Summary:

The secondary ticket market in the U.S. is estimated at $5 billion a year and growing. The dominant player is StubHub (owned by eBay), with about 50 percent of the market share, according to Billboard. Ticketmaster’s TM+ system is catching up, however, with an aggressive marketing plan for lower fees and cheaper tickets.

But what does it mean for consumers, especially on the concert side? Does the secondary market mean people spend very different amounts of money to see the same acts? Is there more variance by venue size, location or even type of audience (ie geezer rock vs. pop vs. new acts)? Can consumers apply a strategy in the timing of their ticket purchases (or sales) for certain acts and venues? And is there evidence of repeated, large ticket purchases that hit the secondary market, suggesting a wholesale business for some venues/artists?

Data:

A daily scrape of StubHub’s website for all concerts over the next few months. Currently there are probably several hundred artists in the ‘Rock, Pop and HipHop’ category, and there are probably several hundred venues that will be in play over the next few months. I will time the scrape to timeout every 10-20 minutes so we don’t get shut down, but we will be collecting information on thousands of tickets daily. Eventually, I may need to cull some artists or venues, but for now I will attempt to scrape them all.

Methodology:

The data will be scraped into compressed XML files; I’ll need to write a second Python script to get them into either a large .csv file or database. The initial script will need to search for the first time a seat is offered for a given date.artist.venue [d.a.v], as well as the last time, in order to get the final purchased price and the first offered price, gradually filling in the secondary price paid for each seat for each [d.a.v.] by the time of the concert. I will also need to scrape the list price for each [d.a.v.] as a baseline.

Variables:

artist scrape date

seat (section/number) price

event date outlet (if more than StubHub is scraped)

artist type (rock, hip hop, etc.)\* appeal (young, old tours, new, etc.)\*

venue venue size (large, medium, etc.)\*

city area type (suburban, urban, etc.)\*

Secondary scrape:

venue price for each [d.a.v]

seats/per section section category\* (prime, tiers 1-4, etc)

Known assumptions:

* Scraping only 50% of the market; will not capture seats sold other ways
* Cannot be sure seat is sold (could be withdrawn)
* Will not capture tickets sold in < a day

Challenges:

* Will need more data than can be captured by end of class
* Scraper may be too big