Linear Regression Project Report

Marketing Analytics

MAR 6668

Venkata Sai Satvik Reddy Tanuboddi

Date

March, 19, 2024

Requirement 1: ***Identify the set of objects (e.g., cars) chosen and the segment in mind. Then, present descriptive statistics for the variables in the dataset such as range of prices, average price, range of mileage, etc.***

The dataset was collected through web scraping from Carstory.com, focusing specifically on affordable sedan-type cars. It includes 431 observations (n = 431) of used vehicles from popular brands like Toyota, Honda, Hyundai, Kia, Nissan, Chevrolet, Dodge, Ford, and Volkswagen. The study targets the family-oriented, compact to midsize used car segment, deliberately omitting luxury or high-end models.

Here are the descriptive statistics for the numerical variables:

- **Price:** The average price of the cars is $26,494.03 with a standard deviation of $6,709.007. The range is from $14,589.00 to $60,589.00.
- **Years Old:** The average age of the cars is 3.61 years with a standard deviation of 2.12 years. The range is from 1 to 12 years old.
- **Mileage Miles:** The average mileage is 32,743.87 miles with a standard deviation of 24,290.19 miles. The range is from 61 miles to 113,497 miles.
- **City MPG:** The average city miles per gallon (MPG) is 28.07657 with a standard deviation of 7.097735. The range is from 15.0 to 53.0 MPG.
- **Highway MPG:** The average highway MPG is 36.63341 with a standard deviation of 5.185418. The range is from 24.0 to 56.0 MPG.

Requirement 2: ***Present the variables you used to predict the price of the object. Give your theoretical justification for your model, and for non-obvious variables, spend a little more time justifying the use of the variable in the model.***

The variables used to predict the price of the object (cars) in the regression analysis include:

- **Years Old**: The age of the car, calculated by subtracting the year of manufacture from the current year (2024).
- **Mileage**: The mileage of the car, indicating its usage.
- **City MPG**: The miles per gallon (MPG) of fuel consumed while driving in city conditions.
- **Highway MPG**: The miles per gallon (MPG) of fuel consumed while driving on highways.
- **Make**: Categorical variables representing the make or brand of the car.
- **Engine Type**: Categorical variables representing the type of engine (e.g., 4 Cyl, 6 Cyl, 8 Cyl).
- **Transmission Type**: Categorical variables representing the type of transmission (e.g., Automatic, Manual).
- **Drive Type**: Categorical variables representing the drive type of the car (e.g., AWD, FWD, RWD).
- **Fuel Type**: Categorical variables representing the type of fuel used by the car (e.g., Flexible-Fuel, Gasoline, Hybrid).

Theoretical Justification for the Model:

We are implementing the linear regression model using XLSTAT. This approach allows us to quantitatively assess the influence of various attributes on used car prices. By inputting data on the age of the car in years, mileage, fuel efficiency (City MPG and Highway MPG), and categorical variables like Make, Engine Type, Transmission Type, Drive Type, and Fuel Type into XLSTAT, we can analyze how these factors contribute to the market value of used cars. This process involves using statistical techniques to understand the relationship between car characteristics and their pricing, providing insights into the factors that consumers value when purchasing a used vehicle.

Justification for Non-Obvious Variables:

Variables like Make, Engine Type, Transmission Type, Drive Type, and Fuel Type may not be immediately obvious choices for predicting car prices in a regression model. However, these variables capture important aspects of consumer preferences and market segmentation that significantly influence car prices. For example:

- **Make**: Different car brands have different brand images, reputations for reliability, and target markets, all of which can affect prices.
- **Engine Type**: Engine size and performance are key factors in determining a car's price, with larger engines typically commanding higher prices due to increased power and performance.
- **Transmission Type**: Automatic and manual transmissions appeal to different types of drivers, with automatic transmissions often associated with convenience and manual transmissions with performance. This choice can affect prices accordingly.
- **Drive Type**: All-wheel drive (AWD), front-wheel drive (FWD), and rear-wheel drive (RWD) offer different driving experiences and capabilities, with AWD often associated with enhanced traction and stability, particularly in adverse weather conditions, which can influence pricing.
- **Fuel Type**: With growing concerns about environmental impact and fuel efficiency, the type of fuel a car uses (e.g., gasoline, flexible-fuel, hybrid) can significantly influence consumer preferences and, consequently, prices.

Requirements 3: ***Report the regression results and indicate which variables play a role in influencing the price and which don't? Make sure to include the relevant regression output which includes the adjusted R^2 and the model coefficients along with standard errors*****.**

Here are the key findings based on the provided regression output:

Model Coefficients and Standard Errors:
The model coefficients represent the effect of each predictor variable on the price of cars, holding other variables constant. The standard errors indicate the precision of these estimates.

1. **Years Old**: For each additional year of age (which typically means an older model), the price is expected to decrease by $1,260.908, holding all other factors constant. Older used cars tend to be priced lower due to more wear and tear and older features.

2. **Mileage_miles**: For each additional mile on the car, the price is expected to decrease by $0.058, holding all other factors constant. This captures the depreciation effect, where higher mileage usually indicates more wear and could lead to lower prices.
3. **Engine-4 Cyl**: Compared to 8-cylinder engines, having a 4-cylinder engine is associated with a decrease in price by $16,405, holding all other factors constant. This reflects a preference for more powerful engines.
4. **Engine-6 Cyl**: Compared to 8-cylinder engines, having an 6-cylinder engine is associated with a decrease in price by $13,662, holding all other factors constant. This significantly larger increase indicates a premium for high-power engines.
5. **Fuel-Gasoline**: Compared to fuel Hybrid, having a gasoline engine is associated with a decrease in price by $14,169, holding all other factors constant. This indicates a price premium for gasoline vehicles over flexible-fuel ones within the dataset.
6. **Fuel-flexible-fuel**: Compared to Hybrid vehicles, having a Fuel Flexible engine is associated with a decrease in price by $20,741, holding all other factors constant. This decrease signifies the value placed on hybrid technology for its fuel efficiency and environmental benefits.
7. **(City MPG*Highway MPG)**: For each one-unit increase in the product of City MPG and Highway MPG, the price is expected to increase by $8.952, holding all other factors

constant. This interaction term shows the additional value that cars with good fuel efficiency in both city and highway conditions command in the market. For example if the City MPG increases from 30 to 31 and Highway MPG increased from 38 to 39, the price of the used increases by $8.952

8. **make-dodge**: Dodge vehicles are expected to be priced $6045.947 lower than Volkswagen vehicles, holding all other factors constant. This indicates that Dodges are perceived to have a lower value or desirability than Volkswagens among the cars included in the analysis.
9. **make-ford**: Ford vehicles are expected to be priced $4174.035 lower than Volkswagen vehicles, holding all other factors constant. This might reflect market preferences or differences in the models of Ford cars compared to Volkswagen cars in the dataset.
10. **make-honda**: Honda vehicles are expected to be priced $4285.978 higher than Volkswagen vehicles, holding all other factors constant. Hondas may have a higher resale value or might be more in demand in the used car market compared to Volkswagens.
11. **make-toyota**: Toyota vehicles are expected to be priced $2050.500 higher than Volkswagen vehicles, holding all other factors constant. This premium might be attributed to Toyota's reputation for reliability and higher demand in the resale market

Requirement 4: Report the sale price elasticity of the significant variables in the regression. Note that this can only be done for continuous variables.

To calculate the sales price elasticity for the significant continuous variables in the regression analysis (years old, mileage, city MPG, highway MPG), we use the following formula:

Elasticity = Coefficient of Variable × Mean of Variable / Mean of Price

Let's calculate and analyse the sales price elasticity for each significant variable:

Elasticity for Years old = (−1,260.908) ×3.61/26,494.03 = **-0.172**

Elasticity for Mileage = (−0.058) × 32,743.87 / 26,494.03 = **−0.071677**

Elasticity for City MPG*Highway MPG = 8.952 × (28.07657 × 36.63341) / 26,494.03 = **0.348**

Analysis:

- **Years old**: For each 1% increase in the age of the car (measured in 'Years Old'), there is an expected decrease in the price of about 0.172%, holding all other factors constant. This negative elasticity reflects the depreciation effect of cars as they age.
- **Mileage**: The elasticity for mileage is -0.071677, implying that a 1% increase in mileage results in a -0.071677% decrease in the price of the car. This negative elasticity confirms the intuitive notion that higher mileage leads to lower prices due to increased wear and tear.
- **City MPG*Highway MPG**: The price elasticity for the interaction term City MPG*Highway MPG is approximately 0.348, meaning that a 1% increase in the product of City MPG and Highway MPG is associated with a 0.348% increase in the price of the car. This suggests a moderate level of elasticity**.**