

...

Webscraping Project

Trip Advisor

*By : Joseph Sabaybay
Maryjane Tela*



Agenda



Motivation



Data/Visuals



Challenges/ Next Steps

Motivation

www.tripadvisor.ca

Founded in February
2000

Largest travel guidance
platform

Helps users guide them
on where to stay/eat/see

Curious as to what
Vancouver has to offer

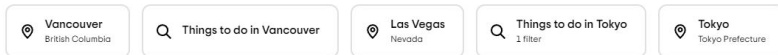
Top attractions





Website





https://www.tripadvisor.ca/Attractions-g154943-Activities-c41-a_allAttractions.true-Vancouver_British_Columbia.html





Data

1375 links/rows
6 attributes



Code

```
# grabs all the links from all the pages
num = 0
links = [] # appends it to this List
for _ in range(46):
    get_page(num)
    sl(2) # had to manipulate the order of the sleep function. It skipped the first 2 pages and here is the position where it work
    soup = driver.page_source
    soup = BeautifulSoup(soup, 'lxml')
    print(soup)
    sl(2)
    scroll_dwn()
    for div in soup.find_all('div', class_ = 'a1PVI eNWhq PgLKC tnGGX'):
        a = div.find('a').get('href')
        new_a = 'https://tripadvisor.ca' + a
        links.append(new_a)
    #print(a)
    num += 30 # the url changes by 30 when going to the next page
```

Dataframe

	attraction_names	address	category	rating	num_reviews	suggested_duration(hr)	main_category	sub_cat1	sub_cat2
0	Stanley Park	Stanley Park Causeway, Vancouver, British Columbia	[Parks]	4.5	29062	3	Parks	None	None
1	Granville Island	NaN	[Neighbourhoods Points of Interest & Landmarks]	4.5	19830	2	Neighbourhoods	Points of Interest & Landmarks	Islands
2	Vancouver Aquarium	845 Avision Way, Stanley Park, Vancouver, British...	[Aquariums]	4.5	8522	2	Aquariums	None	None
3	Queen Elizabeth Park	4600 Cambie St, Vancouver, British Columbia V5L 4V5	[Parks, Gardens]	4.5	2367	2	Parks	Gardens	None
4	VanDusen Botanical Garden	5251 Oak St, Vancouver, British Columbia V6M 4L4	[Gardens]	4.5	2077	3	Gardens	None	None
...
1034	Taxi2Airport Vancouver	Wateringschans 28, Vancouver, British Columbia	[Taxis & Shuttles]	NaN	0	NaN	Taxis & Shuttles	None	None
1035	Heartspace Massage and Healing Arts	925 Granville St Suite 609, Vancouver, British...	[Spas]	NaN	0	2	Spas	None	None
1036	Vancouver Orthotics for Flat Feet, Foot Pain, ...	NaN	[Sports Camps & Clinics]	NaN	0	NaN	Sports Camps & Clinics	None	None
1037	A Stylish Arrival	4155 Beaufort Pl, Vancouver, British Columbia	[Taxis & Shuttles]	NaN	0	NaN	Taxis & Shuttles	None	None
1038	Sams Original Art	1736 2nd Ave W, Vancouver, British Columbia V5Z 1V5	[Art Galleries]	NaN	0	NaN	Art Galleries	None	None

1039 rows x 9 columns

```
data = []
# once i get all the links
for link in links:
    driver.get(link) # gets to the link
    sl(3) # pause for 3 seconds before scrolling down
    soup = driver.page_source
    soup = BeautifulSoup(soup, 'lxml') # updates the page source every single current page
    scroll_dwn()
    sl(2)
    # get all attributes on your page
    try:
        t = soup.find('h1', class_ = 'biGQs _P fiohW eIegn').get_text()
        data.append(t)
    except:
        data.append('N/a')

    try:
        a = soup.find_all('button', class_ = 'UiKNM _G B- _S _T c G_ P0 wSSLS wMNQs raEKE')[1].get_text()
        data.append(a)
    except:
        data.append('N/a')

    try:
        c = soup.find('div', class_ = 'firGe _T bgWZj').get_text()
        data.append(c)
    except:
        data.append('N/a')

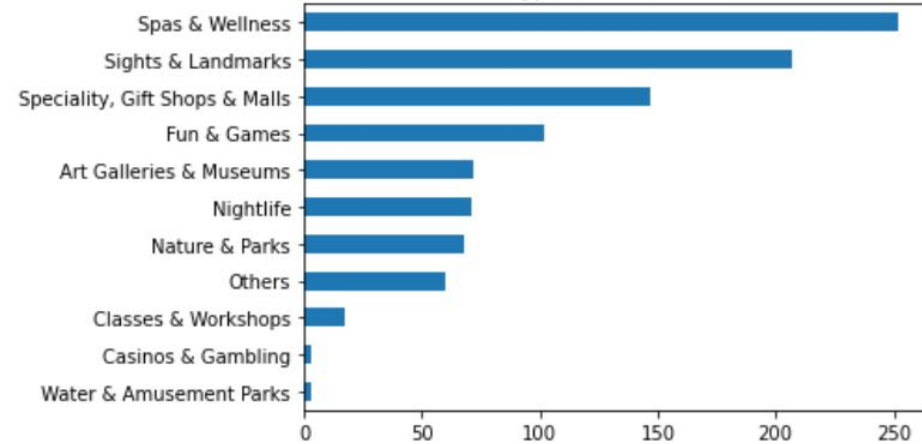
    try:
        r = soup.find('div', class_ = 'biGQs _P fiohW hzzSG uuBRH').get_text()
        data.append(r)
    except:
        data.append('N/a')

    try:
        num_review = soup.find('span', class_ = 'biGQs _P pZUb8 bikBZ KxBGd').get_text()
        data.append(num_review)
    except:
        data.append('N/a')

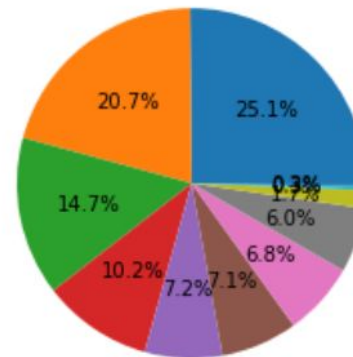
    try:
        suggested_duration = soup.find('div', class_ = '_c').get_text().split(' ')[0].split('.')[-1]
        data.append(suggested_duration)
    except:
        data.append('N/a')
```

Visuals

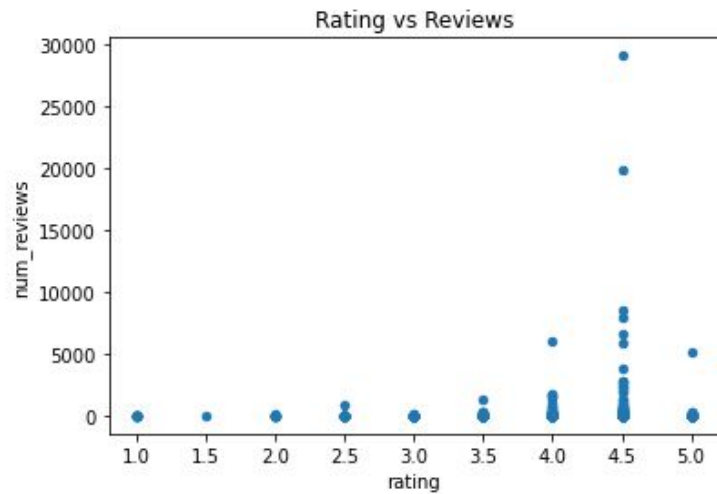
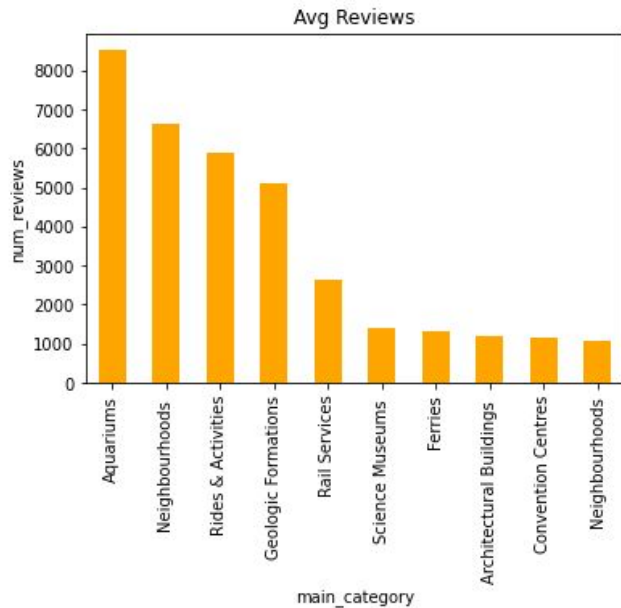
Type of Attractions



Type of Attractions



Visuals



Challenges



Writing the code for
scraping



Handling lots of errors



Errors in the dataframe



Analyzing the data

Next Steps



Comparison

**What Vancouver vs
Toronto looks like**

Scrape Hotels

**Find the best price or
rated hotel**

Scrape

**Restaurants
Find the best price or
rated restaurants**

Make an algorithm



Thank you!

