Buy Used not New

# **Jontavius Caston Semester GitHub Portfolio URL**

# Which Domain?

What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you’re doing with these data.

The data domain will be the used car market. Kelly blue box, Car fax, concerns that new cars have on the market, Car Guru, craigslist

# Which Data?

What is the dataset you’ll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.

Vehicles

<https://www.kaggle.com/austinreese/craigslist-carstrucks-data>

# Research Questions? Benefits? Why analyze these data?

How are you proposing to analyze this dataset? This is about your approach. Here, you’ll be proposing your research questions as well as justifications for why you’d offer these data in this way.

1. What model of vehicles retains the best value?
2. Most available model of vehicle in later years.
3. Least available model of vehicle in later years.
4. How does the mileage defer to say the number of possible seats available on a car ie. 2 door sports car compared to suv.
5. How does geography affect car choices?

# What Method?

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

1. Two tailed Hypothesis test to determine if geography affects the car choices.
2. K means and mean shift clustering methods to see the distribution of cars by geography
3. Linear and Logistic regression, to start, to see the decrease in value over time by mileage.
4. Feature Reduction may be needed here as well.

# Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

I do not clearly see many things that can go wrong with this project. I am assuming this as if the dataset is clearly laid out with a lot of good information. For the methods that I want to use I believe the only thing that can slow me down is by not having enough features. If there are too many features then I think that feature reduction would be a great fit to help us efficiently use more machine learning.

# Concluding Remarks

Tie it all together. Think of this section as your final report’s abstract.

1. I will be doing my project on the used car market. A couple of references that I will be looking at is Kelly blue book, Car fax, all concerns that new cars may have on the market, Car Guru, and Craigslist. My dataset will come from Kaggle and it is known as Vehicles. I have a couple questions that I am hoping to get answers from through my analysis. Two of the bigger question I am looking to answer are: 1.What model of vehicles retains the best value? 2. Most available model of vehicle in later years. I have a couple of methods that I will be using for this assignment. One of the models I am most excited to try is K means and mean shift clustering methods to see the distribution of cars by geography. I do not foresee many issues coming from this assignment. There is still a chance that some issues may arise as I get deeper into the project.