

Analyzing the Impact of Car

Project Description: This data analysis project focuses on assisting a car manufacturer in optimizing pricing and product development decisions to maximize profitability and meet consumer demand. Using the "Car Features and MSRP" dataset, it aims to analyze car features, market categories, and pricing relationships. The project involves data cleaning, preprocessing, and advanced Excel skills to answer key questions. Insights from this analysis will enable the manufacturer to adapt to changing consumer preferences, enhance competitiveness, and improve profitability in the dynamic automotive industry.

Approach: The following steps have been implemented to analyze the dataset:

1) The first step involves Data Cleaning.

	A	B	C	D	E	F	G	H	I
1	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors
1909	Acura	ZDX	2011	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4
1910	Acura	ZDX	2011	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4
1911	Acura	ZDX	2012	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4
1912	Acura	ZDX	2012	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4
1913	Acura	ZDX	2012	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4
1914	Acura	ZDX	2013	premium unleaded (recommended)	300	6	AUTOMATIC	all wheel drive	4
1915	Lincoln	Zephyr	2006	regular unleaded	221	6	AUTOMATIC	front wheel drive	4
1916	0		0		0.02518046	0.582524	0.252440256	0	0.050386295
1917									
1918									
Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors	Market Category
BMW	Z3	2001	premium unleaded (required)	225	6	MANUAL	rear wheel drive	2	Luxury,Performance
BMW	Z3	2001	premium unleaded (required)	184	6	MANUAL	rear wheel drive	2	Luxury,Performance
BMW	Z3	2001	premium unleaded (required)	225	6	MANUAL	rear wheel drive	2	Hatchback,Luxury
BMW	Z3	2002	premium unleaded (required)	225	6	MANUAL	rear wheel drive	2	Luxury,Performance
BMW	Z3	2002	premium unleaded (required)	184	6	MANUAL	rear wheel drive	2	Luxury,Performance
BMW	Z3	2002	premium unleaded (required)	225	6	MANUAL	rear wheel drive	2	Hatchback,Luxury
BMW	Z4 M	2007	premium unleaded (required)	330	6	MANUAL	rear wheel drive	2	Factory Tuner
BMW	Z4 M	2007	premium unleaded (required)	330	6	MANUAL	rear wheel drive	2	Factory Tuner
BMW	Z4 M	2008	premium unleaded (required)	330	6	MANUAL	rear wheel drive	2	Factory Tuner
BMW	Z4 M	2008	premium unleaded (required)	330	6	MANUAL	rear wheel drive	2	Factory Tuner
BMW	Z4	2014	premium unleaded (required)	240	4	MANUAL	rear wheel drive	2	Luxury,Performance
BMW	Z4	2014	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxury,High-Performance
BMW	Z4	2014	premium unleaded (required)	335	6	AUTOMATED_MANUAL	rear wheel drive	2	Luxury,High-Performance
BMW	Z4	2015	premium unleaded (required)	240	4	MANUAL	rear wheel drive	2	Luxury,Performance
BMW	Z4	2015	premium unleaded (required)	300	6	AUTOMATED_MANUAL	rear wheel drive	2	Luxury,High-Performance
BMW	Z4	2015	premium unleaded (required)	335	6	AUTOMATED_MANUAL	rear wheel drive	2	Luxury,High-Performance
BMW	Z4	2016	premium unleaded (required)	300	6	AUTOMATED_MANUAL	rear wheel drive	2	Luxury,High-Performance
BMW	Z4	2016	premium unleaded (required)	240	4	MANUAL	rear wheel drive	2	Luxury,Performance
BMW	Z4	2016	premium unleaded (required)	335	6	AUTOMATED_MANUAL	rear wheel drive	2	Luxury,High-Performance
BMW	Z8	2001	premium unleaded (required)	394	8	MANUAL	rear wheel drive	2	Exotic,Luxury,High-Performance
BMW	Z8	2002	premium unleaded (required)	394	8	MANUAL	rear wheel drive	2	Exotic,Luxury,High-Performance
BMW	Z8	2003	premium unleaded (required)	394	8	MANUAL	rear wheel drive	2	Exotic,Luxury,High-Performance
Acura	ZDX	2011	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4	Crossover,Hatchback
Acura	ZDX	2011	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4	Crossover,Hatchback
Acura	ZDX	2011	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4	Crossover,Hatchback
Acura	ZDX	2012	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4	Crossover,Hatchback
Acura	ZDX	2012	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4	Crossover,Hatchback
Acura	ZDX	2012	premium unleaded (required)	300	6	AUTOMATIC	all wheel drive	4	Crossover,Hatchback
Acura	ZDX	2013	premium unleaded (recommended)	300	6	AUTOMATIC	all wheel drive	4	Crossover,Hatchback
Lincoln	Zephyr	2006	regular unleaded	221	6	AUTOMATIC	front wheel drive	4	Luxury
	0		0		0	0	0	0	0

I used MODE to replace NULL values with Qualitative data and MEDIAN for Quantitative data.

- 2) Popularity of the Car Model in different Market Categories: A Pivot Table has been created to summarize the number of car models in each market category and their corresponding popularity scores. Then, a combo chart has been created to visualize the relationship.

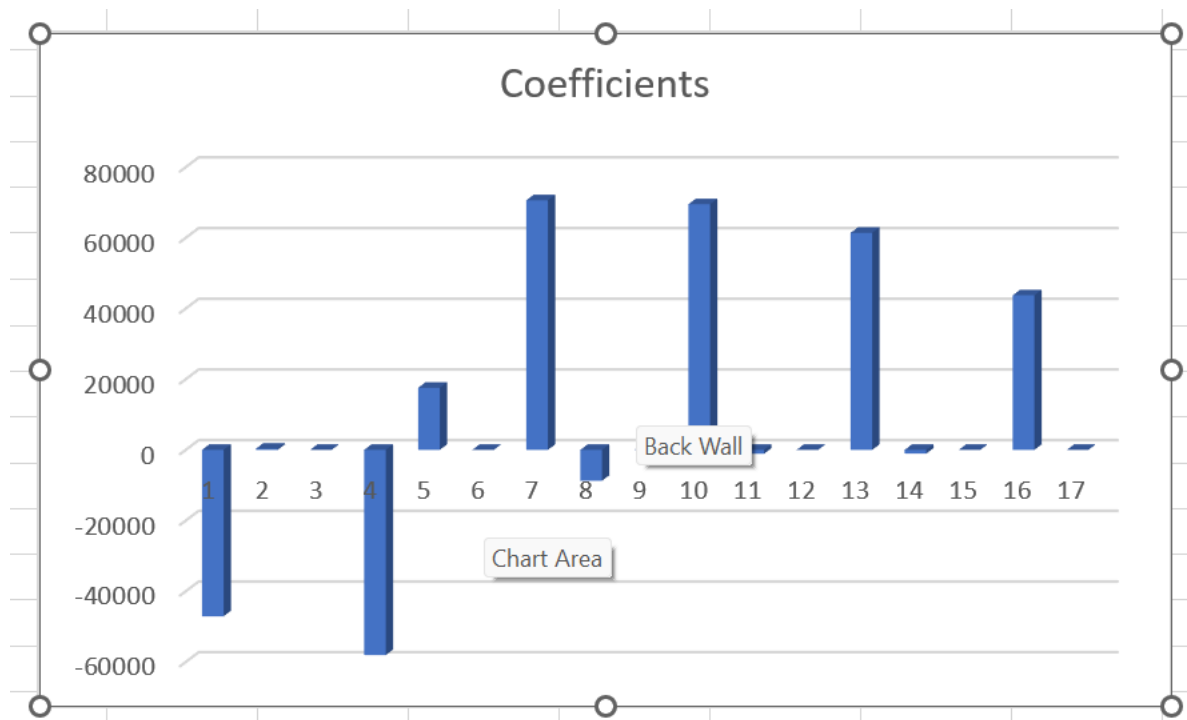
	A	B	C
1	Row Labels	Sum of Popularity	Count of Model
2	Crossover	1715242	1110
3	Crossover,Diesel	6111	7
4	Crossover,Exotic,Luxury,High-Performance	238	1
5	Crossover,Exotic,Luxury,Performance	238	1
6	Crossover,Factory Tuner,Luxury,High-Performance	47410	26
7	Crossover,Factory Tuner,Luxury,Performance	13037	5
8	Crossover,Factory Tuner,Performance	840	4
9	Crossover,Flex Fuel	132720	64
10	Crossover,Flex Fuel,Luxury	11732	10
11	Crossover,Flex Fuel,Luxury,Performance	9744	6
12	Crossover,Flex Fuel,Performance	33942	6
13	Crossover,Hatchback	120650	72
14	Crossover,Hatchback,Factory Tuner,Performance	12054	6
15	Crossover,Hatchback,Luxury	1428	7
16	Crossover,Hatchback,Performance	12054	6
17	Crossover,Hybrid	107662	42
18	Crossover,Luxury	362665	410
19	Crossover,Luxury,Diesel	73080	34
20	Crossover,Luxury,High-Performance	9335	9
21	Crossover,Luxury,Hybrid	15142	24
22	Crossover,Luxury,Performance	151968	113
23	Crossover,Luxury,Performance,Hybrid	7832	2
24	Crossover,Performance	178431	69
25	Diesel	145396	84
26	Diesel,Luxury	116025	51
27	Exotic,Factory Tuner,High-Performance	21974	21
28	Exotic,Factory Tuner,Luxury,High-Performance	26912	52
29	Exotic,Factory Tuner,Luxury,Performance	1560	3
30	Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	6760	13
31	Exotic,Flex Fuel,Luxury,High-Performance	5720	11
32	Exotic,High-Performance	331818	261
33	Exotic,Luxury	1352	12
34	Exotic,Luxury,High-Performance	36899	79
35	Exotic,Luxury,High-Performance,Hybrid	204	1
36	Exotic,Luxury,Performance	7813	36

	A	B	C
36	Exotic,Luxury,Performance	7813	36
37	Exotic,Performance	13910	10
38	Factory Tuner,High-Performance	205790	106
39	Factory Tuner,Luxury	1234	2
40	Factory Tuner,Luxury,High-Performance	458674	215
41	Factory Tuner,Luxury,Performance	43816	31
42	Factory Tuner,Performance	156004	92
43	Flex Fuel	1933488	872
44	Flex Fuel,Diesel	90512	16
45	Flex Fuel,Factory Tuner,Luxury,High-Performance	258	1
46	Flex Fuel,Hybrid	310	2
47	Flex Fuel,Luxury	29115	39
48	Flex Fuel,Luxury,High-Performance	29004	33
49	Flex Fuel,Luxury,Performance	38642	28
50	Flex Fuel,Performance	146201	87
51	Flex Fuel,Performance,Hybrid	310	2
52	Hatchback	845393	641
53	Hatchback,Diesel	12222	14
54	Hatchback,Factory Tuner,High-Performance	15667	13
55	Hatchback,Factory Tuner,Luxury,Performance	7982	9
56	Hatchback,Factory Tuner,Performance	47499	22
57	Hatchback,Flex Fuel	39599	7
58	Hatchback,Hybrid	152730	72
59	Hatchback,Luxury	63457	46
60	Hatchback,Luxury,Hybrid	1362	3
61	Hatchback,Luxury,Performance	59513	38
62	Hatchback,Performance	261991	252
63	High-Performance	362468	199
64	Hybrid	258985	123
65	Luxury	942772	855
66	Luxury,High-Performance	557118	334
67	Luxury,High-Performance,Hybrid	6826	12
68	Luxury,Hybrid	35029	53
69	Luxury,Performance	869930	Vertical (
70	Luxury,Performance,Hybrid	25665	11
71	N/A	6274920	3742
72	Performance	810673	601
73	Performance,Hybrid	155	1
74	Grand Total	18525212	11914

4) Relationship between Car Features and Price Level: To understand this, I have performed regression analysis to understand the importance of each car feature on the price level and further created a bar chart between each variable's coefficient to visualize their relative importance.

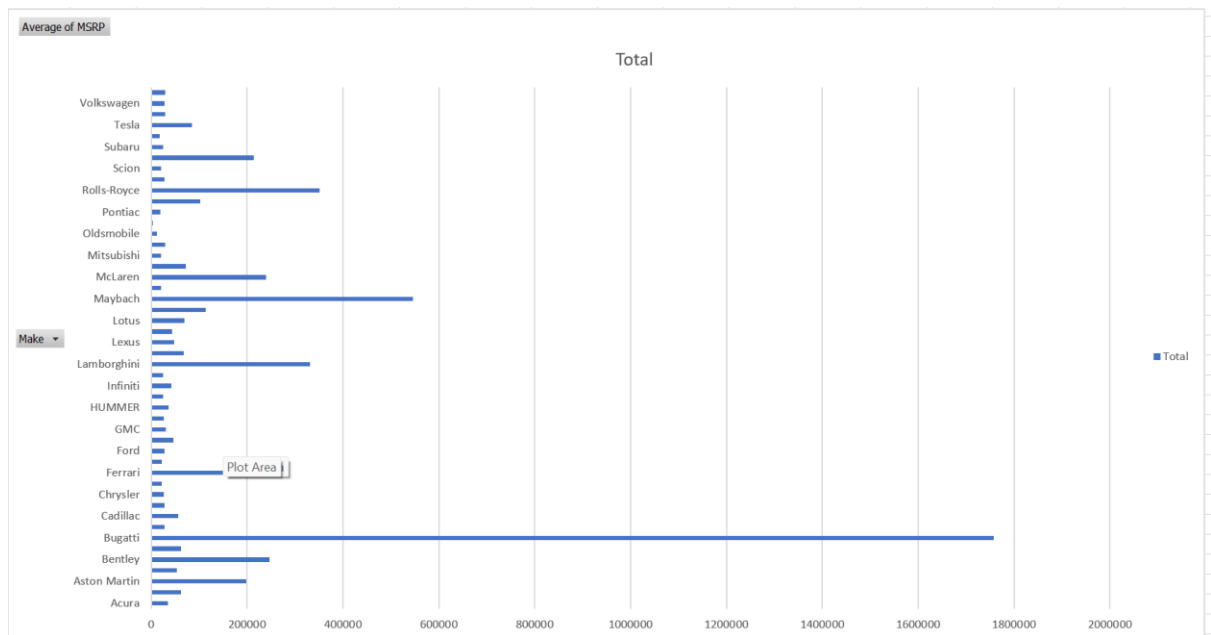
1	SUMMARY OUTPUT for Engine HP								
2									
3	Regression Statistics								
4	Multiple R	0.650095344							
5	R Square	0.422623957							
6	Adjusted R Square	0.422575486							
7	Standard Error	45675.97852							
8	Observations	11914							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	1	1.8191E+13	1.8191E+13	8719.26819	0			
13	Residual	11912	2.48519E+13	2086295014					
14	Total	11913	4.30429E+13						
15									
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	-47080.0158	1027.962929	-45.799332	0	-49094.9908	-45065.041	-49094.9908	-45065.041
18	Engine HP	353.6102858	3.786909018	93.3770218	0	346.1873262	361.033245	346.1873262	361.033245
19									
20									
21	SUMMARY OUTPUT for Engine Cylinders								
22									
23	Regression Statistics								
24	Multiple R	0.526274107							
25	R Square	0.276964436							
26	Adjusted R Square	0.276903738							
27	Standard Error	51113.81652							
28	Observations	11914							
29									
30	ANOVA								
31		df	SS	MS	F	Significance F			
32	Regression	1	1.19214E+13	1.1921E+13	4562.98491	0			
33	Residual	11912	3.11216E+13	2612622239					
34	Total	11913	4.30429E+13						
35									
36		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
37	Intercept	-58048.9261	1533.555224	-37.852518	3.563E-296	-61054.9446	-55042.908	-61054.9446	-55042.908
38	Engine Cylinders	17568.96241	260.0887681	67.5498698	0	17059.14599	18078.7788	17059.14599	18078.7788
39									
40	SUMMARY OUTPUT for No. of Doors								
41									
42	Regression Statistics								
43	Multiple R	0.128626626							
44	R Square	0.016544809							
45	Adjusted R Square	0.016462249							
46	Standard Error	59612.28492							
47	Observations	11914							
48									
49	ANOVA								
50		df	SS	MS	F	Significance F			
51	Regression	1	7.12137E+11	7.1214E+11	200.397298	3.96943E-45			
52	Residual	11912	4.23308E+13	3553624513					
53	Total	11913	4.30429E+13						
54									
55		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
56	Intercept	70616.70992	2189.961808	32.2456354	8.741E-219	66324.02748	74909.3924	66324.02748	74909.3924
57	Number of Doors	-8741.64247	617.5144279	-14.156175	3.9694E-45	-9952.0715	-7531.2134	-9952.0715	-7531.2134

SUMMARY OUTPUT for Highway MPG								
<i>Regression Statistics</i>								
Multiple R	0.160042679							
R Square	0.025613659							
Adjusted R Square	0.02553186							
Standard Error	59336.79348							
Observations	11914							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	1.10249E+12	1.1025E+12	313.13032	3.47759E-69			
Residual	11912	4.19404E+13	3520855061					
Total	11913	4.30429E+13						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	69507.42705	1721.963252	40.3652209	0	66132.09813	72882.756	66132.09813	72882.756
highway MPG	-1085.41364	61.3384385	-17.695489	3.4776E-69	-1205.64699	-965.1803	-1205.64699	-965.180297
SUMMARY OUTPUT for City MPG								
<i>Regression Statistics</i>								
Multiple R	0.157675722							
R Square	0.024861633							
Adjusted R Square	0.024779772							
Standard Error	59359.68696							
Observations	11914							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	1.07012E+12	1.0701E+12	303.702313	3.49385E-67			
Residual	11912	4.19728E+13	3523572436					
Total	11913	4.30429E+13						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	61403.70291	1312.071577	46.7990497	0	58831.82855	63975.5773	58831.82855	63975.5773
city mpg	-1054.51259	60.51007904	-17.427057	3.4939E-67	-1173.12222	-935.90296	-1173.12222	-935.902965
SUMMARY OUTPUT for Popularity								
<i>Regression Statistics</i>								
Multiple R	0.048476232							
R Square	0.002349945							
Adjusted R Square	0.002266193							
Standard Error	60040.95554							
Observations	11914							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	1.01148E+11	1.0115E+11	28.0584821	1.19801E-07			
Residual	11912	4.29418E+13	3604916343					
Total	11913	4.30429E+13						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	43737.07584	809.0095601	54.0624957	0	42151.28511	45322.8666	42151.28511	45322.8666
Popularity	-2.02091207	0.381518263	-5.2970258	1.198E-07	-2.76875012	-1.273074	-2.76875012	-1.27307403

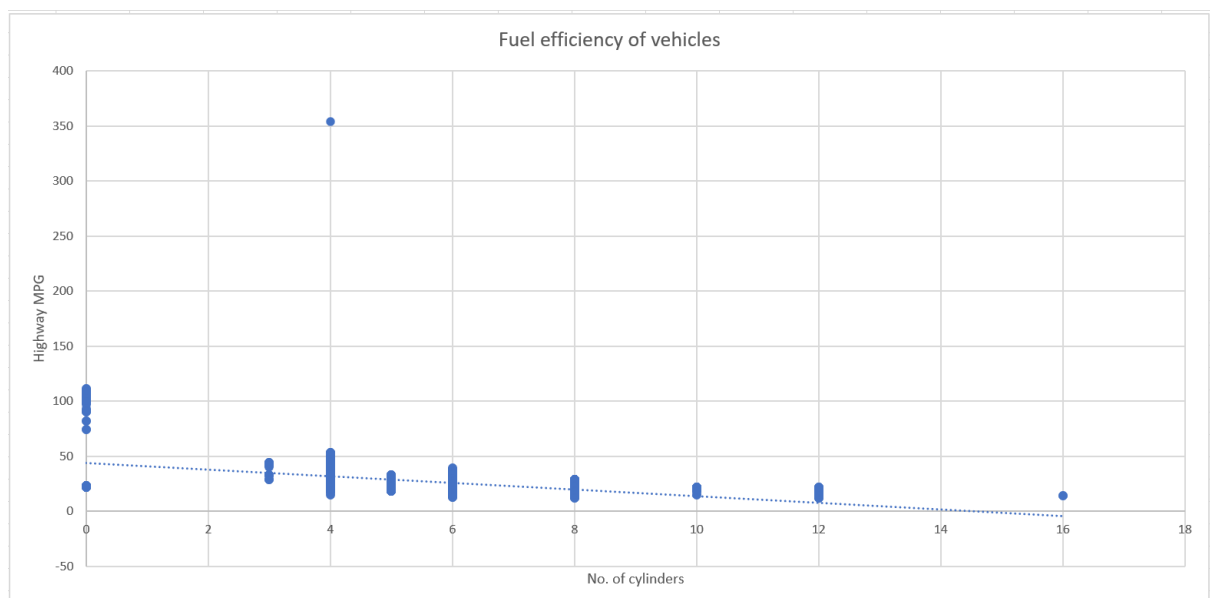


5) Average price of cars across different manufacturers: To understand this, I have used a Pivot Table and a Bar Chart between both variables for visualization purposes.

Row Labels	Average of MSRP
Acura	34887.5873
Alfa Romeo	61600
Aston Martin	197910.3763
Audi	53452.1128
Bentley	247169.3243
BMW	61546.76347
Bugatti	1757223.667
Buick	28206.61224
Cadillac	56231.31738
Chevrolet	28350.38557
Chrysler	26722.96257
Dodge	22390.05911
Ferrari	238218.8406
FIAT	22670.24194
Ford	27399.26674
Genesis	46616.66667
GMC	30493.29903
Honda	26674.34076
HUMMER	36464.41176
Hyundai	24597.0363
Infiniti	42394.21212
Kia	25310.17316
Lamborghini	331567.3077
Land Rover	67823.21678
Lexus	47549.06931
Lincoln	42839.82927
Lotus	69188.27586
Maserati	114207.7069
Maybach	546221.875
Mazda	20039.38298
McLaren	239805
Mercedes-Benz	71476.22946
Mitsubishi	21240.53521
Nissan	28583.4319
Oldsmobile	11542.54
Plymouth	3122.902439
Pontiac	19321.54839
Porsche	101622.3971
Rolls-Royce	351130.6452
Saab	27413.5045
Scion	19932.5
Spyker	213323.3333
Subaru	24827.50391
Suzuki	17907.20798
Tesla	85255.55556
Toyota	29030.01609
Volkswagen	28102.38072
Volvo	28541.16014
Grand Total	40594.73703



6) Relationship between Fuel Efficiency and No. of Cylinders used:
A scatter plot has been created for visualization purposes and a trendline is added to determine the slope. Further, correlation coefficients between the variables are determined to quantify the strength.



Formula:

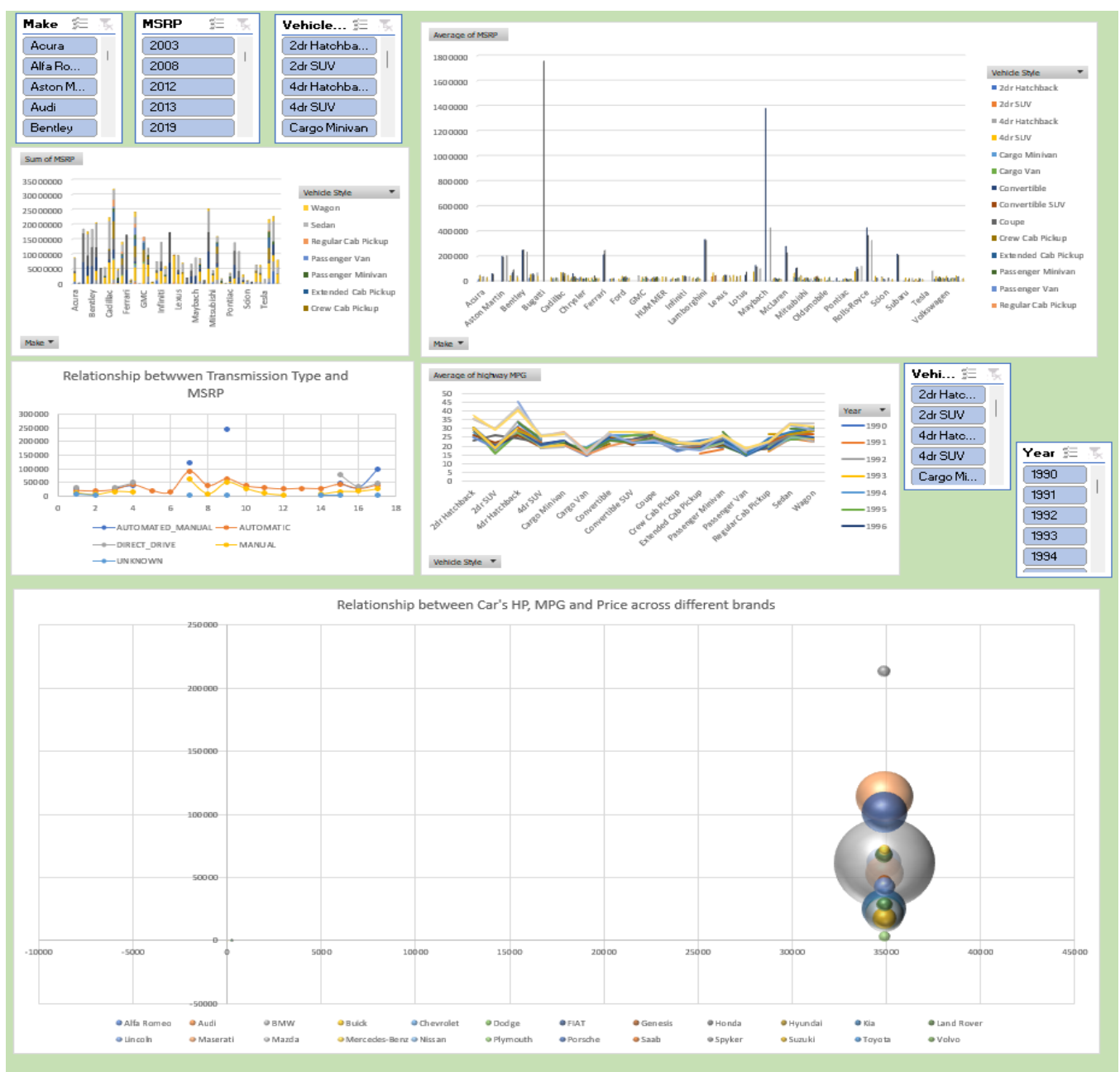
```
=CORREL(Car_data!F2:F11915,Car_data!M2:M11915)
```

Result:

0

7) To understand these further relations, a dashboard has been created.

- Distribution of car prices by their Body and Style
- Body Style and their MSRP
- Different features and their MSRP
- Efficiency of cars with different models
- Distribution of car's HP, MPG, and Price across different brands



Tech-Stack Used: For this project, I have used MS Excel (2019 version) due to its user-friendly interface and my expertise in it.

Insights:

- The maximum number of cars do not belong to any such market category and the popularity of these cars is the highest.
- 42.26% variance of the MSRP can be explained by Engine HP indicating a moderate correlation between these variables.
- A car's Engine HP has the highest relation with its MSRP as compared to other variables.
- The highest average MSRP is observed in "Bugatti" with up to \$1757223.67 and the lowest average MSRP is observed in "Plymouth" with up to \$3122.90.
- There is no predictable relationship found between Highway MPG and the Number of Cylinders used.
- To get further insights, Filters of each chart on the Dashboard can be used to get the required insights.

Result: This project helped me to learn and implement various concepts of Advanced Statistics. The dataset has considered a few of the factors used in determining the price and further market research will be helpful.

Link:

[Car_data.xlsx](#)