Analyzing the Impact of Car Features on Price and Profitability

Trainity Project No. 7

Prepared by:
Nilesh Kulkarni

AGENDA

- Project Description
- Approach
- Tech Stack Used
- Insights
- Results

Prepared By –
Nilesh Kulkarni

Project Description

We have been given metadata around Cars like Make, year, price and technical details like Fuel economy, horse power etc. The aim of the project is to analyse the data and find out factors impacting Cars price as well as Relationship between various features of cars like number of cylinders & fuel economy for example. The project also involve data cleaning and preparation before analysis.

The data analytics tasks are summarized below:

- **Analysis Tasks/Insights needed:** Analyse and derive below listed insights:
  - Pivot table for models in each market category and corresponding popularity scores & combo chart for average popularity score and market category wise count of models
  - Scatter chart that plots engine power on the x-axis and price on the y-axis
  - Regression analysis to identify the variables that have the strongest relationship with a car's price
  - Pivot table for average price of cars for each manufacturer & bar chart for price by manufacturer
  - Find out relationship between fuel efficiency and the number of cylinders in a car's engine using a scatter chart as well as calculate relationship co-efficient
- **Dashboard preparation:** Create a dashboard with below details with slicers/filters:
  - Distribution of car prices variation by brand and body style
  - Car brands with the highest and lowest average MSRPs and variation of MSRP by body type
  - How transmission type affect the MSRP, and how does this vary by body style?
  - How does the fuel efficiency of cars vary across different body styles and model years?
  - How does the car's horsepower, MPG, and price vary across different Brands?

Project Approach

High level steps for the Project approach are as outlined below:

- Data cleaning/preparation – Try to understand data quality and find out missing data using available sources like internet. Format data and cleaning.
- Data Analysis: Understand the given data set and it's attributes
- Insights Analysis: Analyse each insights requirement in detail and prepare MS Excel formula or functions to extract insights. Select optimal and efficient approach.
- Extract insights: Use MS Excel as a tool to extract new insights as required including visuals/charts. Create Dashboard as required using slicers.
- Review: Review and cross check output to verify it matches with the requirements/insights required
- Document: Document the insights and results to be shared across business teams

For link to analysis files folder (Excel file), please follow this link **Analysis-Files**

Data Preparation/Cleaning

Overall data quality was fairly good except few missing values in important features like engine horse power. Based on analysis, I found and corrected below listed data errors/missing values with data from available sources on internet since missing data can impact analysis.

- Fuel type blank set to 'regular unleaded ' for 3 rows for 'Suzuki-Verona' with year 2004
- Horse power was blank for many rows – so corrected as listed below:
  - Engine HP set to 118 for 3 rows with Model 'FIAT – 500e'
  - Engine HP set to 168 for 4 rows with Model 'Ford – Escape'
  - Engine HP set to 335 for 4 rows with Model 'Lincoln-Continental' with year 2017
  - Engine HP set to 123 for 2 rows with Model 'Honda – FitEV' with year 2013/2014
  - Engine HP set to 143 for 3 rows with Model 'Ford – Focus'
  - Engine HP set to 203 for 6 rows with Model 'Ford – Free star' with year 2005
  - Engine HP set to 305 for 6 rows with Model 'Chevrolet – Impala' with year 2015/2016
  - Engine HP set to 147 for 10 rows with Model 'Nissan – Leaf'
  - Engine HP set to 109 for 4 rows with Model 'Kia – Soul EV'
  - 17 rows with blank in engine HP for Model 'Tesla – Model S', set to values from internet sources
  - There are other few rows with blank HP, it has been updated with data from Internet sources
- For few rows with electric fuel type, number of cylinders was blank, it has been set to zero.
- For Model 'Mazda – RX7' and 'Mazda – RX-8' – number of cylinders set to zero.
- Blank value in Number of doors set to 3 for Model 'Ferrari-FF' with year 2013
- Blank value in Number of doors set to 4 for Model 'Tesla – Model S' with year 2016

Tech Stack Used

Data Analytics tool: Microsoft Excel (Office 365)
Car related data has been provided in Excel format and excel is further used for data cleaning, analysis and creating visuals/charts to demonstrate insights. Excel is user friendly and functionally rich tool to analyze, visualize and report the data insights. Excel provides functions like Pivot tables and regression analysis which would help analysis.

Operating System: Microsoft Windows 11 Version 22H2

Documentation: Microsoft office 365 (Power Point) & Acrobat PDF

Prepared By –
Nilesh Kulkarni

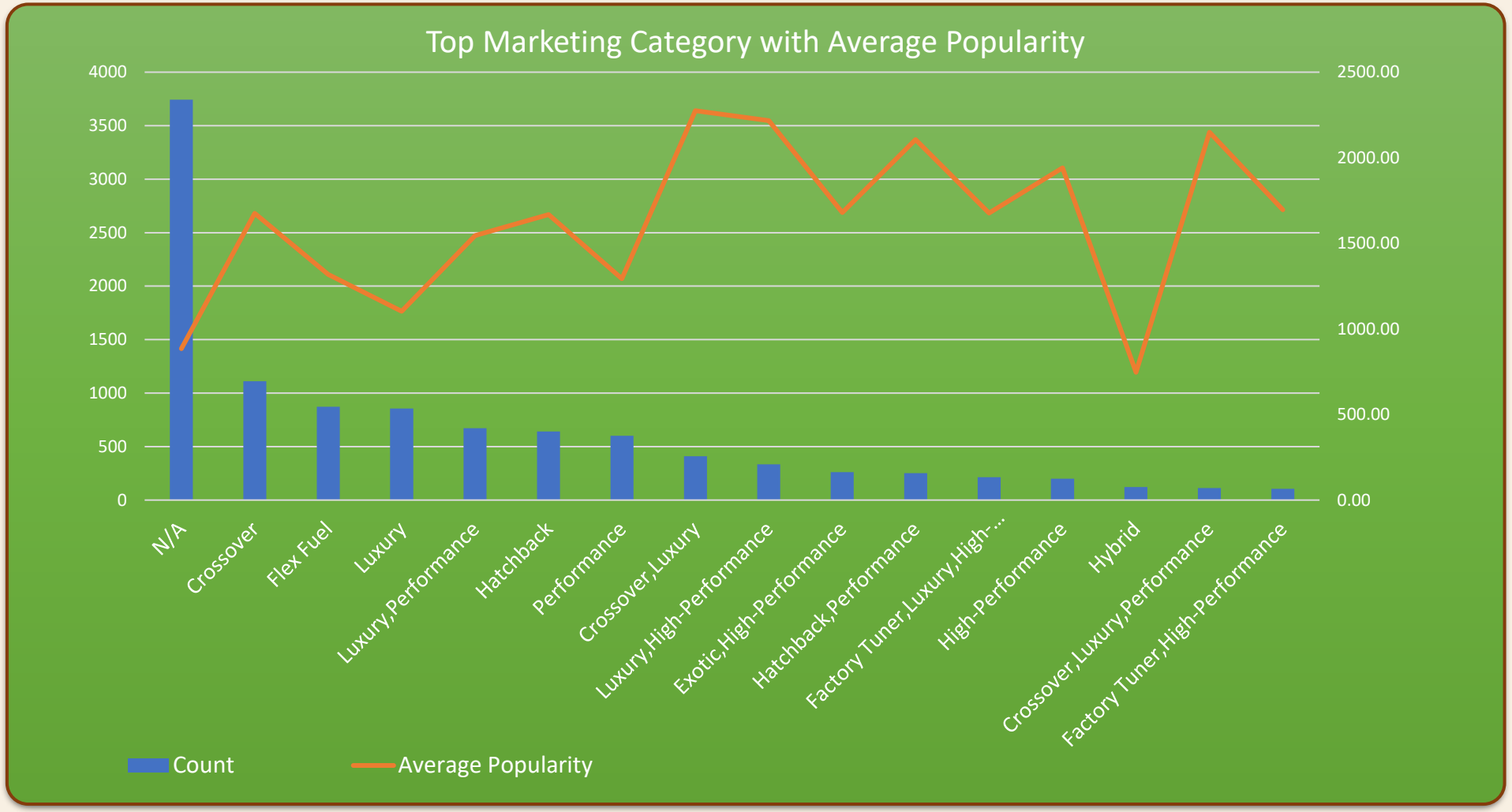
Task 1.A - Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

For pivot table, Please refer to XLS Datasheet at given link. The snip of Pivot table is as shown where we can expand each car market category and check count of models as well as average popularity. The count and average popularity is also summed up for each market category

Pivot Table with Car Market Category with Average Popularity		
Row Labels	Count of Model	Average of Popularity
N/A	3742	1676.89
Crossover	1110	1545.26
Flex Fuel	872	2217.30
Luxury	855	1102.66
Luxury,Performance	673	1292.62
Hatchback	641	1318.87
Performance	601	1348.87
Crossover,Luxury	410	884.55
Luxury,High-Performance	334	1668.02
Exotic,High-Performance	261	1271.33
Hatchback,Performance	252	1039.65
Factory Tuner,Luxury,High-Performance	215	2133.37
High-Performance	199	1821.45
Hybrid	123	2105.57
Crossover,Luxury,Performance	113	1344.85
Factory Tuner,High-Performance	106	1941.42
Factory Tuner,Performance	92	1695.70
Flex Fuel,Performance	87	1680.47
Diesel	84	1730.90
Exotic,Luxury,High-Performance	79	467.08
Hatchback,Hybrid	72	2121.25
Crossover,Hatchback	72	1675.69
Crossover,Performance	69	2585.96
Crossover,Flex Fuel	64	2073.75

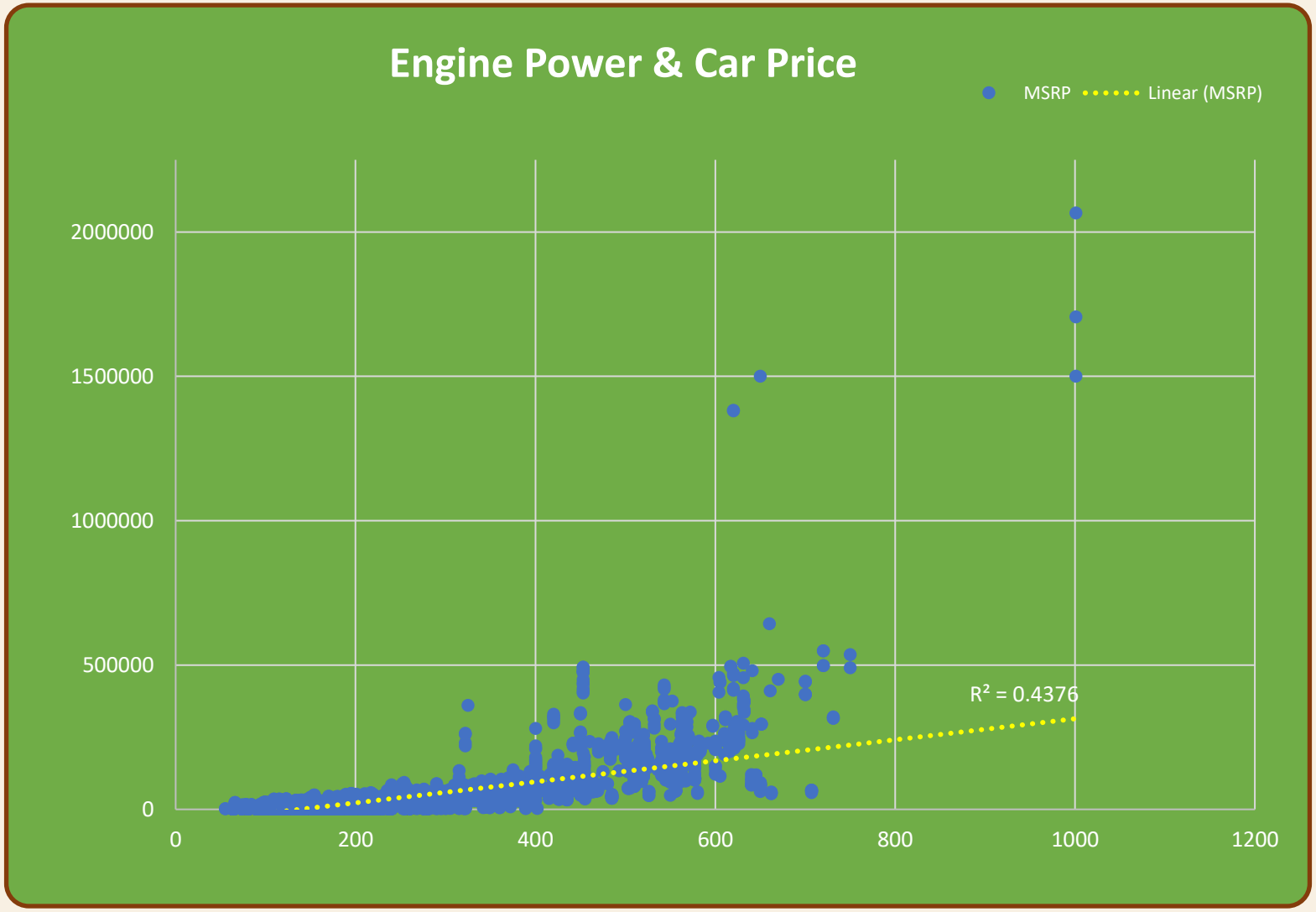
Task 1.B - Create a combo chart that visualizes the relationship between market category and popularity

For simplicity of visualization, only the Car Marketing Categories with count more than 100 are shown in below combo chart – which shows prominent market categories with average popularity



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.

Scatter chart for Engine Power & Car Price is as shown – there is fairly positive relationship between Horse power & price wit R^2 value as 0.43 which denote that 43% variation in care price can be accounted for Variation in engine power.



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.

Regression analysis allows us to analyze the relationship between the dependent variable (car price) and multiple independent variables (car features).

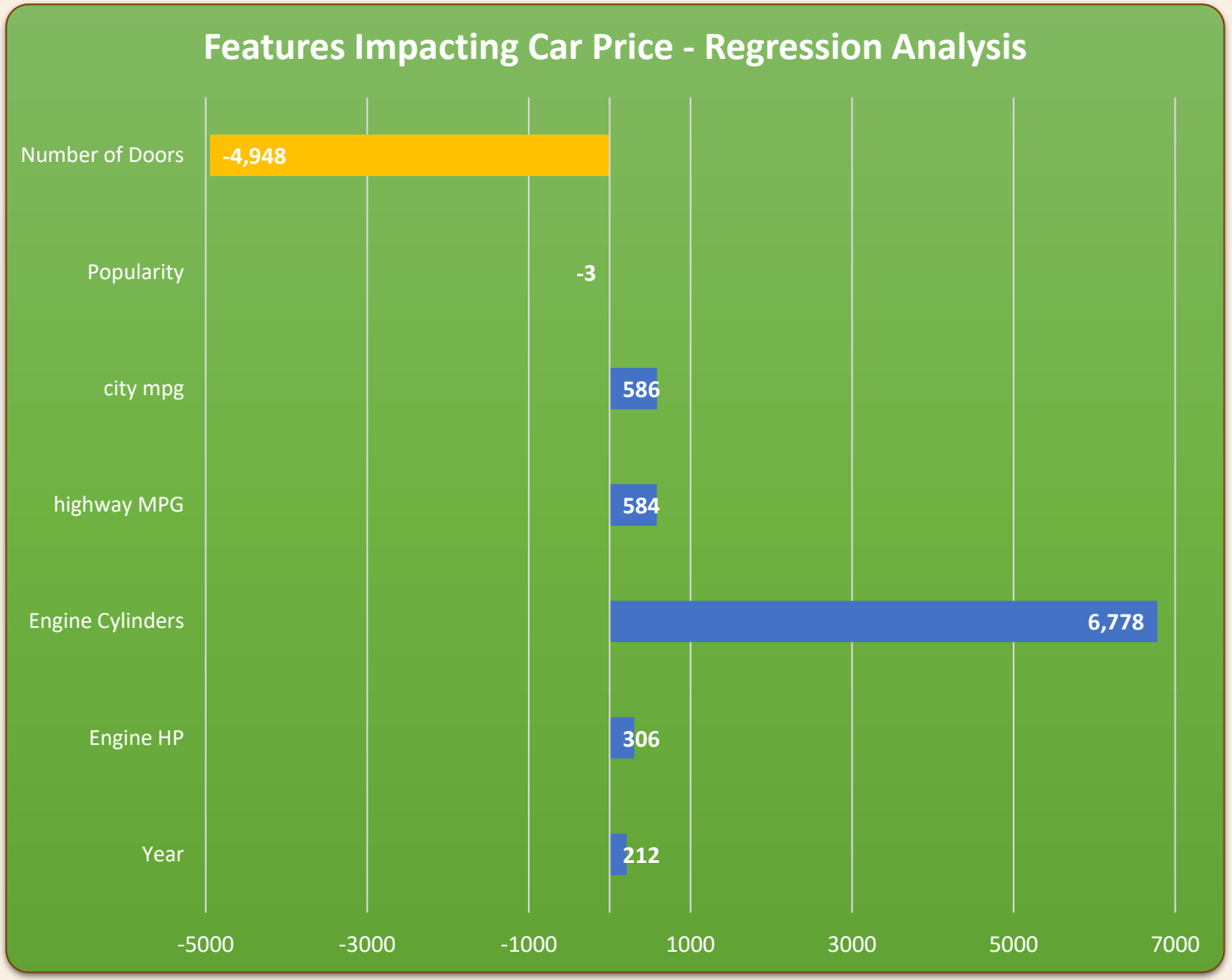
Multiple linear regression Analysis has been performed on 6 numeric features that can impact price. The summarized results of regression analysis are as given below:

Regression Statistics	
Multiple R	0.682515559
R Square	0.465827488
Adjusted R Square	0.465513428
Standard Error	43944.91725
Observations	11914

Column1	*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-505087.8366	142592.837	-3.54217	0.000398	-784593.076	-225582.597	-784593.0761	-225582.5971
Year	212.1361675	71.4638582	2.96844	0.002999	72.0553386	352.2169964	72.05533861	352.2169964
Engine HP	305.8335736	7.150082397	42.77343	0	291.818245	319.8489024	291.8182448	319.8489024
Engine Cylinders	6778.090772	447.0394218	15.16218	1.9E-51	5901.82052	7654.36102	5901.820524	7654.36102
highway MPG	583.6436569	104.2200627	5.600108	2.19E-08	379.35532	787.9319942	379.3553197	787.9319942
city mpg	585.863009	97.95184672	5.981133	2.28E-09	393.861398	777.8646197	393.8613984	777.8646197
Popularity	-3.343984909	0.281874069	-11.8634	2.79E-32	-3.8965041	-2.79146572	-3.896504101	-2.791465717
Number of Doors	-4948.441239	488.6514501	-10.1267	5.26E-24	-5906.27786	-3990.60462	-5906.277855	-3990.604622

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.

Regression Analysis showed that Features like Engine Cylinders, MPG, Year and Horse Power has positive influence on car Price i.e. price increase along with these features.
Features like popularity and number of door has negative impact on car price as shown.



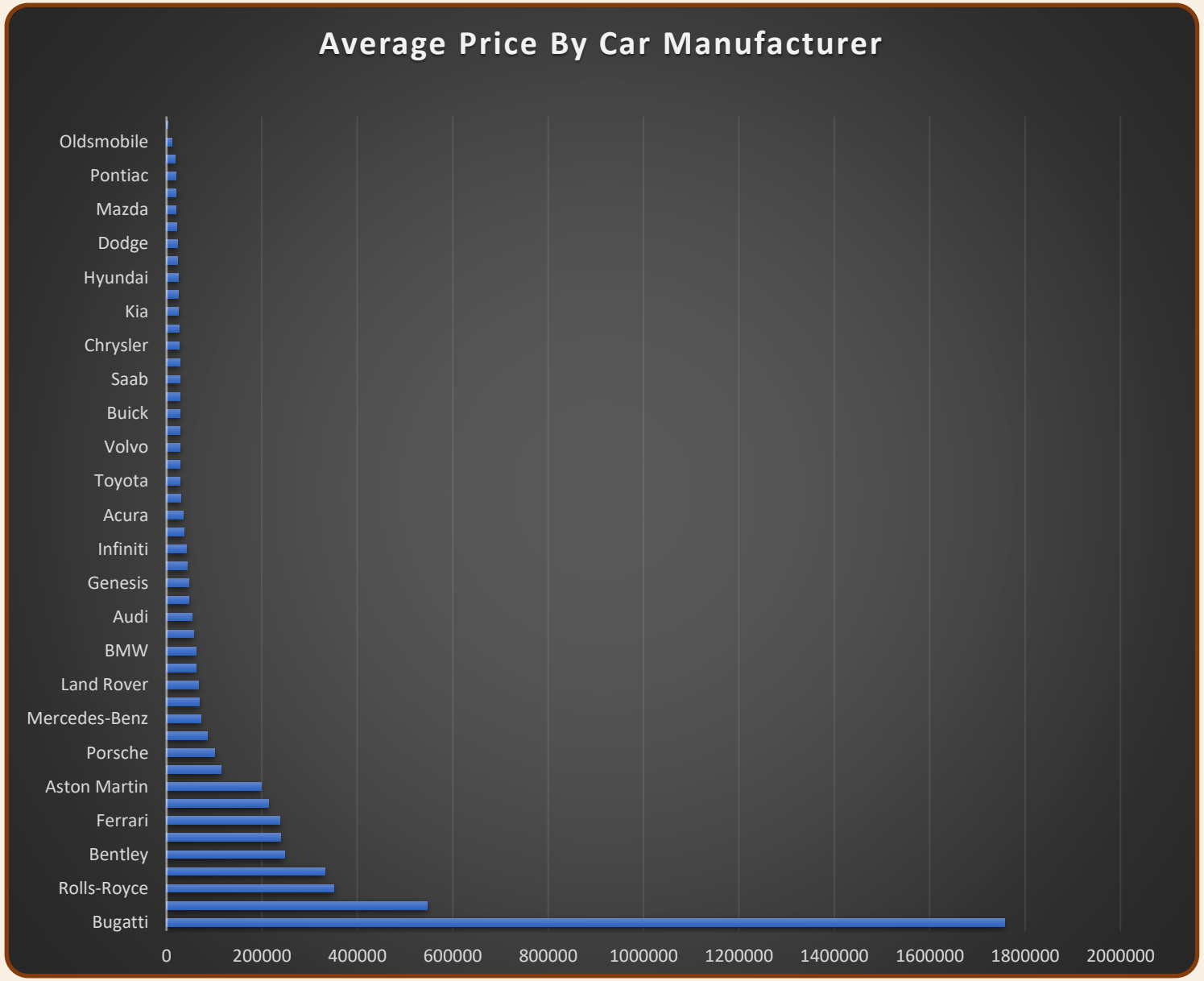
Task 4 A - Create a pivot table that shows the average price of cars for each manufacturer.

Pivot table for average car price by Car manufacturer is created and please refer to XLS for the same on the drive. A snapshot of Pivot table is as shown here.

Pivot Table for Average Price by Manufacters	
Manufacturer	Average of MSRP
Bugatti	1757223.67
Maybach	546221.88
Rolls-Royce	351130.65
Lamborghini	331567.31
Bentley	247169.32
McLaren	239805.00
Ferrari	238218.84
Spyker	213323.33
Aston Martin	197910.38
Maserati	114207.71
Porsche	101622.40
Tesla	85255.56
Mercedes-Benz	71476.23
Lotus	69188.28
Land Rover	67823.22
Alfa Romeo	61600.00
BMW	61546.76
Cadillac	56231.32
Audi	53452.11
Lexus	47549.07
Genesis	46616.67
Lincoln	42839.83
Infiniti	42394.21

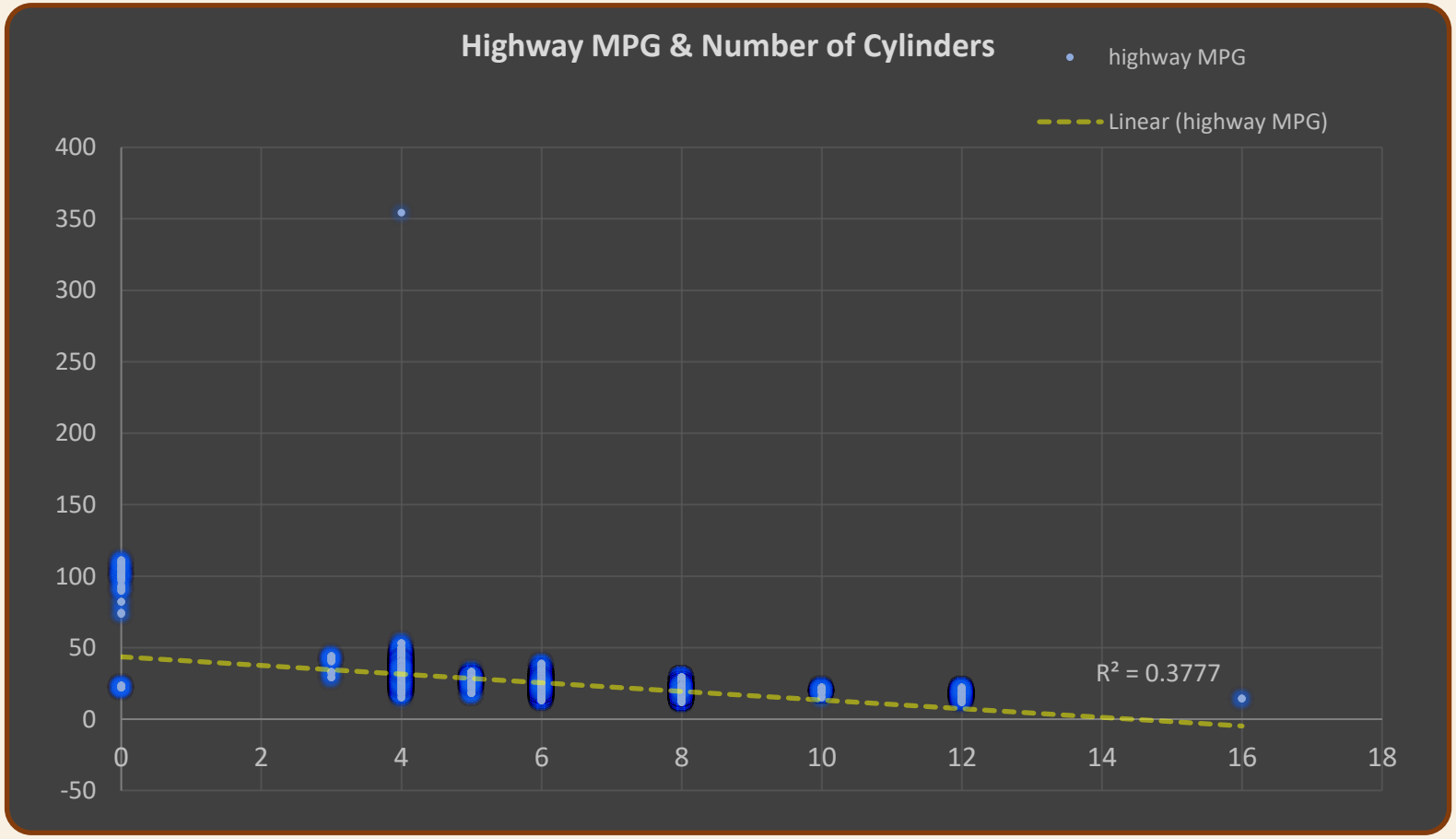
Task 4 B - **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.

Bar chart for average price by Manufacturers is as shown –it shows how premium brand and popular brands are positioned in terms of price. We can clearly see that premium brands charge much higher price compared to other brands



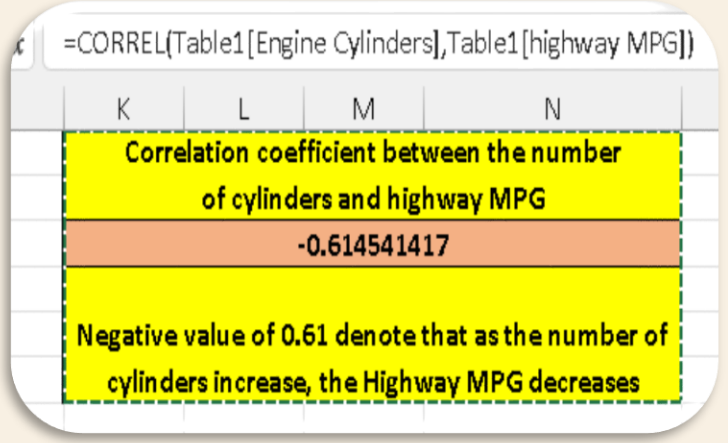
TASK 5A - Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship & assess its significance.

Scatter plot for highway MPG & Number of Cylinders shows that as the number of cylinders increase, the highway MPG decreases. R2 value 0.3777 denote that 37% decrease in highway MPG can be accounted for number of cylinders increase



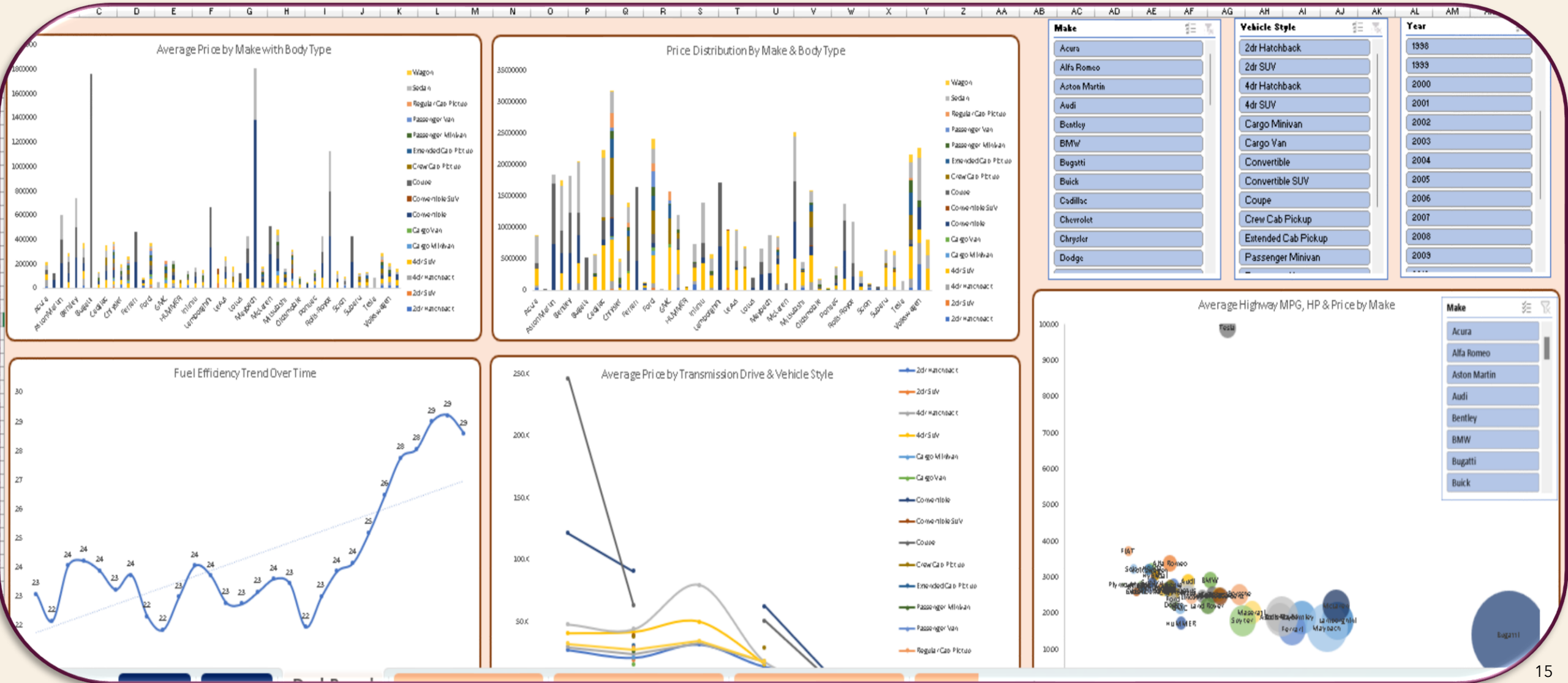
TASK 5 B - Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

Corelation co-efficient between Highway MPG & number of cylinders calculated using CORREL function in Excel as Shown below:



Dashboard

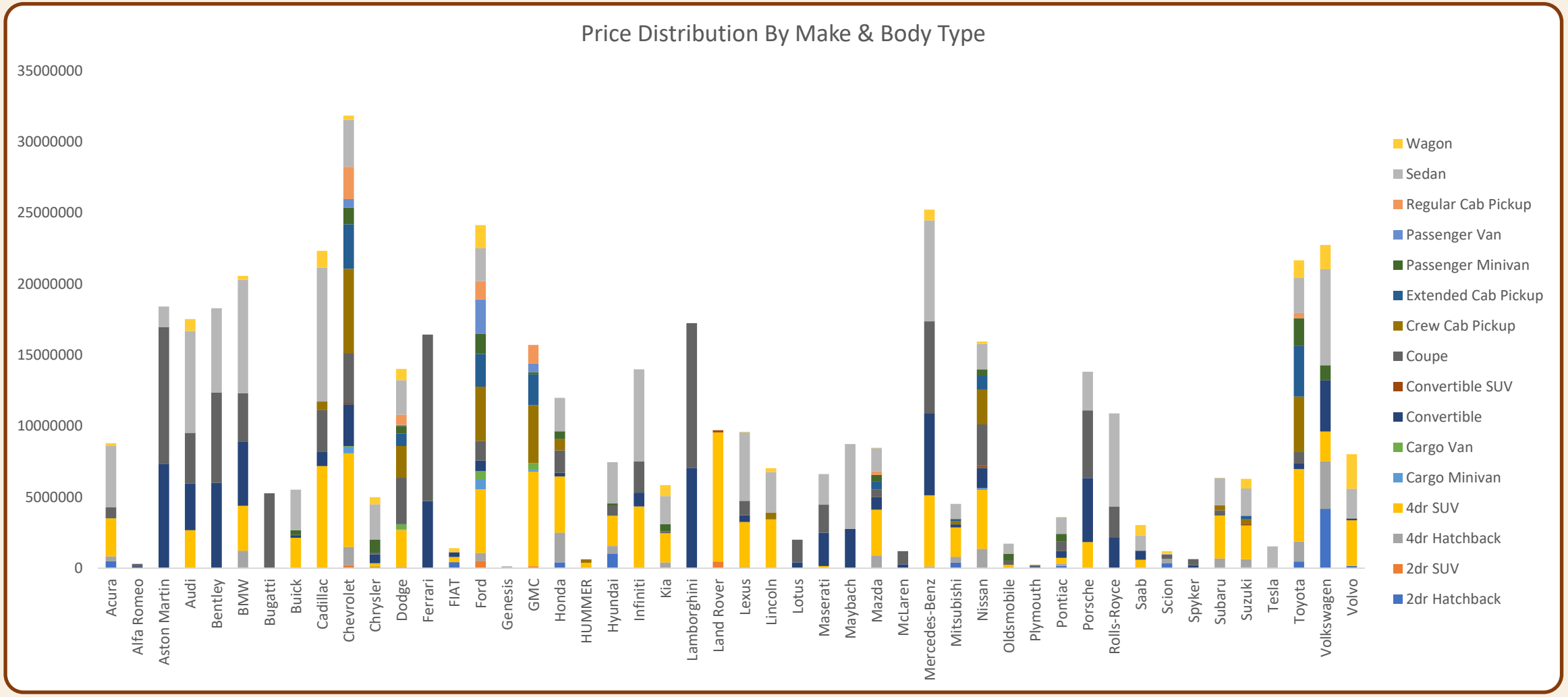
Please refer to excel datasheet for detailed dashboard, a snapshot of dashboard is as shown below



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

Total Price distribution by Make and Body type is as shown in stacked column chart below

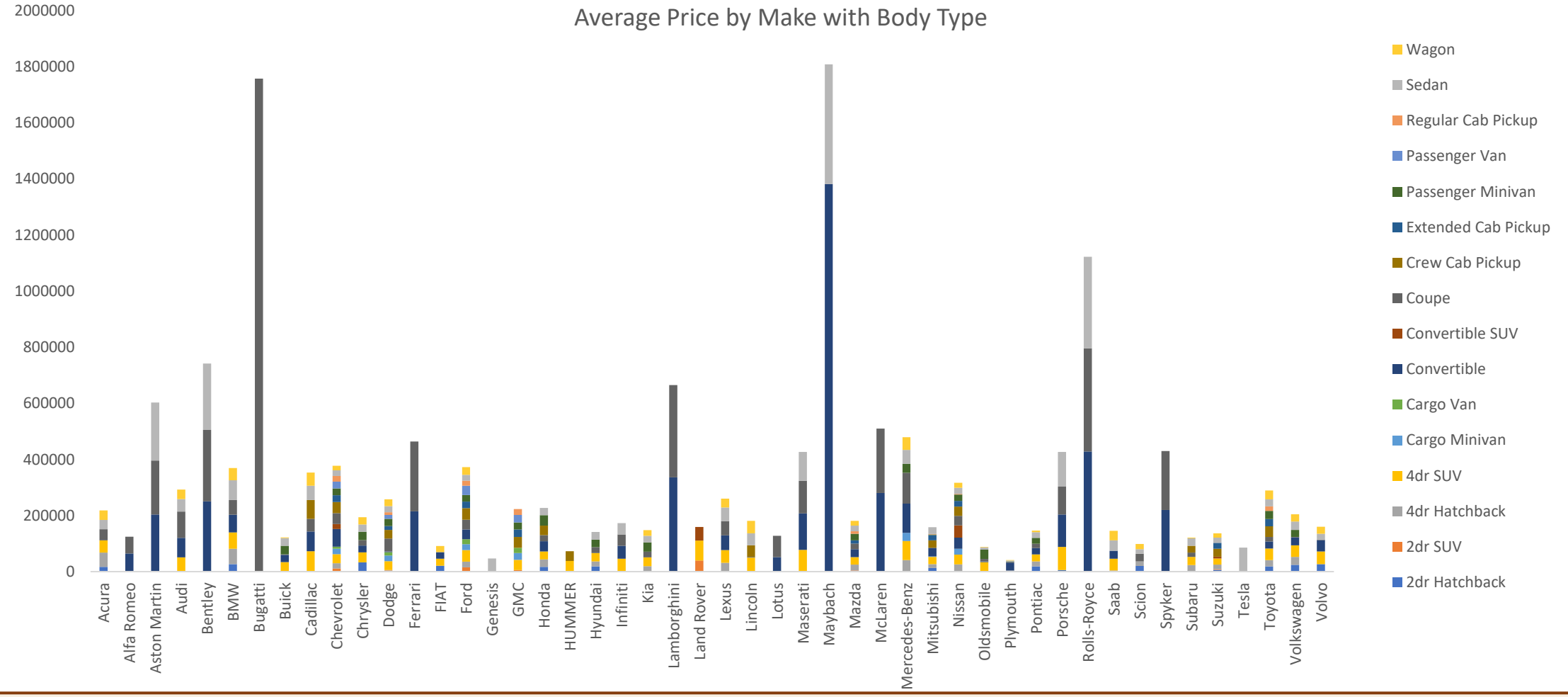
Chevrolet & Mercedes-Benz has highest total price (might be due to more number of models in given data)



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

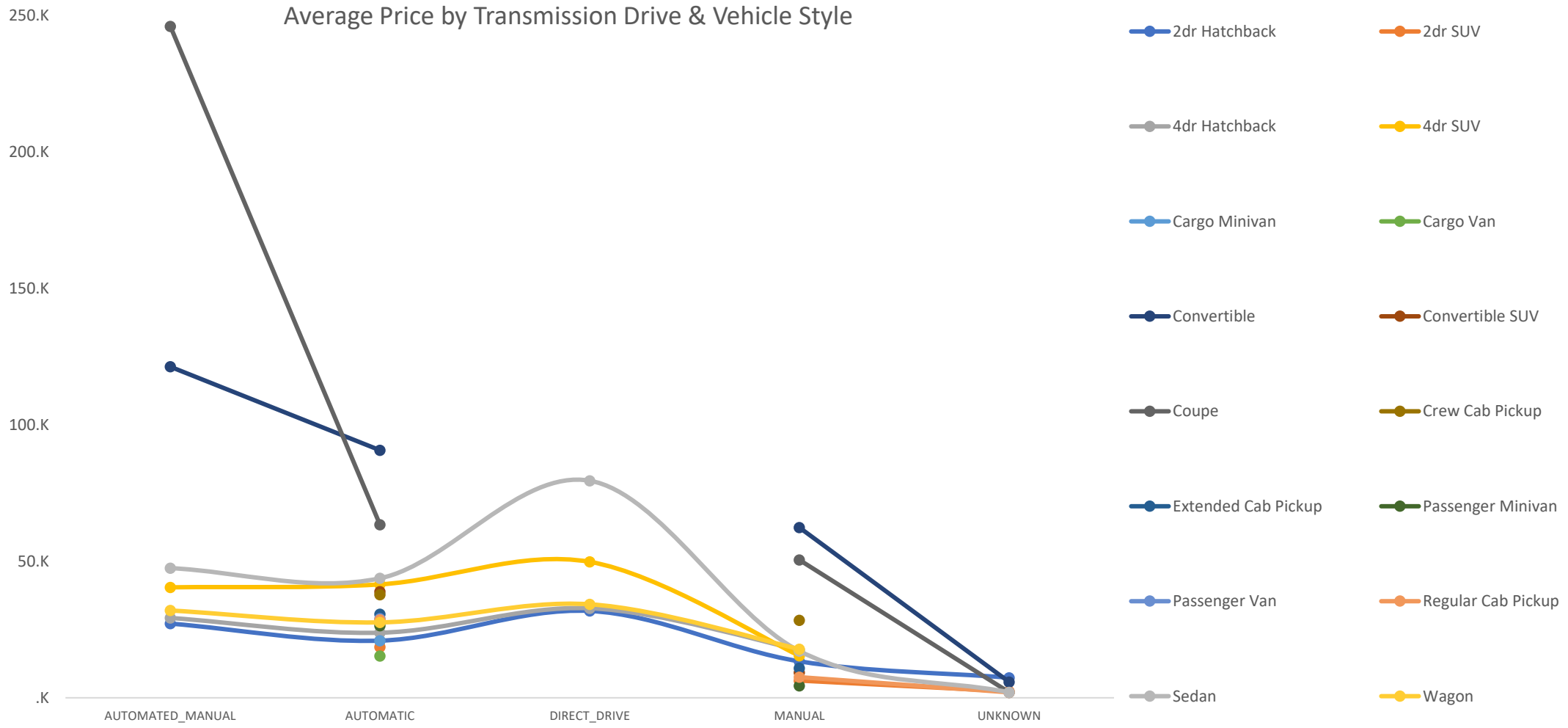
Average Price distribution by Make and Body type is as shown in stacked column chart below

Brand ‘Dodge’ & “Mazda’ has lowest average price and ‘Maybach’ & ‘Bagatti’ has highest average price



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

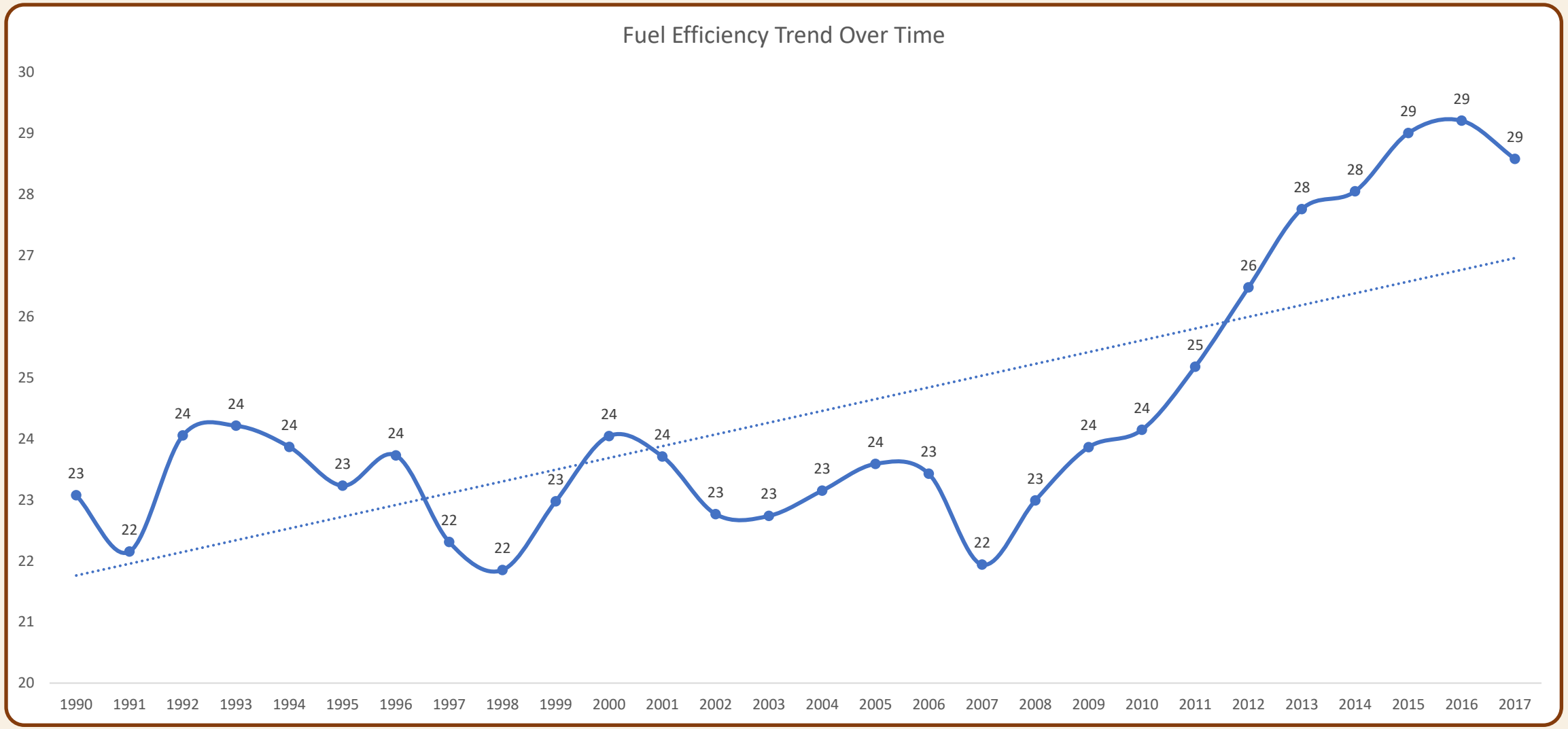
As shown in chart below, it looks like MANUAL transmission type has range of least MSRP(price) and DIRECT-DRIVE as well as Automated-Manual has generally high MSRP.



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

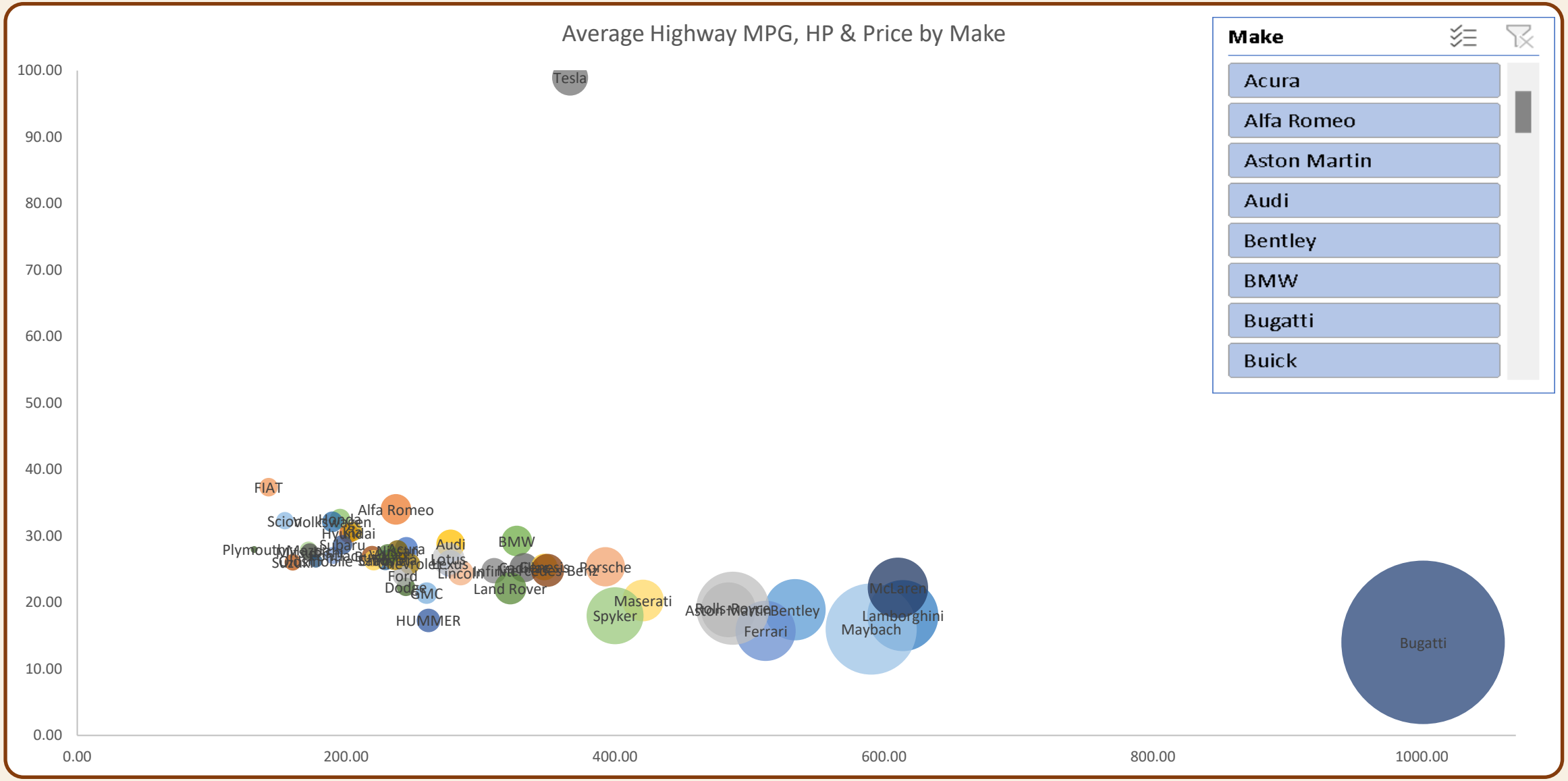
Fuel efficiency data for all the years with trendline is as shown below.

Fuel efficiency is generally shows increasing trend over 2-3 decades primarily might be due to technological advancement and focus of market on fuel economy improvements



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

Bubble chart with MPG & Horsepower is as shown with a slicer to enable selection by Make – as expected ‘Bugatti’ make showing highest price with highest Horsepower & low MPG.



Insights – Summary/Conclusion

With the data analysis and insights extraction, we can draw below listed summarized insights which can help business to understand critical factors about car market and gain actionable insights:

- The input data has many missing values in Engine HP, number of door and cylinders. So there is certain scope for improving data quality
- Cross-over market category in general has high popularity with high model count
- Approximately 43% variation in car price can be accounted for increase in engine power
- In-general, car price increases with increase in number of cylinders, MPG, year of make and deceases with popularity & number of doors
- MPG has been seen as decreasing with increase in number of cylinders in car engine
- Fuel efficiency is generally showing an increasing trend over 2-3 decades primarily might be due to technological advancement and focus of market on fuel economy improvements
- Premium brands like 'Bugatti' and others typically has low MPG and high power
- Data analysis shows that MANUAL transmission type has range of least MSRP(price) and DIRECT-DRIVE as well as Automated-Manual has generally high MSRP

In summary, the analysis shows various factors impacting car price as well as inter-related factors like MPG variation due to engine power and cylinders. The insights would be very helpful to business in future strategies and improving the overall car market dynamics.

Results – Personal Up-Skilling

The project helped me to get exposure to various new skills and areas in data analytics, specifically as listed below:

- Learned more about automobile and retail domain basics
- Got exposure to various excel functions like regression analysis which also has experience with using statistics' concepts
- Practical hands-on experience with creating dashboards with slicers
- Learned more about creating visuals in Excel for required insights

Overall, the project has been great learning and professionally enriching experience with Trainity.

THANK YOU

Nilesh Kulkarni

inileshkulkarni@gmail.com