

Project: Analyzing the Impact of Car Features on Price and Profitability

Final Project-3

Excel Workbook Link:

[ImpactOfCarFeatures.xlsx](#)

Project Description:

The automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition among manufacturers and a changing consumer landscape, it has become more important than ever to understand the factors that drive consumer demand for cars.

In recent years, there has been a growing trend towards electric and hybrid vehicles and increased interest in alternative fuel sources such as hydrogen and natural gas. At the same time, traditional gasoline-powered cars remain dominant in the market, with varying fuel types and grades available to consumers.

For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?

This problem could be approached by analyzing the relationship between a car's features, market category, and pricing, and identifying which features and categories are most popular among consumers and most profitable for the manufacturer. By using data analysis techniques such as regression analysis and market segmentation, the manufacturer could develop a pricing strategy that balances consumer demand with profitability, and identify which product features to focus on in future product development efforts. This could help the manufacturer improve its competitiveness in the market and increase its profitability over time.

Dataset Description:

Here is a brief overview of the dataset:

- **Number of observations:** 11,159
- **Number of variables:** 16
- **File type:** CSV (Comma Separated Values)

The variables in the dataset are:

- **Make:** the make or brand of the car
- **Model:** the specific model of the car
- **Year:** the year the car was released
- **Engine Fuel Type:** the type of fuel used by the car (gasoline, diesel, etc.)
- **Engine HP:** the horsepower of the car's engine
- **Engine Cylinders:** the number of cylinders in the car's engine
- **Transmission Type:** the type of transmission (automatic or manual)
- **Driven_Wheels:** the type of wheels driven by the car (front, rear, all)
- **Number of Doors:** the number of doors the car has
- **Market Category:** the market category the car belongs to (Luxury, Performance, etc.)
- **Vehicle Size:** the size of the car
- **Vehicle Style:** the style of the car (Sedan, Coupe, etc.)
- **Highway MPG:** the estimated miles per gallon the car gets on the highway
- **City MPG:** the estimated miles per gallon the car gets in the city
- **Popularity:** a ranking of the popularity of the car (based on the number of times it has been viewed on Edmunds.com)
- **MSRP:** the manufacturer's suggested retail price of the car

Approach:

The given tasks below based on the business problem would require advanced Excel skills and knowledge of data analysis techniques such as regression analysis, pivot tables, sensitivity analysis, optimization, and time series analysis.

However, by answering these questions and building an interactive dashboard, a data analyst could provide valuable insights to a car manufacturer and help them optimize their pricing and product development decisions to maximize profitability while meeting consumer demand.

Starting with cleaning our data before analysis using EDA technique, we handle missing values, outliers and duplicates.

We have created ten separate worksheets for five analysis tasks and five dashboards each, with two additional sheets for original and cleaned dataset.

Tech-Stack Used:

Softwares Used:

WPS Office
Sheets

Hardware Used:

HP Pavilion
i5 Processor

Microsoft Edge
OBS Studio
Google Drive
Google Docs

Insights:

Before diving into the analysis of the given dataset, it is important to perform thorough data cleaning to ensure accurate and reliable results.

We perform data cleaning in following steps:

Step 1: Handling Missing Values.

We make use of the countblank function to count the number of cells having missing values in each column.

No. of Missing Values	Make	Model	Year	Engine Fuel	Engine HP	Engine Cylinders	Transmission	Driven Wheels	Number of Doors	Market Category	Vehicle Size	Vehicle Size	Highway MPG	City MPG	Popularity	MSRP
0	BMW	1 Series M	2011	premium u	335	6	MANUAL	rear wheel	2	Factory Tu	Compact	Coupe	26	19	3916	46135
0	BMW	1 Series	2011	premium u	300	6	MANUAL	rear wheel	2	Luxury,Per	Compact	Convertible	28	19	3916	40650
0	BMW	1 Series	2011	premium u	300	6	MANUAL	rear wheel	2	Luxury,Hig	Compact	Coupe	28	20	3916	36350
0	BMW	1 Series	2011	premium u	230	6	MANUAL	rear wheel	2	Luxury,Per	Compact	Coupe	28	18	3916	29450
0	BMW	1 Series	2011	premium u	230	6	MANUAL	rear wheel	2	Luxury	Compact	Convertible	28	18	3916	34500
0	BMW	1 Series	2012	premium u	230	6	MANUAL	rear wheel	2	Luxury,Per	Compact	Coupe	28	18	3916	31200
0	BMW	1 Series	2012	premium u	300	6	MANUAL	rear wheel	2	Luxury,Per	Compact	Convertible	26	17	3916	44100
0	BMW	1 Series	2012	premium u	300	6	MANUAL	rear wheel	2	Luxury,Hig	Compact	Coupe	28	20	3916	39300
0	BMW	1 Series	2012	premium u	230	6	MANUAL	rear wheel	2	Luxury	Compact	Convertible	28	18	3916	36900
0	BMW	1 Series	2013	premium u	230	6	MANUAL	rear wheel	2	Luxury	Compact	Convertible	27	18	3916	37200
0	BMW	1 Series	2013	premium u	300	6	MANUAL	rear wheel	2	Luxury,Hig	Compact	Coupe	28	20	3916	39600
0	BMW	1 Series	2013	premium u	230	6	MANUAL	rear wheel	2	Luxury,Per	Compact	Coupe	28	19	3916	31500
0	BMW	1 Series	2013	premium u	300	6	MANUAL	rear wheel	2	Luxury,Per	Compact	Convertible	28	19	3916	44400
0	BMW	1 Series	2013	premium u	230	6	MANUAL	rear wheel	2	Luxury	Compact	Convertible	28	19	3916	37200
0	BMW	1 Series	2013	premium u	230	6	MANUAL	rear wheel	2	Luxury,Per	Compact	Coupe	28	19	3916	31500
1	BMW	1 Series	2013	premium u	320	6	MANUAL	rear wheel	2	Luxury,Hig	Compact	Convertible	25	18	3916	48250
0	BMW	1 Series	2013	premium u	320	6	MANUAL	rear wheel	2	Luxury,Hig	Compact	Coupe	28	20	3916	43550
0	Audi	100	1992	regular unl	172	6	MANUAL	front whee	4	Luxury	Midsize	Sedan	24	17	3105	2000
0	Audi	100	1992	regular unl	172	6	MANUAL	front whee	4	Luxury	Midsize	Sedan	24	17	3105	2000
0	Audi	100	1992	regular unl	172	6	AUTOMAT	all wheel d	4	Luxury	Midsize	Wagon	20	16	3105	2000
0	Audi	100	1992	regular unl	172	6	MANUAL	front whee	4	Luxury	Midsize	Sedan	24	17	3105	2000
0	Audi	100	1992	regular unl	172	6	MANUAL	all wheel d	4	Luxury	Midsize	Sedan	21	16	3105	2000
0	Audi	100	1993	regular unl	172	6	MANUAL	front whee	4	Luxury	Midsize	Sedan	24	17	3105	2000
0	Audi	100	1993	regular unl	172	6	AUTOMAT	all wheel d	4	Luxury	Midsize	Wagon	20	16	3105	2000
0	Audi	100	1993	regular unl	172	6	MANUAL	front whee	4	Luxury	Midsize	Sedan	24	17	3105	2000
0	Audi	100	1993	regular unl	172	6	MANUAL	front whee	4	Luxury	Midsize	Sedan	24	17	3105	2000
0	Audi	100	1993	regular unl	172	6	MANUAL	all wheel d	4	Luxurv	Midsize	Sedan	21	16	3105	2000

As we can see, Engine HP has the highest number of missing values which is not possible to manually enter.

In this case we can replace them with mean or median values, but here we have chosen to delete the rows.

Other columns like Engine Fuel, Engine Cylinders and No of Doors have a manageable number of missing values.

These values we replace by searching for them online.

Step 2: Remove Duplicate Values

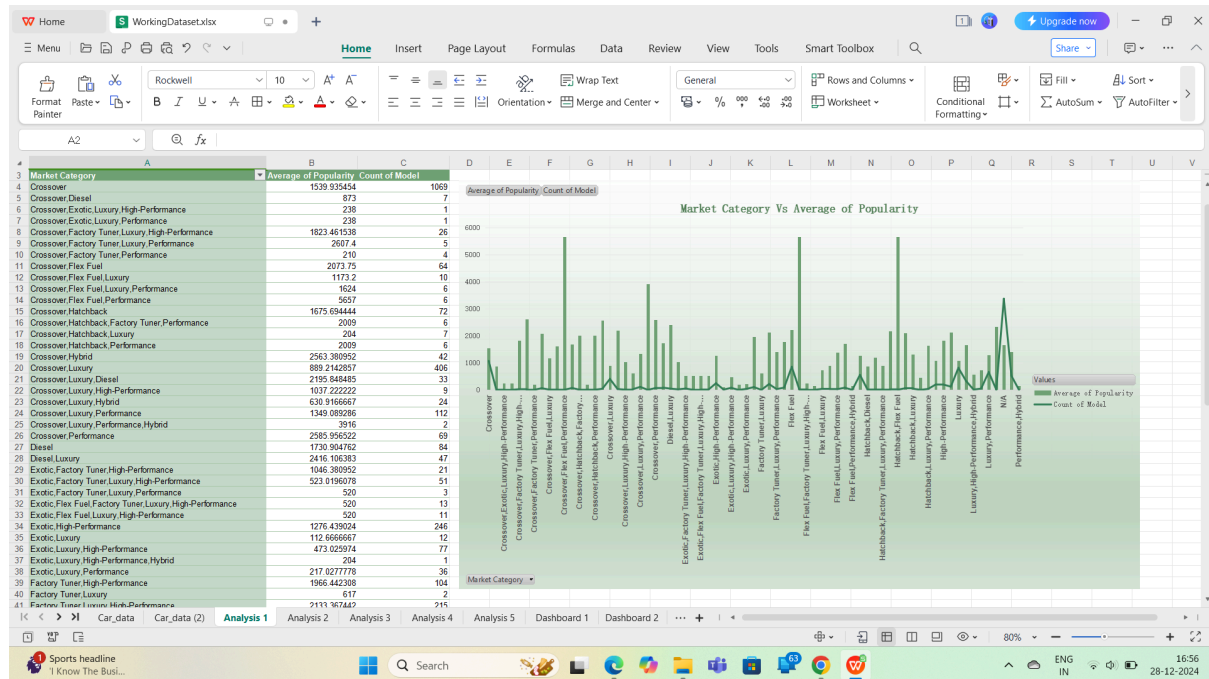
After you insert your data in a table, you can delete the duplicate values from the table tools ribbon tab.

We make use of Pivot Table and Column Chart for this task.

First column of the table shows Market Category. Average of Popularity and Count of Model are derived using Σ Values field of Pivot Table.

A trendline has been added to show Count of Model while columns show Average of Popularity.

Result:



Most Popular market Category: Flex Fuel, Diesel |Hatchback, Flex Fuel | Crossover, Flex Fuel, Performance

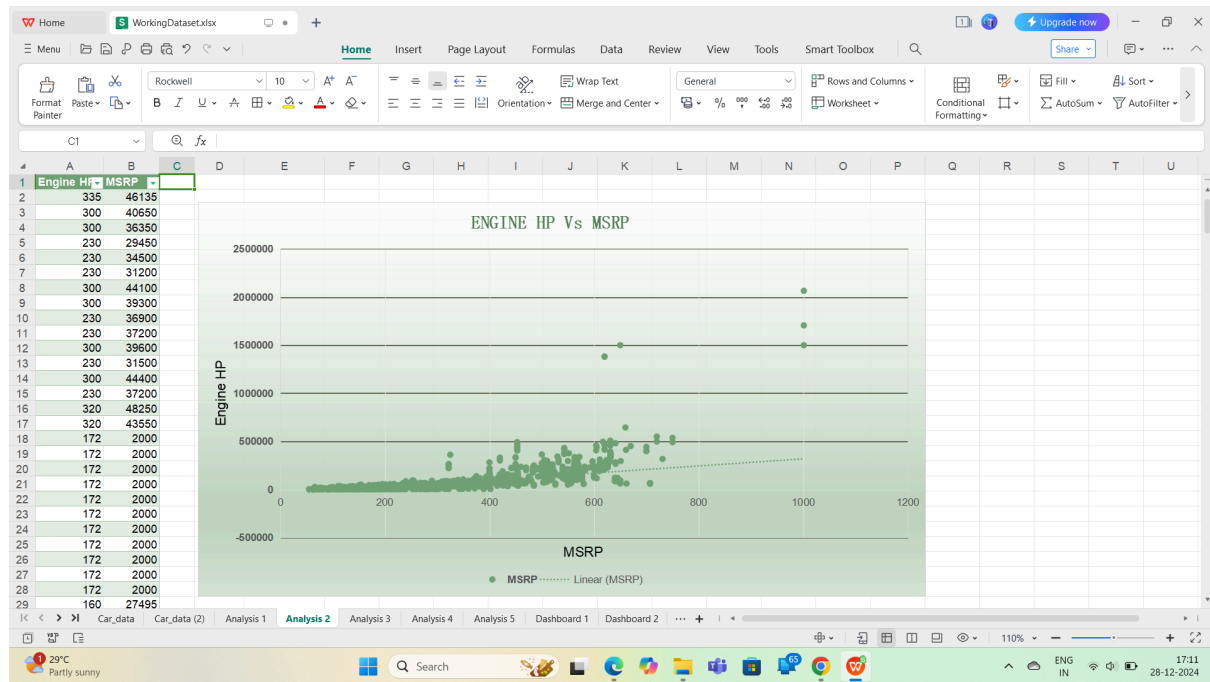
Least Popular Market Category: Exotic, Luxury | Flex Fuel, Hybrid

Insight Required: What is the relationship between a car's engine power and its price?

- **Task 2:** Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.

We copy paste required columns from the table on a separate sheet and create a new table. We use this table to create a scatter plot chart. Also we add a linear trendline to show a relationship between the two variables.

Results:



The Price will increase with the increasing number of Engine Power.

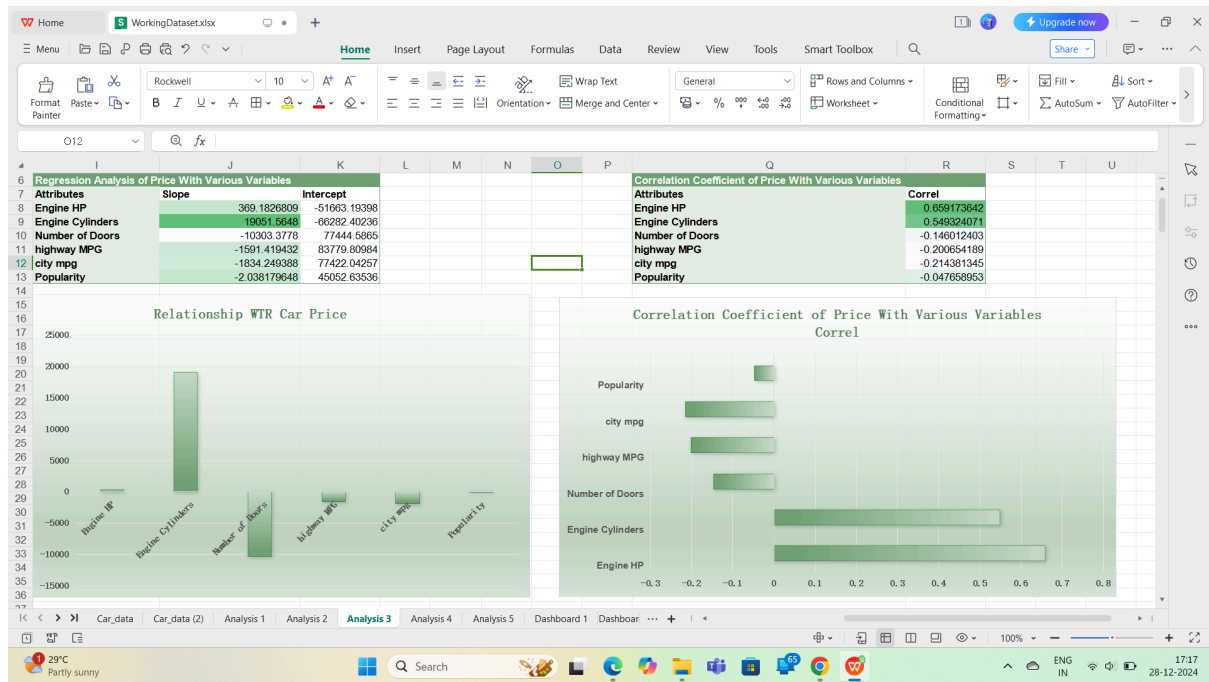
Insight Required: Which car features are most important in determining a car's price?

- **Task 3:** Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.

To perform Regression Analysis we use the LINEST function of Excel. It gives us two values as Slope and Intercept as a result.

To find the Correlation Coefficient we use the CORREL function.

Result:



Engine HP and Engine Cylinders are having a positive relationship with Price whereas Highway MPG, City MPG, Number of Doors, and Popularity are having a negative relationship with Price.

Insight Required: How does the average price of a car vary across different manufacturers?

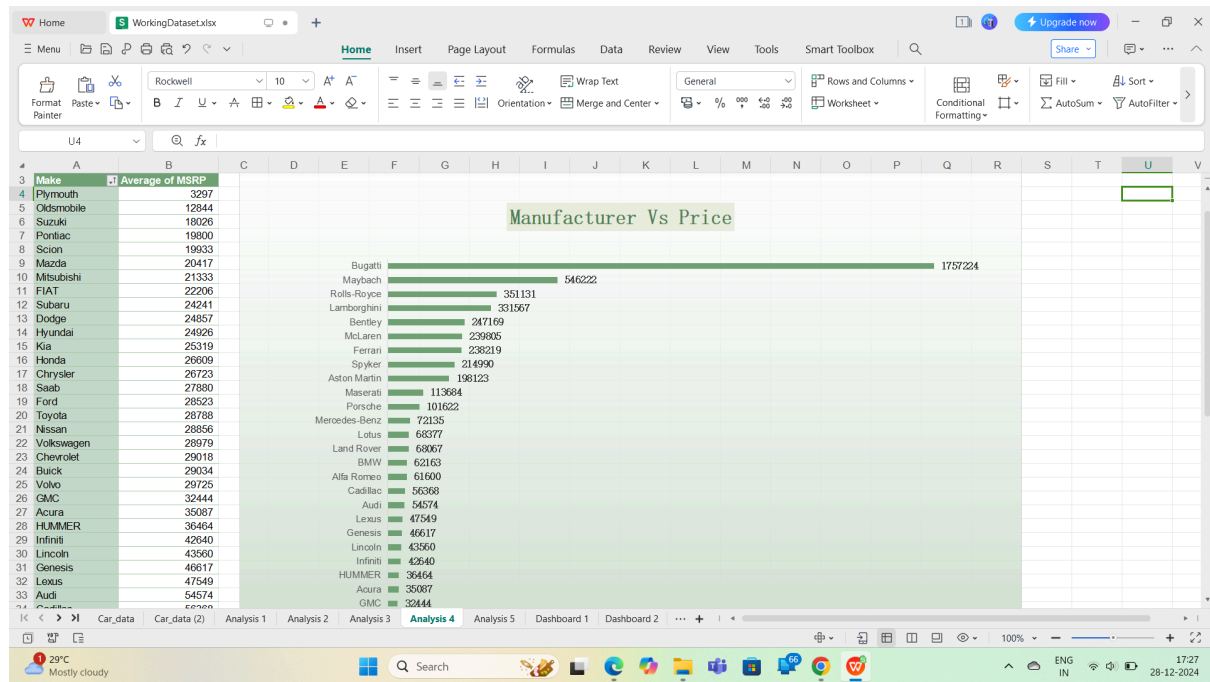
- **Task 4.A:** Create a pivot table that shows the average price of cars for each manufacturer.
- **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.

We copy paste required columns on another sheet and insert it in a new table.

We sort the table in descending prices.

Using this table we plot a bar chart to showcase Manufacturer Vs Prices of cars.

Result:



Highest Average Price of Cars Manufacturer: Bugatti | Maybach | Rolls-Royce | Lamborghini

Lowest Average Price of Cars Manufacturer: Plymouth | Oldsmobile | Suzuki

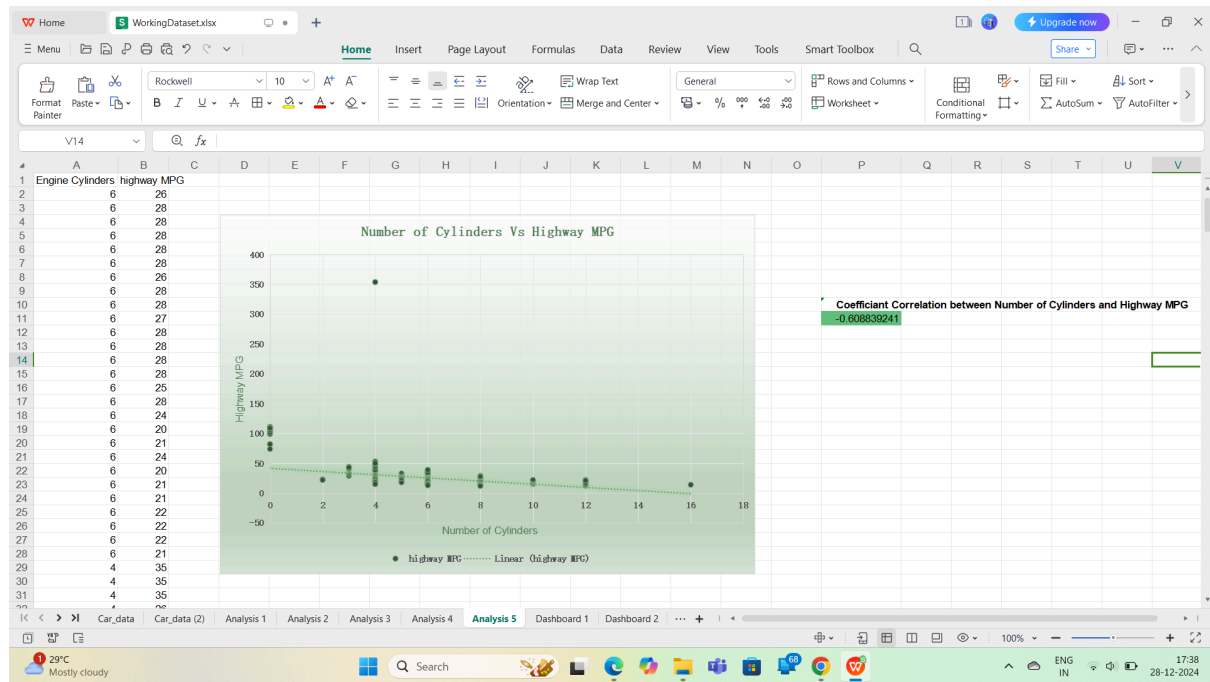
Insight Required: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

- **Task 5.A:** Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.
- **Task 5.B:** Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

Here we have used a scatter plot chart to show the relationship between Engine Cylinders and Highway MPG. A linear trendline has been added to the chart.

Also, we use the CORREL function to find Correlation Coefficient between the two variables.

Result:



The Trendline for the relationship between the number of Cylinders and Highway MPG is negative which means a lesser number of cylinders will give more highway mpg.

Building the Dashboard:

Now for the Next portion of the Project, you need to create the Interactive Dashboard. The client has requested these questions given below:

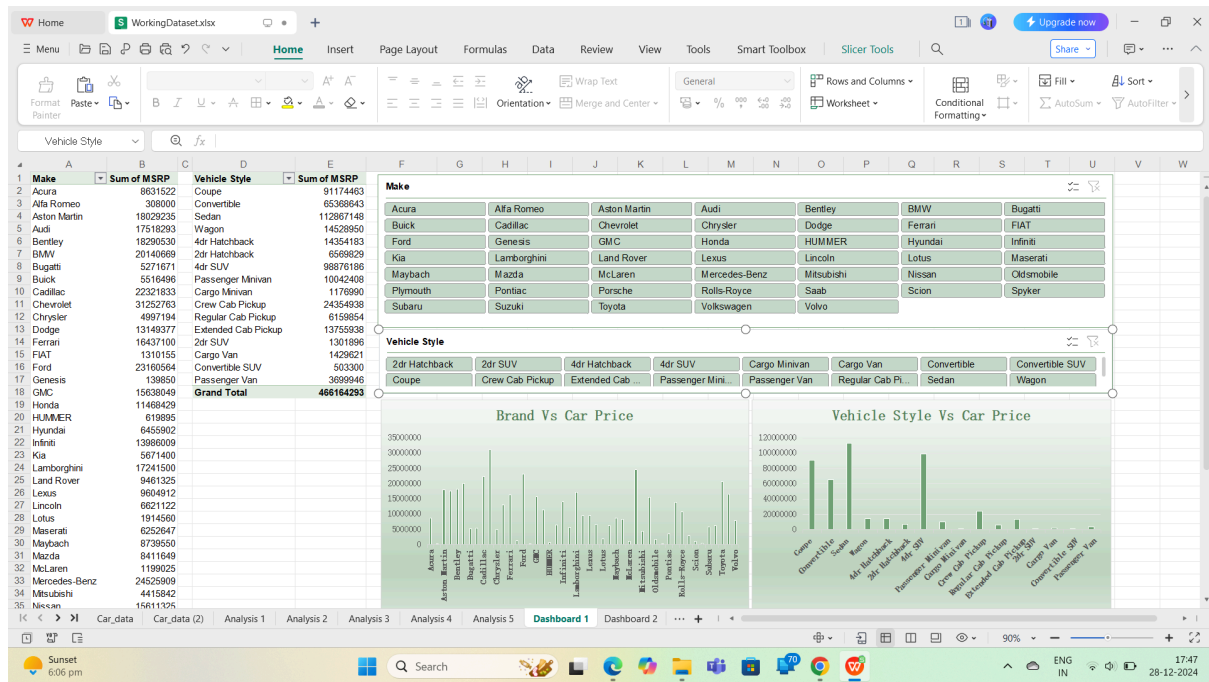
Task 1: How does the distribution of car prices vary by brand and body style?

- **Hints:** Stacked column chart to show the distribution of car prices by brand and body style. Use filters and slicers to make the chart interactive. Calculate the total MSRP for each brand and body style using SUMIF or Pivot Tables.

We create two Pivot Tables for Total MSRP Vs Make and Total MSRP Vs Vehicle Style. We use these tables to plot two separate column charts for the same. Then we add a slicer for each table.

We connect both slicers to each using the Report Connections button in the Slicer Tools ribbon.

Result:



Chevrolet Brand and Sedan Vehicle Types will likely have greater MSRP.

Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

- **Hints:** Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculate the average MSRP for each brand and body style using AVERAGEIF or Pivot Tables.

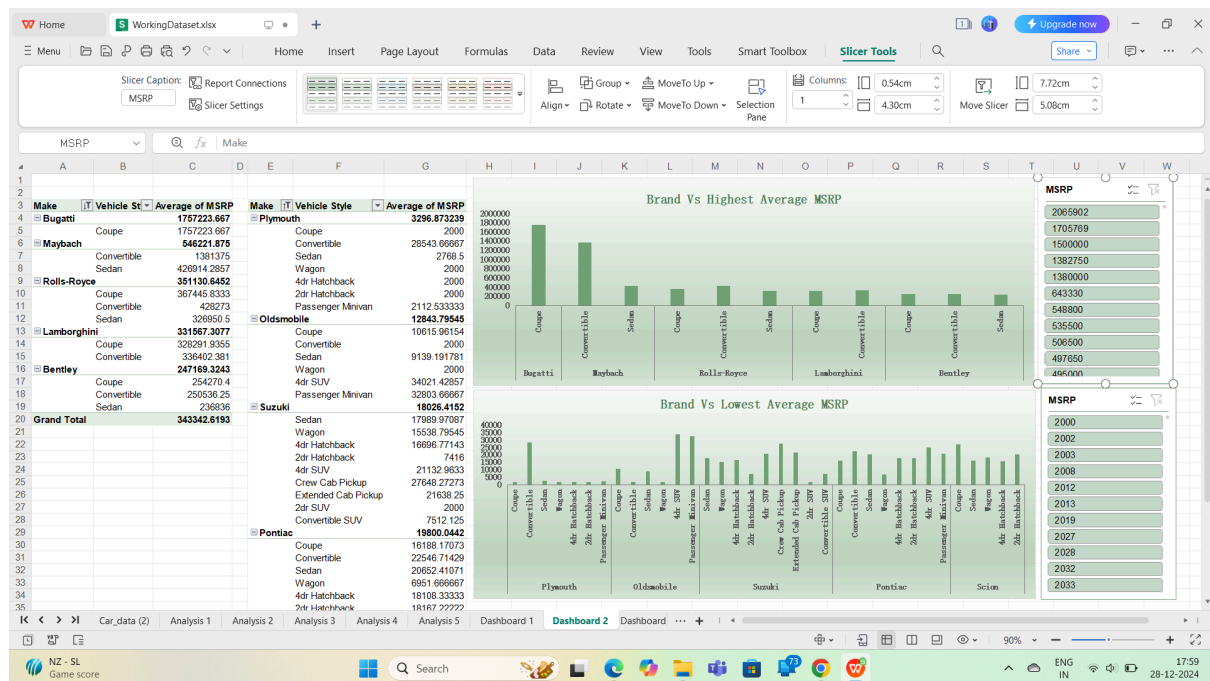
We create two Pivot Tables for Average MSRP Vs Make Vs Vehicle Style and Average MSRP Vs Vehicle Style Vs Make.

Then we sort the first table in Descending order of Average MSRP, while the second table is sorted in ascending order.

We use these tables to plot two separate column charts for the same. Then we add a slicer for each table.

We connect both slicers to each using the Report Connections button in the Slicer Tools ribbon.

Result:



The **Highest Average Car Price Brand with Vehicle Style** is **Bugatti's Coupe** and **Maybach's Convertible**.
The **Lowest Average Car Price Brand with Vehicle Style** is **Plymouth's Coupe & 2dr hatchback** and **Oldsmobile's Wagon & Convertible**.

Task 3: How do the different features such as transmission type affect the MSRP, and how does this vary by body style?

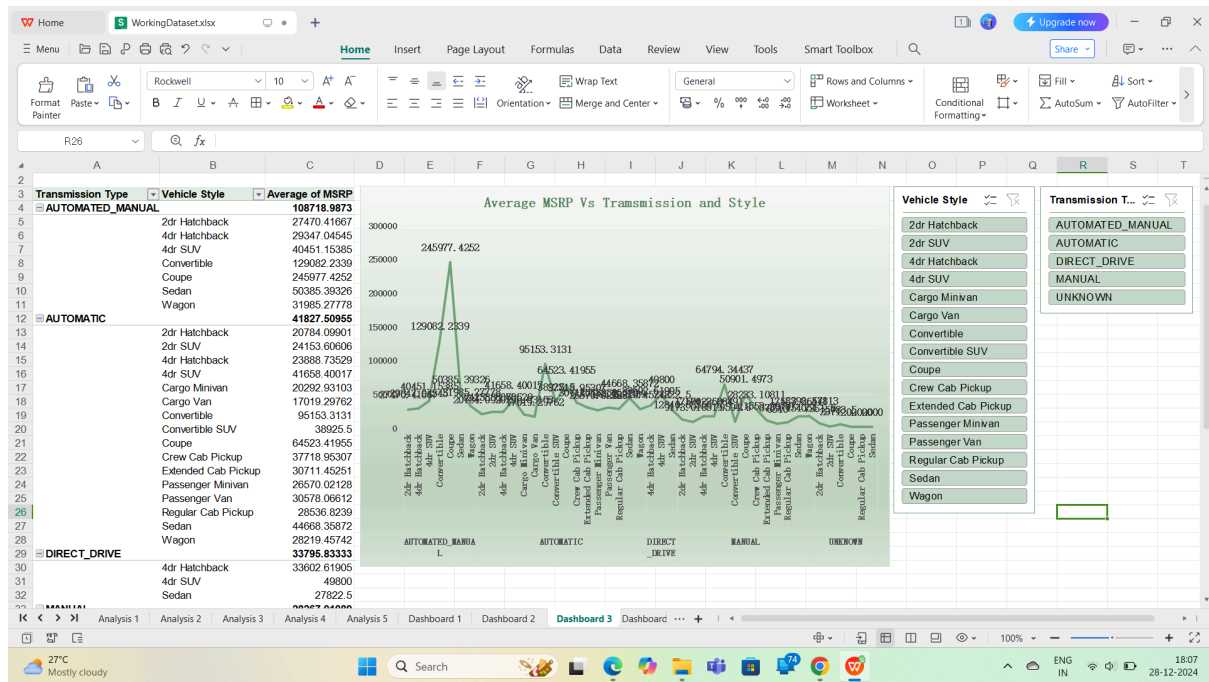
- **Hints:** Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculate the average MSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.

We create a Pivot Table for Transmission Type|Vehicle Style|Average MSRP.

We use this table to plot a line chart for the same. Then we add a slicer for Transmission Type and Vehicle Style.

We connect both slicers to each using the Report Connections button in the Slicer Tools ribbon.

Result:



Upon analysis, we found out that Automated/Manual has the highest Average MSRP while Manual is having lowest Average MSRP.

Task 4: How does the fuel efficiency of cars vary across different body styles and model years?

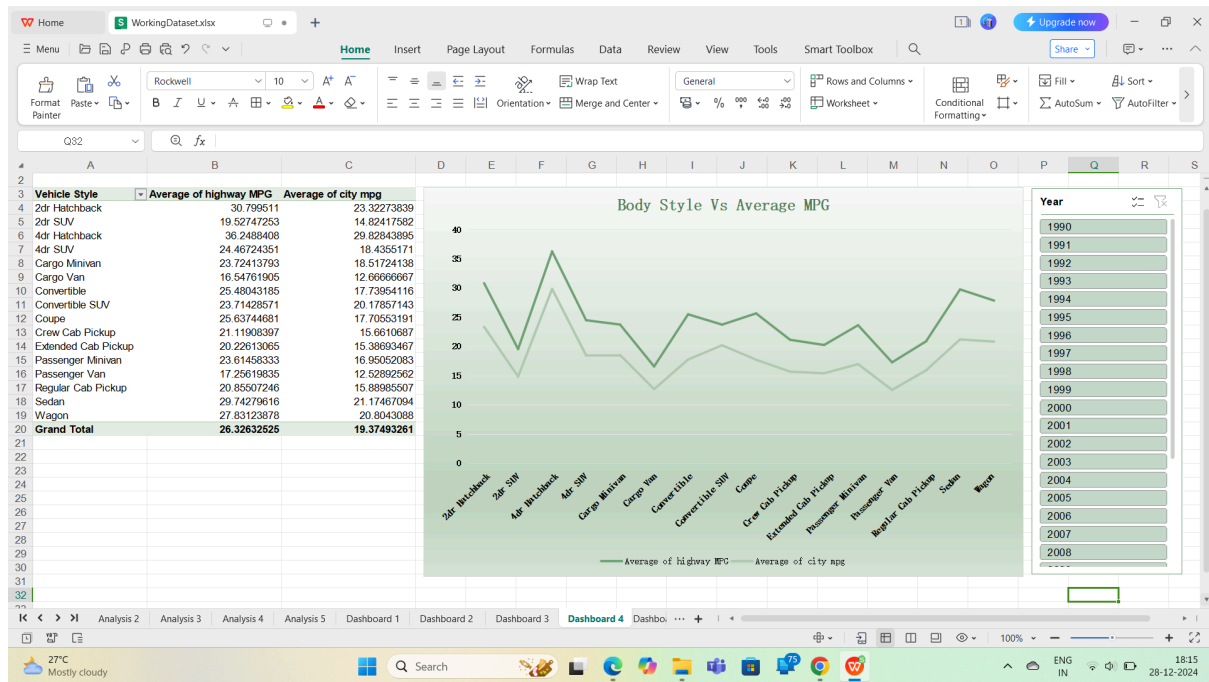
- **Hints:** Line chart to show the trend of fuel efficiency (MPG) over time for each body style. Calculate the average MPG for each combination of body style and model year using AVERAGEIFS or Pivot Tables.

We create a Pivot Table for |Vehicle Style|Average Highway MPG|Average City MPG.

We use this table to plot a line chart for the same.

Then we add a slicer for Year.

Result:



4dr Hatchback's fuel efficiency is the highest whereas Cargo Van and Passenger Van's fuel efficiency is the lowest.

Task 5: How does the car's horsepower, MPG, and price vary across different Brands?

- **Hints:** Bubble chart to visualize the relationship between horsepower, MPG, and price across different car brands. Assign different colors to each brand and label the bubbles with the car model name. Calculate the average horsepower, MPG, and MSRP for each car brand using AVERAGEIFS or Pivot Tables.

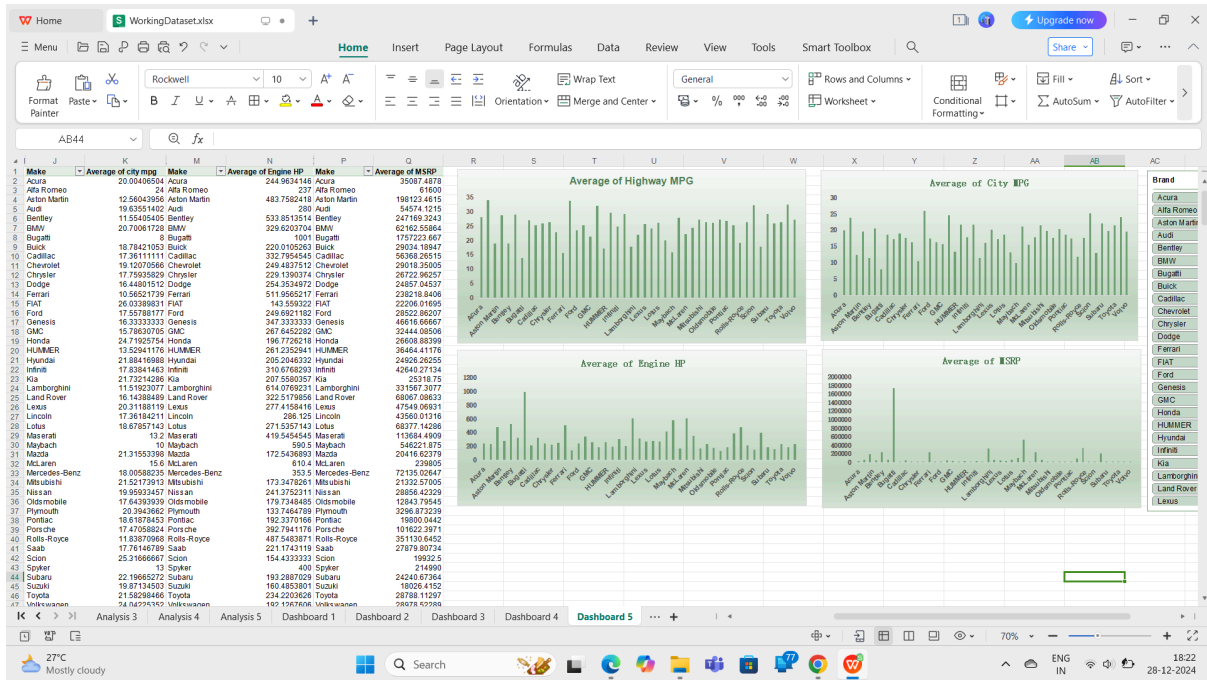
We have created four pivot charts, namely

- 1) Make Vs Average Highway MPG
- 2) Make Vs Average City MPG
- 3) Make Vs Average Engine HP
- 4) Make Vs Average MSRP

We have made use of column charts.

Lastly we add a slicer for Make and interconnect it to the charts.

Result:



Bugatti has the highest Engine HP and Car Price while Tesla has the highest MPG (both on Highway and in City).

Excel Workbook Link:
[ImpactOfCarFeatures.xlsx](#)