics form the basis for understanding the dataset's central tendencies and dispersion. Key metrics such as mean, median, and standard deviation are calculated to provide a comprehensive summary.

The descriptive statistics module extends beyond the basics, incorporating advanced statistical measures such as skewness and kurtosis to capture the distribution's shape accurately.

* **Mean (Average):** This is the sum of all values divided by the total number of values. It gives a central measure of the dataset.
* **Median:** The middle value in a dataset when it's ordered. It's less sensitive to extreme values than the mean.
* **Standard Deviation:** A measure of how spread out the values in a dataset are. A low standard deviation indicates that the data points tend to be close to the mean, while a high standard deviation indicates that they are spread out over a large range.

**Correlation Analysis** Correlation matrices are deployed to delve into relationships between different features. This analytical step proves pivotal in identifying factors influencing used car prices, paving the way for nuanced interpretations.

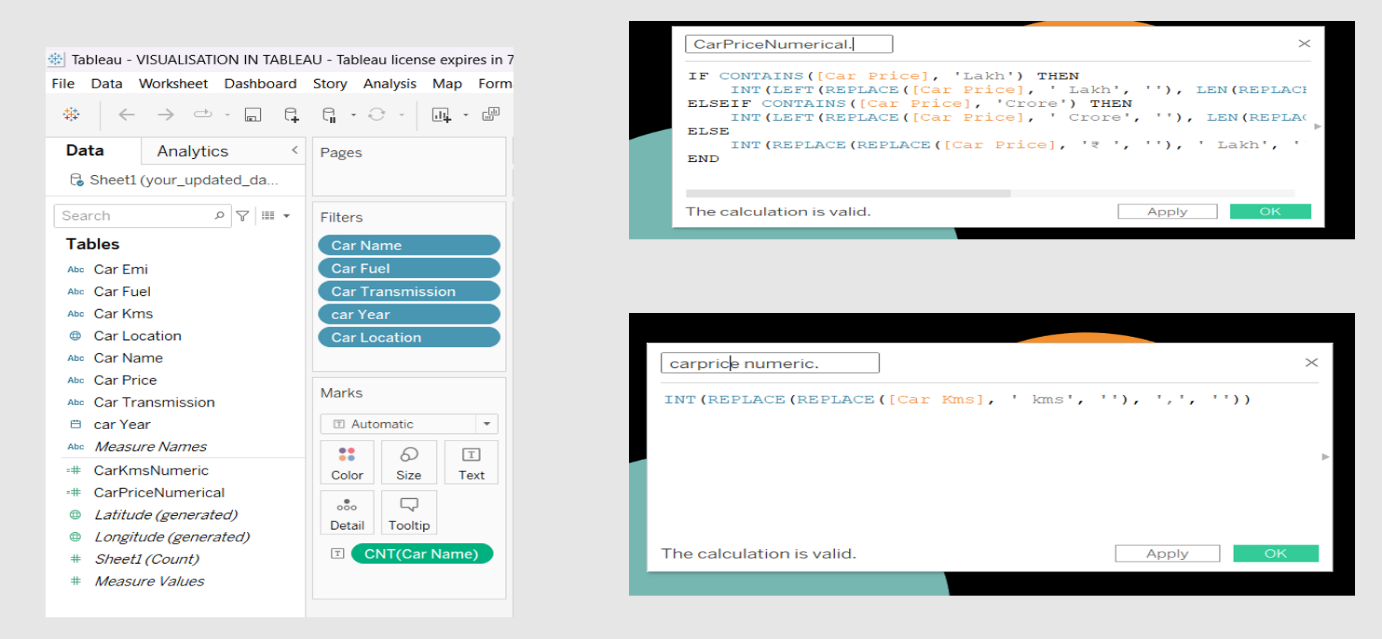
The correlation analysis extends to a multivariate approach, examining not only pairwise correlations but also exploring the impact of multiple variables on price trends.

* **Correlation Coefficient:** A statistical measure that describes the extent to which two variables change together. A positive correlation indicates a direct relationship (both variables increase or decrease together), while a negative correlation indicates an inverse relationship (one variable increases as the other decreases).
* **Correlation Matrix:** A table showing correlation coefficients between many variables. It helps identify patterns and relationships in the dataset.
* **Multivariate Approach:** Analyzing the relationships among more than two variables simultaneously. This provides a more comprehensive understanding of how different factors influence each other.

**3.2 Observation of visualization graphs**

**Data Visualization** Harnessing the power of Tableau the project employs a variety of visualization techniques, including histograms, line chart,tree map ,area chart scatter plots, and box plots,bar graph etcc... These visualizations serve to bring out patterns and trends in the dataset, enhancing interpretability. The data visualization module is designed to be interactive, allowing stakeholders to explore the dataset dynamically.

**In Tableau:-**



Here , I had imported the excel file and and the data all in dimension soo first thing I want to convert them into required form like car\_kms and car \_price in to measures .As per my data imported in this tableau the car\_kms and car\_price contain string values so I had changed that to numerical (decimal) by table calculation .After this iam draging the dimension and measure to get good meaningful insight.

**Explanation for CarPriceNumerical:- CODE**

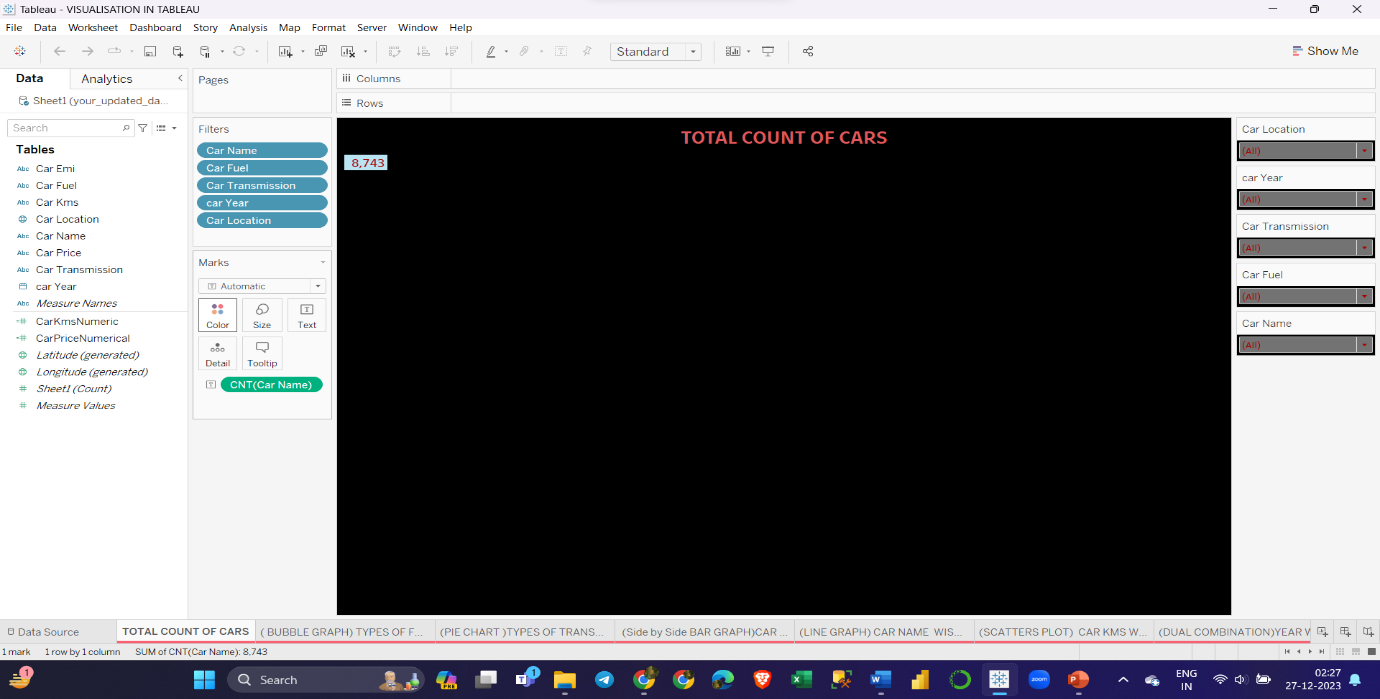
* The code is checking if the text in the "Car Price" field contains the word 'Lakh' or 'Crore'.
* If it contains 'Lakh', it removes 'Lakh', converts the remaining number to an integer, and multiplies it by 100,000 to get the actual price.
* If it contains 'Crore', it does a similar operation but multiplies by 10,000,000 instead.
* If it doesn't contain 'Lakh' or 'Crore', it assumes the price is already in numerical format, removes any currency symbols or extra text, and converts it to an integer.

**Explanation for CarkmsNumerical:-CODE**

* The code removes the 'kms' substring from the "Car Kms" field.
* It also removes any commas in the number (often used for thousand separators).
* The remaining text is then converted to an integer, representing the numerical value of kilometers.

**Observation of visualization graphs**

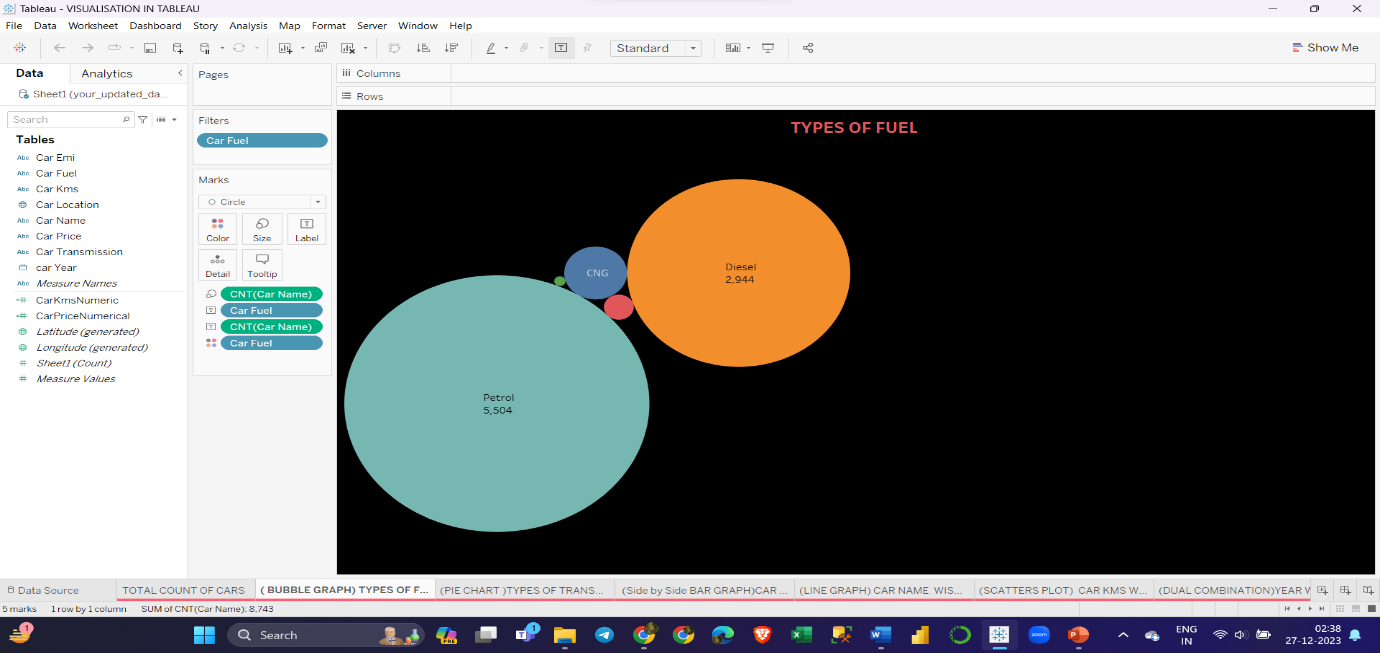
**Checking count of cars:-**



**OBSERVATION :-**Here, The screenshot shows a Tableau visualization dashboard with various data fields and visual representations of car-related informations.

**The total count of used cars are 8,743**

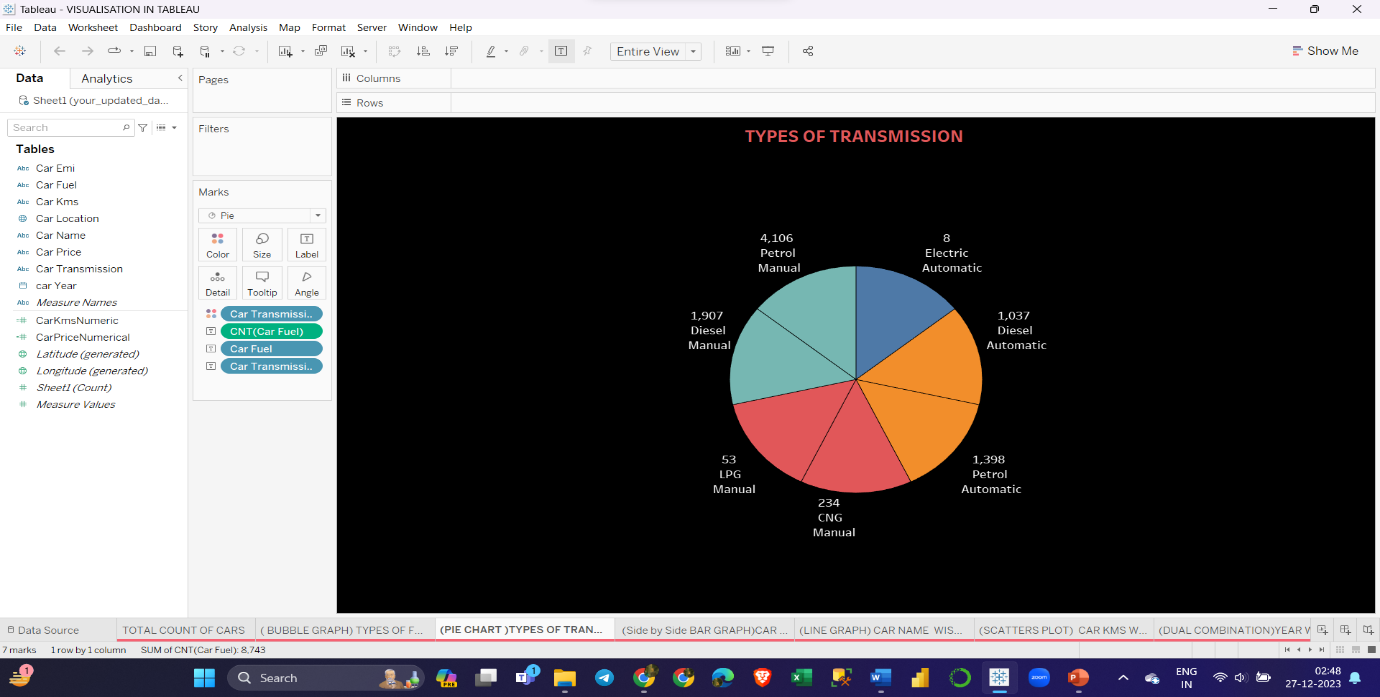
**“BUBBLE GRAPH”**



**OBSERVATION :-**It includes various data categories such as car fuel types and there count of cars

* Here,**Petrol cars are**:-**5,504** and **Diesel cara are**:-**2,944** and **Cng**:-**234** and **Lpg**:-**54** and **Electric:-8**

**By the observation of graph used petrol car are more compared to other fuel types on cardekho as per our analysis on scraped data**

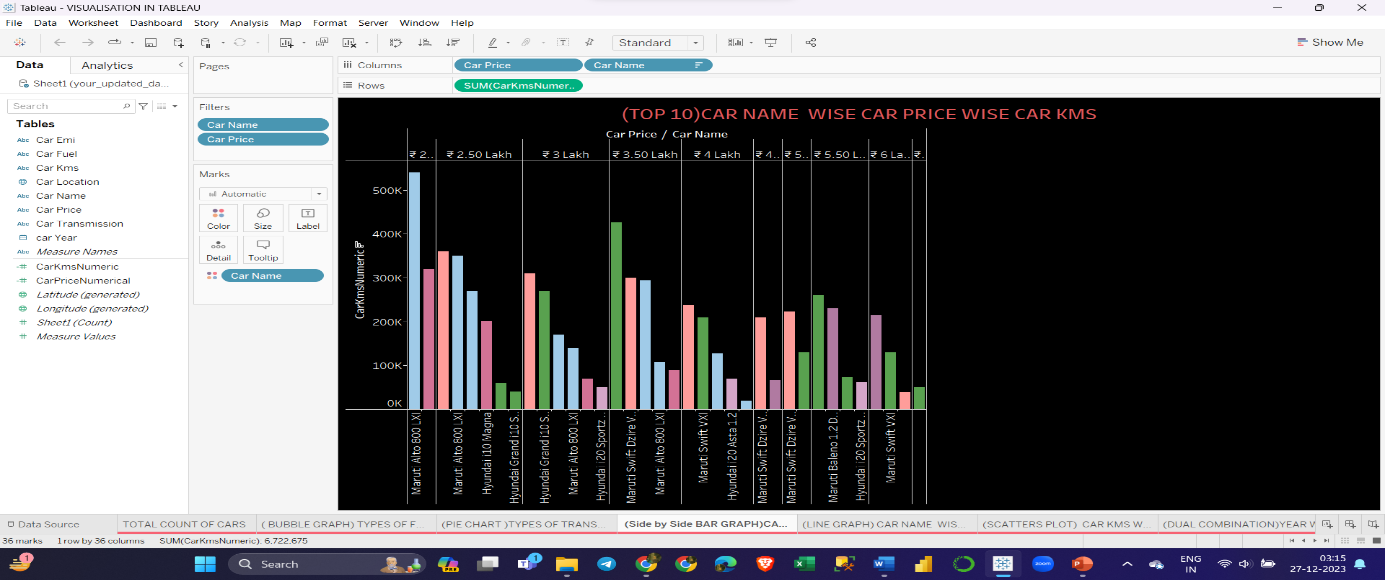
**“PIE CHART”** 

**OBSERVATION :-**In this chart it showing types of transmission and also including the fuel type and there count of cars

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **AUTOMATIC** | | | | | **MANUAL** | | | | |
| PETROL  1,398 | DIESEL  1,037 | LPG  0 | CNG  0 | ELECTRIC  8 | PETROL  4,106 | DIESEL  1,907 | LPG  53 | CNG  234 | ELECTRIC  8 |

**HERE , By the observation of graph Mostly the used manual car are more compared to automatic**

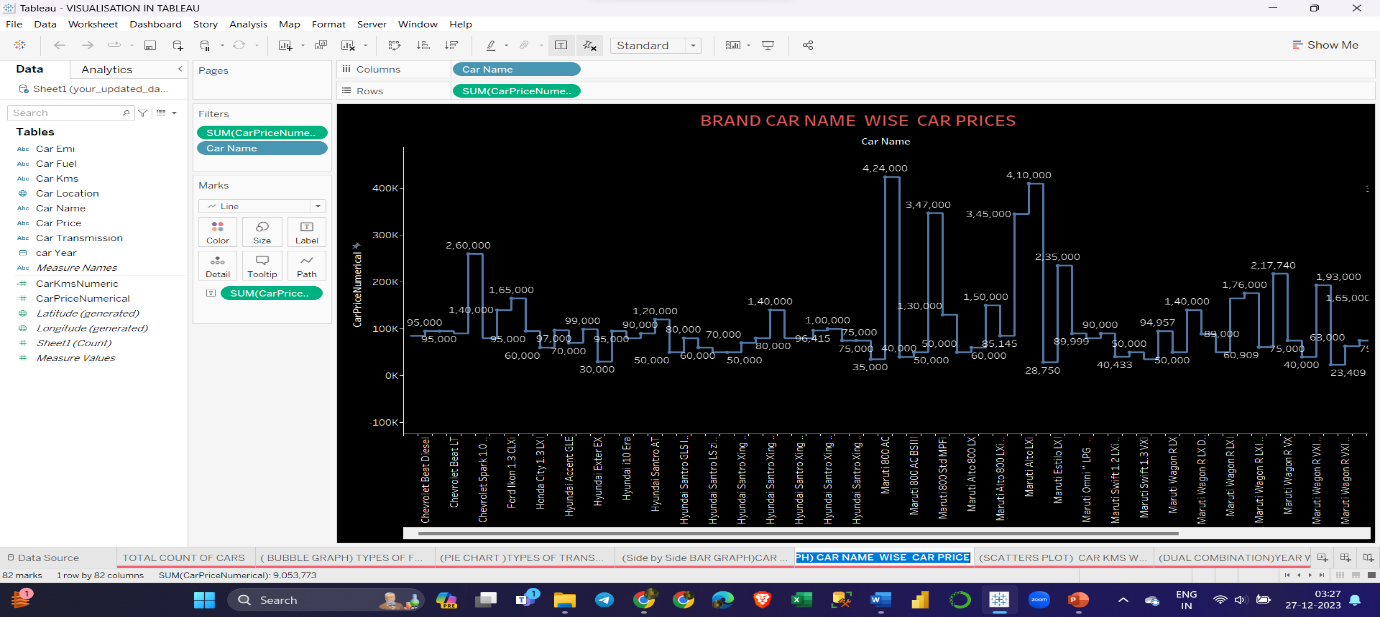
**“SIDE BY SIDE BAR GRAPH”**



**OBSERVATION :-**In this Graph it showing top 10 car names and car price along with their kilometers

**Here ,By the observation of graph for the top used car Maruti alt 800 LXI had more used cras with low cost and it driven around 540000kilometers ,price 2 lakhs**

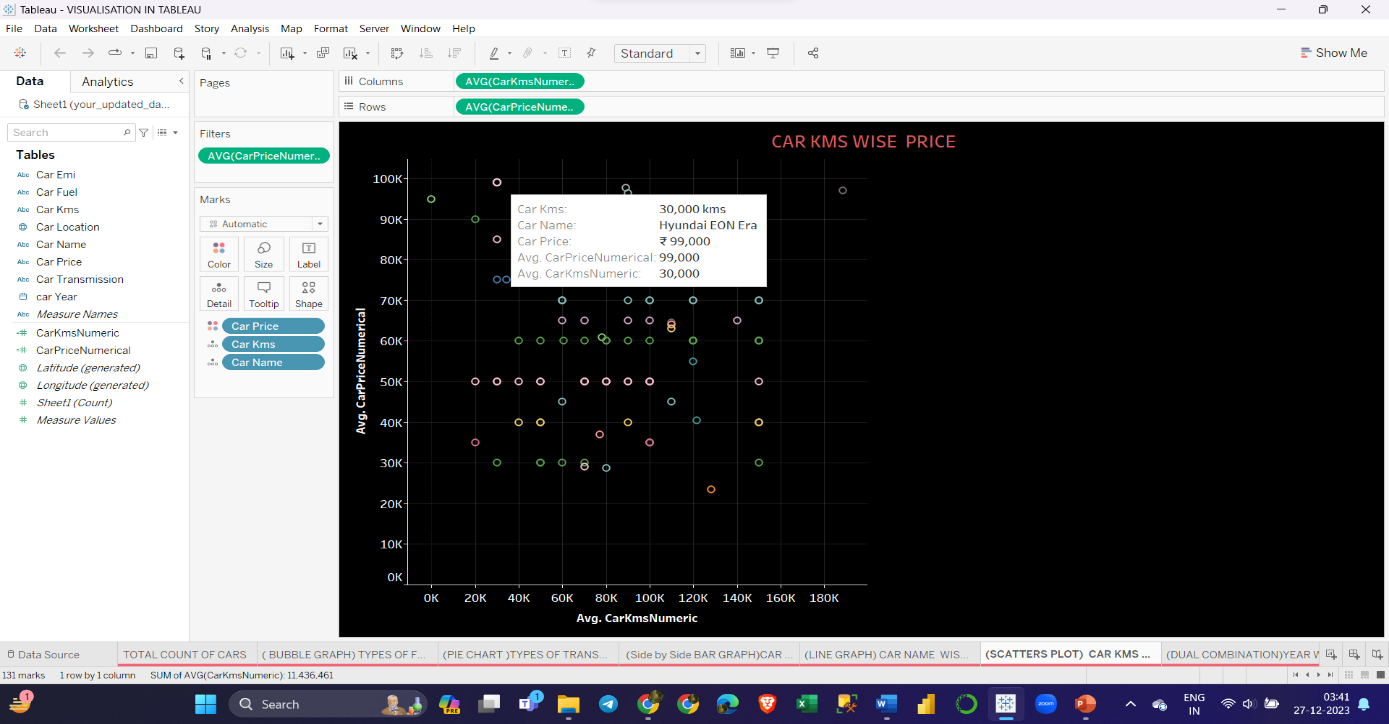
**“Line graph”**

****

**OBSERVATION :-**In this Graph it showing brand and car name and their sum of prices

**Here ,by the observation finaly states that different car and brand are selling for different prices on cardekho as per our analysis on scraped data**

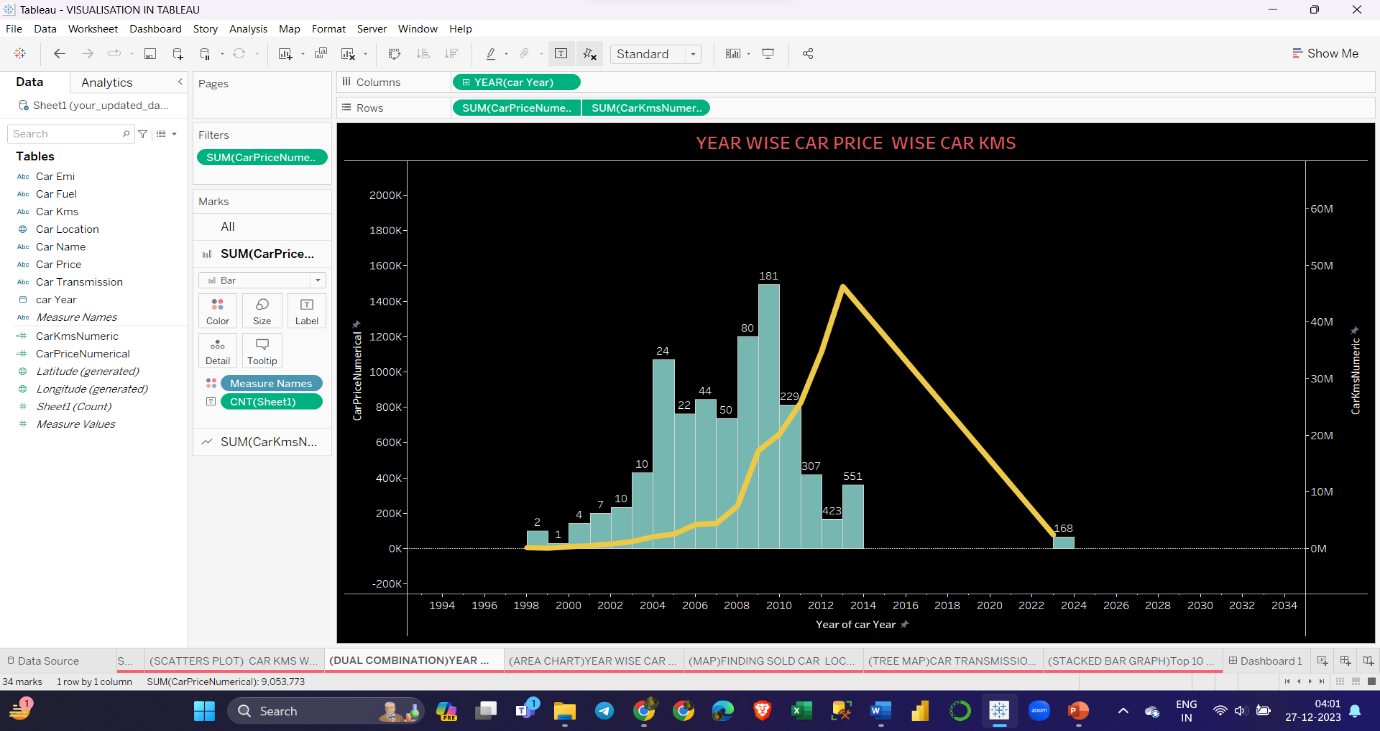
**“SCATTER PLOT”**

****

**OBSERVATION :-**In this Graph it showing car prices and their kilometers

**By the observation oF this graph it telling about avg car price and avg kilometers for example HYUNDAI EON ERA selling avg price of used car is 99,000 and their avg kilometers 30,000 like knowing the price and kilometers we can estimate the sales price of cars by the kms.**

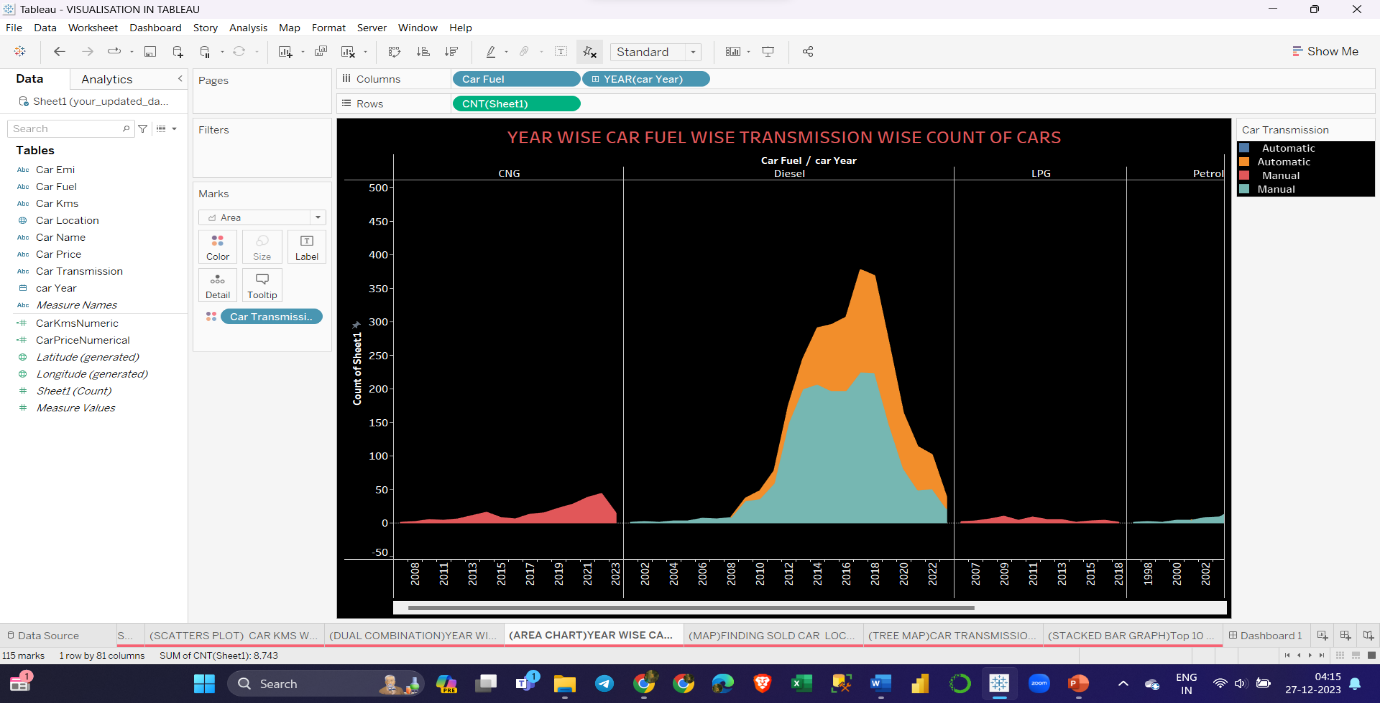
**“DUAL COMBINATION”**

****

**OBSERVATION :-**In this Graph it showing YEAR WISE CAR PRICE AND THEIR KMS

**Hence the observation states the showing the distribution of car prices and kilometers over different years. The visualization also includes various elements such as filters, tables, and other types of charts for analyzing car data. And year to see the more sales of used cars present 2023 there less usage of used car compared to below years.**

**“AREA CHART”**

****

**OBSERVATION :-**In this Graph it showing YEAR WISE CAR FUEL WISE TRANSMISSION WISE COUNT OF CARS

**By the observation there is distribution of cars year wise fuel wise transmission and their count and compared to this analysis 2015 to 2019 there more used cars in petrol automatic**

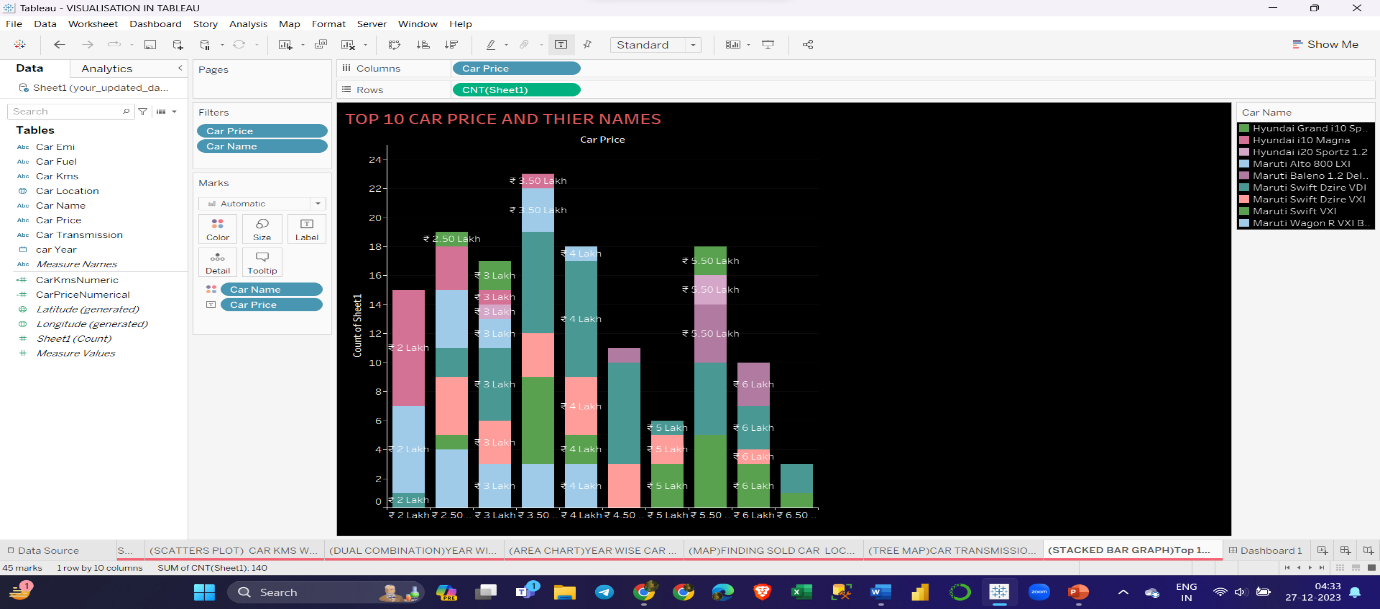
**“TREE MAP”**

****

**OBSERVATION:-**CAR TRANSMISSION WISE FUEL WISE COUNT OF NO EMI OPTION

**By this observation of tree map there is more no emi options start on petrol ,manual used car around 962 and second diesel has no emi option car around 363 hre the petrol car has more no emi options**

**“SATCKED BAR GRAPH”**

****

**OBSERVATION:-**HERE THERE IS TOP 10 CAR PRICES and their names

**By the Obseervation this states the car price and the car names only and here the mostly HYUNDAI 10 MAGNA has more used car and their price is around 2-5 lakhs and rest of the otherv car and their priceses in this stacked barchart**

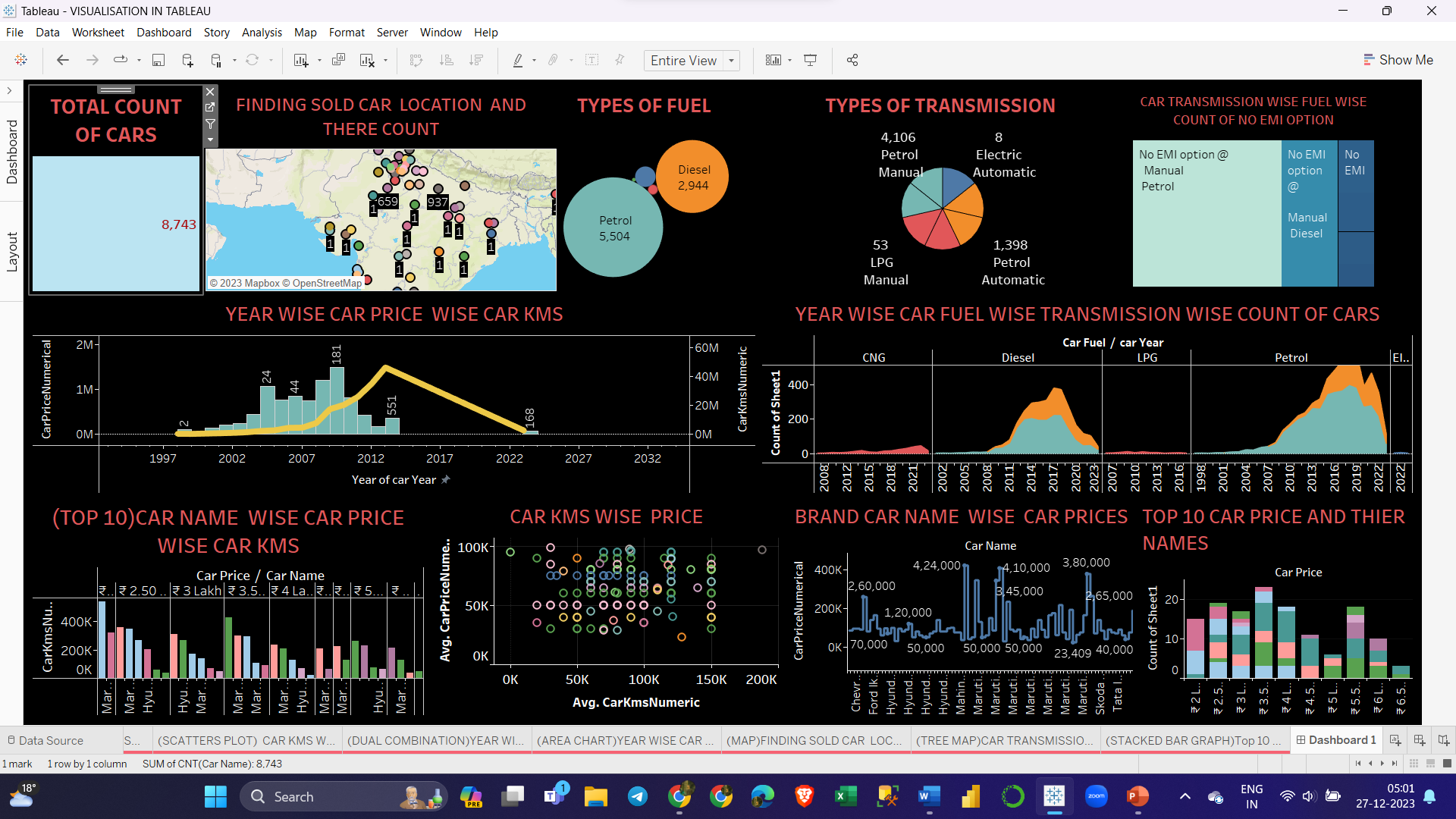
**MAPS:-**



**OBSERVATION:-** Here the find the sold car location and ther count

By this analysis of map we can see which location has sold more used cras and their count. Jaipur,and lucknow, delhii are having more sold used cars

# DASHBOARD



**THIS IS THE FINAL VISUALIZATIONS OF DASHBORAD IN TABLEAU WITH GAIN BEAUTIFUL INSIGHT AND MEANINGFUL VISULIZATION**

**3.3 Result and challanges**

**CHALLENGES** :- Documenting challenges faced during data collection, cleaning, and analysis processes, along with proposing potential solutions. This is important for transparency and learning from the project.

**Web Scraping Challenges:\_Dynamic Content:** Selenium may encounter challenges with websites that heavily rely on dynamic content loading, requiring you to implement mechanisms to wait for elements to load.

**Anti-Scraping Measures:** Some websites implement anti-scraping measures, such as CAPTCHAs or IP blocking. Strategies for handling these challenges need to be incorporated into your web scraping process.

**Data Preprocessing Challenges:**

**Missing Data:** Dealing with missing data is a common challenge. Decide whether to impute missing values, remove incomplete records, or use other strategies based on the nature of the data.

**Data Cleaning:** Addressing inconsistencies, outliers, and errors in the scraped data is crucial for accurate analysis. This may involve standardizing formats, handling outliers, and correcting data entry errors.

**Normalization and Scaling**: scaled to ensure that they contribute equally to the analysis.

**Data Visualization Challenges in Tableau:**

**Data Integration:** Combining data from different sources or merging datasets with varying granularity can be challenging. Ensure that the data in Tableau is structured appropriately for the visualizations you intend to create.

**Performance Optimization:** Large datasets may impact Tableau's performance. Aggregation, filtering, and other optimization techniques can be employed to enhance performance.

**Custom Calculations:** Creating custom calculations or calculated fields may be necessary to derive insights not directly available in the raw data. This requires a good understanding of the underlying mathematical relationships.

**Reproducibility and Documentation:**

**Reproducibility:** Document your entire process, from web scraping scripts to data preprocessing steps and Tableau visualizations. This ensures that others can reproduce your analysis.

**Documentation:** Clearly document any assumptions made, choices of models, and reasons for preprocessing decisions. This documentation is crucial for the transparency and replicability of your work.Throughout these challenges, maintaining a neutral and objective approach is essential. The goal is to derive insights from the data and communicate them effectively through visualizations and analyses.

**RESULT:-**

Sucessfully completion of project with webscraping and data preprocessing and data visualization with meaningfull manner and beautiful visualizations\

Chapter 4 Conclusions

* 1. **In Scope and out of Scope**

**In Scope:**

1. **Web Scraping:** The task involves collecting data from CarWale using appropriate web scraping tools or libraries. This step is crucial for obtaining raw data for analysis.
2. **Data Cleaning and Manipulation:** After scraping, the collected data needs processing to remove inconsistencies, handle missing values, and make it ready for analysis. This step ensures that the data is accurate and reliable.
3. **Visualization and Analysis:**. The focus is on specific factors such as brand, name,price, EMIs, fuel efficiency, and Transmission.
4. **Identifying Trends:** Analyzing the data to identify trends, correlations, and patterns that can lead to meaningful conclusions. This step aims to extract valuable insights from the collected information.
5. **Challenges:** Documenting challenges faced during data collection, cleaning, and analysis processes, along with proposing potential solutions. This is important for transparency and learning from the project.

**Out of Scope:**

1. **Real-time Data:** The project does not involve the collection and analysis of real-time data beyond the specified Cardekho link. The focus is on a snapshot of data rather than continuous monitoring.
2. **Predictive Modeling:** Building predictive models or forecasting future car prices based on historical data is not within the project scope. The emphasis is on descriptive analysis and identifying current trends rather than predicting future outcomes.

# 4.2 Conclusion & Reference

**Conclusion:-**

* **Recap of key points covered:-** Summarize the essential concepts discussed in this presentation about CarDekho and web scraping and data visualization.
* **Importance of CarDekho website and web scraping and data visualization for data analysis:-** Highlight the significance of web scraping and data visualization in the process of data analysis**.**

**Refrence:-**

I have also used few external resources that helped me to complete this project successfully. Below are the external resources that were used to create this project.

1. <https://www.google.com/>
2. [www.ripublication.com](http://www.ripublication.com)
3. [www.irjet.net](http://www.irjet.net/)
4. [www.ijcseonline.org](http://www.ijcseonline.org/)
5. <https://towardsdatascience.com/>
6. <https://www.analyticsvidhya.com/>

**THANK YOU**