# **MARKETING ANALYTICS PROJECT**

AURELIA SHEARS

TREVOR WHITMILL

ARPITA NAYAK

**Dataset**

Our project deals with predicting prices of used cars and what factors influence the most. We chose the segment of family crossover vehicles. We used the standard definition of a crossover, a vehicle built on a car chassis platform, in unibody fashion. To keep it family oriented, we limited our data collection to five-seater models in the mid-price range and intentionally excluded all luxury makes and models.

We have taken the data from cargurus.com focusing on predominantly on three specific models of varying years: *the Subaru Forester, the Ford Escape, and the Honda CR-V.*

**Descriptive statistics:**

Make: the make of all the cars were either Subaru, Ford, or Honda.

Model: The models were Subaru- Forester, Ford-Escape and Honda CR-V.

Price. The range of prices are $1,395-$27,200 and the average price is $16,016.

Mileage: The range of mileage on the cars is 1,668-151,413 and the average mileage is 47,094.

Number of Reported Accidents: The range of reported accidents is 0-2 and the average number of accidents is .3067 ~1.

Number of owners: The range of number of owners is 1-4 and the average number of owners is 1.315 ~ 2. There are two entries missing data in this set.

Dealership: All of the cars are sold by dealerships.

Year: The range of years for the cars is 2006-2018 and the average year is 2015.36 ~ 2016. The median year is 2016.

Age of Vehicle: We added one more column called age of the vehicle by subtracting the Year from current year i.e. 2019. The range of the age of the cars is 1-13 years old and the average age of the cars is 3.64 years.

Highway Gas Mileage: The range of highway gas mileage on the vehicles is 24 mpg- 34 mpg and the average gas mileage is 30.986. There is one incident of missing data for this variable.

Exterior Color: There are 10 different exterior colors identified in this data set. There are two incidents of missing data for this variable.

Interior Color: There are four different interior colors identified in this data set. There are three incidents of missing data for this variable.

Engine Type: There are three different engine types observed in this data set.

Leather Seats: In this dataset, there are 58 vehicles that do not have leather seats and 17 vehicles that do have leather seats.

Sunroof: In this dataset, there are 38 vehicles that do not have sunroof and 37 vehicles that do have sunroof.

Dummy coding had to be used in Leather seats and Sunroof cases since both variables were present in yes/no form.

In a nutshell,

Leather seats -No -0; Leather seats -Yes -1

Sunroof -No -0; Sunroof -Yes -1

**Core Hypothesis**

Now after cleaning the data and the variables required are obtained, the core hypothesis can be formulated. Our dependent variable which we created the model to explain is the listed price of a used vehicle. We used a variety of quantitative and qualitative variables in our model to accurately predict the price of a used car in our segment.

The hypotheses are:

* **H1a: βmileage<0** - We strongly believe that the mileage would have an effect of the price of the vehicle. We believe that the price should decrease in increase in mileage. The mileage a car has is a good representation of how much it has been used and likely experienced wear and tear.
* **H2a: βage <0** - We believe that as the age of a vehicle increases, the price of the car decreases. We included the year of the car which we translated into age for a more accurate measurement. Logically it makes sense that newer cars would be more expensive.
* **H3a: βhwy\_gas\_mileage <0** - Further, we added fuel efficiency which was represented in our model by highway gas mileage. The more fuel efficient the vehicle is the cheaper it will be to use and therefore would affect the selling price.
* **H4a: βnumber\_of\_accidents <0** - The number of accidents the car has been in was added because we believe it will reduce the selling price. Having been in more accidents means the car has had repairs and is not in its original condition and could also have lingering damage.
* **H5a: βnumber\_of\_owners <0** - The last quantitative variable we used in our model was number of owners. We theorized that how many owners a car has previously had could impact its selling price.
* **H6a: βleatherseats\_yes >0**- Some customers find leather seats to be a luxury amenity and therefore it would result in a price increase.
* **H7a: βsunroof\_yes >0**- We believe having a sunroof can increase price of a car.
* Another qualitative variable we used was the model of the vehicle which is also attached to the vehicles make. We thought that the **brand of the vehicle** would play a role in its listing price. We didn’t find any interaction term required to build this model.

**Results and Interpretations**

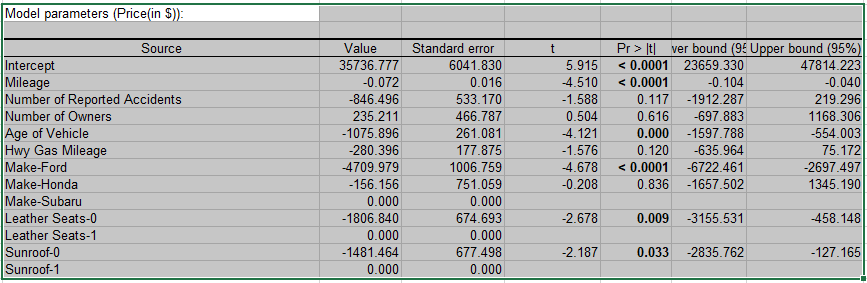


Figure I: Results of the regression

Our model was fairly accurate and explains about 85.2% of the variance in the price of used cars in our selected segment. We used a 95% confidence interval to evaluate the statistical significance of each variable in respect to our dependent variable. The variable that measured the mileage of the vehicle was statistically significant in our model. Based on our model, for every 1 unit increase in mileage the used cars listing price dropped on average $0.72 dollars keeping all other variables constant.

Further, the age of the vehicle was also significant and resulted on average in a decrease in $1076 dollars for every additional year increase in the age of the vehicle keeping all other independent variables constant. The model and therefore the make of the car was also statistically significant for the Ford Escape resulting on average in a $4710 reduction in used car price as compared to the Subaru Forester. Having leather seats also influenced the price and resulted on average in an increase of $1807 compared to vehicles that did not have leather seats keeping all other independent variables constant. Additionally, having a sunroof also played a role in the affecting the price of the cars. Having a sunroof on average increased the price $1482 compared to vehicles that did not have a sunroof keeping all other independent variables constant.

The number of previous owners did not play a statistically significant role in our model for determining the end listing price. Furthermore, the number of reported accidents did not influence price in our model significantly. Surprisingly highway gas mileage did not affect the price in a statistically significant way based on our model either. While previously stated the make of the Ford Escape had a statistically significant effect in our model, the Honda C-RV was not statistically significant in our model compared to the Subaru Forester.

**Sale Price Elasticity**

Elasticity can only be found for continuous variables. So, for our model, the variables that are significant and are continuous in nature are age of the vehicle and mileage.

So, **elasticity of price for change in age**= % change in price/% change in age

= (-1075) \* 3.640/16016

= (-0.244)

0.244 is the percent change in price if age differs by 1% that implies 24.4% decrease in price if age changes by 100%.

So, it’s an **inelastic change** since 1% change in independent variable gives less than 1% change in dependent variable.

So, **elasticity of price for change in mileage**= % change in price / % change in mileage

= (-0.072) \* 47094.3/16016

= (-0.211)

0.211 is the percent change in price if mileage differs by 1% that implies 21.1% decrease in price if mileage changes by 100%.

So, it’s an **inelastic change** since 1% change in independent variable gives less than 1% change in dependent variable.