Individual Case Study Plan

Bihan Zhuang

11/9/2018

Motivation

As I am doing research on a potential Spring break trip to Japan, I find the Booking.com a great resource. Booking.com is a travel fare aggregator website and travel metasearch engine for lodging reservations. Hotels in Japanese major tourist cities post listings on Booking.com, making it really convenient for people, like me, who are not proficient in Japanese to book directly through a well-organized English site (The Japanese sites are really fun to look at though! Very different from American ones, and usually contain lots of discount information).

One thing currently not on Booking.com is summary statistics about all the listings in a city at a given time. Personally I think this information could be very useful for foreign visitors. For example, if I have a week in Japan, which part of the country should I visit? Depending on the average hotel price of a region, I might choose the Kansai (Osaka & Kyoto) area over Hokkaido. I might even first visit one region and then another based on the price. Moreoever, what types of room are available and which should I book? Japanese style Ryokan with Tatami or Western room with beds? Will I have my own bathroom or need to share one with others? As foreign travelers, we might not be familiar with all the lodging options and customs in Japan, and sometimes could find ourselves surprised by some rules (e.g. traditional ryokan hotels have late checkin time and early checkout time due to the thoroughness of cleaning that is required). Summary statistics and visualizations that provide an overview could come to our rescue.

With this in mind, I plan to web scrape the Booking.com, aiming to produce useful summary statistics and visualizations for lodging in a few Japanese major tourist cities – Tokyo, Osaka, Kyoto, and Sapporo.

Data Source, Methods and Expected End Product

From the landing page of Booking.com, users can search lodging information with a very simple query, which consists of required fields – destination and checkin/checkout date, and optional fields – number of rooms, travelers information, and whether the trip is for work. The search query URL can be constructed by filling in the names of Japanese cities, and any date that I am interested in researching.

Once the search result page is generated, I will scrape the distance from city center as well as the hyperlinks to go further down to each hotel listing page. On any hotel's page, I will scrape the following information:

- Address
- Ratings
- Room type
- Price per night
- Free cancellation or not
- Checkin time, checkout time
- Facilities

With the scraped information, I present some of my initial plans regarding statistics and visualizations:

- Table of room types
- Average room price per night
- Average ratings
- Table of checkin time, checkout time
- Table of noteworthy/interesting facilities

- Number of listings allowing free cancellation

I plan to show these stats and visualizations in a Shiny app or web app. The app will be interactive, allowing the users to choose one of the four cities, the time of travel, and to toggle parameters. Potentially the app will also allow comparison between dates and cities. Ideally the app will be able to scrape the Booking.com and generate stats and visualizations on the fly, so that the information is most up-to-date. However this depends on how much overhead would be experienced.