

## Final Report

### Data Sources:

Cars.com, TrueCar.com (aggregate used car websites)

### Detailing the process of the extraction, transformation, and loading steps

- Manually enter search parameters to create a sample of cars.
- Use splinter to navigate between search results.
  - Scraping each individual page added time to the extraction process. For TrueCar we decided to scrape the search page to speed up the process, although there is less specific information. We were still able to scrape the URL to each individual car result.
- Scrape information from each website

### What data sources you chose, and why

Sites such as: AutoTrader, Kelly Blue Book and MarketCheck all lack serious functionality and usability when applying filtering options. These vehicle database sites also have large inventory listing for vehicle searches due to large visibility. Unfortunately, the available vehicles are not consolidated within one location or 100% accurate.

We chose examples of sites which have limited functionality when searching for specific vehicle features. The examples we chose display the VIN (vehicle identification number) which is composed of 17 alpha numeric characters that act as a unique identifier for each vehicle since 1981.

### Explication why you have performed the types of transformations you did

- We stored all search results in a non-relational database (MongoDB)
- Collections are grouped by website.
- Each search yields different types of data. Cars.com gives a link to Google Maps, while TrueCar simply gives the city as a location.
- Any useful table join will be implied by search parameters (i.e, wanting to see all the cars of a single color, make, body style, etc. will be returned by the search results during extraction).

### Hypothetical use cases for your database:

Application designed to assist the user in locating a vehicle available for sale within designated US location. This application will allow the user to gather used car results across multiple platforms, eliminating the redundancy of searching for the same criteria on each site. Using the unique vehicle identifier (VIN) allows us to provide the features of the vehicle determined by the manufacture avoiding user input errors.

### This VIN number can be decoded and provide:

- world manufacturer identifier
- the vehicle attributes - vehicle type (VDS)
  - automobile platform
  - model
  - body style
- model year
- the manufacturing plant code
- the vehicle identifier section (VIS)

- engine
  - transmission
  - fuel type
- manufacturer identifier
- sequential number

Utilizing this decoded VIN we can offer the actual vehicle statistics in conjunction with the listing sites, providing a simple compiled source, a checks/balance of relevant information per the user's interest. Compounded with the functionality to allow the user to define a custom search and group each search into a collection, linking collections by a search id.