EXPLORATORY DATA ANALYSIS ON CAR PRICING

AMENA MEHREEN

- **Contents**

- Introduction
- Executive Summary
- Descriptive Statistical Analysis
- Plots With Findings And Implications
- Correlation And Causation
- Conclusion

INTRODUCTION

- Explore features or characteristics to predict price of a car.

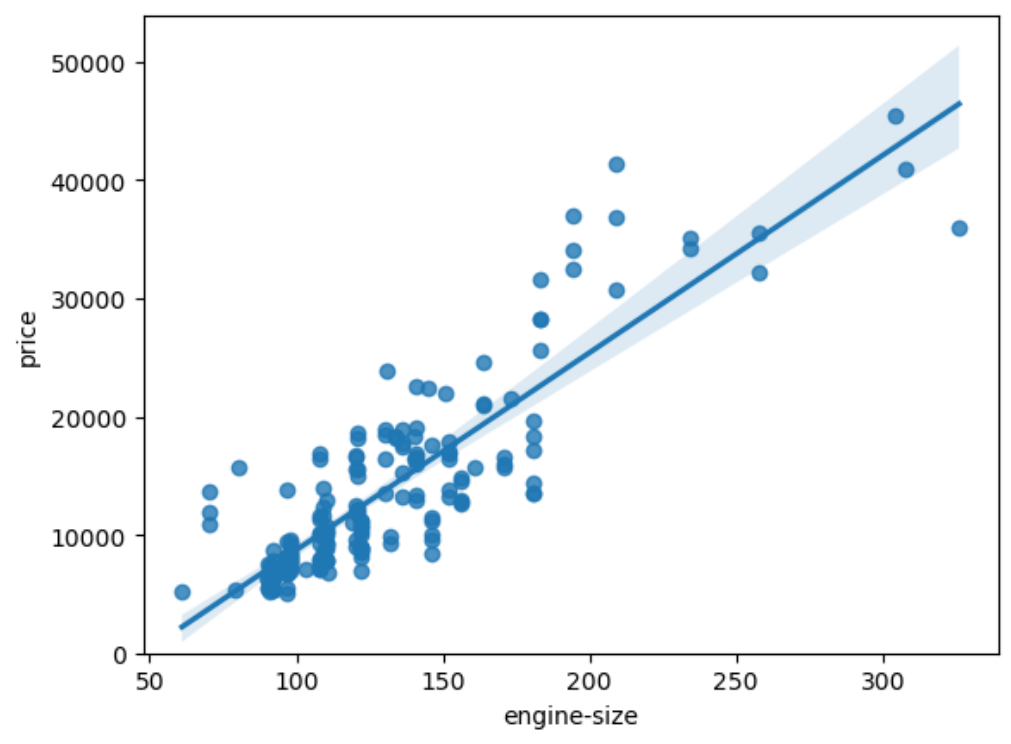
EXECUTIVE SUMMARY

- Analyze patterns and run descriptive statistical analysis
- Group data based on identified parameters and create pivot tables
- Identify the effect of independent attributes on price of cars

DESCRIPTIVE STATISTICAL ANALYSIS

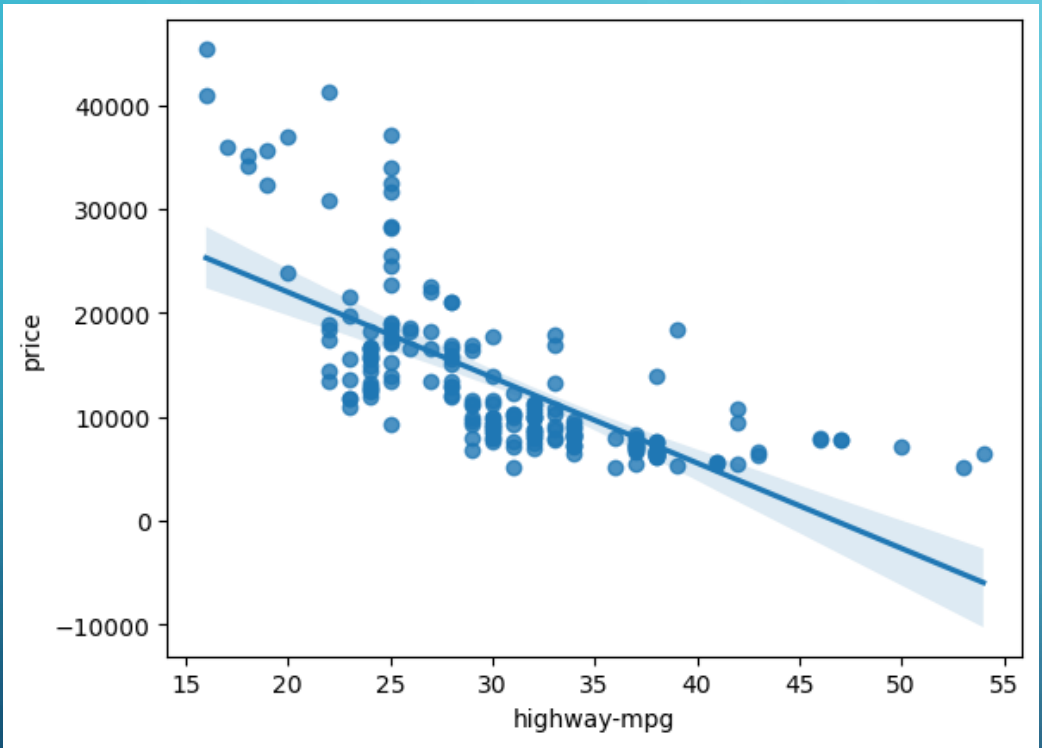
	symboling	normalized-losses	wheel-base	length	width	height	curb-weight	engine-size	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg
count	201.000000	201.00000	201.000000	201.000000	201.000000	201.000000	201.000000	201.000000	201.000000	197.000000	201.000000	201.000000	201.000000	201.000000	201.000000
mean	0.840796	122.00000	98.797015	0.837102	0.915126	53.766667	2555.666667	126.875622	3.330692	3.256904	10.164279	103.405534	5117.665368	25.179104	30.686567
std	1.254802	31.99625	6.066366	0.059213	0.029187	2.447822	517.296727	41.546834	0.268072	0.319256	4.004965	37.365700	478.113805	6.423220	6.815150
min	-2.000000	65.00000	86.600000	0.678039	0.837500	47.800000	1488.000000	61.000000	2.540000	2.070000	7.000000	48.000000	4150.000000	13.000000	16.000000
25%	0.000000	101.00000	94.500000	0.801538	0.890278	52.000000	2169.000000	98.000000	3.150000	3.110000	8.600000	70.000000	4800.000000	19.000000	25.000000
50%	1.000000	122.00000	97.000000	0.832292	0.909722	54.100000	2414.000000	120.000000	3.310000	3.290000	9.000000	95.000000	5125.369458	24.000000	30.000000
75%	2.000000	137.00000	102.400000	0.881788	0.925000	55.500000	2926.000000	141.000000	3.580000	3.410000	9.400000	116.000000	5500.000000	30.000000	34.000000
max	3.000000	256.00000	120.900000	1.000000	1.000000	59.800000	4066.000000	326.000000	3.940000	4.170000	23.000000	262.000000	6600.000000	49.000000	54.000000

REGPLOT OF ENGINE-SIZE AND PRICE



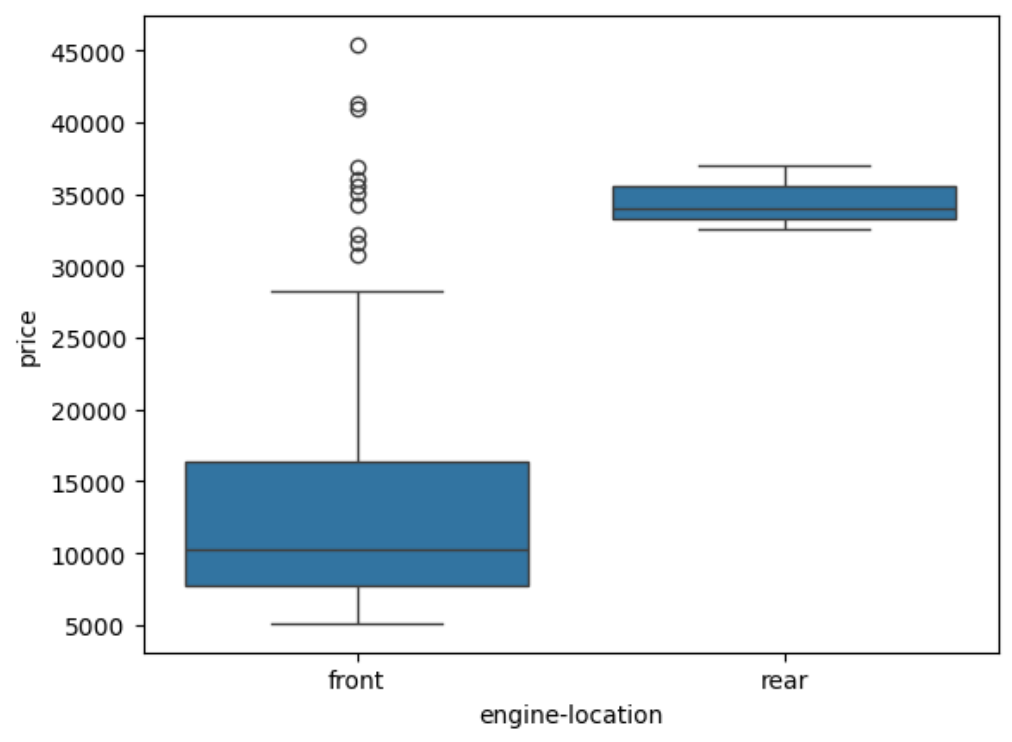
As the engine-size goes up, the price goes up: This indicates a positive direct correlation between these two variables. Engine size seems like a pretty good predictor of price since the regression line is almost a perfect diagonal line.

REGPLOT OF HIGHWAY-MPG AND PRICE



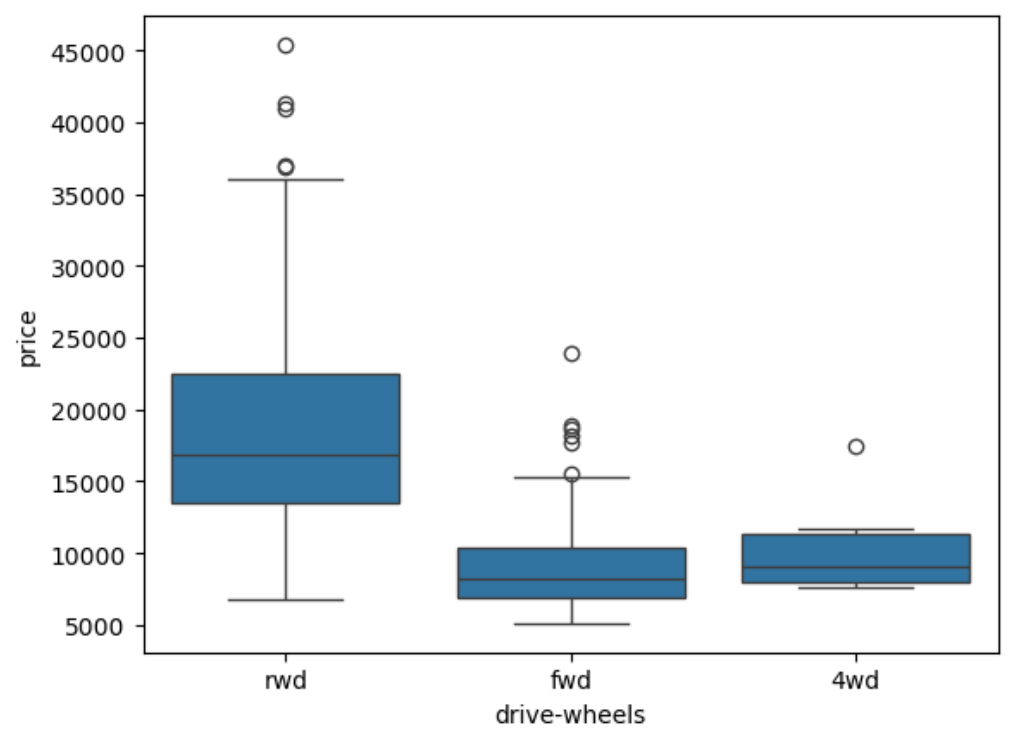
As highway-mpg goes up, the price goes down: This indicates an inverse/negative relationship between these two variables. Highway mpg could potentially be a predictor of price.

BOX PLOT OF ENGINE-LOCATION AND PRICE.



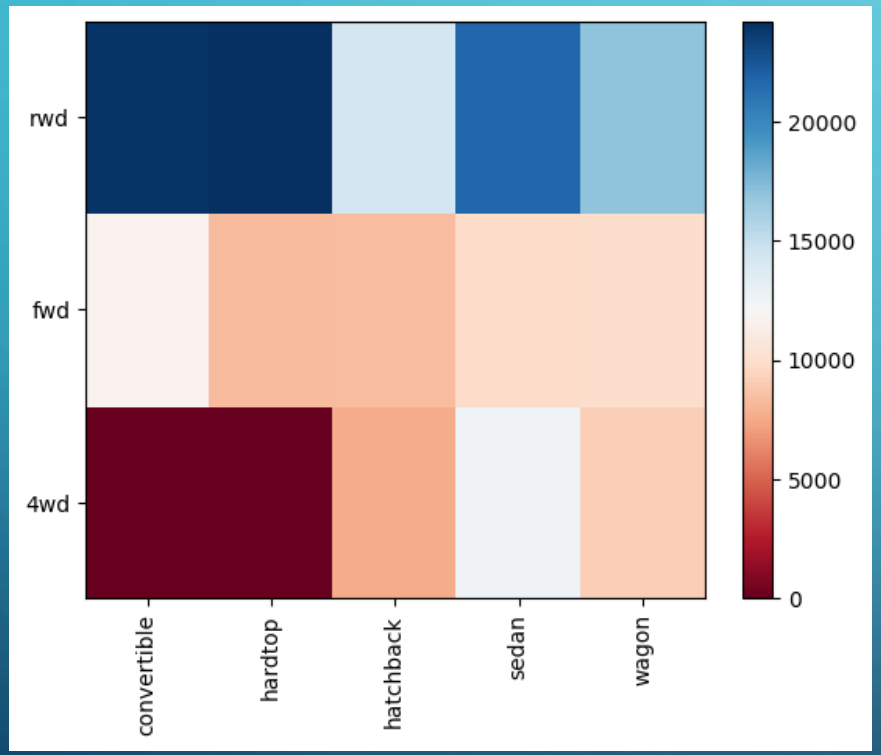
Here we see that the distribution of price between these two engine-location categories, front and rear, are distinct enough to take engine-location as a potential good predictor of price.

BOX PLOT OF DRIVE-WHEELS AND PRICE



Here we see that the distribution of price between the different drive-wheels categories differs. As such, drive-wheels could potentially be a predictor of price.

HEAT MAP TO VISUALIZE THE RELATIONSHIP BETWEEN BODY STYLE VS PRICE.



The heatmap plots the target variable (price) proportional to colour with respect to the variables 'drive-wheel' and 'body-style' on the vertical and horizontal axis, respectively. This allows us to visualize how the price is related to 'drive-wheel' and 'body-style'.

CORRELATION AND CAUSATION

Horsepower vs. Price

- The Pearson Correlation Coefficient is 0.809574567003655 9 with a P-value of P = 6.3690 5742825956e-48
- Conclusion: Since the p-value is 0.001, the correlation between horsepower and price is statistically significant, and the linear relationship is quite strong (~0.809, close to 1).

Engine-Size vs. Price

- The Pearson Correlation Coefficient is 0.872335167445518 8 with a P-value of P = 9.2654 9162219582e-64
- Conclusion: Since the p-value is 0.001, the correlation between engine-size and price is statistically significant, and the linear relationship is very strong (~0.872).

CORRELATION AND CAUSATION

Width vs. Price

- The Pearson Correlation Coefficient is 0.7512653440522663 with a P-value of P = 9.200355510485071e-38
- Conclusion: Since the p-value is < 0.001, the correlation between width and price is statistically significant, and the linear relationship is quite strong (~0.751).

Curb-Weight vs. Price

- The Pearson Correlation Coefficient is 0.8344145257702845 with a P-value of P = 2.189577238893965 4e-53
- Conclusion: Since the p-value is 0.001, the correlation between curb-weight and price is statistically significant, and the linear relationship is quite strong (~0.834).

CONCLUSION:

Important Variables

- The important variables to take into account when predicting the car price are as follows:

Continuous Numerical Variables:

- Length
- Width
- Curb-weight
- Engine-size
- Horsepower
- City-mpg
- Highway-mpg
- Wheel-base
- Bore

Categorical variables:

- Drive-wheels