

Problem Statement – 2. Acquire list of *possibly* all tourist destinations in the world

GitHub Repo

<https://github.com/ladynsnoopy/TouristDestinations>

Selected Data Source

Trip Advisor (<https://www.tripadvisor.com.sg/>)

Reason

Allows for web scraping, has a comprehensive list of attractions for almost every country and has similar webpage format for most of the country's information.

Completeness

Out of 196 countries, tourist destination data for 142 countries was obtained. The trip advisor webpages for the rest had varying formats hence, web scraping was difficult.

Names of countries where tourist destination data was not obtained

'Afghanistan', 'Angola', 'Antigua and Barbuda', 'Armenia', 'The Bahamas', 'Barbados', 'Bolivia', 'Burundi', 'Cabo Verde', 'Comoros', 'Congo, Democratic Republic of the', 'Congo, Republic of the', 'Côte d'Ivoire', 'Cuba', 'Denmark', 'Djibouti', 'Dominica', 'Dominican Republic', 'Eswatini', 'The Gambia', 'Georgia', 'Grenada', 'Haiti', 'Iran', 'Jamaica', 'Kiribati', 'Korea, North', 'Lebanon', 'Lesotho', 'Liberia', 'Libya', 'Madagascar', 'Maldives', 'Malta', 'Mauritania', 'Mauritius', 'Micronesia, Federated States of', 'Mongolia', 'Montenegro', 'Nauru', 'Netherlands', 'Palau', 'Saint Kitts and Nevis', 'Saint Lucia', 'Saint Vincent and the Grenadines', 'San Marino', 'Sao Tome and Principe', 'Singapore', 'Sudan, South', 'Syria', 'Trinidad and Tobago', 'Turkmenistan', 'Tuvalu', 'Vatican City'

Approach

- 1) Navigate from trip advisor homepage and search for a country's attractions
- 2) Obtain the hyperlink for each category of attractions from the main attraction page
- 3) Saved these hyperlinks along with the corresponding country names in weblinks.csv
- 4) Repeat this process for all the countries in countrynames.csv

Things to Do in Japan

When are you travelling? Start Date End Date

Types of Things to Do

Top Attractions

Shop Tours & Tickets

Types of Attractions

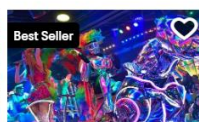
- ☐ Sights & Landmarks (262)
- ☐ Nature & Parks (287)
- ☐ Museums (92)
- ☐ Zoos & Aquariums (39)
- ☐ Water & Amusement Parks (18)
- ☐ Fun & Games (23)
- ☐ Spas & Wellness (36)
- More ▾

Explore Japan (29)

See all



Universal Studios Jap...
130 Reviews
from S\$101.50*



Tokyo Robot Restaur...
492 Reviews
from S\$68.25*



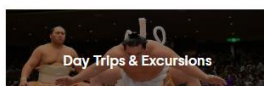
Official Street Go-Kar...
164 Reviews
from S\$110.60*



Mt Fuji, Hakone, Lake...
4,609 Reviews
from S\$201.69*

Shop Tickets And Tours (4949)

Shop Now



Step 2

- 5) Built a web scraper to obtain tourist destination data from all the hyperlinks obtained in step 3
- 6) Upon visiting each of the hyperlink, the web scraper will obtain the name of the attractions located at the bottom part of the page which are sorted by traveller ranking
- 7) The name of the attractions and the hyperlink it was obtained from is saved in resultdata.csv

The screenshot displays four attraction listings from Trip Advisor. Yellow circles highlight the attraction names: 'Dotonbori', 'Fushimi Inari-taisha Shrine', 'Kinkakuji Temple', and 'Kiyomizu-dera Temple'. Yellow arrows point from each of these circles to a central point labeled 'Step 6', indicating the step in the data engineering process where these specific attractions are being processed.

Attraction Name	Location	Reviews	Ranking	Experiences
Dotonbori	Osaka	12,412	#1 of 944 Sights & Landmarks in Osaka	See 61 Experiences
Fushimi Inari-taisha Shrine	Kyoto	24,306	#1 of 1,162 Sights & Landmarks in Kyoto	See 128 Experiences
Kinkakuji Temple	Kyoto	16,935	#2 of 1,162 Sights & Landmarks in Kyoto	See 117 Experiences
Kiyomizu-dera Temple	Kyoto	11,432	#3 of 1,162 Sights & Landmarks in Kyoto	See 106 Experiences

Explanation of Approach

I chose to use the traveller ranked attractions for each category as Trip Advisor does completely display all the attractions data in their webpages. Based on the category one clicks, Trip Advisor applies a filter. Hence, I decided that the web crawler had to click on each of the attraction category independently and obtain the traveller ranked attractions for the category.

Further Improvements

Instead of searching in Trip Advisor by country names, searching by state names may yield more tourist destination data.