# Finding a location for a restaurant in New York

# Introduction

Restaurants are extremely popular with people of all ages. It's a lucrative business. But starting a business is not an easy task.

Success or failure depends on a lot of factors and the 2 of the important ones are:

- Picking the right location
- Knowing your customers

### In order to be profitable, your restaurant:

- Should be easily reachable and accessible
- Should be located in a part of the city that is frequently visited by people matching your customer profile
- Should have a menu and prices that attract customers

# **Business Problem**

How can a potential restaurant owner use location data and user ratings to choose a place for their restaurant that will be the most profitable?

# Target Audience

The target audience for this project could be:

- An entrepreneur or investor who wishes to setup a restaurant in New York
- Existing restaurant owners expand their business in other locations

### Data

The data for this project was collected from 3 main sources:

- Wikipedia List of New York neighbourhoods
- Geopy Geographic co-ordinates for the neighbourhoods
- Foursquare API Restaurants ratings, likes, attributes, etc.

# Methodology

#### Data collection

- Wikipedia data was scraped using BeautifulSoup
- Geographic co-ordinates for Manhattan neighbourhoods were fetched using geopy
- Restaurants from Manhattan were fetched using Foursquare API's venues explore endpoint
- Restaurant details were fetched using Foursquare API's venue details endpoint
- Closest venues around restaurants were fetched using Foursquare API's search endpoint
- Data was stored in pandas dataframes

### List of NY neighborhoods from Wikipedia

Area km² *	Pop. Census \$ 2010	Pop./ km² \$	Neighborhoods						
7.17	91,497	12,761	Melrose, Mott Haven, Port Morris						
5.54	52,246	9,792	Point, Longwood						
4.07	79,762	19,598	ont, Concourse Village, Crotona Park, Morrisania						
5.28	146,441	27,735	course, High Bridge						
3.55	128,200	36,145	dham, Morris Heights, Mount Hope, University Heights						
4.01	83,268	20,765	hgate, Belmont, East Tremont, West Farms						
4.84	139,286	28,778	ord Park, Norwood, University Heights						
8.83	101,731	11,521	dston, Kingsbridge, Kingsbridge Heights, Marble Hill, Riverdale, Spuyten Duyvil, Van Cortlandt Village						
12.41	172,298	13,884	onx River, Bruckner, Castle Hill, Clason Point, Harding Park, Parkchester, Soundview, Unionport						
16.76	120,392	7,183	City Island, Co-op City, Locust Point, Pelham Bay, Silver Beach, Throgs Neck, Westchester Square						
9.32	113,232[2]	12,149	Allerton, Bronxdale, Indian Village, Laconia, Morris Park, Pelham Gardens, Pelham Parkway, Van Nest[3][4]						
	7.17 5.54 4.07 5.28 3.55 4.01 4.84 8.83 12.41 16.76	Area km²	Area km² + Census + 2010   Nem² + 2010   Nem						

#### Borough Community Board Neighbourhood

0	Bronx	Bronx CB 1	Melrose
1	Bronx	Bronx CB 1	Mott Haven
2	Bronx	Bronx CB 1	Port Morris
3	Bronx	Bronx CB 2	Hunts Point
4	Bronx	Bronx CB 2	Longwood

# Geographic co-ordinates fetched using geopy

Battery Park City, Manhattan, New York 40.7110166 -74.0169369
Financial District, Manhattan, New York 40.7076124 -74.009378
TriBeCa, Manhattan, New York 40.7153802 -74.0093063
Chinatown, Manhattan, New York 40.7164913 -73.9962504
Greenwich Village, Manhattan, New York 40.7319802 -73.9965658
Little Italy, Manhattan, New York 40.7192728 -73.9982152
Lower East Side, Manhattan, New York 40.7159357 -73.9868057
NoHo, Manhattan, New York 40.7258746 -73.9939566

	Borough	Community Board	Neighbourhood	Latitude	Longitude
0	Manhattan	Manhattan CB 1	Battery Park City	40.711017	-74.016937
1	Manhattan	Manhattan CB 1	Financial District	40.707612	-74.009378
2	Manhattan	Manhattan CB 1	TriBeCa	40.715380	-74.009306
3	Manhattan	Manhattan CB 2	Chinatown	40.716491	-73.996250
4	Manhattan	Manhattan CB 2	Greenwich Village	40.731980	-73.996566
5	Manhattan	Manhattan CB 2	Little Italy	40.719273	-73.998215

# Foursquare response for restaurants in Manhattan neighbourhoods

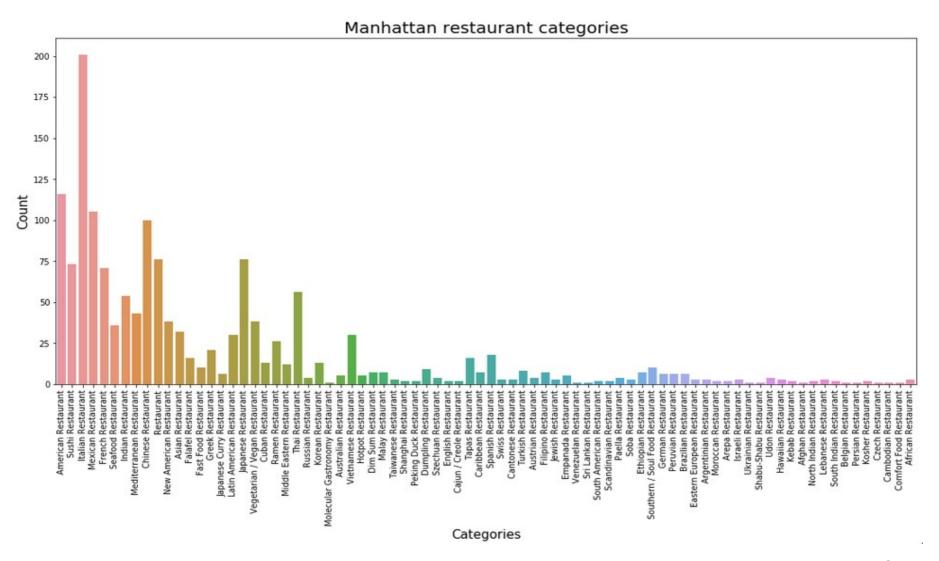
```
{'meta': {'code': 200, 'requestId': '5c6587e2351e3d4e704c2e54'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
  'filters': [{'name': 'Open now', 'key': 'openNow'},
   {'name': '$-$$$', 'key': 'price'}]},
  'headerLocation': 'Battery Park City',
  'headerFullLocation': 'Battery Park City, New York',
  'headerLocationGranularity': 'neighborhood',
  'query': 'restaurants',
  'totalResults': 71,
  'suggestedBounds': {'ne': {'lat': 40.7155166045, 'lng': -74.0110113729468},
  'sw': {'lat': 40.7065165955, 'lng': -74.02286242705321}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
        'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '5362a2ae498e3b18c22334be',
       'name': 'Hudson Eats',
       'location': {'address': '225 Liberty St',
        'crossStreet': 'South End Avenue',
        'lat': 40.712802562960505,
        'lnq': -74.01610221841605,
        'labeledLatLngs': [{'label': 'display',
         'lat': 40.712802562960505
          'lng': -74.01610221841605}],
        'distance': 210,
        'postalCode': '10281',
        'cc': 'US',
        'city': 'New York',
        'state': 'NY',
        'country': 'United States',
        'formattedAddress': ['225 Liberty St (South End Avenue)',
         'New York, NY 10281',
         'United States']},
       'categories': [{'id': '4bf58dd8d48988d120951735',
         'name': 'Food Court',
         'pluralName': 'Food Courts',
         'shortName': 'Food Court',
         'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shops/food_foodcourt_',
          'suffix': '.png'},
         'primary': True}].
```

	Address	Categories	DistanceToNeighbourhood	Latitude	Longitude	Neighbourhood	Venue
0	225 Liberty St (at Hudson Eats),New York, NY 1	American Restaurant	216	40.712835	-74.016013	Battery Park City	Dig Inn
1	200 Vesey St,New York, NY 10281,United States	Sushi Restaurant	205	40.712742	-74.016065	Battery Park City	Blue Ribbon Sushi Bar
2	250 Vesey St (West St),New York, NY 10080,Unit	Italian Restaurant	393	40.714514	-74.016264	Battery Park City	Parm
3	101 Liberty St (at Greenwich St),New York, NY	Italian Restaurant	376	40.710554	-74.012519	Battery Park City	Osteria della Pace
4	259 Vesey St (North End Avenue),New York, NY 1	Mexican Restaurant	458	40.714940	-74.015300	Battery Park City	El Vez

### Dataframe with restaurant details

Likes	On Lists	Price Message	Price Tier	Rating	Rating Signals	Tips	Verified	Beer	Breakfast	Brunch	Cocktails	Credit Cards	Delivery	Dessert	Dinner	Full Bar
239	172	Moderate	2	8.5	335.0	67	1	1	0	1	1	1	1	1	1	1
41	18	Moderate	2	7.9	51.0	3	0	0	0	0	0	0	0	0	1	0
305	251	Expensive	3	8.3	503.0	140	1	1	0	1	1	1	1	0	1	1
33	13	Very Expensive	4	8.2	52.0	17	0	0	0	0	1	1	0	0	1	1
140	131	Very Expensive	4	8.2	217.0	50	1	0	1	1	1	1	1	0	1	1

# Majority of the restaurants in Manhattan were Italian



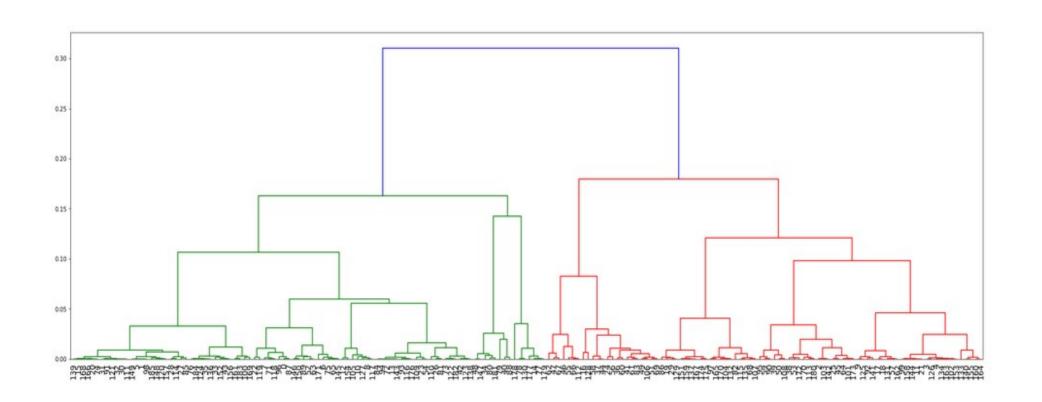
### Feature selection and scaling

- Likes, Ratings, Rating Signals, Tips, Lists and Price Tier were chosen for modeling
- Feature values were scaled between the range 0-1 using min max scaling

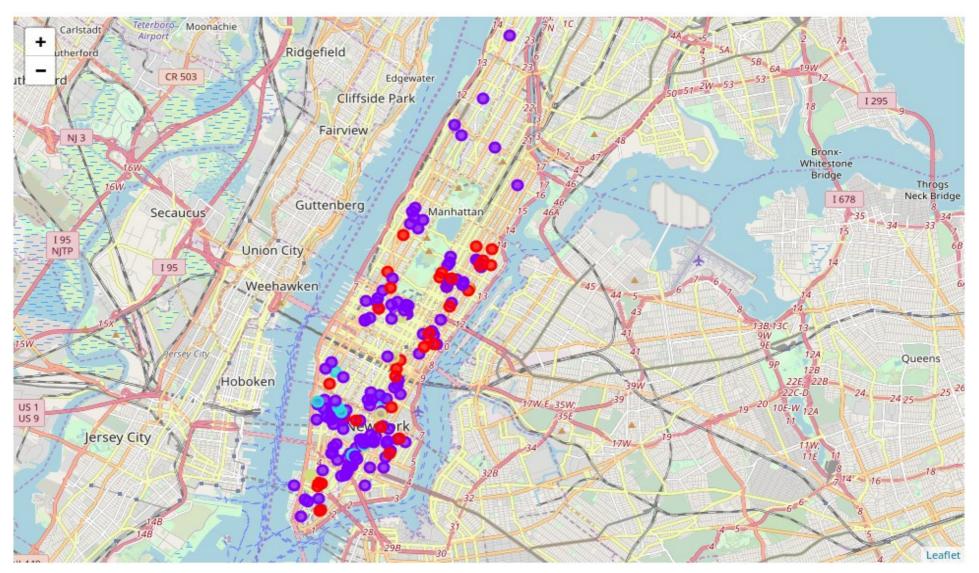
### Modeling

- The agglomerative clustering algorithm was used as the number of clusters in the dataset could be determined without being explicitly set
- A distance matrix was constructed and used to create a dendrogram to determine the number of clusters
- Agglomerative clustering model was created to segment the restaurants into 4 clusters using the chosen features

### Dendrogram visualizing clusters



#### 4 clusters visualized



# Results

- The 1st cluster had 38 restaurants. All of them served dinner and had no music or take-out. The median rating of this cluster was 7.7 with a median rating signal of 47.5.
- The 2nd cluster had 131 restaurants. The median rating of this cluster was 8.3 with a median rating signal of 179.
- The 3rd cluster had 9 restaurants. All of them served dinner, cocktails and wine. They also accepted credit cards. The median rating of this cluster was 9.2 with a median rating signal of 1420.
- The 4th cluster has 8 restaurants. All of them served lunch and none of them served cocktails nor had music. All were in the same price tier 1. The median rating of this cluster was 8.4 with a median rating signal of 189.5.

# Conclusion

- The 3rd cluster had the highest median rating, rating signals, tips and likes
- 8 / 9 restaurants in this cluster had other Italian restaurants as their nearest neighbours, which did not seem to affect their ratings
- There were 4 restaurants located in West Village, 2 in SoHo, and 1 each in Chelsea, Gramercy Park and NoHo
- Since West Village already had more than one Italian restaurant,
   SoHo, Chelsea, Gramercy Park and NoHo neighbourhoods were short-listed to set-up a new restaurant
- Since the restaurants in the short-listed neighborhoods are highly rated and reviewed, setting up another restaurant in similar surroundings could be profitable

# Future research

This analysis was based on user ratings and reviews. It could be improved with data like:

- Population density in each of the neighbourhoods
- Type of customers and their frequency of visits
- Menu details