Scraping yellowpages.com to create csv file of vegan restaurants in New York

1. Using request library to download the web page

In [21]:
```
!pip install requests --upgrade --quiet
import requests
```

In [27]:
```
url='https://www.yellowpages.com/search?search_terms=vegan+restaurant&geo_location_terms=New+York%2C+NY'
response=requests.get(url)
response.status_code
```

Out[27]:
```
200
```

In [29]:
```
page_content=response.text
```

2. Using beautiful soup to parse and extract information

In [30]:
```
!pip install beautifulsoup4 --upgrade --quiet
from bs4 import BeautifulSoup
```

In [39]:
```
document=BeautifulSoup(page_content,'html.parser')
title=document.find_all('a',{'class':'business-name'})
```

Retrieving names of the restaurants

In [38]:
```
title_tags=document.find_all('a',{'class':'business-name'})
```

```
['Little Beet', 'Ramen Ishida', 'Liquiteria Chelsea', 'Pita Grill', 'EN Japanese Brasserie', 'Vatan Indian Restaurant', 'Pure Food & Wine', 'Candle 79', 'Darbar Grill', 'Taim West
Village', 'Cafe Fiorello', "Alice's Tea Cup Chapter 2", 'Caravan Of Dreams', 'Saravana Bhavan', 'Good Enough to Eat', "Kate's Joint", 'Bite', 'Franchia Teahouse & Restaurant', 'The
Organic Grill', 'Quartino Bottega Organica', 'Ayurveda Cafe', "Nonna's L.E.S. Pizzeria", 'Balade', 'Bite', 'Strictly Roots Restrnt', 'Mudspot', "Alice's Tea Cup", 'Massawa Restaura
nt', 'Tandoor Palace', 'Company']
```

In [43]:
```
topic_title=[]
for title in  title_tags:
    topic_title.append(title.text.strip())
print(topic_title)
len(topic_title)
```

```
['Little Beet', 'Ramen Ishida', 'Liquiteria Chelsea', 'Pita Grill', 'EN Japanese Brasserie', 'Vatan Indian Restaurant', 'Pure Food & Wine', 'Candle 79', 'Darbar Grill', 'Taim West
Village', 'Cafe Fiorello', "Alice's Tea Cup Chapter 2", 'Caravan Of Dreams', 'Saravana Bhavan', 'Good Enough to Eat', "Kate's Joint", 'Bite', 'Franchia Teahouse & Restaurant', 'The
Organic Grill', 'Quartino Bottega Organica', 'Ayurveda Cafe', "Nonna's L.E.S. Pizzeria", 'Balade', 'Bite', 'Strictly Roots Restrnt', 'Mudspot', "Alice's Tea Cup", 'Massawa Restaura
nt', 'Tandoor Palace', 'Company']
```

Out[43]:
```
30
```

Retrieving contact number of the restaurants

In [41]:
```
contact_tags=document.find_all('div',{'class':'phones phone primary'})
```

In [45]:
```
contacts=[]
for contact in  contact_tags:
    contacts.append(contact.text.strip())
print(contacts)
len(contacts)
```

```
['(917) 338-7111', '(646) 590-3276', '(212) 929-0002', '(212) 717-2005', '(212) 647-9196', '(212) 689-5666', '(212) 477-1010', '(212) 537-7179', '(212) 751-4600', '(212) 691-1287',
'(212) 595-5330', '(212) 486-9200', '(212) 254-1613', '(212) 679-0204', '(212) 496-0163', '(212) 777-7059', '(212) 431-0301', '(212) 213-1001', '(212) 477-7177', '(212) 529-5133',
'(212) 932-2400', '(212) 673-3450', '(212) 529-6868', '(212) 677-3123', '(800) 728-3375', '(212) 228-9074', '(212) 799-3006', '(212) 663-0505', '(212) 349-7643', '(212) 243-1105']
```

Out[45]:
```
30
```

Retrieving address of the restaurants

- Retrieving street address

In [73]:
```
street_address_tags=document.find_all('div',{'class':'street-address'})
```

In [72]:
```
street_address=[]
for street in street_address_tags:
    street_address.append(street.text.strip())
print(street_address)
len(street_address)
```

```
['111 W 40th St', '122 Ludlow St', '102 8th Avenue Corner of 8th Ave. and 15th St.', '1603 2nd Ave', '435 Hudson St', '409 3rd Ave', '54 Irving Pl', '154 E 79th St', '157 E 55th S
t', '222 Waverly Pl', '1900 Broadway', '156 E 64th St', '405 E 6th St', '81 Lexington Ave', '520 Columbus Ave', '58 Avenue B', '333 Lafayette St', '12 Park Ave', '123 1st Ave', '11
Bleecker St', '706 Amsterdam Ave', '105 Clinton St', '208 1st Ave Frnt A', '211 E 14th St', '2058 Adam Clayton Powell Jr Blvd', '307 E 9th St', '102 W 73rd St', '1239 Amsterdam Av
e', '88 Fulton St', '230 9th Ave']
```

Out[72]:
```
30
```

- Retrieving locality of the restaurants

In [71]:
```
locality_address_tags=document.find_all('div',{'class':'locality'})
```

Out[71]:
```
[<div class="locality">New York, NY 10018</div>,
 <div class="locality">New York, NY 10002</div>,
 <div class="locality">New York, NY 10011</div>,
 <div class="locality">New York, NY 10028</div>,
 <div class="locality">New York, NY 10014</div>]
```

In [74]:
```
locality_address=[]
for locality in locality_address_tags:
    locality_address.append(locality.text.strip())
print(locality_address)
len(locality_address)
```

```
['New York, NY 10018', 'New York, NY 10002', 'New York, NY 10011', 'New York, NY 10028', 'New York, NY 10014', 'New York, NY 10016', 'New York, NY 10003', 'New York, NY 10075', 'Ne
w York, NY 10022', 'New York, NY 10014', 'New York, NY 10023', 'New York, NY 10065', 'New York, NY 10009', 'New York, NY 10016', 'New York, NY 10024', 'New York, NY 10009', 'New Yo
rk, NY 10012', 'New York, NY 10016', 'New York, NY 10003', 'New York, NY 10012', 'New York, NY 10025', 'New York, NY 10002', 'New York, NY 10009', 'New York, NY 10003', 'New York,
NY 10027', 'New York, NY 10003', 'New York, NY 10023', 'New York, NY 10027', 'New York, NY 10038', 'New York, NY 10001']
```

Out[74]:
```
30
```

Retreiving url of each restaurant

In [78]:
```
link_url_tags=document.find_all('a',{'class':'business-name'})
len(link_url_tags)
topicUrl='https://yellowpages.com'+link_url_tags[0]['href']
print(topicUrl)
```

```
https://yellowpages.com/new-york-ny/mip/little-beet-483865664
```

In [79]:
```
link_to_the_restaurant=[]
base='https://yellowpages.com'

for link in link_url_tags:
    link_to_the_restaurant.append(base+link['href'])
link_to_the_restaurant
```

Out[79]:
```
['https://yellowpages.com/new-york-ny/mip/little-beet-483865664',
 'https://yellowpages.com/new-york-ny/mip/ramen-ishida-536625242',
 'https://yellowpages.com/new-york-ny/mip/liquiteria-chelsea-480731271',
 'https://yellowpages.com/new-york-ny/mip/pita-grill-468793878',
 'https://yellowpages.com/new-york-ny/mip/en-japanese-brasserie-11487346',
 'https://yellowpages.com/new-york-ny/mip/vatan-indian-restaurant-29635107',
 'https://yellowpages.com/new-york-ny/mip/pure-food-wine-12304005',
 'https://yellowpages.com/new-york-ny/mip/candle-79-468493961',
 'https://yellowpages.com/new-york-ny/mip/darbar-grill-8908203',
 'https://yellowpages.com/new-york-ny/mip/taim-west-village-11487910',
 'https://yellowpages.com/new-york-ny/mip/cafe-fiorello-2981486',
 'https://yellowpages.com/new-york-ny/mip/alices-tea-cup-chapter-2-2147582',
 'https://yellowpages.com/new-york-ny/mip/caravan-of-dreams-2408805',
 'https://yellowpages.com/new-york-ny/mip/saravana-bhavan-2286094',
 'https://yellowpages.com/new-york-ny/mip/good-enough-to-eat-481122116',
 'https://yellowpages.com/new-york-ny/mip/kates-joint-18227806',
 'https://yellowpages.com/new-york-ny/mip/bite-7447248',
 'https://yellowpages.com/new-york-ny/mip/franchia-teahouse-restaurant-383021',
 'https://yellowpages.com/new-york-ny/mip/the-organic-grill-3951895',
 'https://yellowpages.com/new-york-ny/mip/quartino-bottega-organica-10689248',
 'https://yellowpages.com/new-york-ny/mip/ayurveda-cafe-7779083',
 'https://yellowpages.com/new-york-ny/mip/nonnas-l-e-s-pizzeria-533468580',
 'https://yellowpages.com/new-york-ny/mip/balade-452130033',
 'https://yellowpages.com/new-york-ny/mip/bite-1501521',
 'https://yellowpages.com/new-york-ny/mip/strictly-roots-restrnt-473152538',
 'https://yellowpages.com/new-york-ny/mip/mudspot-525431847',
 'https://yellowpages.com/new-york-ny/mip/alices-tea-cup-7275840',
 'https://yellowpages.com/new-york-ny/mip/massawa-restaurant-5529423',
 'https://yellowpages.com/new-york-ny/mip/tandoor-palace-3355173',
 'https://yellowpages.com/new-york-ny/mip/company-471045147']
```

3. Creating pandas data frame

In [81]:
```
!pip install pandas --quiet
import pandas as pd
```

In [83]:
```
topics_dict={'Name of the restaurant':topic_title,'Contact Number':contacts,'Street Address':street_address,
             'Locality':locality_address,'Link to the restaurant':link_to_the_restaurant}
```

In [85]:
```
topic_df=pd.DataFrame(topics_dict)
topic_df
```

Out[85]:

	Name of the restaurant	Contact Number	Street Address	Locality	Link to the restaurant
0	Little Beet	(917) 338-7111	111 W 40th St	New York, NY 10018	https://yellowpages.com/new-york-ny/mip/little...
1	Ramen Ishida	(646) 590-3276	122 Ludlow St	New York, NY 10002	https://yellowpages.com/new-york-ny/mip/ramen-...
2	Liquiteria Chelsea	(212) 929-0002	102 8th Avenue Corner of 8th Ave. and 15th St.	New York, NY 10011	https://yellowpages.com/new-york-ny/mip/liquit...
3	Pita Grill	(212) 717-2005	1603 2nd Ave	New York, NY 10028	https://yellowpages.com/new-york-ny/mip/pita-g...
4	EN Japanese Brasserie	(212) 647-9196	435 Hudson St	New York, NY 10014	https://yellowpages.com/new-york-ny/mip/en-jap...
5	Vatan Indian Restaurant	(212) 689-5666	409 3rd Ave	New York, NY 10016	https://yellowpages.com/new-york-ny/mip/vatan-...
6	Pure Food & Wine	(212) 477-1010	54 Irving Pl	New York, NY 10003	https://yellowpages.com/new-york-ny/mip/pure-f...
7	Candle 79	(212) 537-7179	154 E 79th St	New York, NY 10075	https://yellowpages.com/new-york-ny/mip/candle...
8	Darbar Grill	(212) 751-4600	157 E 55th St	New York, NY 10022	https://yellowpages.com/new-york-ny/mip/darbar...
9	Taim West Village	(212) 691-1287	222 Waverly Pl	New York, NY 10014	https://yellowpages.com/new-york-ny/mip/taim-w...
10	Cafe Fiorello	(212) 595-5330	1900 Broadway	New York, NY 10023	https://yellowpages.com/new-york-ny/mip/cafe-f...
11	Alice's Tea Cup Chapter 2	(212) 486-9200	156 E 64th St	New York, NY 10065	https://yellowpages.com/new-york-ny/mip/alices...
12	Caravan Of Dreams	(212) 254-1613	405 E 6th St	New York, NY 10009	https://yellowpages.com/new-york-ny/mip/carava...
13	Saravana Bhavan	(212) 679-0204	81 Lexington Ave	New York, NY 10016	https://yellowpages.com/new-york-ny/mip/sarava...
14	Good Enough to Eat	(212) 496-0163	520 Columbus Ave	New York, NY 10024	https://yellowpages.com/new-york-ny/mip/good-e...
15	Kate's Joint	(212) 777-7059	58 Avenue B	New York, NY 10009	https://yellowpages.com/new-york-ny/mip/kates-...
16	Bite	(212) 431-0301	333 Lafayette St	New York, NY 10012	https://yellowpages.com/new-york-ny/mip/bite-7...
17	Franchia Teahouse & Restaurant	(212) 213-1001	12 Park Ave	New York, NY 10016	https://yellowpages.com/new-york-ny/mip/franch...
18	The Organic Grill	(212) 477-7177	123 1st Ave	New York, NY 10003	https://yellowpages.com/new-york-ny/mip/the-or...
19	Quartino Bottega Organica	(212) 529-5133	11 Bleecker St	New York, NY 10012	https://yellowpages.com/new-york-ny/mip/quarti...
20	Ayurveda Cafe	(212) 932-2400	706 Amsterdam Ave	New York, NY 10025	https://yellowpages.com/new-york-ny/mip/ayurve...
21	Nonna's L.E.S. Pizzeria	(212) 673-3450	105 Clinton St	New York, NY 10002	https://yellowpages.com/new-york-ny/mip/nonnas...
22	Balade	(212) 529-6868	208 1st Ave Frnt A	New York, NY 10009	https://yellowpages.com/new-york-ny/mip/balade...
23	Bite	(212) 677-3123	211 E 14th St	New York, NY 10003	https://yellowpages.com/new-york-ny/mip/bite-1...
24	Strictly Roots Restrnt	(800) 728-3375	2058 Adam Clayton Powell Jr Blvd	New York, NY 10027	https://yellowpages.com/new-york-ny/mip/strict...
25	Mudspot	(212) 228-9074	307 E 9th St	New York, NY 10003	https://yellowpages.com/new-york-ny/mip/mudspo...
26	Alice's Tea Cup	(212) 799-3006	102 W 73rd St	New York, NY 10023	https://yellowpages.com/new-york-ny/mip/alices...
27	Massawa Restaurant	(212) 663-0505	1239 Amsterdam Ave	New York, NY 10027	https://yellowpages.com/new-york-ny/mip/massaw...
28	Tandoor Palace	(212) 349-7643	88 Fulton St	New York, NY 10038	https://yellowpages.com/new-york-ny/mip/tandoo...
29	Company	(212) 243-1105	230 9th Ave	New York, NY 10001	https://yellowpages.com/new-york-ny/mip/compan...

4. Creating csv file

In [87]:
```
topic_df.to_csv('restaurants.csv',index=None)
```