



# Capstone Project – Chinese Restaurant in NYC

# Introduction

- ◆ In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Chinese restaurant in Manhattan, New York**.
- ◆ Since there are lots of restaurants in NYC we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Chinese restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.
- ◆ We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.



# Analytical approach

- ◆ Our problem is clearly a clustering problem. We will therefore rely on a clustering model to solve it. Clustering models are numerous, with the two most popular being K-means clustering and hierarchical clustering.
- ◆ Fortunately, most clustering algorithms are already implemented in open source libraries for the language we will use (Python), therefore we won't have to do much coding. The most critical and the most tedious part of this project, as with most data science projects, will be to collect and clean the data.

# Data

- ◆ Based on definition of our problem, factors that will influence our decision are:
  - number of existing restaurants in the neighborhood (any type of restaurant)
  - number of and distance to Italian restaurants in the neighborhood, if any
  - distance of neighborhood from city center
- ◆ We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.
- ◆ Following data sources will be needed to extract/generate the required information:
  - centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
  - number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
  - coordinate of NYC center will be obtained using **Google Maps API geocoding** of well known NYC location

# Methodology:

## K-Means Clustering Algorithm



# Result:

- ❖ Turtle Bay is the most promising neighborhood shareholder should consider for a new Chinese restaurant.
- ❖ Cluster 1 is the most visited venues that covered with light blue.

```
no_ch_neighborhood = manhattan_grouped_sorted[manhattan_grouped_sorted['Chinese Restaurant']==0].reset_index(drop=True)
```

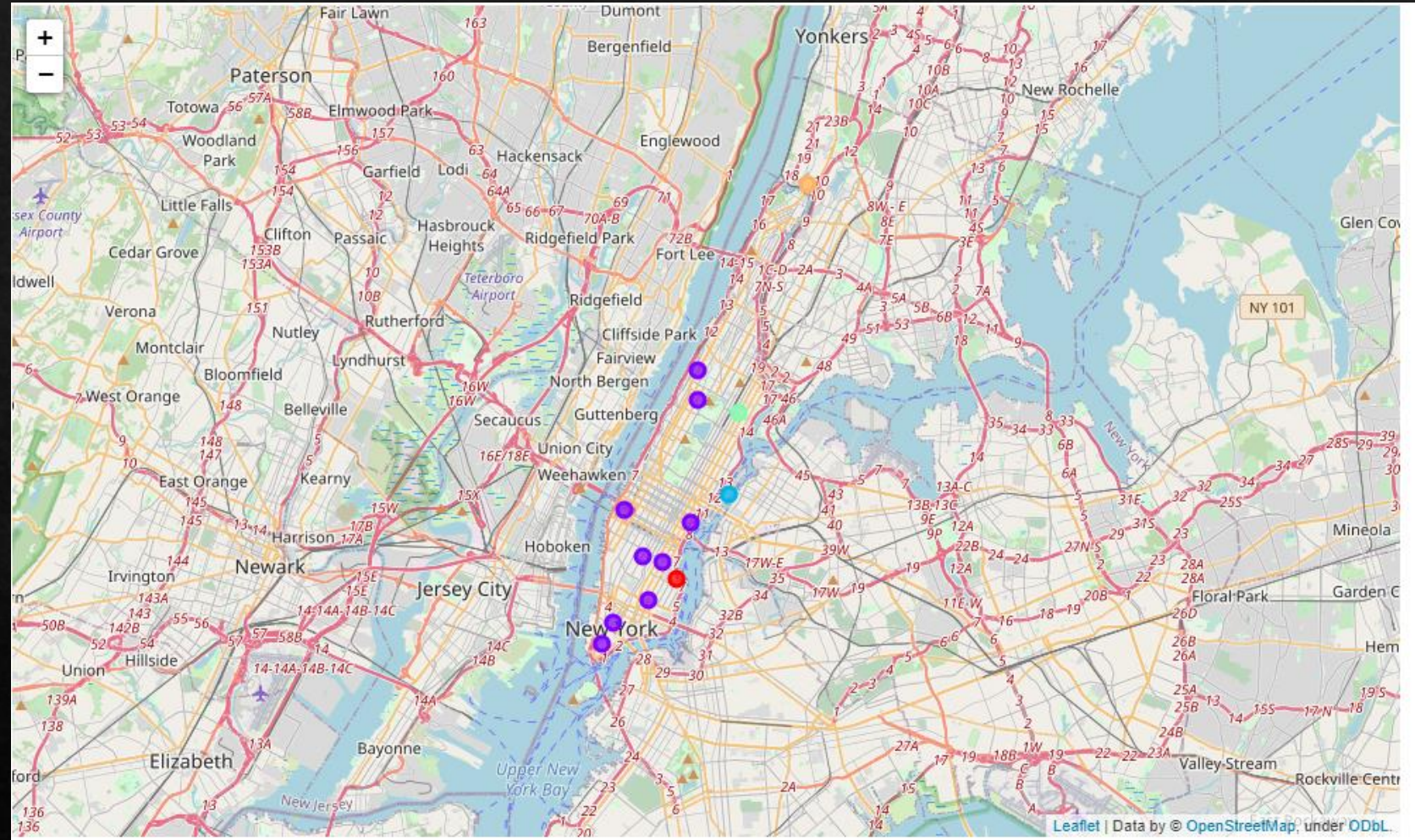
```
food_neighborhoods = no_ch_neighborhood.drop(columns='Total Visited Frequency')  
food_neighborhoods.head(10)
```

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	...	Video Store	Vietnamese Restaurant	Volleyball Court	W
0	Turtle Bay	0.0	0.0	0.0	0.0	0.020000	0.00	0.0	0.00	0.000000	...	0.0	0.000000	0.0	
1	East Harlem	0.0	0.0	0.0	0.0	0.000000	0.00	0.0	0.00	0.000000	...	0.0	0.000000	0.0	
2	Manhattan Valley	0.0	0.0	0.0	0.0	0.000000	0.00	0.0	0.00	0.000000	...	0.0	0.021739	0.0	
3	Noho	0.0	0.0	0.0	0.0	0.020000	0.00	0.0	0.01	0.040000	...	0.0	0.000000	0.0	
4	Gramercy	0.0	0.0	0.0	0.0	0.044444	0.00	0.0	0.00	0.011111	...	0.0	0.000000	0.0	
5	Civic Center	0.0	0.0	0.0	0.0	0.020000	0.01	0.0	0.00	0.010000	...	0.0	0.000000	0.0	
6	Flatiron	0.0	0.0	0.0	0.0	0.030000	0.00	0.0	0.00	0.010000	...	0.0	0.000000	0.0	
7	Financial District	0.0	0.0	0.0	0.0	0.040000	0.00	0.0	0.00	0.000000	...	0.0	0.000000	0.0	
8	Morningside Heights	0.0	0.0	0.0	0.0	0.073171	0.00	0.0	0.00	0.000000	...	0.0	0.000000	0.0	
9	Hudson Yards	0.0	0.0	0.0	0.0	0.053571	0.00	0.0	0.00	0.000000	...	0.0	0.000000	0.0	



# Result:

- Cluster 1 is the most visited venues that covered with light blue.





❖ Conclusion:

❖ Manhattan has totally 40 neighborhoods

❖ Top 10 neighborhoods with most food venues are:

East Village *freq*=0.59,

Upper West Side *freq*=0.56,

Manhattanville *freq*=0.53,

Turtle Bay *freq*=0.52,

Greenwich Village *freq*=0.52,

Central Harlem *freq*=0.51,

Chinatown *freq*=0.48,

Hamilton Heights *freq*=0.46,

Inwood *freq*=0.46,

West Village *freq*=0.45

❖ Turtle Bay is the best location with most food venues and no Chinese restaurant

```
result=manhattan_grouped_sorted.loc[:,['Neighborhood','Total Visited Frequency']]
result.head(10)
```

	Neighborhood	Total Visited Frequency
0	East Village	0.590000
1	Upper West Side	0.555556
2	Manhattanville	0.533333
3	Turtle Bay	0.520000
4	Greenwich Village	0.520000
5	Central Harlem	0.511111
6	Chinatown	0.480000
7	Hamilton Heights	0.457627
8	Inwood	0.456140
9	West Village	0.450000