Matthew Wolfgram

SI 206 003

3/25/2018

Project Proposal

I plan to crawl the Autotrader website (<https://www.autotrader.com/> ,) using the drop-down menu to cycle through different car manufacturers, then displaying a list of an automaker’s available models once a user selects a brand on the command line. After a specific model is chosen, the first 100 listings are crawled, and the HTML for the listings are cached. After the caching has been executed, the HTML stored in it is parsed and compiled into two tables. The first table is generated as soon as the user selects a brand—it contains the brand name, the models produced by the brand, as well as the model ID affiliated with each specific model. Once a specific model of car is chosen from the brand, the 100 listings are put into a second table that compiles data about the mileage, body style, drive type, engine, transmission, and price, as well as the interior/exterior color of each car. The second table is linked to the first via the model ID associated with each car. Once this displayable data is compiled, the user can choose to generate Plotly graphs of certain items in the table—if specified, a graph could be generated about the distribution of interior colors of a specific model, showing how often certain colors were selected over the range of 100 listings sampled. The user could generate similar distribution graphs for mileage, body style, drive type, engine, transmission, price, and exterior color.

***Challenge Score:*** Crawling [and scraping] multiple pages in a site you haven’t used before ✣ (8) = ***8 points***