IBM Data Science Capstone

Battle of The Neighborhoods

Thomas Vitalis

Coursera Capstone Project

Geolocating Neighborhoods

Use of Nominatim workframe

In [3]:
```
address = 'Georgiou Papandreou, Thessaloniki, Greece'

geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geo coords of current home are {}, {}.'.format(latitude, longitude))
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/ipykernel/__main__.py:3: DeprecationWarning: Using Nominatim with the defa
ult "geopy/1.18.1" `user_agent` is strongly discouraged, as it violates Nominatim's ToS https://operations.osmfoundation.org/po
licies/nominatim/ and may possibly cause 403 and 429 HTTP errors. Please specify a custom `user_agent` with `Nominatim(user_age
nt="my-application")` or by overriding the default `user_agent`: `geopy.geocoders.options.default_user_agent = "my-applicatio
n"`. In geopy 2.0 this will become an exception.
  app.launch_new_instance()
```

```
The geo coords of current home are 40.5997456, 22.9515123.
```

In [4]:
```
#Initial Coords Data (Existing Residence example)
neighborhood_latitude=40.5997456
neighborhood_longitude=22.9515123
```

In [5]:
```
i=1
LIMIT = 500
radius = 1500*i

url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)

results = requests.get(url).json()
```

Importing datasets

Data Formatting & Data Reading (csv's & pdf's)

NYC Neighbourhood data

In [10]:
```
# Read csv file with clustered neighborhoods with geodata
nyc_data  = pd.read_csv('https://raw.githubusercontent.com/tvitalis/Coursera-Capstone/master/nynta.csv')
nyc_geojson = 'https://raw.githubusercontent.com/tvitalis/Coursera-Capstone/master/NeighborhoodAreas.geojson'
nyc_neigh = nyc_data['NTAName']
nyc_neigh.drop_duplicates
nyc_data.head()
```

Out[10]:

	BoroName	the_geom	CountyFIPS	BoroCode	NTACode	NTAName	Shape_Leng	Shape_Area
0	Brooklyn	MULTIPOLYGON (((-73.97604935657381 40.63127590...	47	3	BK88	Borough Park	39247.228028	5.400502e+07
1	Queens	MULTIPOLYGON (((-73.80379022888246 40.77561011...	81	4	QN51	Murray Hill	33266.904995	5.248828e+07
2	Queens	MULTIPOLYGON (((-73.8610972440186 40.763664477...	81	4	QN27	East Elmhurst	19816.712293	1.972685e+07
3	Queens	MULTIPOLYGON (((-73.75725671509139 40.71813860...	81	4	QN07	Hollis	20976.335574	2.288777e+07
4	Manhattan	MULTIPOLYGON (((-73.94607828674226 40.82126321...	61	1	MN06	Manhattanville	17040.685413	1.064708e+07

In [11]:
```
nyc_data.tail()
```

Out[11]:

	BoroName	the_geom	CountyFIPS	BoroCode	NTACode	NTAName	Shape_Leng	Shape_Area
190	Brooklyn	MULTIPOLYGON (((-73.93213397515774 40.72815960...	47	3	BK76	Greenpoint	29047.573201	3.533358e+07
191	Manhattan	MULTIPOLYGON (((-73.96236596889439 40.72420906...	61	1	MN50	Stuyvesant Town-Cooper Village	12021.790416	5.582283e+06
192	Bronx	MULTIPOLYGON (((-73.8312915777183 40.855434104...	5	2	BX37	Van Nest-Morris Park-Westchester Square	42870.392803	3.630238e+07
193	Bronx	MULTIPOLYGON (((-73.90958727269663 40.84275637...	5	2	BX14	East Concourse-Concourse Village	27223.847106	1.822240e+07
194	Bronx	MULTIPOLYGON (((-73.9119181232027 40.843257886...	5	2	BX63	West Concourse	28499.044417	1.937982e+07

In [12]:
```
unidata = pd.read_csv('https://raw.githubusercontent.com/tvitalis/Coursera-Capstone/master/COLLEGE_UNIVERSITY.csv')
unidata
```

IBM Data Science Capstone

Parsing Craiglist's listings

-Querying CL Classifieds using zip codes

-Selenium Driver

-Chrome Automation

-Storing Results

-Results Quality

-Dropping false entries

-Creating a dataframe for data analysis

In [17]:
```
# Generate Craigslist Links

base_links = []
for i in range(0, 6):
    link = "https://newyork.craigslist.org/search/aap?postal={}".format(unidata['ZIP'].iloc[i])
    base_links.append(link)
```

In [18]:
```
base_links = list(dict.fromkeys(base_links))
base_links
```

Out[18]:
```
['https://newyork.craigslist.org/search/aap?postal=11369',
 'https://newyork.craigslist.org/search/aap?postal=11201',
 'https://newyork.craigslist.org/search/aap?postal=10027',
 'https://newyork.craigslist.org/search/aap?postal=10011']
```

In [19]:
```
def getZipListings(link):
    driver = webdriver.Chrome(ChromeDriverManager().install())
    driver.get(link)

    titles = []
    dates = []
    prices = []
    bedrooms = []
    links = []

    items = driver.find_elements_by_class_name('result-info')
    for item in items:
        try:
            titles.append(item.find_element_by_class_name('result-title').get_attribute('innerText'))
        except:
            titles.append("")

        try:
            dates.append(item.find_element_by_class_name('result-date').get_attribute('datetime'))
        except:
            dates.append("")

        try:
            prices.append(item.find_element_by_class_name('result-price').get_attribute('innerText'))
        except:
            prices.append("")

        try:
            bedrooms.append(item.find_element_by_class_name('housing').get_attribute('innerText'))
        except:
            bedrooms.append("")

        try:
            links.append(item.find_element_by_class_name('result-title').get_attribute('href'))
        except:
            links.append("")

    driver.close()
    data = [titles, dates, prices, bedrooms, links]
    df = pd.DataFrame(data).transpose()
    df.columns = ['Title', 'Date', 'Price', 'Bedrooms', 'Link']
    df['Zipcode'] = int(link[-5:])
```

Created Dataframe

Out[21]:

	Unnamed: 0	Zipcode	Date	Price	Bedrooms	Title	Link
349	110	10011	2019-06-25 15:56	$3775	1br -	The Elusive West Village 1 bedroom. Renoed & B...	https://newyork.craigslist.org/mnh/fee/d/new-y...
350	111	10011	2019-06-24 19:38	$4500	1br -	SPECTACULAR CHELSEA 1BR-LAUNDRY-PRIV TERRACE-S...	https://newyork.craigslist.org/mnh/fee/d/new-y...
351	112	10011	2019-06-24 18:23	$2995	1br - 500ft2 -	Large Studio Alcove on 19th & 9th No-Fee July ...	https://newyork.craigslist.org/mnh/nfb/d/new-y...
352	113	10011	2019-06-24 17:05	$2900	NaN	**BEAUTY IN THE HEART OF CHELSEA** **STUDIO W/...	https://newyork.craigslist.org/mnh/fee/d/new-y...
353	114	10011	2019-06-24 17:04	$7695	1br -	WEST VILLAGE- CHRISTOPHER ST-LUXURY LIFESTYLE	https://newyork.craigslist.org/mnh/nfb/d/new-y...
354	115	10011	2019-06-23 19:13	$5995	3br -	AMAZING/SPACIOUS/RENOVATED/PRIVATE BALCONY 3 B...	https://newyork.craigslist.org/mnh/nfb/d/new-y...
355	116	10011	2019-06-23 10:36	$2950	NaN	77 W.15St, SS Kit, Lg STUDIO, Balc, SS Kit, Gy...	https://newyork.craigslist.org/mnh/nfb/d/new-y...
356	117	10011	2019-06-22 10:19	$2950	NaN	77 W.15St, SS Kit, Lg STUDIO, Balc, SS Kit, Gy...	https://newyork.craigslist.org/mnh/nfb/d/new-y...
357	118	10011	2019-06-19 10:31	$1895	NaN	!!GREAT DEAL!! !!STUDIO HEART OF CHELSEA!!	https://newyork.craigslist.org/mnh/fee/d/new-y...
358	119	10011	2019-06-18 16:08	$2995	1br -	2 x's as Big as any others. 700 Squarefeet in ...	https://newyork.craigslist.org/mnh/fee/d/new-y...

Visualization with Folium

Map Creation – Geodata Referencing – Using Nominatim with Folium

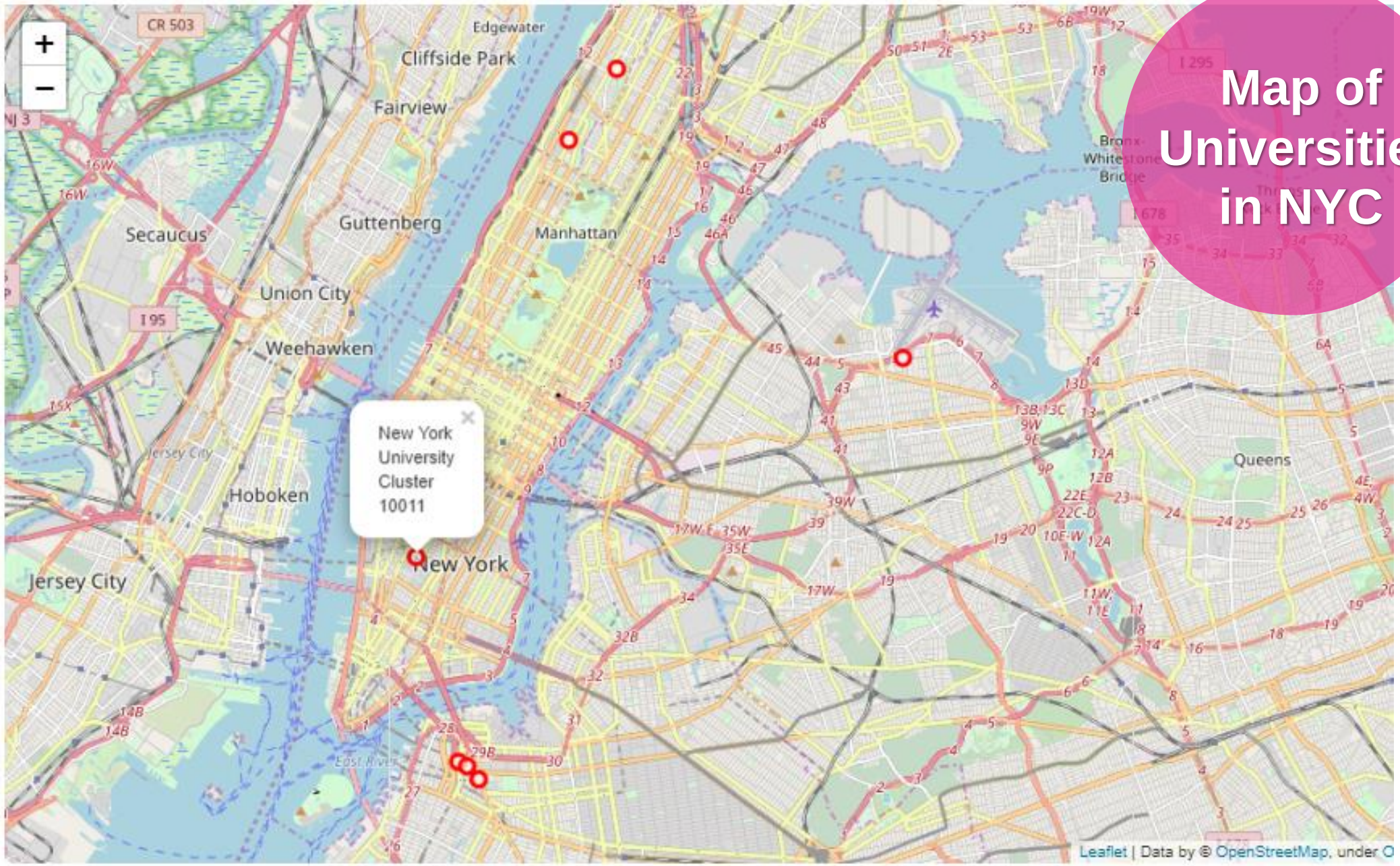
In [25]:
```
latitude= 40.7308619
longitude= -73.9871558
map_points = folium.Map(location=[latitude, longitude], zoom_start=14)

#folium.GeoJson(
#    nyc_geojson,
#    name='nyc_geojson'
#).add_to(map_points)

for lat, lng, label in zip(mtastations['GTFS Latitude'], mtastations['GTFS Longitude'], mtastations['Stop Name']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=7,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_points)

map_points
```

MTA Subway Stations

Out[25]:

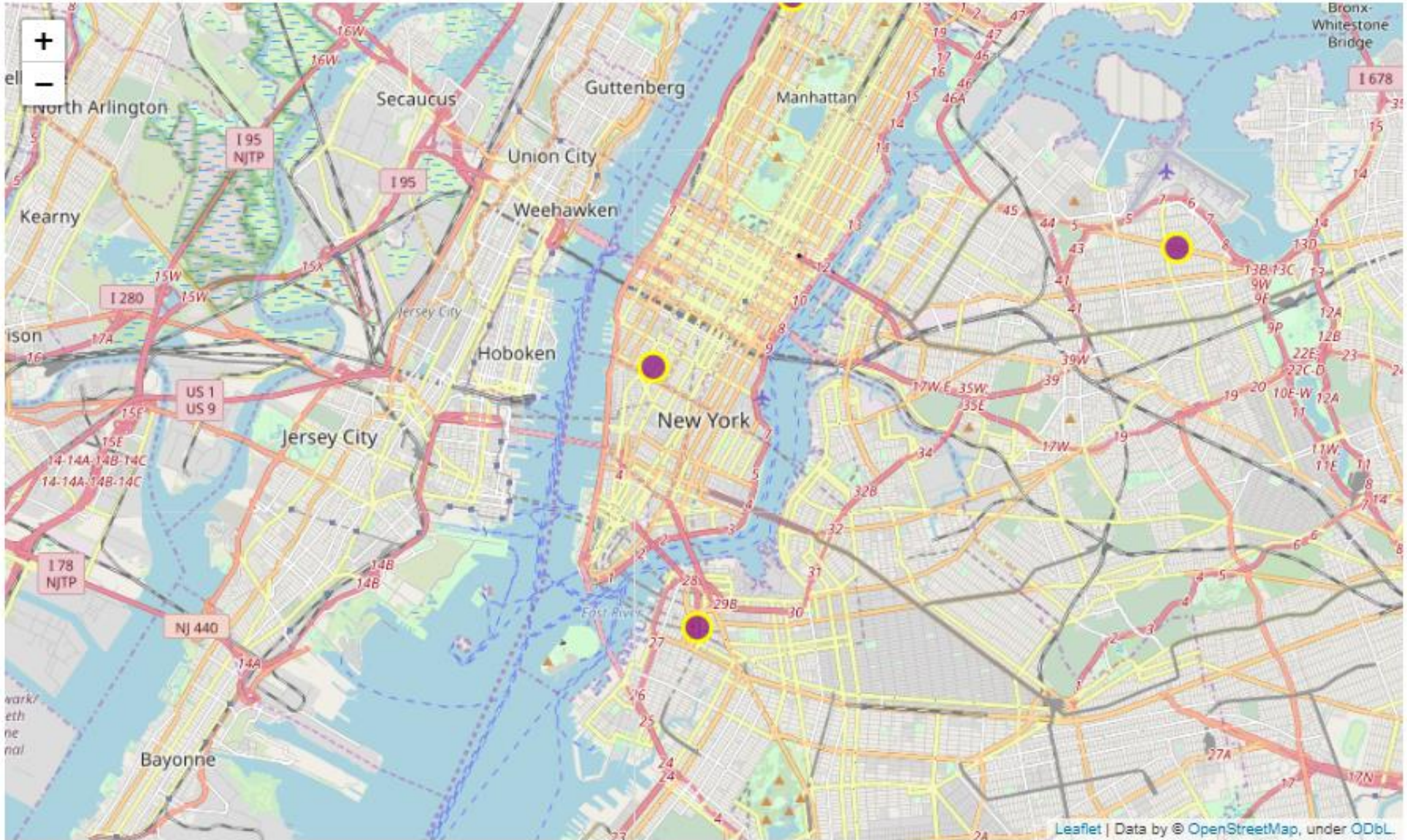
Map of Universities in NYC

6]:

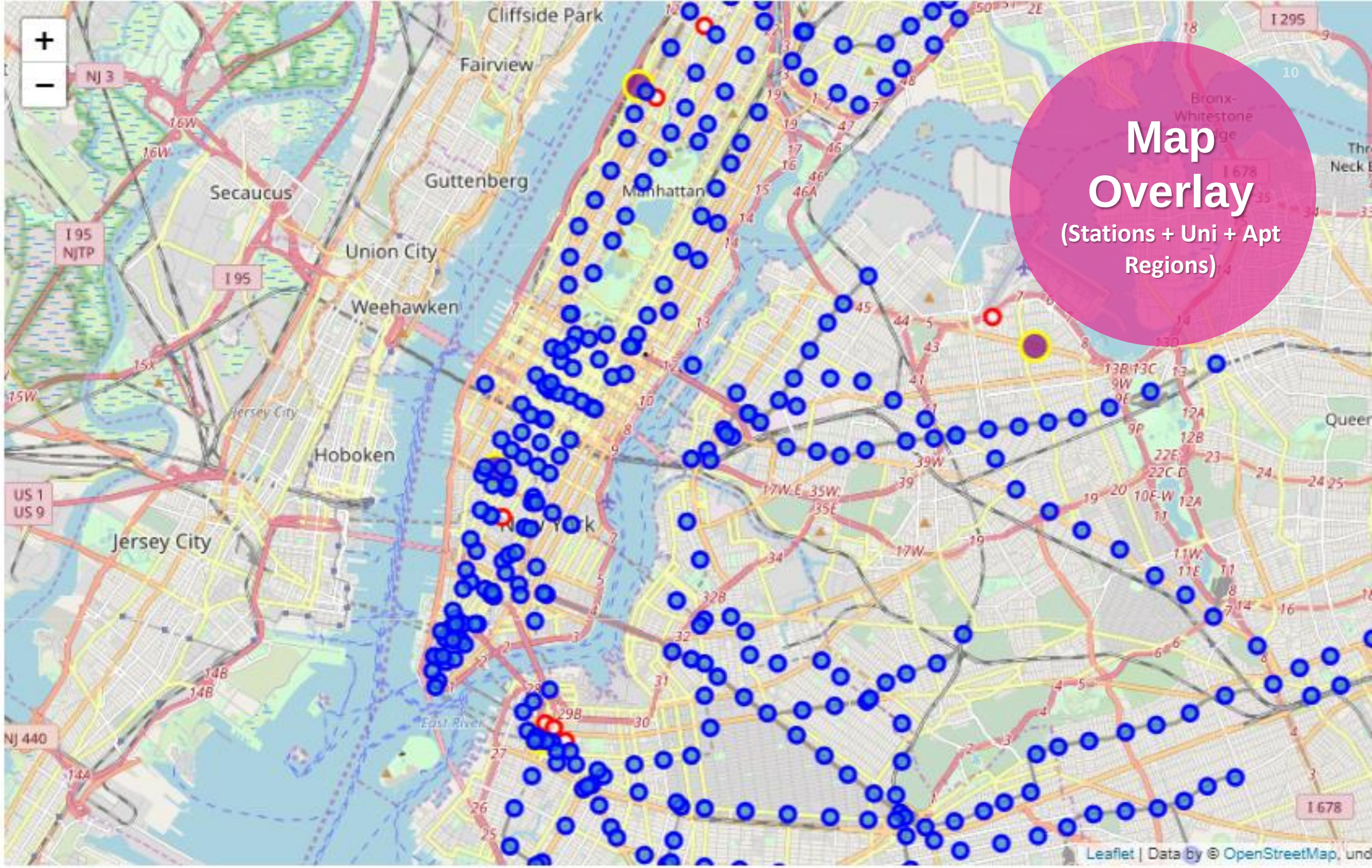


Queried ZIP Locations

Out[28]:



IBM Data Science Capstone



Neighborhood Characteristics ZIP 11201

	name	categories	lat	lng
0	SoulCycle Brooklyn Heights	Cycle Studio	40.692253	-73.991042
1	Shake Shack	Burger Joint	40.692127	-73.988658
2	New York Transit Museum	History Museum	40.690469	-73.989963
3	Perelandra Natural Foods	Grocery Store	40.693380	-73.991341
4	Heatwise	Yoga Studio	40.693450	-73.991788
5	Equinox Brooklyn Heights	Gym	40.692530	-73.991587
6	Orangetheory Fitness	Gym	40.693967	-73.991519
7	Xtend Barre Brooklyn Heights	Gym / Fitness Center	40.693599	-73.992376
8	Brooklyn Historical Society	History Museum	40.694942	-73.992333
9	Queen	Italian Restaurant	40.691319	-73.991647

```
nearby_venues_all_11201['categories'].value_counts().nlargest(12)
```

```
Coffee Shop              7
Yoga Studio              6
Cocktail Bar             5
Grocery Store            4
Italian Restaurant       4
Bar                      4
Gym / Fitness Center     4
Pizza Place              3
Ice Cream Shop           3
Bakery                   3
Japanese Restaurant      2
History Museum           2
Name: categories, dtype: int64
```

	name	categories	lat	lng
0	Riverside Park	Park	40.806809	-73.968651
1	Book Culture	Bookstore	40.806629	-73.964940
2	Columbia Greenmarket	Farmers Market	40.807195	-73.964335
3	Riverside Park @ 115th St.	Park	40.806640	-73.966514
4	Shake Shack	Burger Joint	40.807933	-73.964013
5	Alma Mater Statue	Outdoor Sculpture	40.807726	-73.962252
6	Book Culture	Bookstore	40.805104	-73.964980
7	sweetgreen	Salad Place	40.807284	-73.964753
8	Arts and Crafts Beer Parlor	Pub	40.806689	-73.961094
9	Community Food & Juice	American Restaurant	40.805823	-73.965483

```
nearby_venues_all_10027['categories'].value_counts().nlargest(12)
```

```
Park                    6
Italian Restaurant      6
Coffee Shop             5
American Restaurant     4
Mexican Restaurant      3
Bakery                  3
Seafood Restaurant      3
Bookstore               3
Grocery Store           2
Indian Restaurant       2
Tennis Court            2
Burger Joint            2
Name: categories, dtype: int64
```

Neighborhood Characteristics ZIP 10027

Decision Making Based on Best Price, Proximity to Campus and proximity to transit

Our Verdict:

Best combination provided by the neighborhood of Columbia University

THANK YOU!

- https://github.com/tvitalis/Coursera-Capstone
- https://www.linkedin.com/in/tvitalis/