Analyzing the Impact of Car Features on Price and Profitability

Project Description:

This project dives into the world of car data, exploring its potential to empower a car company's decision-making. By analyzing the relationship between car features, market categories, and pricing, it aims to uncover hidden insights. These insights will reveal the features most desired by customers and the categories offering the highest profitability. Ultimately, the project seeks to equip the company with a data-driven roadmap for optimizing pricing strategies and product development, leading to increased profitability and customer satisfaction in the competitive automotive landscape.

Approach:

step 1: Understanding the dataset and choosing the right tool for the analysis.

step 2: Prepare the data for analysis. Check for data quality, completeness and accuracy.

step 3: Identifying the business tasks and check if the data you have is sufficient or not.

step 4: Clean, modify and rearrange the dataset, and add new columns if needed to suit your analysis.

step 5: Using the data, analyze and answer the key questions and solve the business tasks.

step 6: Provide insights to the stakeholders, supported by data visualization if needed.

Tech-Stack Used:

Microsoft Excel : Features like pivot table, charts, etc were used mainly.

Python(Jupyter Lab) for Data analysis

Notion: It is a versatile tool that can be used for a variety of tasks, including project management, data analysis, and report generation.

Dataset:

The dataset contains information on over 11,000 car models and their specifications, including details on the car's make, model, year, fuel type, engine power, transmission, wheels, number of doors, market category, size, style, estimated miles per gallon, popularity, and manufacturer's suggested retail price (MSRP).

This dataset could be a valuable resource for data analysts interested in exploring various aspects of the automotive industry and could provide insights that could inform decisions related to product development, marketing, and pricing.

Data Cleaning:

715 duplicate values were removed using excel's remove duplicate option.

Some datatypes were changed to it's proper form, like car model numbers to text, and year(integer) to custom date format(year).

There were a lot of missing values in the column market category.
Using python, box plots were drawn to check the outliers and all those values were cross checked from other sources to see if values were acceptable.

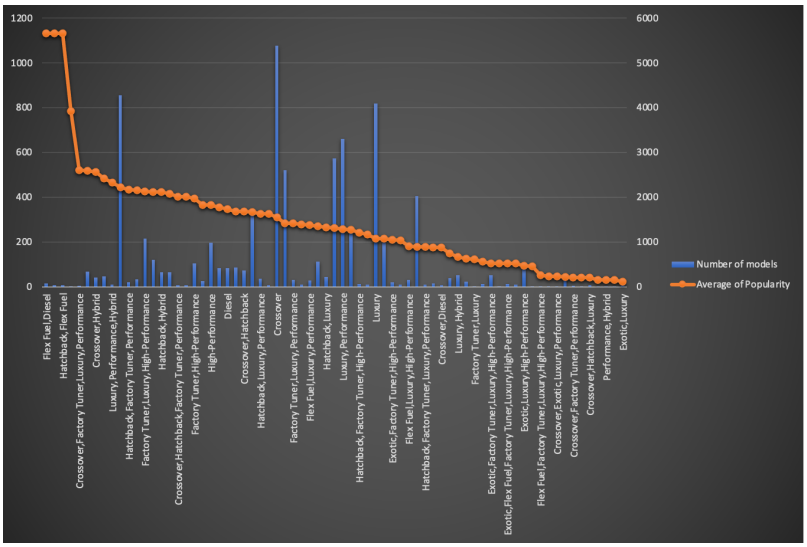
In the highway MPG column, an Audi model had 354 which was wrong. Correct value 35 was added referring the internet.

Insights:

Business tasks:

Task 1: How does the popularity of a car model vary across different market categories?

- **A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.
- **B:** Create a combo chart that visualizes the relationship between market category and popularity.



Market Category	Number of models	Average of Popularity
Crossover	1075	1556
Flex Fuel	855	2226
Luxury	819	1079
Luxury,Performance	659	1293
Hatchback	574	1309
Performance	520	1415
Crossover,Luxury	406	889
Luxury,High-Performance	334	1668
Exotic,High-Performance	254	1280
Factory Tuner,Luxury,High-Performance	215	2133

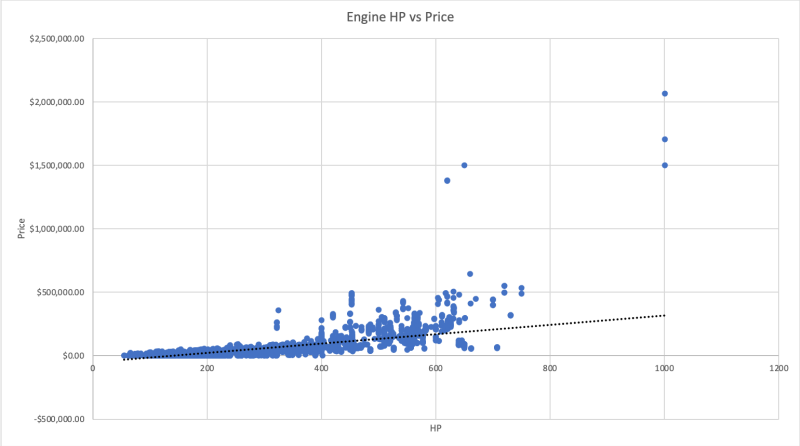
Market Category	Number of models	Average of Popularity
Flex Fuel,Diesel	16	5657
Crossover,Flex Fuel,Pe	6	5657
Hatchback,Flex Fuel	7	5657
Crossover,Luxury,Perfc	2	3916
Crossover,Factory Tun	5	2607
Crossover,Performanc	69	2586
Crossover,Hybrid	42	2563
Diesel,Luxury	47	2416
Luxury,Performance,H	11	2333
Flex Fuel	855	2226

The most number of car models belong to the Cross over, flex fuel and luxury market category

Flex fuel combined with other market categories like crossover,hatchback and diesel has the most popularity points in average. This can be fuelled by a very few high end car models too so it's difficult to make a conclusion here.

Task 2: What is the relationship between a car's engine power and its price?

- 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.



The correlation coefficient of Engine HP and Price is 0.71 which is a fairly good correlation. More the engine horsepower, greater chances that it'd be an expensive car.

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

- **:** 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.

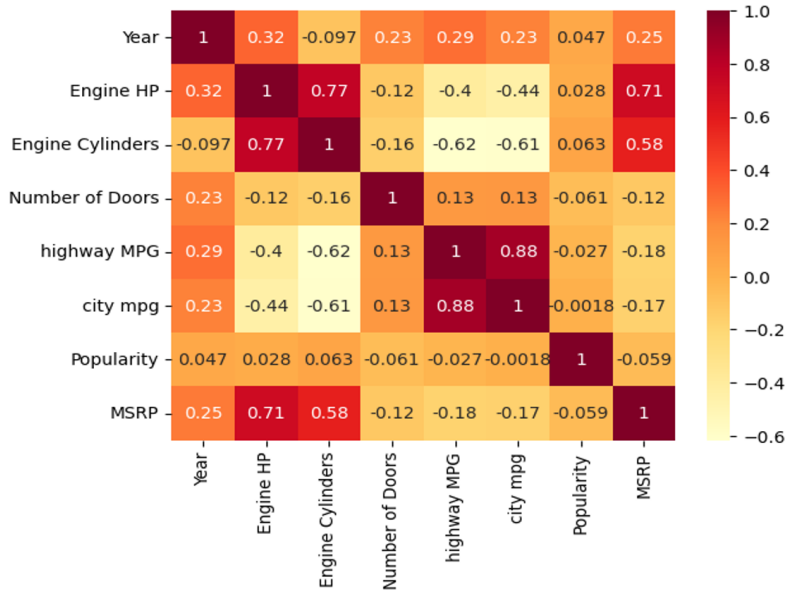
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.18656886
R Square	0.03480794
Adjusted R S	0.03454929
Standard Err	60462.7081
Observations	11199

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1.4759E+12	4.9197E+11	134.575941	1.1852E-85
Residual	11195	4.0926E+13	3655739074		
Total	11198	4.2402E+13			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	83212.2166	2470.42128	33.6834116	5.321E-237	78369.7563	88054.6769	78369.7563	88054.6769
highway MP	-1710.717	203.307029	-8.4144507	4.4268E-17	-2109.2345	-1312.1994	-2109.2345	-1312.1994
city mpg	392.011113	186.65851	2.1001513	0.03573785	26.1275981	757.894628	26.1275981	757.894628
Popularity	-2.2758891	0.39604011	-5.7466128	9.3425E-09	-3.0521974	-1.4995809	-3.0521974	-1.4995809



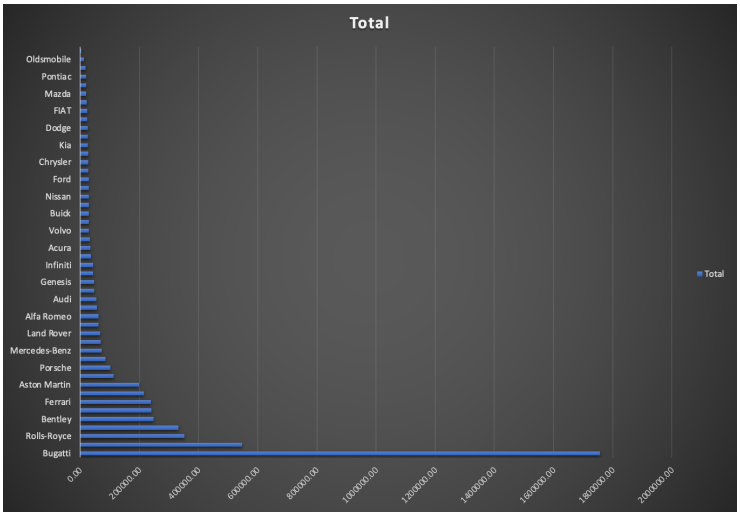
The following features have high correlations:

- Highway MPG- City MPG
- Engine HP - MSRP
- Engine HP - Engine Cylinders

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

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

Car manufacturers	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	214990.00
Aston Martin	198123.46
Maserati	113684.49
Porsche	101622.40
Tesla	85255.56
Mercedes-Benz	72069.53
Lotus	68377.14
Land Rover	68067.09
BMW	62162.56

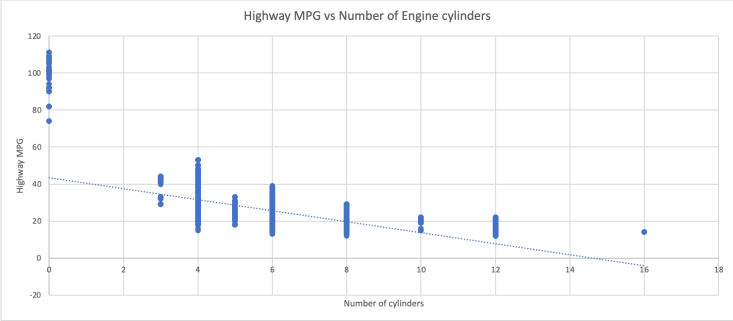


The above table and chart shows the average price of car for each manufacturer.

Buggati costs an average of thrice the amount of the 2nd most expensive Maybach which is a significant price gap.

Task 5: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

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



From the heatmap, correlation coeff is -0.62 which indicates a weak negative correlation

DASHBOARD:

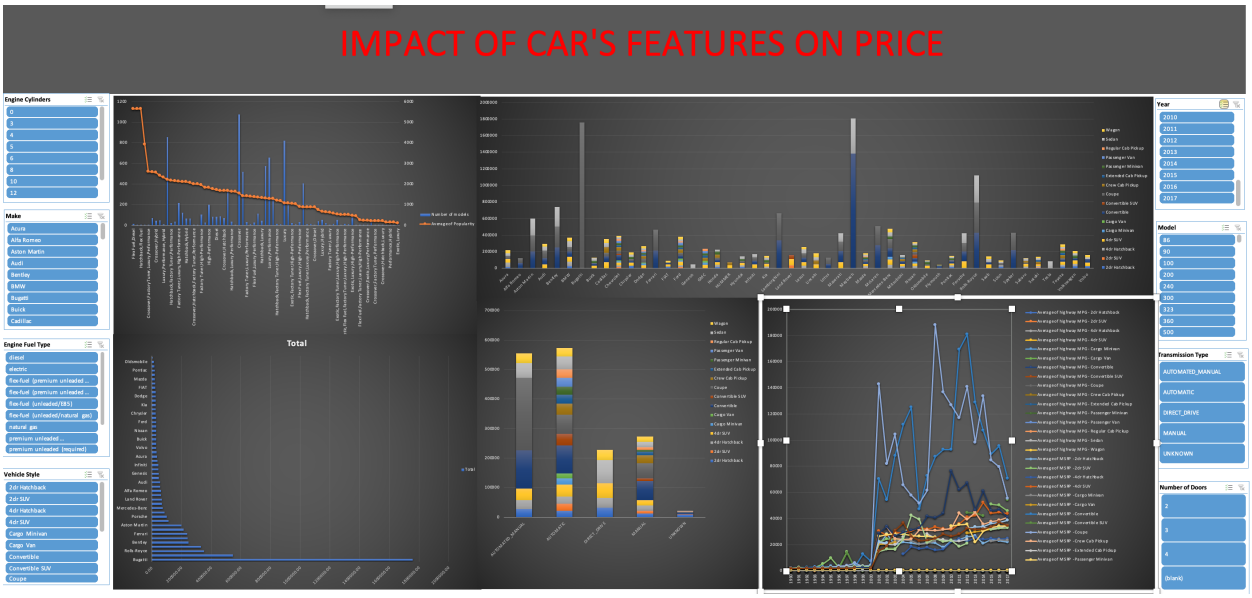
Task 1: How does the distribution of car prices vary by brand and body style?

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

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

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

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



Link to the Excel Dashboard:
https://1drv.ms/x/c/103032cf9edd842e/EQnGadQAxahFkhHG8GSL_TkB1FlLHVdATFBERlfeujlfqw

Result:

These findings provide valuable insights for the car manufacturer to optimize their pricing and development strategies. The correlations between features and price offer a data-driven basis for informed decision-making regarding which features to prioritize in new models and how to price them competitively. Additionally, understanding the popularity of different car categories and market segments equips the manufacturer to tailor their offerings to customer preferences, fostering greater market success.

A comprehensive dashboard was created that translates insights into actionable visualizations. This interactive tool empowers intuitive exploration of key findings, offering a clear view of the relationship between features and price, fuel efficiency trade-offs, and the popularity of various car categories and market segments.