**Names:** Michael Fonseca and Jatin Chittoor

**Databases:**

1. **Craigslist listings:** <https://www.kaggle.com/austinreese/craigslist-carstrucks-data>
2. **Car features & MSRP:** <https://www.kaggle.com/CooperUnion/cardataset>
3. **Edmunds database** (From blog csv from 2008 for car ratings): <https://kavita-ganesan.com/entity-ranking-data/#.Xi-J8xNKhQL>

**Thesis:** Correlation between year/make/model/car ratings for a set of cars and their current pricing on Craigslist

– Information we need to cross reference:

* Choose one year (2008) for inflation equal comparison
* Cooper Union = craigslist headings to align:
* Make = manufacturer
* Model = model
* Vehicle Style = type
* Tie in Edmunds 2008 database for each make / model to get letter rating
* Manipulate the make and model underscores from docid heading and add overall\_rating value to the table

– Scrub in the following selections:

* Condition = excellent
* Title Status = clean

– Final values we want:

* Manufacturer
* Model
* Type
* Overall Rating
* Condition
* Title Status
* Miles
* Price

– Steps:

* **Extract:** Download all csvs from kaggle to Postgres (Jatin lead)
* **Transform:** Manipulate with SQL - join tables and create new table (Mike lead)
* **Load:** Load up the new table - back to Postgres and thereafter Github (Mike lead)
  + (Note obviously we will be in contact on Slack / text as needed on all)

– Information to find a relevant correlation **(Optional)**:

* Miles and a distribution for miles with equal condition and title as it relates to price
* Price as it relates to rating in each equivalent class
* Price as it relates to make/manufacturer in each equivalent class