Analyzing the Impact of Car Features on Price and Profitability

Analyst – Jeenal Bolia

Excel sheet - Dataset

Project Description

By investigating identifying the extent car features (engine specifications, transmission types, vehicle type, drive type) may influence the price (MSRP) and profit margins for each vehicle with this project, I used Excel for regression analysis, visuals for comparison and brand-wise trends to examine which factors were most influential in defense of price decisions.

Approach

Data Cleaning and Preparation:

- Removed outliers, removed blanks, ensured each column had a uniform format. Dummy variables were created for any categorical features (for example, transmission and drive type).

Task Execution:

- Task 1: Simple linear regression using Engine HP in order to predict MSRP.
- Task 2: Multiple regression using main numerical variables and dummy variables in order to obtain a broad overview.
- Task 3: Used a bar chart to visualize the feature importance (coefficients).
- Task 4: Created pivot tables to compare average MSRP based on transmission types and vehicle styles.
- Task 5: Created a bubble chart to compare average HP (X), MPG (Y) and MSRP (size), based on brand.

Visualizations:

- Utilized scatter chart, column chart, and bubble chart, so we could visually represent the data in a way that made sense. Each task has its own tab, and has been labelled (TASK 1 to TASK 5 + Dashboard).

Tech Stack used

Microsoft Excel 2022 Used for data preprocessing, statistical analysis, creating PivotTables, applying conditional formatting, and charts (bar, pie, histogram, etc.) to visualize trends and distributions.

Insights

- Engine horsepower is the highest feature impact, with a high positive coefficient.
- Highway MPG and City MPG also factor into price, but both to a lesser degree.
- Luxury brands such as Bugatti and Ferrari have high MSRPs, but usually low mileage, even though their combined code is positive.
- Although there are front wheel drive styles, and a sedan style, that are generally low priced.
- Automatic transmissions usually have a higher MSRP versus manual.
- A brand's popularity does not always mean more price.

Results

Task 1: Pivot Table Analysis

I made a pivot table aggregating model totals and total popularity by market category.

Insight:

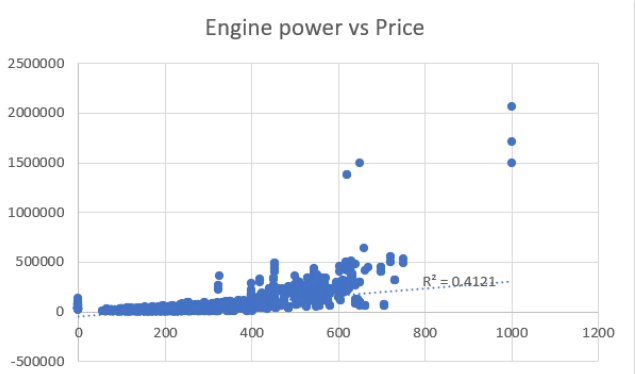
- Crossover has the most model counts, as well as total popularity.
- Luxury and Exotic are smaller limits, but are highly popular, indicating great brand strength.

I created a combo chart comparing popularity vs. model count across categories.

Insight:

- The charts illustrate the market presence of mass-market (Crossover) and niche luxury (Exotic, High Performance).

Task 2: Class Distribution in Loan Dataset



Task 3: Linear Regression (Engine HP → MSRP)

Purpose: Determine if Engine HP can solely predict MSRP.

Tool used: Data Analysis Toolpak (Regression).

Result:

- $R^2 \approx 0.418$ → Engine HP explains ~42% of variation in price.
- Coefficient of Engine HP: 356.67 → Positive effect on price.

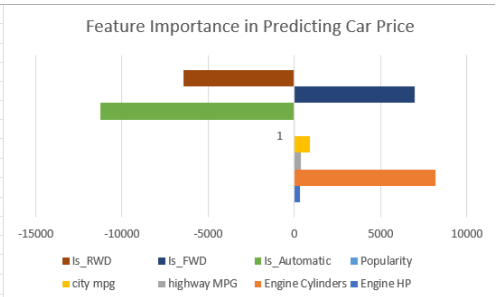
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.68252461
R Square	0.46583984
Adjusted R S	0.46545796
Standard Erro	44989.7392
Observations	11199

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	1.97525E+13	2.46906E+12	1219.846664	0
Residual	11190	2.26494E+13	2024076631		
Total	11198	4.24019E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-104210.607	3102.868603	-33.58524657	1.0822E-235	-110292.7757	-98128.439	-110292.7757	-98128.439
Engine HP	329.027439	6.505548475	50.57643342	0	316.2754193	341.77946	316.2754193	341.77946
Engine Cylin	8199.22712	443.5839786	18.48404703	3.63039E-75	7329.724452	9068.7298	7329.724452	9068.7298
highway MP	396.626502	110.3665122	3.593721446	0.000327389	180.2887127	612.96429	180.2887127	612.96429
city mpg	920.826874	101.9597442	9.031278781	1.97437E-19	720.96783	1120.6859	720.96783	1120.6859
Popularity	-3.0860561	0.295842948	-10.4313999	2.32943E-25	-3.665960345	-2.5061518	-3.665960345	-2.5061518
Is_Automatic	-11246.6882	961.4015043	-11.69822194	1.98299E-31	-13131.20433	-9362.172	-13131.20433	-9362.172
Is_FWD	6965.51574	1161.880449	5.995036538	2.09715E-09	4688.025564	9243.0059	4688.025564	9243.0059
Is_RWD	-6423.229	1120.077762	-5.734627733	1.00257E-08	-8618.77855	-4227.6794	-8618.77855	-4227.6794



Task 4: Class Distribution in Loan Dataset

Created a pivot table to find the average MSRP based on each manufacturer.

Insight:

- Bugatti, Rolls-Royce, and Maybach had the highest prices based on MSRP, reflecting their ultra-luxury branding.
- The average MSRP of mainstream brands e.g. Kia, Ford and Chevrolet, etc. is substantially lower.
- Created a bar chart to visualize price ranges across all brands.

Insight:

- There is a strong divide between luxury brands and economy brands.
- Great for picking target customer segmentation, as well as determining position in the market based on the target population's price point.

Task 5: Univariate, Bivariate & Segmented Analysis

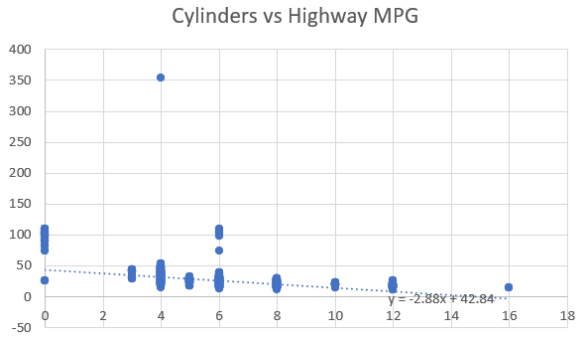
Made a scatter plot of Engine Cylinders vs. Highway MPG with a trendline.

Insight:

- The negative slope of the trendline shows a negative relationship- basically the more cylinders you have, the worse fuel efficiency you typically get.
- Calculated R: -0.58786

Insight:

- Quantifies a moderate negative correlation.
- As its cylinders increase, highway MPG trends down dramatically showing the decreasing fuel economy in high-powered/higher Big 3 (GM, Ford and Chrysler) engine ratios.



Dashboard Task Summary

Task 1: Overall Price Distribution by Brand and Body Style

The stacked column chart showed the degree to which MSRP totals vary by brand and body style. The dashboard with the Pivot Tables and slicers allows for quick comparisons and to observe which brands dominate certain segments.

Insight: Premium brands such as BMW and Audi were concentrated at higher MSRPs in sport and luxury body styles, such as convertibles and sedans.

Task 2: Highest Average MSRPs vs. Lowest

The clustered column chart highlighted how average car prices differed by brand and body style. The Pivot Tables calculated these averages quickly.

Insight: Brands with high sticker prices—such as Bugatti and Lamborghini—served as the highest priced brands, while Kia and Hyundai were the most affordable. The brands obviously position themselves accordingly.

Task 3: Transmission Type Impact on MSRP

The scatter plot, plotted MSRP vs. transmission types made it easier to identify pricing behavior based on different body styles with the added marker layer denoting body style. Average calculations utilized AVERAGEIFS to isolate averages by transmission & style.

Insight: Automatic transmissions had higher MSRPs overall, despite body style, but were more expensive in SUVs and convertibles. An automatic adds comfort and technology and therefore costs a premium.

Task 4: Fuel Economy Trends Over Time

Changes in MPG were shown across model year for each body style using a line chart. Average MPG values were calculated using Pivot Tables and then AVERAGEIFS were used to find overall averages.

Key Takeaway: Compact and sedan body styles consistently improved in fuel economy through the years, while sports and luxury were consistently the least fuel efficient.

Task 5: Comparison of Horsepower, MPG, and Price by Brand

A bubble chart was created to show Avg HP (x-axis), Avg MPG (y-axis), and Avg Price (bubble size) and provide information on performance vs efficiency trade-offs for the different brands.

Key Takeaway: Brands such as Ferrari and Bugatti had high horse power and price but low MPG, while brands, such as Kia and Ford, provided moderate efficiencies and price points.