Herkese Açık | Public

PROJECT: BATTLE OF NEIGHBORHOODS

WHERE TO OPEN AN ITALIAN RESTAURANT IN MANHATTAN ?

Serkan ONAR, Nov 2020

AGENDA

- **Introduction (WK4)**
- **Data (WK4)**
- **Methodology (WK5)**
- **Results (WK5)**
- **Discussion (WK5)**
- **Conclusion (WK5)**

Herkese Açık | Public

INTRODUCTION (WK4)

With its metropolitan property New York City -mostly spelled as New York- is one of the most popular cities in the world. As a metropolitan, New York hosts people from several nations. About 65.2 millions tourists every year visit New York and its magnificient monuments.

Over 1.3 million Italians and Italian-Americans live in the greater New York metro area, with about 800,000 living within one of the five New York City boroughs.There is even a neighborhood called Little Italy (Italian: Piccola Italia) is in Lower Manhattan in New York City, once known for its large Italian population(*Wikipedia:Italian Americans in New York City,Little Italy, Manhattan*)

Italian cuisine such as Pizza,Pasta,Fettuccine,Rizotto,Lasagna with aromatic hause wine and a delicious slice of Tiramisu are very famous actors in gastronomy world.Then what about opening a new fine-dine Italian Restaurant in New York , especially in Manhattan, among skyscrappers or near Central Park ? As an investor, where to choose the best location in Manhattan ? **(Business problem)**

DATA (WK4)

Based on the business problem , we need following data ;

1-New York City data including the boroughs, neighborhoods, geographical coordinates.

Data source: https://geo.nyu.edu/catalog/nyu_2451_34572

Description : We will get New York districts (Borough-Neighborhood) table and get geographical coordinates.

2-Polygon file of Manhattan boundaries.

Data source:

https://raw.githubusercontent.com/codeforamerica/click_that_hood/master/public/data/manhattan.geojson (I have modified it a little bit)

Description : We will get Manhattan neighborhood map to visualize potential places.

3-Venue information in each neighborhood of New York:

Data source: Foursquare developer-API

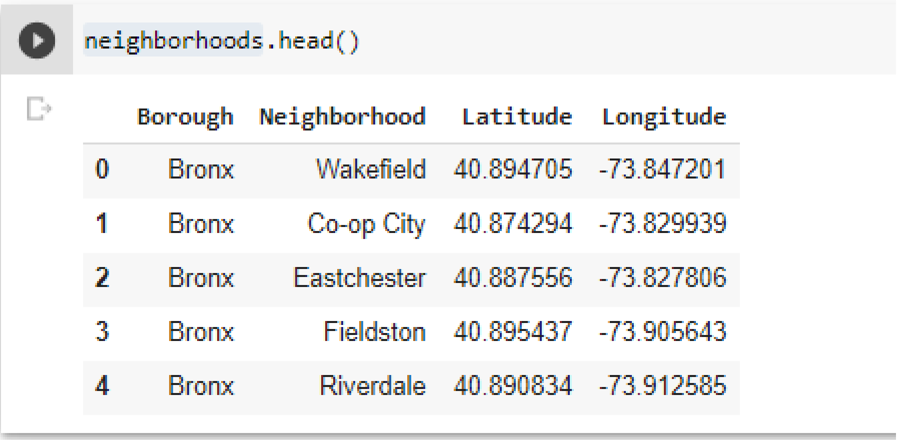
Description : By using this API we will collect all venue information in each neighborhood.

METHODOLOGY (WK5)

- Borough, neighborhood data to be collected from NewYork Dataset file. (Web scrapping)
- Pandas and Numpy library to process the data. (Data wrangling)
- Geopy library to be used to collect geographical coordinates , data will be collected, cleaned and processed into a dataframe. (Map visualization)
- Foursquare to be used to locate all venues and then filtered by Italian restaurants. Venue , venue coordinates, venue categories to be extracted and added to the dataframe. (Working with API)
- Scikit-learn library to cluster venues data. (Machine learning)
- Folium library to be used to visualize the neighborhood,venue,clustering data. (Map visualization)
- Matplotlib library to be used to visualize restaurants data-statistical approach. (Data visualization)

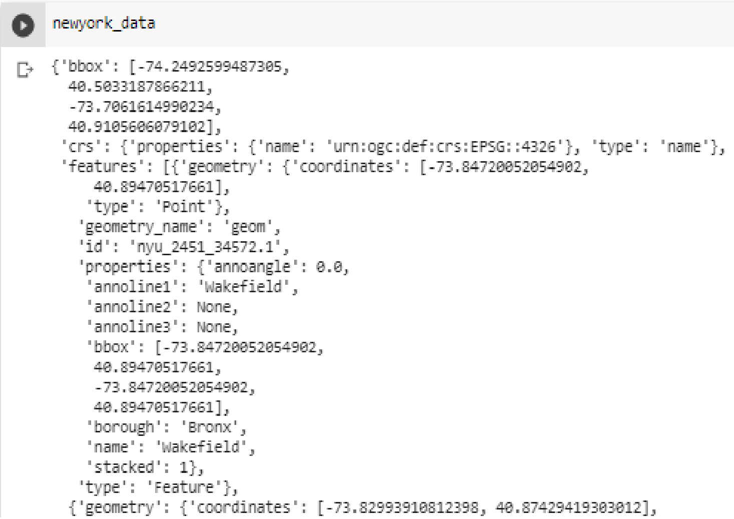
Herkese Açık | Public

RESULTS (WK5)



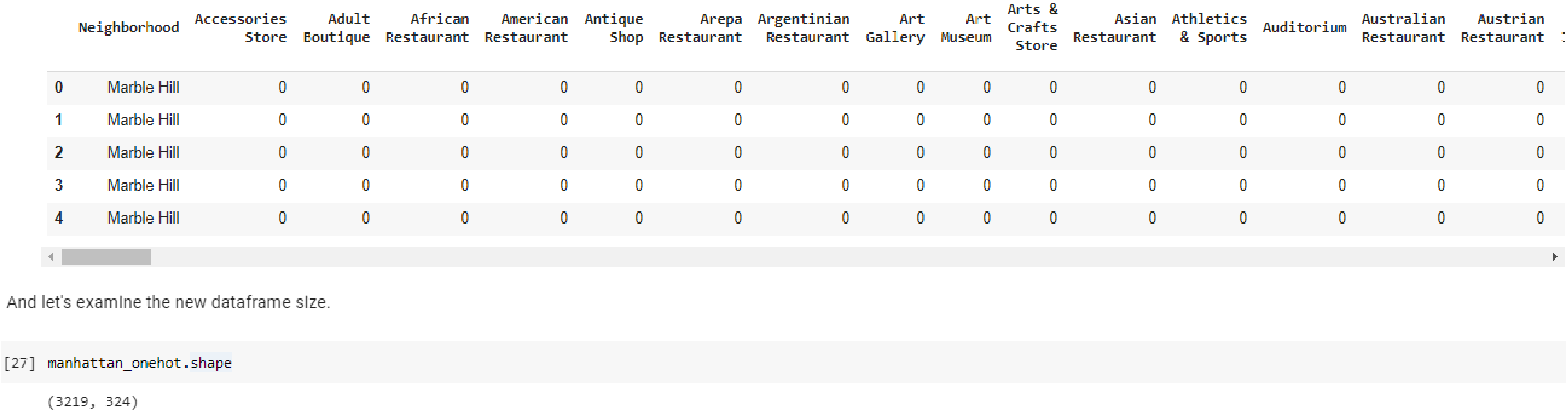
```
neighborhoods.head()
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585



```
newyork_data
```

```
{'bbox': [-74.2492599487305,
  40.5033187866211,
  -73.7061614990234,
  40.9105606079102],
 'crs': {'properties': {'name': 'urn:ogc:def:crs:EPSG::4326'}, 'type': 'name'},
 'features': [{'geometry': {'coordinates': [-73.84720052054902,
     40.89470517661],
    'type': 'Point'},
   'geometry_name': 'geom',
   'id': 'nyu_2451_34572.1',
   'properties': {'annoangle': 0.0,
    'annoline1': 'Wakefield',
    'annoline2': None,
    'annoline3': None,
    'bbox': [-73.84720052054902,
     40.89470517661,
     -73.84720052054902,
     40.89470517661],
    'borough': 'Bronx',
    'name': 'Wakefield',
    'stacked': 1},
   'type': 'Feature'},
  {'geometry': {'coordinates': [-73.82993910812398, 40.87429419303012],
```



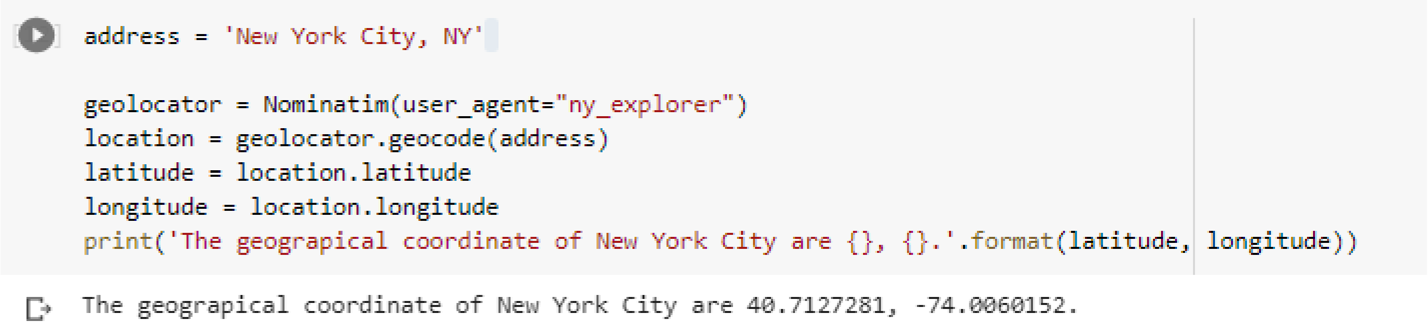
	Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auditorium	Australian Restaurant	Austrian Restaurant
0	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

And let's examine the new dataframe size.

```
[27] manhattan_onehot.shape

(3219, 324)
```

Borough, neighborhood data to be collected from NewYork Dataset file. (Web scrapping)

RESULTS (WK5)



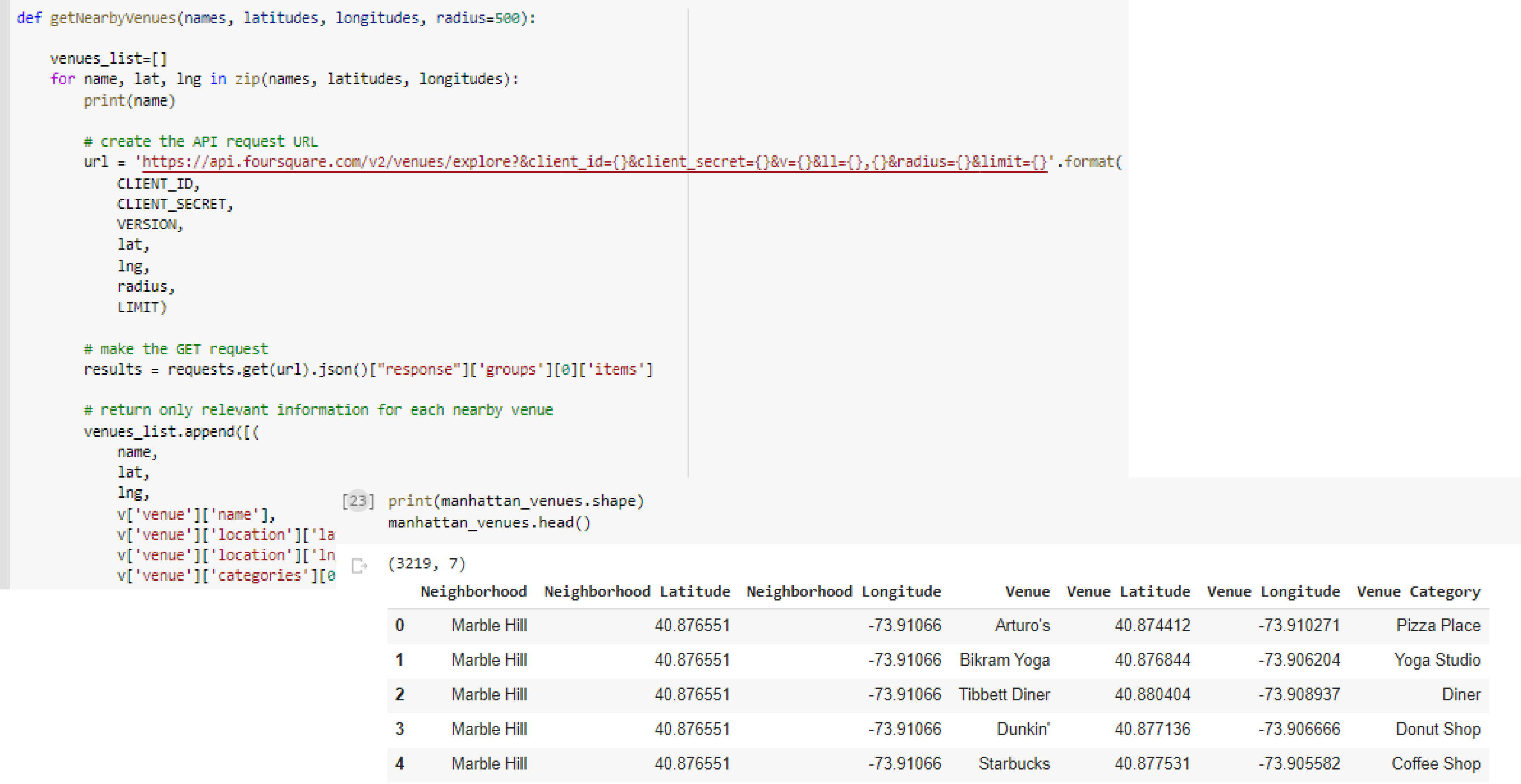
```
address = 'New York City, NY'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York City are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of New York City are 40.7127281, -74.0060152.
```

Geopy library is very useful collect geographical coordinates

Herkese Açık | Public

RESULTS (WK5)



```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['la
            v['venue']['location']['ln
            v['venue']['categories'][0
```

```
[23] print(manhattan_venues.shape)
     manhattan_venues.head()
```

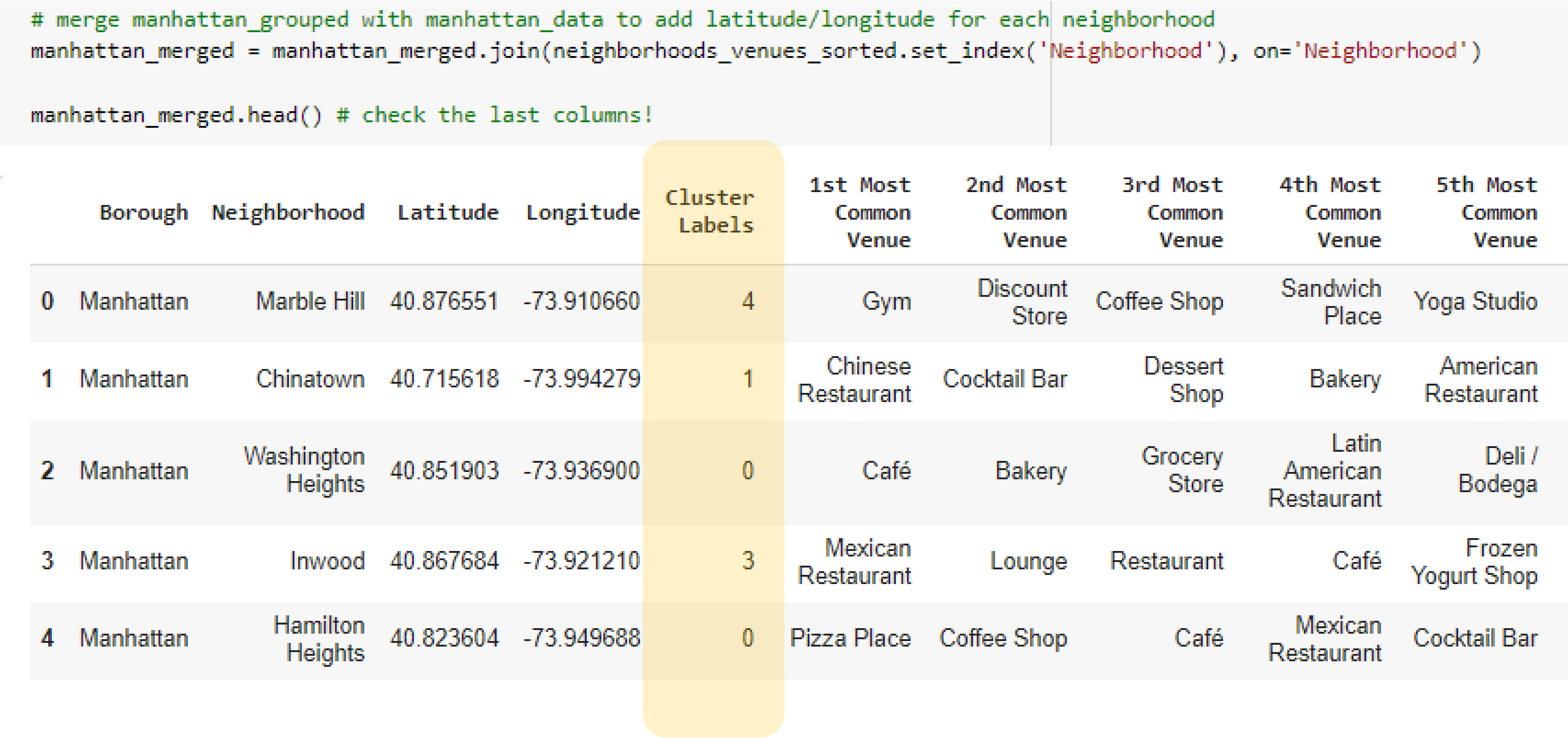
```
(3219, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

Foursquare API usage.

Herkese Açık | Public

RESULTS (WK5)

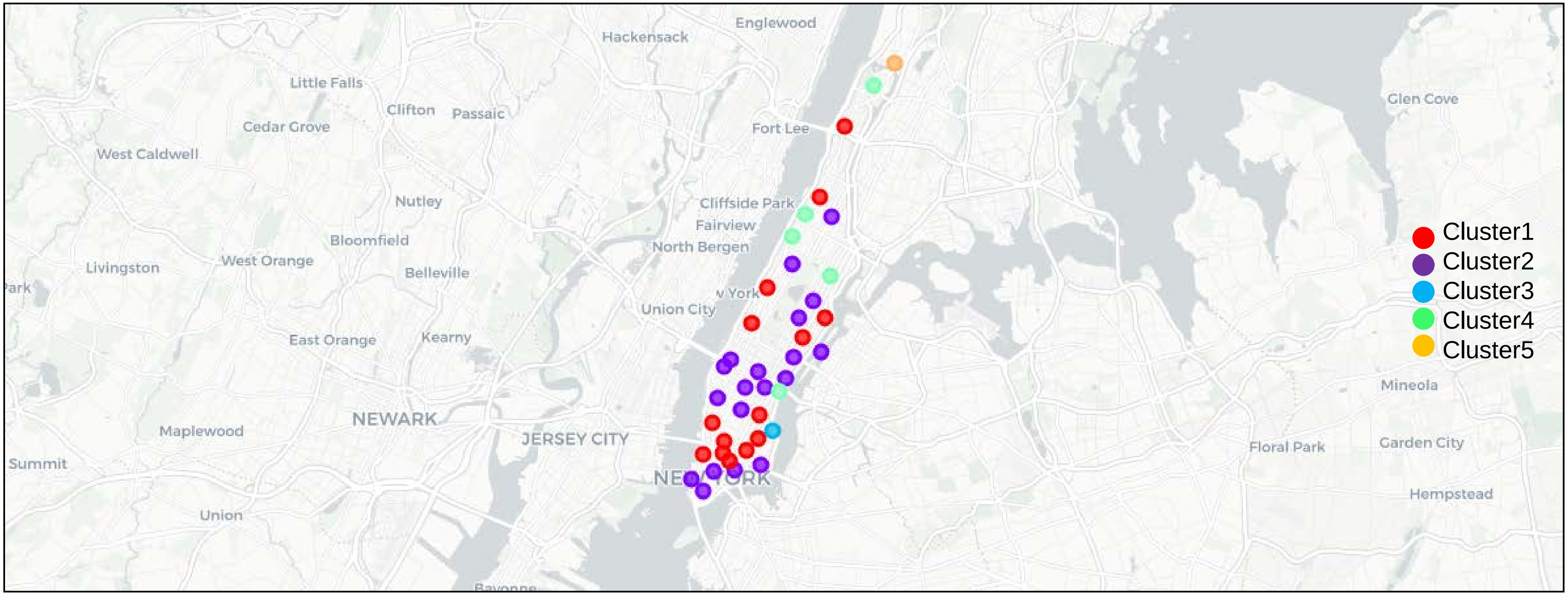


```
# merge manhattan_grouped with manhattan_data to add latitude/longitude for each neighborhood
manhattan_merged = manhattan_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

manhattan_merged.head() # check the last columns!
```

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	4	Gym	Discount Store	Coffee Shop	Sandwich Place	Yoga Studio
1	Manhattan	Chinatown	40.715618	-73.994279	1	Chinese Restaurant	Cocktail Bar	Dessert Shop	Bakery	American Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	0	Café	Bakery	Grocery Store	Latin American Restaurant	Deli / Bodega
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Lounge	Restaurant	Café	Frozen Yogurt Shop
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Pizza Place	Coffee Shop	Café	Mexican Restaurant	Cocktail Bar

Scikit-learn library to cluster venues data. (Machine learning). Clustering of neighborhoods which have similar characteristics.

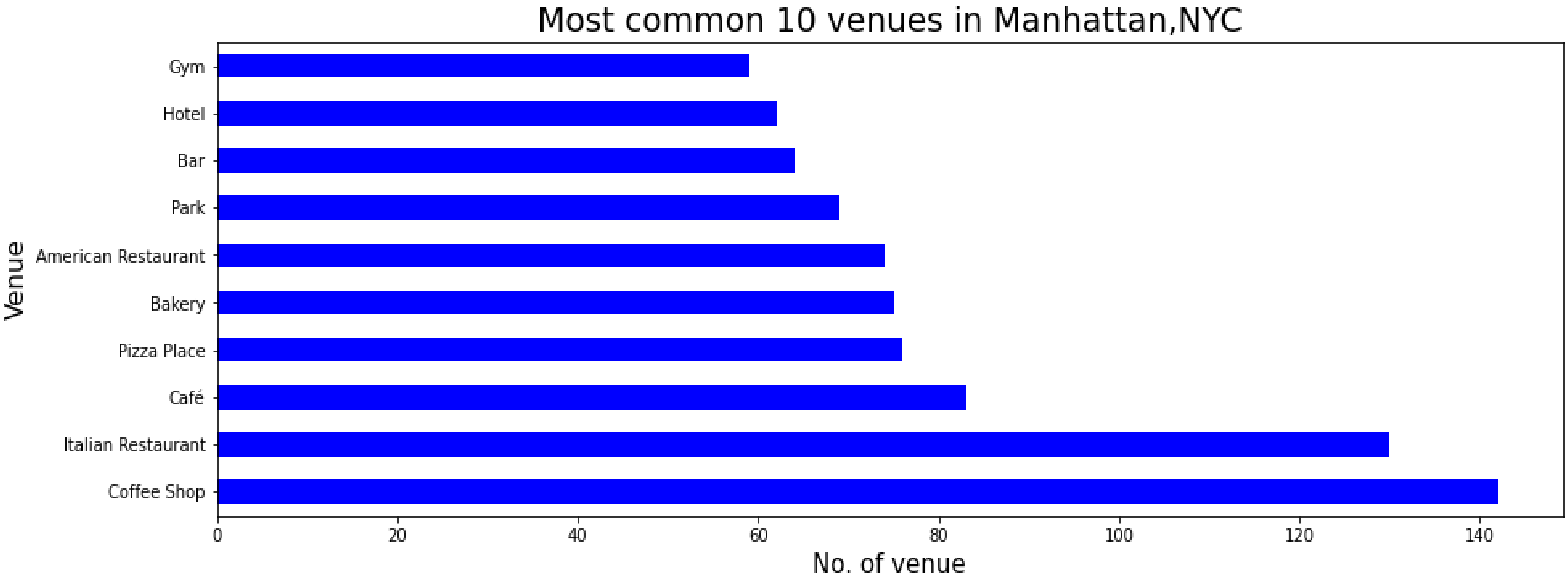
RESULTS (WK5)



Machine learning algorithm does process clustering of neighborhoods which have similar characteristics. Folium to visualize cluster data.
Based on the defining categories, I assign Cluster 1 as 'Restaurants Area'.

Herkese Açık | Public

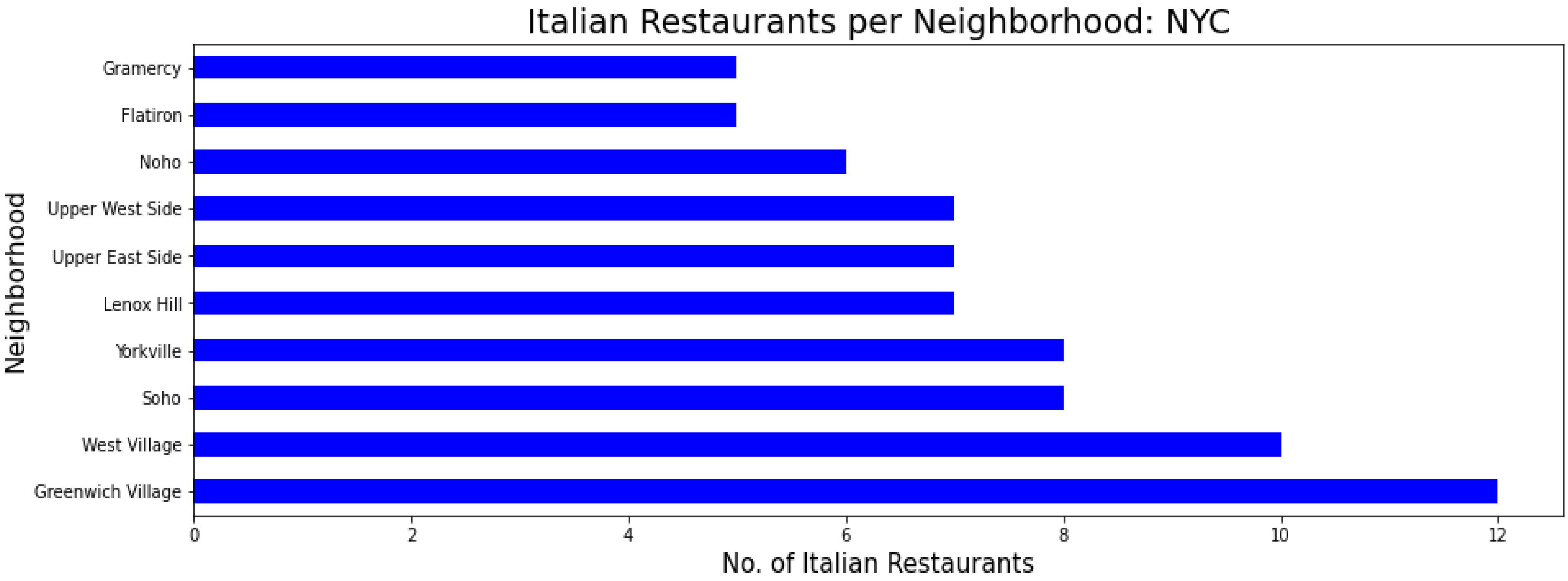
RESULTS (WK5)



This chart gives us the idea of Italian Restaurant investment is a good option. It's the 1st most common in restaurants category and 2nd most common among all venue categories in Manhattan.

Herkese Açık | Public

RESULTS (WK5)



This chart support our decision for Cluster1 area. Most of Italian restaurants are located in Cluster 1 areas (Greenwich Village, West Village, Soho, Yorkville)

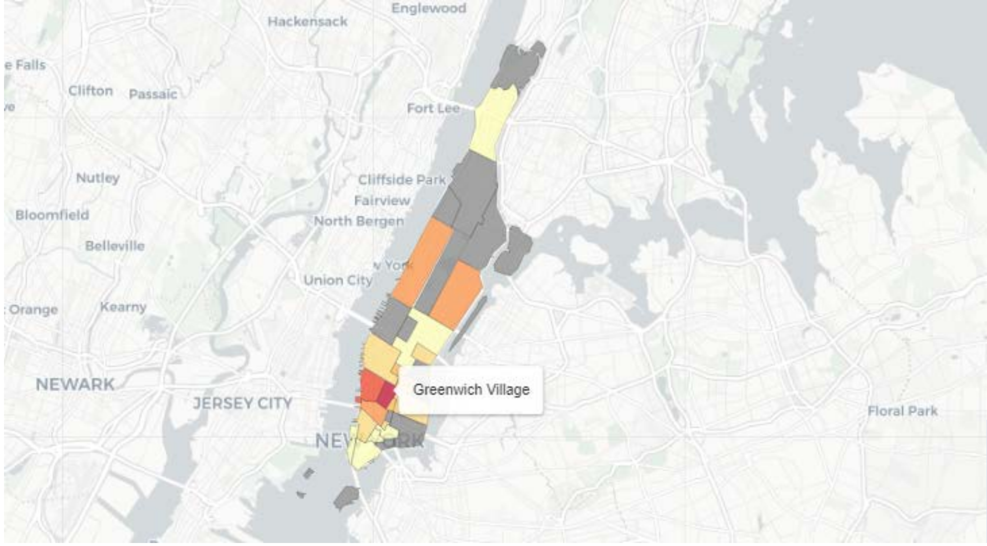
DISCUSSION (WK5)

Based on result of my analysis, I would strongly suggest following neighborhoods for new investment : **Greenwich Village, West Village, Soho, Yorkville**

Which one of the neighborhood to choose can be chosen according to other criterias such as rental prices, transportation opportunities etc. In this project rental prices or other investment costs is not considered.

South regions would be a better option as tourist population is higher in these areas. Customer rating and other foursquare features can be useful in selection.

Although I have modified the geojson file, it still needs further studies. Some of neighborhoods are not visible to user or does't match with JSON file. (For ex: Lenox Hill)



Potential restaurant locations are with darker fill color.

Herkese Açık | Public

CONCLUSION (WK5)

In this project I have studied an investment opportunity in Manhattan by using data and data science methods. To provide a better visibility I have reflected the data to map.

With much more data from other external databases or data sources , study can be expanded with other criterias specified in discussion section Whole process can be combined with decision making techniques.

At the end of the day, playing with data and learning different techniques is very exciting. Thanks to data providers and reviewers of this project.

Please feel free to give your feedback.