Automotive Sales

Transportation is an important factor for day-to-day living. We utilize public roadways, bus systems, airplanes, trains, and maybe self-driving vehicles one day. These means of travel have one thing in common: money. When it comes to transportation there is very little the consumer can do to negotiate prices; public roadways force citizens to pay taxes, busses charge a set fare, and airports overcharge for tickets. So instead of focusing this semester project on rates that are non-negotiable, we’d like to focus on a data set that could benefit the average consumer of one particular type of transportation. We’d like to focus our efforts on personal vehicles and begin to answer questions that will help consumers choose the right car based on future resale value.

The goal of this project is to help provide consumers with useful information before making a decision to purchase a vehicle. Information would depend on the interests of the consumer i.e. does the purchaser plan to resell, buy to own, car features ect.. One way to evaluate our model for the consumer would be to develop a decision tree. Using the preferences of the buyer, a decision tree could distinguish the best match. Another interesting way to evaluate the data would be through clustering. The data could be clustered into vehicle types (sedan, suv, sport) and then be used to compare between vehicle cost, manufacturer, and 4-year resale value.

We plan to use a lot of the same tools we’ve grown familiar with in class. Jupyter Notebook will be the main environment we will be using to clean, analyze, model, and plot most, if not all of our data. Numpy and Pandas will be helpful for cleaning the set of a few null values that reside in one of our columns. While matplotlib and seaborn will allow us to visualize the information so we can get clearer predictions for our project.

The dataset takes a look at 157 different models of cars, spanning 30 different makes. There are 16 different categories included in the data set as well, from monetary values like price and sales, to more technical measurements like curb length and wheelbase. The data set comes from Analytixlabs.

<https://deepblue.lib.umich.edu/bitstream/handle/2027.42/112123/lycumich.pdf?sequence=1&isAllowed=y>

This is a honors thesis done by an individual at the University of Michigan, Yuchen Lin. Lin takes car sales data across 3 decades to try and examine the relationship between the changing prices and the sales of automobiles.

<https://www.inderscienceonline.com/doi/abs/10.1504/IJEHV.2009.029035>

This paper explores the growing car market in India, focusing on 11 different factors that go into the purchasing of a car, including resale value. This data can help manufacturers in their decisions, and, in turn,helps consumers understand the market better, leading to them being better buyers.

<https://www.sciencedirect.com/science/article/pii/S0301421514003024>

This paper focuses its sight on the newer EV market, looking at where this new direction of car sales can go. This data can help us try and compare and contrast the resale values between traditional gas powered vehicles and electric and hybrid vehicles.