

Analyzing the Impact of Car Features on Price and Profitability

Final Project-3

by Nazireen Sania

Introduction

The dataset contains information on various car models and their specifications, and is titled "Car Features and MSRP". It was collected and made available on Kaggle by Cooper Union, a private college located in New York City.

- Number of observations: 11,159
- Number of variables: 16

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

Data Analysis Tasks

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

Row Labels	Count of Popularity
Crossover	1103
Crossover,Luxury	410
Exotic,High-Performance	252
Flex Fuel	872
Hatchback	614
Hatchback,Performance	252
Luxury	851
Luxury,High-Performance	334
Luxury,Performance	673
N/A	3728
Performance	584
Grand Total	9673

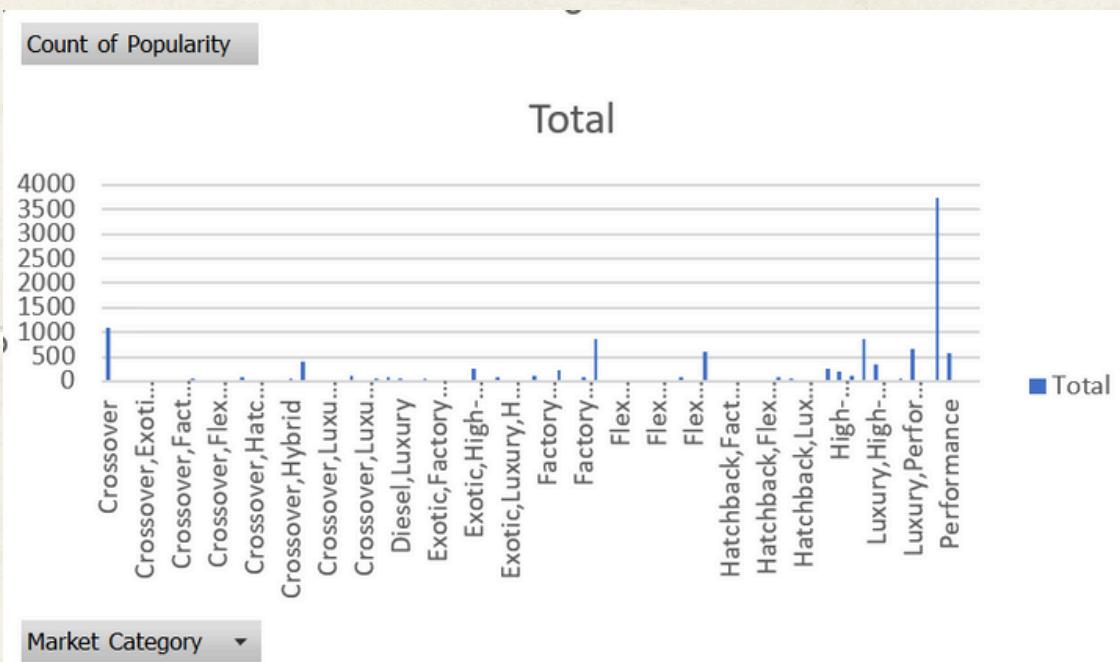
Creating a pivot table :

Market Category ----> Rows

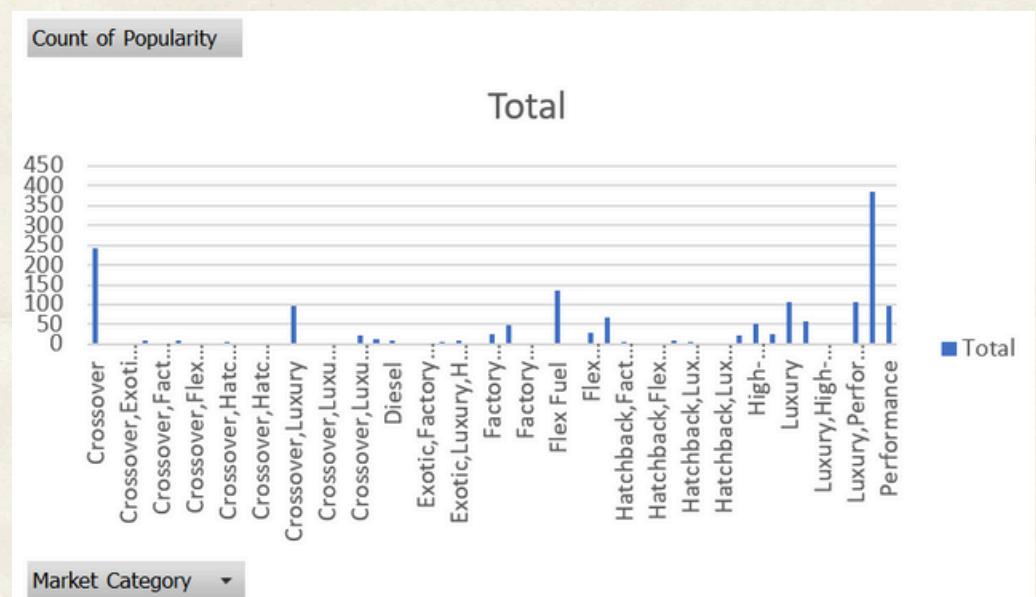
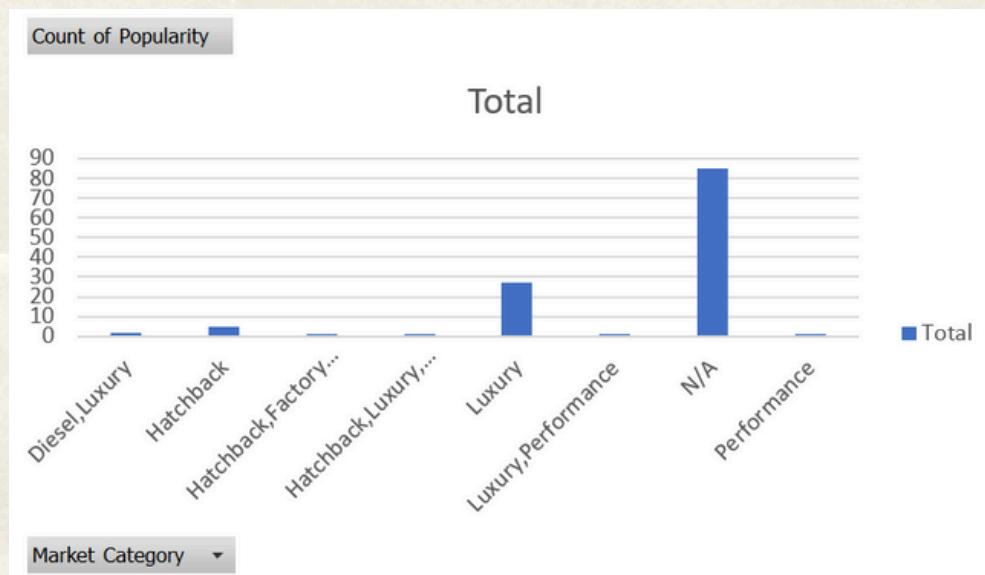
Count of Popularity -----> Values

Filtering the top 10 values using value filters

Creating a Column Chart for the pivot table.



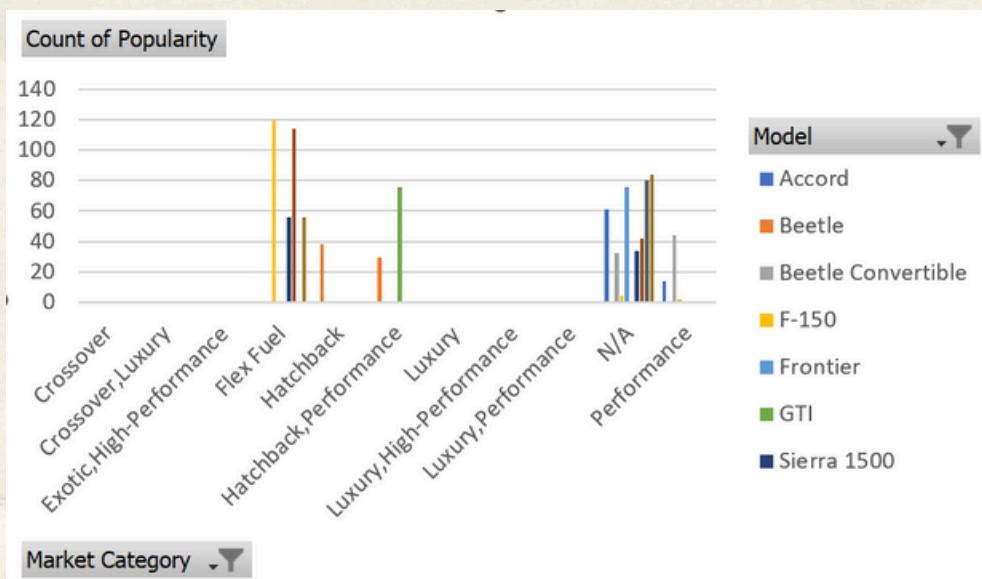
Adding a slicer and comparing market category popularity for the years 1990 and 2017



Count of Popularity	Column Labels	Accord	Beetle	Beetle Convertible	F-150	Frontier	GTI	Sierra 1500	Silverado 1500	Tacoma	Tundra	Grand Total		
Row Labels														
Crossover														
Crossover,Luxury														
Exotic,High-Performance														
Flex Fuel								120				346		
Hatchback			38									38		
Hatchback,Performance			29						76			105		
Luxury														
Luxury,High-Performance														
Luxury,Performance														
N/A		61			32	4	76			34		413		
Performance		14			44	2					42	80	84	60
Grand Total		75	67		76	126	76	76		90	156	80	140	962

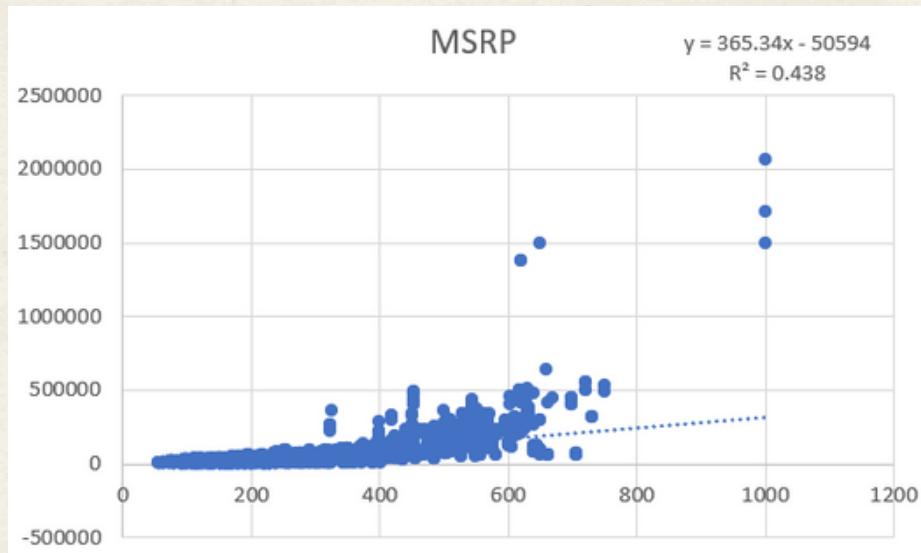
A pivot table that shows the number of car models in each market category and their corresponding popularity scores.

Filtering top 10 values based on Count of Popularity using Value Filter and creating a column chart for the same.



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 select the Engine HP column and the MSRP column and insert a cscatter plot with a trendline and equation.



- The trendline equation $y = 365.34x - 50,594$ suggests that as engine power (x-axis) increases, the MSRP (y-axis) also increases.
- Each additional unit of engine power (likely measured in horsepower) increases the price by \$365.34 on average.
- The R^2 value of 0.438 means that about 43.8% of the variation in car prices can be explained by engine power alone.
- This indicates a moderate correlation, but other factors (brand, features, fuel efficiency, etc.) also influence the price.
- A few points are significantly higher than the rest (e.g., cars with over 1000 horsepower costing more than \$2 million). These are likely luxury or supercars.
- The majority of the data points are concentrated in the 100-600 HP range with prices mostly below \$100,000. This suggests that most standard consumer cars fit in this range.

Task 3: Which car features are most important in determining a car's price

Performing Regression Analysis in Excel where Y(Independent Variable) is MSRP and X(Dependent Variables) are Engine HP, Highway MPG , Engine Cylinder

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.67483							
R Square	0.4553956							
Adjusted R	0.4552572							
Standard Er	44493.726							
Observation	11812							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	3	1.95471E+13	6.51569E+12	3291.263957	0			
Residual	11808	2.33762E+13	1979691685					
Total	11811	4.29233E+13						
Coefficients Standard Error								
	t Stat		P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	-110519.1	3253.13347	-33.9731102	1.8627E-241	-116895.7	-104142	-116896	-10414
Engine HP	319.07395	6.050053931	52.73902525		0	307.21484	330.9331	307.2148
highway MPG	1294.3034	70.83621995	18.27177402	1.42754E-73	1155.4527	1433.154	1155.453	1433.15
Engine Cylinders	6620.3785	437.6904253	15.12571016	3.30508E-51	5762.4331	7478.324	5762.433	7478.32

1. Model Performance (R^2 & Adjusted R^2)

- $R^2 = 0.455 \rightarrow 45.5\% \text{ of the variation in car price (MSRP) is explained by the included features (Engine HP, Highway MPG, Engine Cylinders).}$
- Adjusted $R^2 = 0.455 \rightarrow$ Similar to R^2 , confirming a good fit for the chosen variables.
- This means other factors (brand, body type, features, etc.) also influence price but were not included in this model.

2. Feature Importance & Impact on Price:

- Engine HP - 319.07 ---> Strong positive effect → Higher HP increases price significantly.
- Highway MPG - 1294.30 ---> Positive effect → More fuel-efficient cars have a higher MSRP.
- Engine Cylinders - 6620.38 ---> Strong positive effect → More cylinders increase price significantly

3. All variables are statistically significant (P-value < 0.05), indicating a strong impact on price.

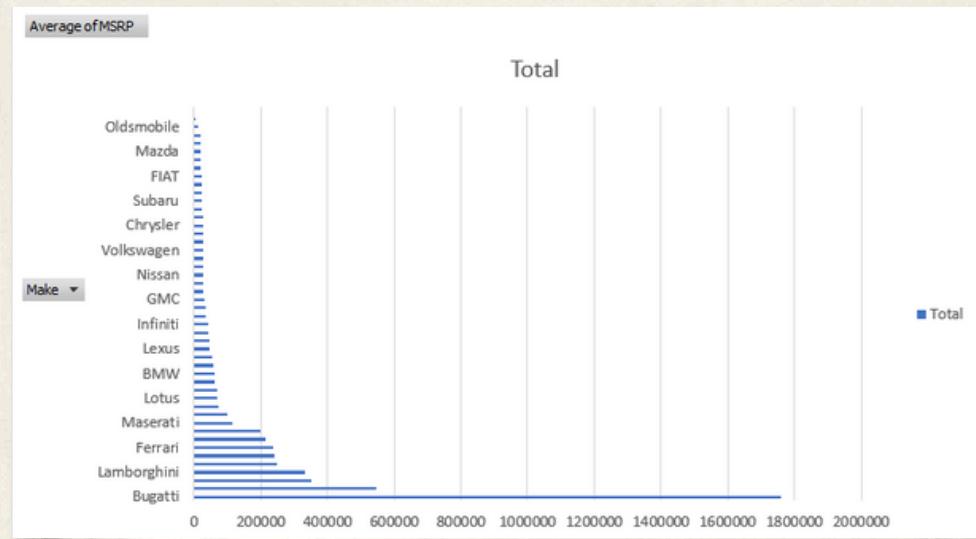
4. Engine HP & Engine Cylinders have the strongest influence on price (MSRP).

5. Higher fuel efficiency (MPG) is associated with higher prices. This suggests premium or hybrid vehicles may be priced higher.

Task 4: How does the average price of a car vary across different manufacturers?

Row Labels	Average of MSRP
Bugatti	1757223.667
Maybach	546221.875
Rolls-Royce	351130.6452
Lamborghini	331567.3077
Bentley	247169.3243
McLaren	239805
Ferrari	237383.8235
Spyker	213323.3333
Aston Martin	197910.3763
Maserati	114207.7069
Porsche	101622.3971
Mercedes-Benz	71537.80966
Lotus	69188.27586
Land Rover	67823.21678
Alfa Romeo	61600
BMW	61546.76347
Cadillac	56231.31738
Audi	53452.1128
Lexus	47549.06931
Genesis	46616.66667
Lincoln	42494.37179
Infiniti	42394.21212
HUMMER	36464.41176
Acura	34887.5873
GMC	30493.29903
Toyota	28946.15343
Volvo	28541.16014
Nissan	28513.36679
Chevrolet	28273.35695
Buick	28206.61224
Volkswagen	28076.2
Saab	27413.5045
Ford	27393.42051
Chrysler	26722.96257
Honda	26629.81879
Kia	25112.38938
Subaru	24827.50391
Hyundai	24597.0363
Dodge	22390.05911
FIAT	22206.01695
Mitsubishi	21215.47143
Scion	19932.5
Mazda	19719.05707
Pontiac	19321.54839
Suzuki	17900.9569
Oldsmobile	11542.54
Plymouth	3122.902439
(blank)	
Grand Total	40559.93532

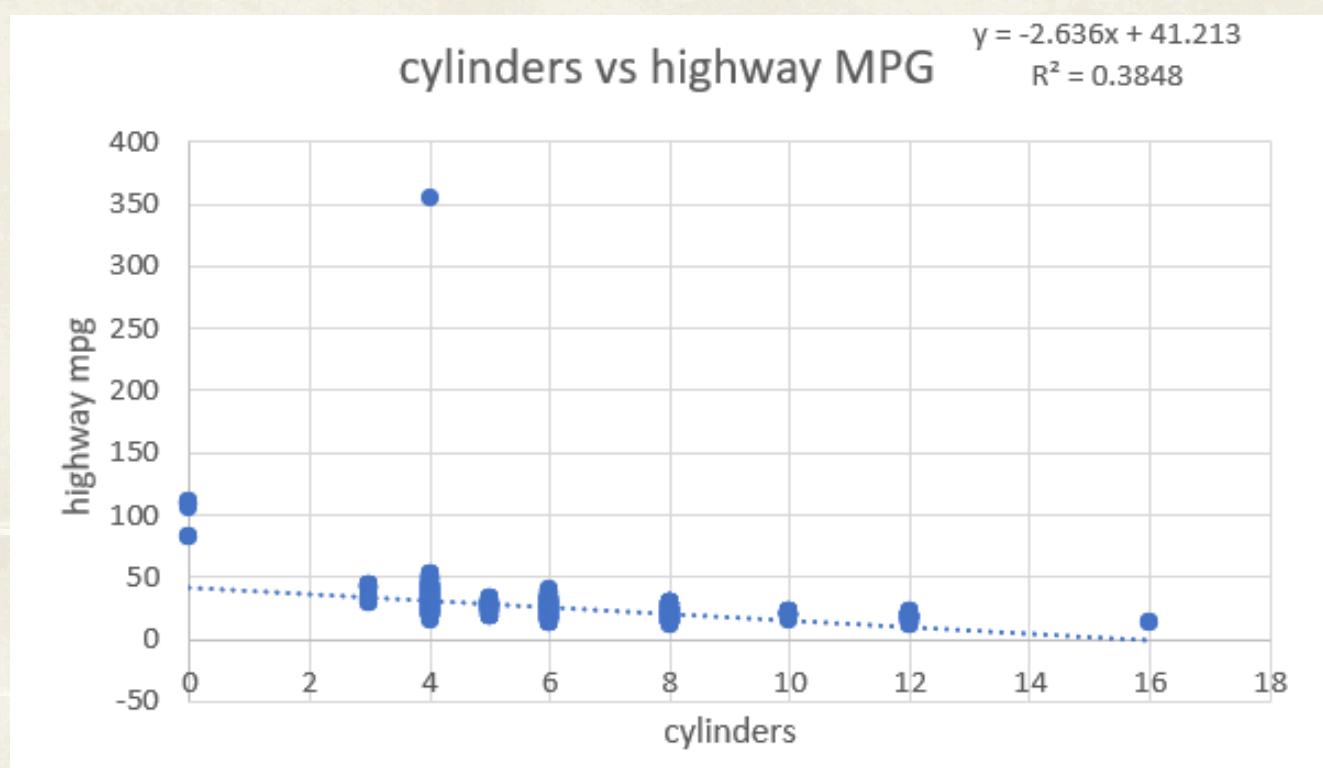
We create a pivot table by placing Make (manufacturer) under Rows and MSRP under values as Average. Then we sort the values in descending order and create a bar chart for the same



- Bugatti, Maybach, Rolls-Royce, Lamborghini, Bentley are the top 5 makers with highest average MSRP. These are the luxury brands, including other brands.
- Economy brands (e.g., Toyota, Honda, Ford) have lower average MSRPs.
- Some mid-tier brands (e.g., Audi, BMW) may have a higher average price than expected due to premium models.

Task 5: Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

Creating a scatter plot between cylinders (X axis) and highway MPG (Y axis). Also, showing equation and R^2 along with the trendline.



- Moderate negative correlation (-0.6 to -0.7 range).
- More cylinders generally lead to lower highway MPG, but other factors (engine efficiency, weight, aerodynamics) also play a role.
- Some outliers (e.g., a very high MPG car with 4 cylinders) might affect the trend.

Calculating the correlation between Cylinders and Highway MPG using the correl function in Excel. $r = -0.62031$

- This indicates a moderate negative correlation between engine cylinders and highway MPG.
- As the number of cylinders increases, highway MPG decreases.
- The relationship isn't perfectly linear (since $r \neq -1$), meaning other factors (engine efficiency, weight, and aerodynamics) also influence fuel economy.
- Moderate negative correlation (-0.62) → More cylinders reduce fuel efficiency.
- $R^2 = 38.48\%$ → Cylinders explain a significant portion of MPG variation but not all.
- Outliers exist (e.g., an extremely high-MPG 4-cylinder car).