# Proposal for Capstone Project

**The most popular 15 used-car model price performance prediction on Craigslist**

**PROBLEM:**

More and more people who plan to sell and buy their used cars tend to choose online platforms since it helps consumers save time and cost efficiently. Craigslist, as the world's largest used car platform, has a great number of trading volumes every year which also provides a large amount of selling data.

Online used-car prices are affected by several factors, including the year, mileage, brand, model, etc. We try to analyze these influential data, build a machine learning model, and make price performance predictions on the most popular 15 models of used cars on craigslist.

**WHO MIGHT CARE:**

**Buyers** who may not have enough knowledge about the used-car market and tend to make their decisions by following the public’s ideas. They prefer to get a clear and concise standard by observing an average standard.

**Used car franchisees** who could know about market trends by observing used online used car price performance and consumer psychology. They could make up their marketing plans that may be helpful to increase benefits and decrease risks.

**Car manufacturers** who conduct market research. The used car price performance usually reflects the hedging rate directly. Range of price performance may affect manufacturers' future development and sales strategies.

**DATA:**

This data is from Kaggle <https://www.kaggle.com/austinreese/craigslist-carstrucks-data>

It contains most all relevant information that Craigslist provides on car sales including columns like price, condition, manufacturer, latitude/longitude, and 18 other categories. Since I am a beginner of data science, I plan to choose a topic which I could follow other’s steps when I meet troubles.

**APPROACH:**

My first step is to clean and analyze the existing data, in order to conclude a price performance standard for each sample. After that I plan to build the model to predict price performance of certain models. The results should be clarified into three classes: over standard, standard, below standard (Multi Class classification) after changing the price variable.

**DELIVERABLES**:

Codes (notebooks) for

a. data acquisition

b. data cleaning

c. data exploration analysis

d. machine learning model development

Report on the capstone project

Presentation on the capstone project

All materials will be uploaded to my Github repository.