Web Scraping Project Data mining,Data Analyzing and Visualization

Submitted by : Praveen M R
Mobile : 9886226470

INTRODUCTION

Web Scraping using Python

Website address

http://www.thrillophilia.com

Data mining,Data Analyzing & Data Visualization of the collected Data

Aim

- To find the URL for each and every category of Activities Being listed on the entire website.
- For every place, fetching the data following info into a csv File: Category name Number of places on each category Total Bought on each category.
- In addition to above mentioned attributes,I have also fetching Activity vs Location vs User Demographics Information for each and every Activities.

Description:

The python script is written to fetch all the individual categories from the website(http://Thrillophilia.com)
The code is written for fetching the data from the first page and it iterates to each and every pages of website (activities, categories, count of bought),for that I have used statistical techniques and mathematically analysis to presenting the data into visualization .

Some packages:

- bs4 (BeautifulSoup4)
- Requests
- Pandas
- Re (regular expression)
- Matplotlib

Each city url ends with thing to do,total 80 cities urls

https://www.thrillophilia.com/cities/ho-chi-minh/things-to-do
https://www.thrillophilia.com/cities/hanoi/things-to-do
https://www.thrillophilia.com/cities/hoi-an/things-to-do
https://www.thrillophilia.com/cities/ubud/things-to-do
https://www.thrillophilia.com/cities/south-kuta-bali/things-to-do
https://www.thrillophilia.com/cities/nusa-dua/things-to-do
https://www.thrillophilia.com/cities/kintamani/things-to-do
https://www.thrillophilia.com/cities/kuala-lumpur/things-to-do
https://www.thrillophilia.com/cities/langkawi/things-to-do
https://www.thrillophilia.com/cities/penang/things-to-do
https://www.thrillophilia.com/cities/bangkok/things-to-do
https://www.thrillophilia.com/cities/phuket/things-to-do
https://www.thrillophilia.com/cities/pattaya/things-to-do
https://www.thrillophilia.com/cities/chiang-mai/things-to-do
https://www.thrillophilia.com/cities/phi-phi/things-to-do
https://www.thrillophilia.com/cities/krabi/things-to-do
https://www.thrillophilia.com/cities/siem-reap/things-to-do
https://www.thrillophilia.com/cities/phnom-penh/things-to-do
https://www.thrillophilia.com/cities/luang-prabang/things-to-do
https://www.thrillophilia.com/cities/vientiane/things-to-do
https://www.thrillophilia.com/cities/dubai/things-to-do
https://www.thrillophilia.com/cities/abu-dhabi/things-to-do
https://www.thrillophilia.com/cities/sharjah/things-to-do
https://www.thrillophilia.com/cities/mumbai/things-to-do
https://www.thrillophilia.com/cities/bangalore/things-to-do
https://www.thrillophilia.com/cities/manali/things-to-do
https://www.thrillophilia.com/cities/pune/things-to-do
https://www.thrillophilia.com/cities/jaipur/things-to-do
https://www.thrillophilia.com/cities/leh/things-to-do
https://www.thrillophilia.com/cities/munnar/things-to-do
https://www.thrillophilia.com/cities/cochin/things-to-do

LINKS OF EACH CATEGORY

```
current page https://www.thrillophilia.com/cities/hue-city/things-to-do
tags each place https://www.thrillophilia.com/cities/hue-city/tags/sightseeing?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/day-cuts?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/biking?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/attraction-visit?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/adventure-special--2?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/boating?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/cycling?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/trekking?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/nature-and-wildlife?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/aerial-activities?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/walking?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/family-fun--2?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/nightlife?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/walking-biking?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/caving?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/arts-culture?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/cruises-sailing--2?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/village-safari?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/sports-games?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/food-drinks?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/transfers?tour_type=Activity
tags each place https://www.thrillophilia.com/cities/hue-city/tags/wildlife?tour_type=Activity
```

USING PANDAS DATAFRAME

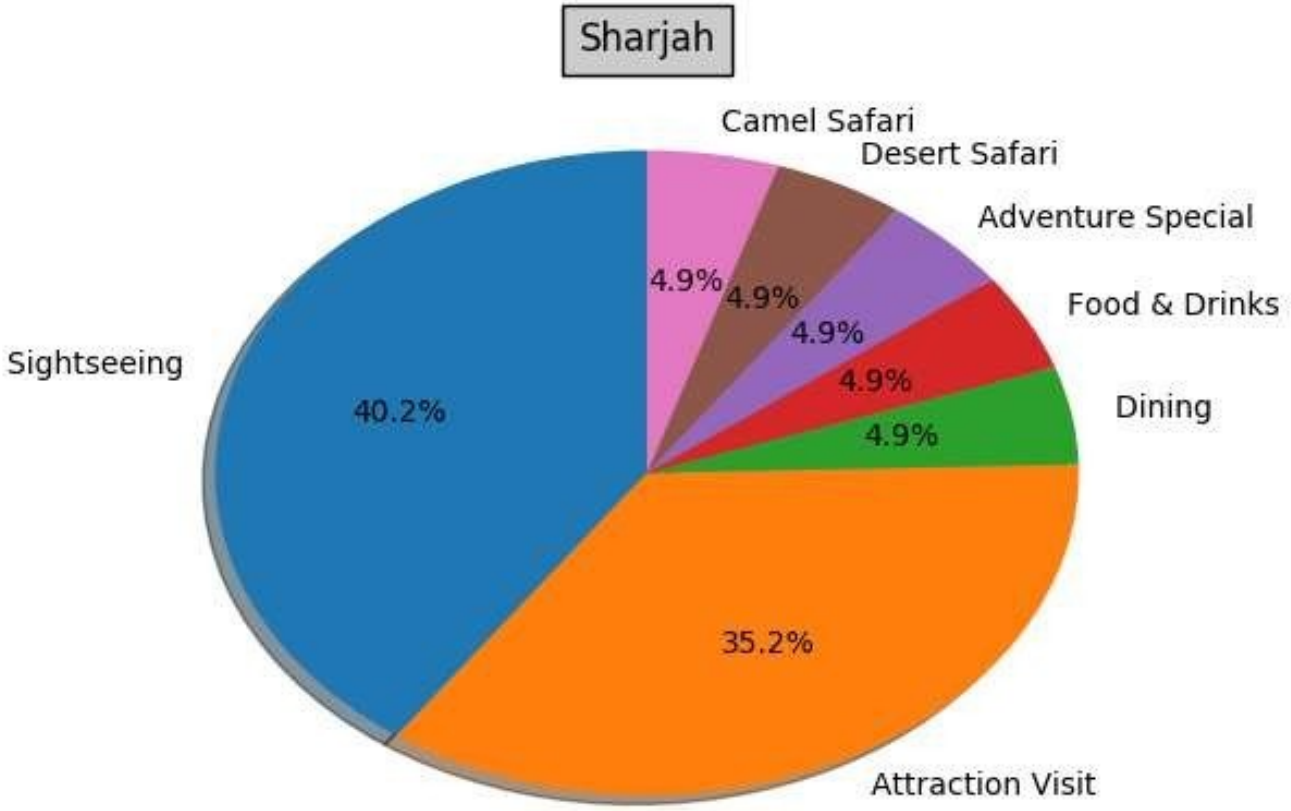
	Activities	Total_Count	Total_Bought	efficieny
0	Sightseeing	26	196	86.734694
1	Day Outs	17	163	89.570552
2	Biking	13	133	90.225564
3	Attraction Visit	8	47	82.978723
4	Adventure Special	8	67	88.059701
5	Boating	6	15	60.000000
6	Cycling	4	5	20.000000
7	Trekking	3	38	92.105263
8	Nature and Wildlife	3	26	88.461538
9	Aerial Activities	3	32	90.625000
10	Walking	2	1	-100.000000
11	Family Fun	2	1	-100.000000
12	Nightlife	2	31	93.548387
13	Walking & Biking	2	1	-100.000000
14	Caving	2	14	85.714286
15	Arts & Culture	1	1	0.000000
16	Cruises & Sailing	1	15	93.333333
17	Village Safari	1	1	0.000000
18	Sports & Games	1	7	85.714286
19	Food & Drinks	1	14	92.857143
20	Transfers	1	10	90.000000
21	Wildlife	1	12	91.666667

CSV Output:

The below mentioned csv file that generated after execution of the python script total 80 cities datasheets and graphs.

	A	B	C	D	E	F	G
1		Activities	Total_Count	Total_Bought	efficieny		
2	0	Sightseeing	26	194	86.59793814		
3	1	Day Outs	17	163	89.57055215		
4	2	Biking	13	133	90.22556391		
5	3	Attraction Visit	8	47	82.9787234		
6	4	Adventure Special	8	67	88.05970149		
7	5	Boating	6	15	60		
8	6	Cycling	4	5	20		
9	7	Trekking	3	38	92.10526316		
10	8	Nature and Wildlife	3	26	88.46153846		
11	9	Aerial Activities	3	32	90.625		
12	10	Walking	2	1	-100		
13	11	Family Fun	2	1	-100		
14	12	Nightlife	2	31	93.5483871		
15	13	Walking & Biking	2	1	-100		
16	14	Caving	2	14	85.71428571		
17	15	Arts & Culture	1	1	0		
18	16	Cruises & Sailing	1	15	93.33333333		
19	17	Village Safari	1	1	0		
20	18	Sports & Games	1	7	85.71428571		
21	19	Food & Drinks	1	14	92.85714286		
22	20	Transfers	1	10	90		
23	21	Wildlife	1	12	91.66666667		
24							
25							

Data visualization :



Data visualization: