

# Homework-4-2

March 31, 2017

## 1 Scraping Hotel Ratings on Tripadvisor

To be completed INDIVIDUALLY and due on April 7 at 3pm.

In this homework we will practice web scraping. Let's get some basic information for each hotel in Boston. On each hotel page, scrape the Traveler ratings. (15 pts)

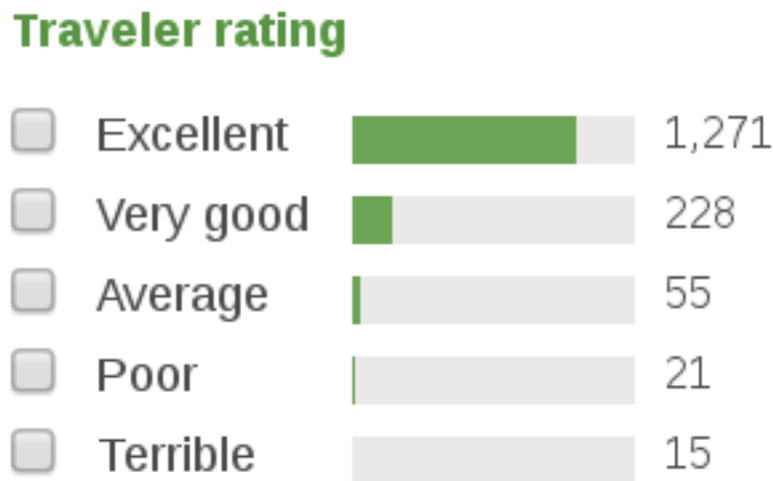


Figure 1: Information to be scraped

Save the data in "traveler\_ratings.csv" in the following format:

hotel\_name, rating, count

To receive credit, you must commit traveler\_ratings.csv to Github.

In [ ]:

---

Next, scrape all the reviews of each hotel for the star ratings of the following attributes: Value, Location, Sleep Quality, Rooms, Cleanliness, Service. Note that some reviews may not have attribute ratings and some may only have some of the attributes. (25 pts)

Save the data in "attribute\_ratings.csv" in the following format:

hotel\_name, review\_id, attribute, star\_value

To receive credit, you must commit attribute\_ratings.csv to Github.

In [ ]:

*“Extraordinary Service and Views...but not the Four Seasons for the price...”*



Reviewed July 20, 2010

Management has evidently done a superb job of teaching staff to offer exemplary, friendly, efficient service in this hotel on the waterfront. Views to the harbor with its yachts, schools, and bridges were beautiful... Shannon, at the reception desk, was helpful to us and Concierge Rob Fournier was ever-ready to assist. Night room service delivery was handled with gracious efficiency. Staff couldn't do enough for us and every request was met. We didn't hear every skipping child and closing door in the hotel. I stayed in a relatively small room, and I preferred the Back Bay room with a view of the Public Garden and the serene, luxurious comfort at a lower price for the same price. My husband enjoyed the boats and I didn't disagree, so we will have something to discuss when we return to Boston!

Stayed July 2010, traveled with family



Value



Location



Sleep Quality



Rooms



Cleanliness



Service

Figure 2: Information to be scraped