RECOMMENDATIONS ON LOCATION AND FOOD TYPE OF A RESTAURANT TARGETING TOURISTS IN NEW YORK

IBM APPLIED DATA SCIENCE CAPSTONE

Boyan Gu

INTRODUCTION

- New York is the most populous city in US, and is an international center of business, finance, arts, and culture. It is recognized as one of the most multicultural and cosmopolitan cities in the world.
- It will be of great chance to open a restaurant successfully and attractive customers, especially tourists.
- The target audiences of this project are investors who want to open a new restaurant targeting tourists.
- The problems:
  1. Where to open a restaurant can attract tourists and is least competitive
  2. Which food types are popular for tourists

DATA

- 1. Scrap the json data including New York neighborhoods from the website:

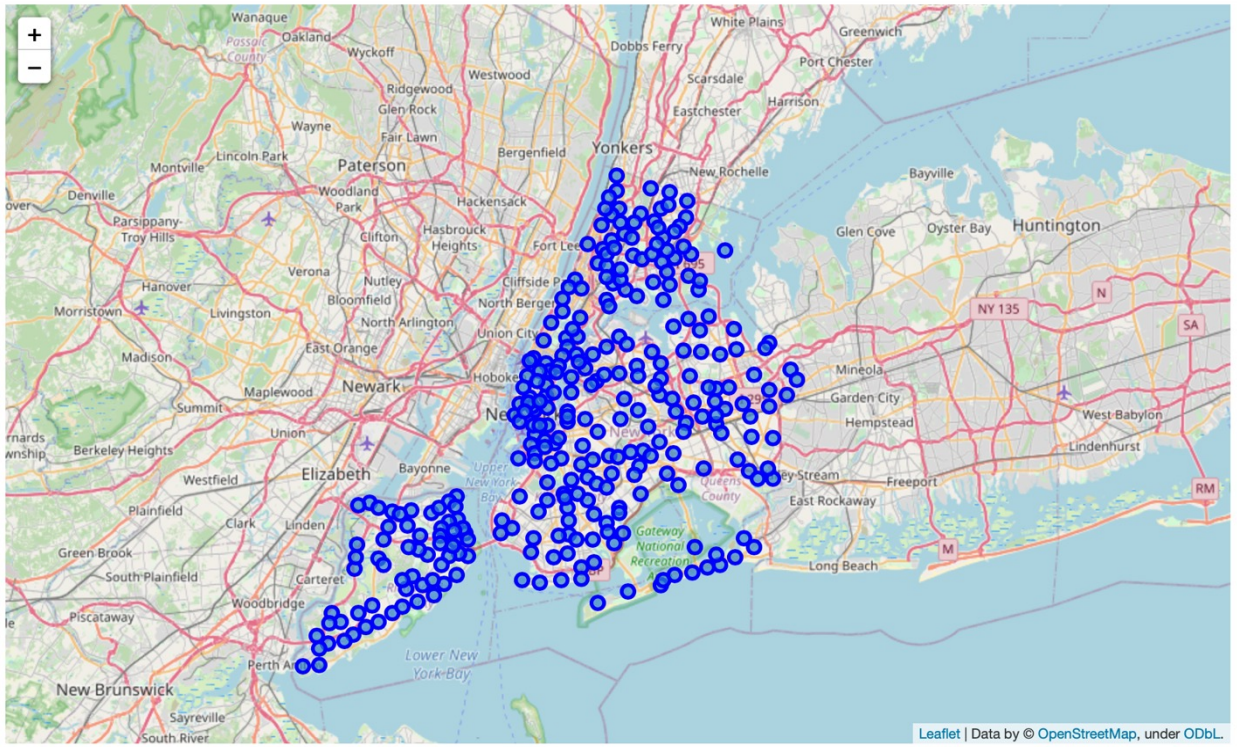
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

- 2. Retrieve venue data from Foursquare API (e.g. Explore call)

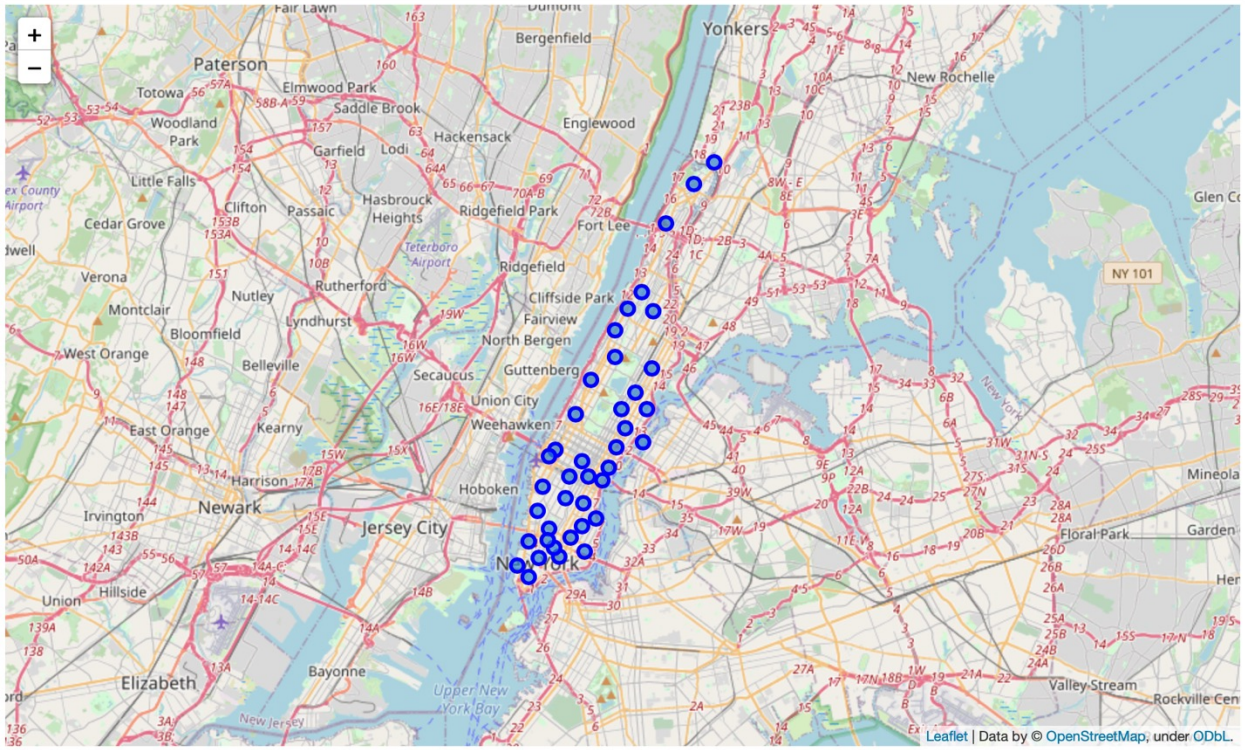
```
# create the API request URL
url = 'https://api.foursquare.com/v2/venues/categories?&client_id={}&client_secret={}&v={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION)

# make the GET request
results = requests.get(url).json()
```

METHODOLOGY

map_newyork

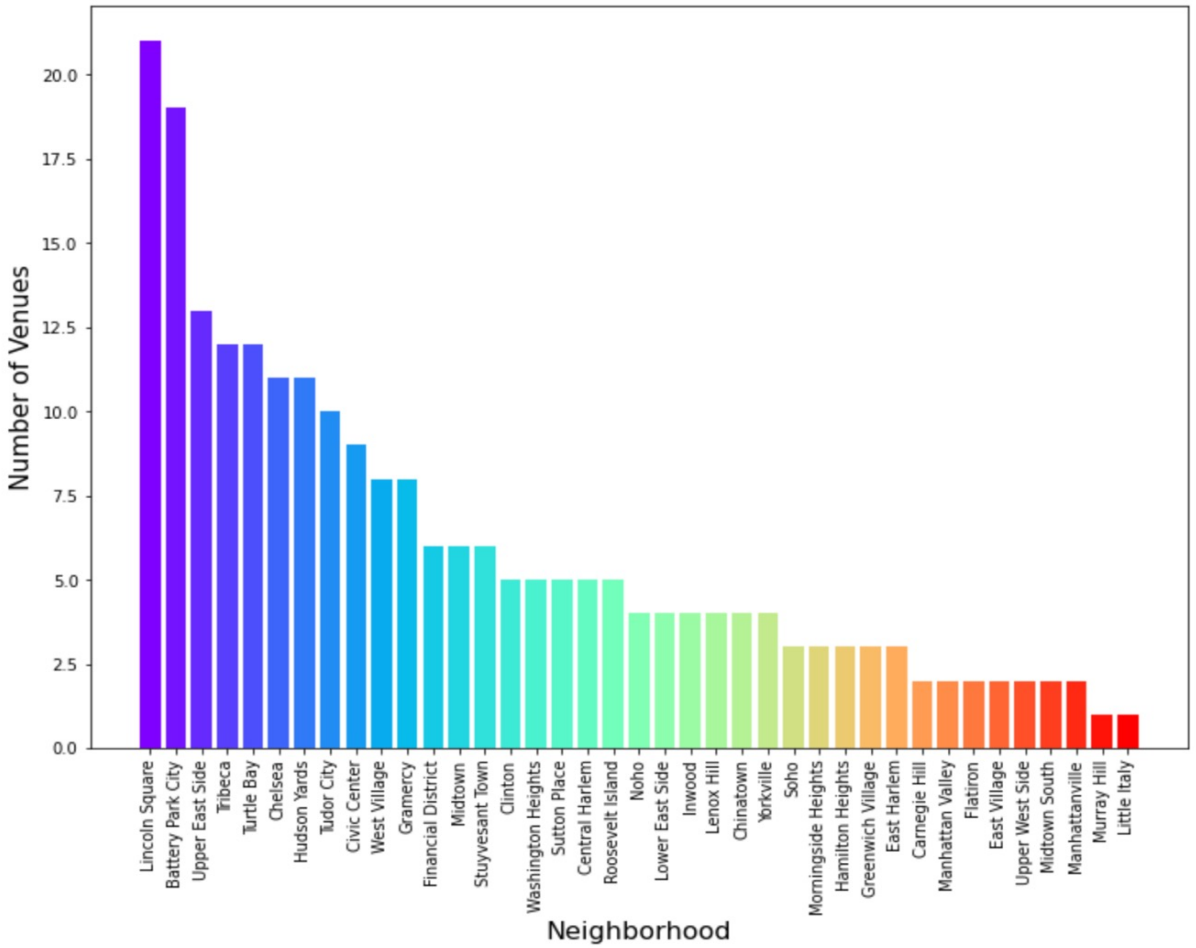


map_manhattan



Map of New York and Manhattan with neighborhoods superposed on top (focus on Manhattan region for rest of presentation)

METHODOLOGY



Foursquare main venue categories

1. Tourism Venues: Arts & Entertainment, Nightlife Spot, and Outdoors & recreations
2. Tourist-related services/business: Food, Shop & Services, and Travel & Transport are relevant tourist services
3. Others: not considered in this project

Fig. 1. The number of nearby tourism venues in each neighborhood

METHODOLOGY

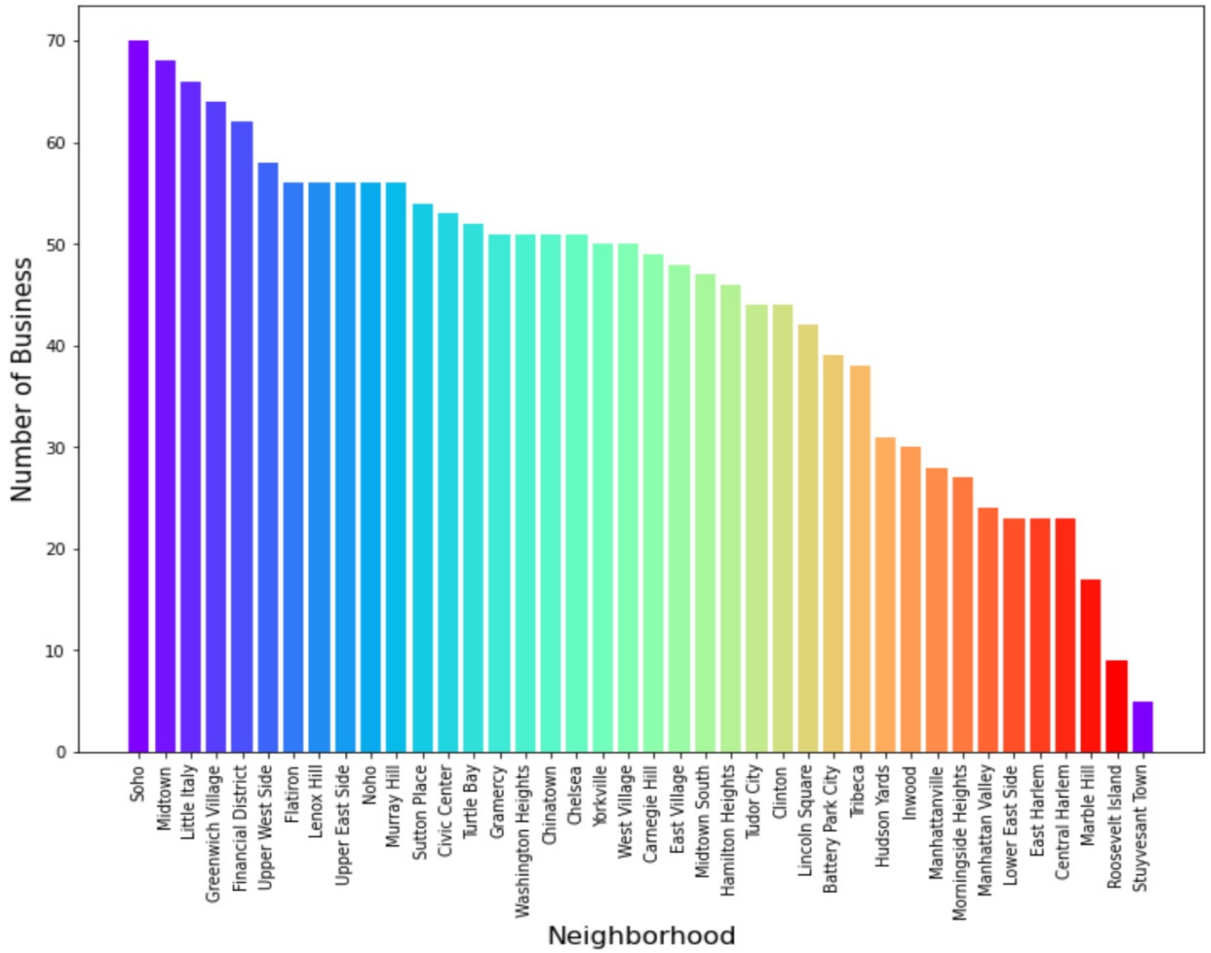


Fig. 2. The number of nearby tourist-related services/businesses in each neighborhood

METHODOLOGY

- One-hot encoding analysis to analyze each neighborhood

```
neighborhoods_venues_sorted.head()
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Battery Park City Esplanade, Hudson River Trai...	9/11 Memorial North Pool, National September 1...	Waterfront Plaza, Brookfield Place, Oculus Plaza	Kowsky Plaza Playground, West Thames Playground	One World Observatory	Winter Garden Atrium	Asphalt Green Battery Park City	Liberty Community Garden		
1	Carnegie Hill	Samuel Seabury Playground	The Jewish Museum								
2	Central Harlem	La masion d'Art, Tatiana Pagès Gallery	Big L Memorial Mural	St. Nicholas Park	Shrine World Music Venue						
3	Chelsea	David Zwirner Gallery, Milk Gallery, Gagosian ...	High Line, Clement Clarke Moore Park	PH-D at Dream Downtown, 1 OAK	London Terrace Gardens Courtyard	High Line 10th Ave Amphitheatre					
4	Chinatown	Sofar HQ	Chinatown Soup	The Crown	Museum at Eldridge Street						

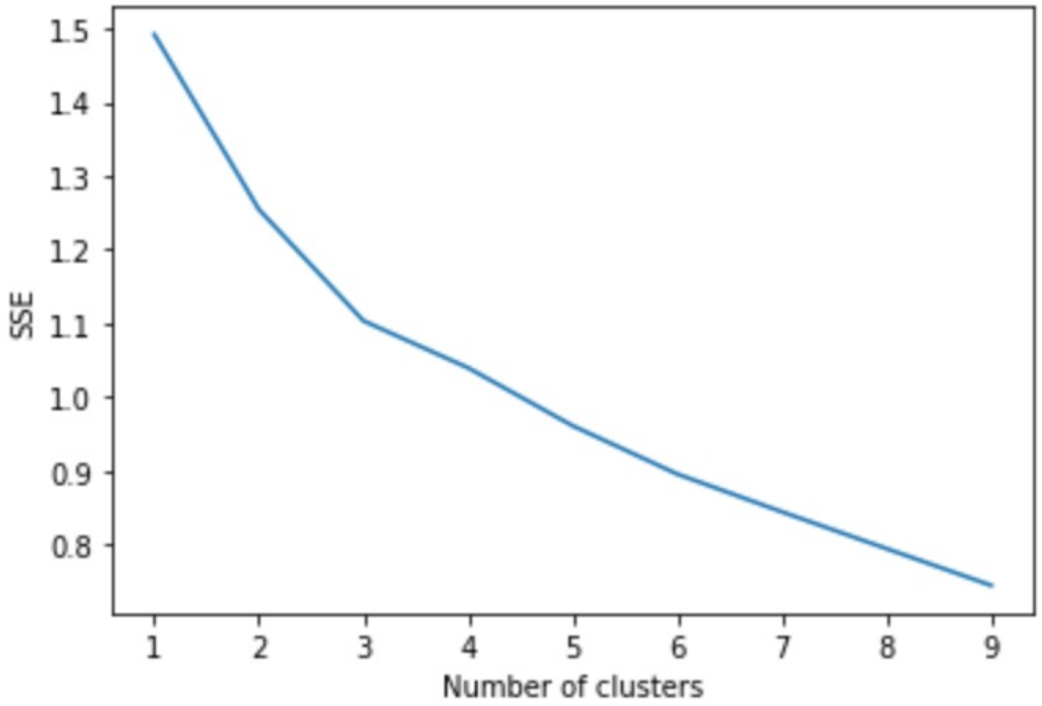
METHODOLOGY

- One-hot encoding analysis to analyze each neighborhood

	Neighborhood	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
0	Battery Park City	Coffee Shop	Clothing Store	Hotel	Food Court	Sandwich Place	Pizza Place	Italian Restaurant	Burger Joint	BBQ Joint	Shopping Mall
1	Carnegie Hill	Coffee Shop	Café	French Restaurant	Bookstore	Cosmetics Shop	Italian Restaurant	Shipping Store	Bakery	Pizza Place	Bank
2	Central Harlem	African Restaurant	French Restaurant	Cosmetics Shop	American Restaurant	Seafood Restaurant	Caribbean Restaurant	Southern / Soul Food Restaurant	Spa	Pizza Place	Café
3	Chelsea	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	French Restaurant	Hotel	Seafood Restaurant	Pet Store	Café	Market
4	Chinatown	Bakery	Salon / Barbershop	Spa	American Restaurant	Dessert Shop	Mexican Restaurant	Optical Shop	Bubble Tea Shop	Asian Restaurant	Dumpling Restaurant
5	Civic Center	Coffee Shop	Spa	French Restaurant	American Restaurant	Hotel	Italian Restaurant	Bakery	Falafel Restaurant	Wings Joint	Bubble Tea Shop

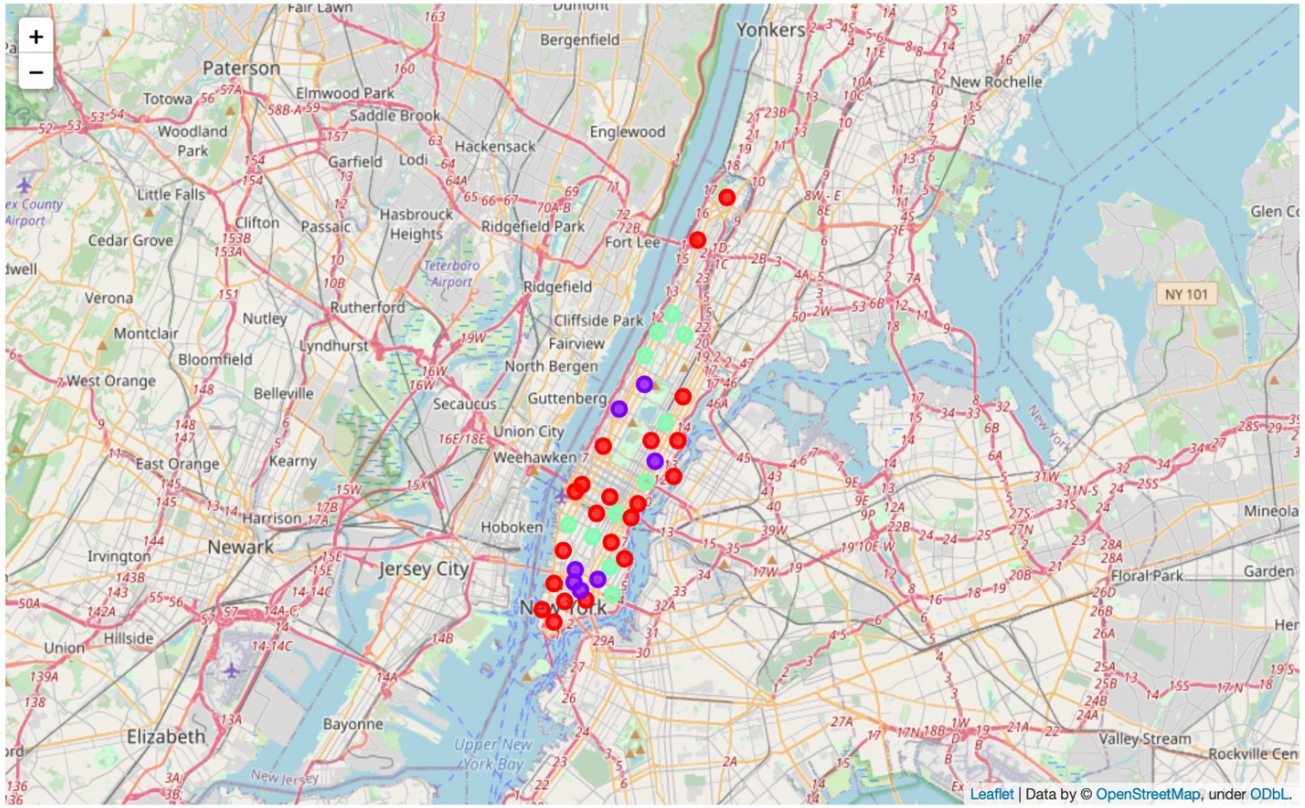
RESULTS

- K-means Clustering is applied to cluster tourism attractions and tourist-related businesses separately in Manhattan



Use elbow method for optimal of k in K-means Clustering:
Choose k = 3 for clustering

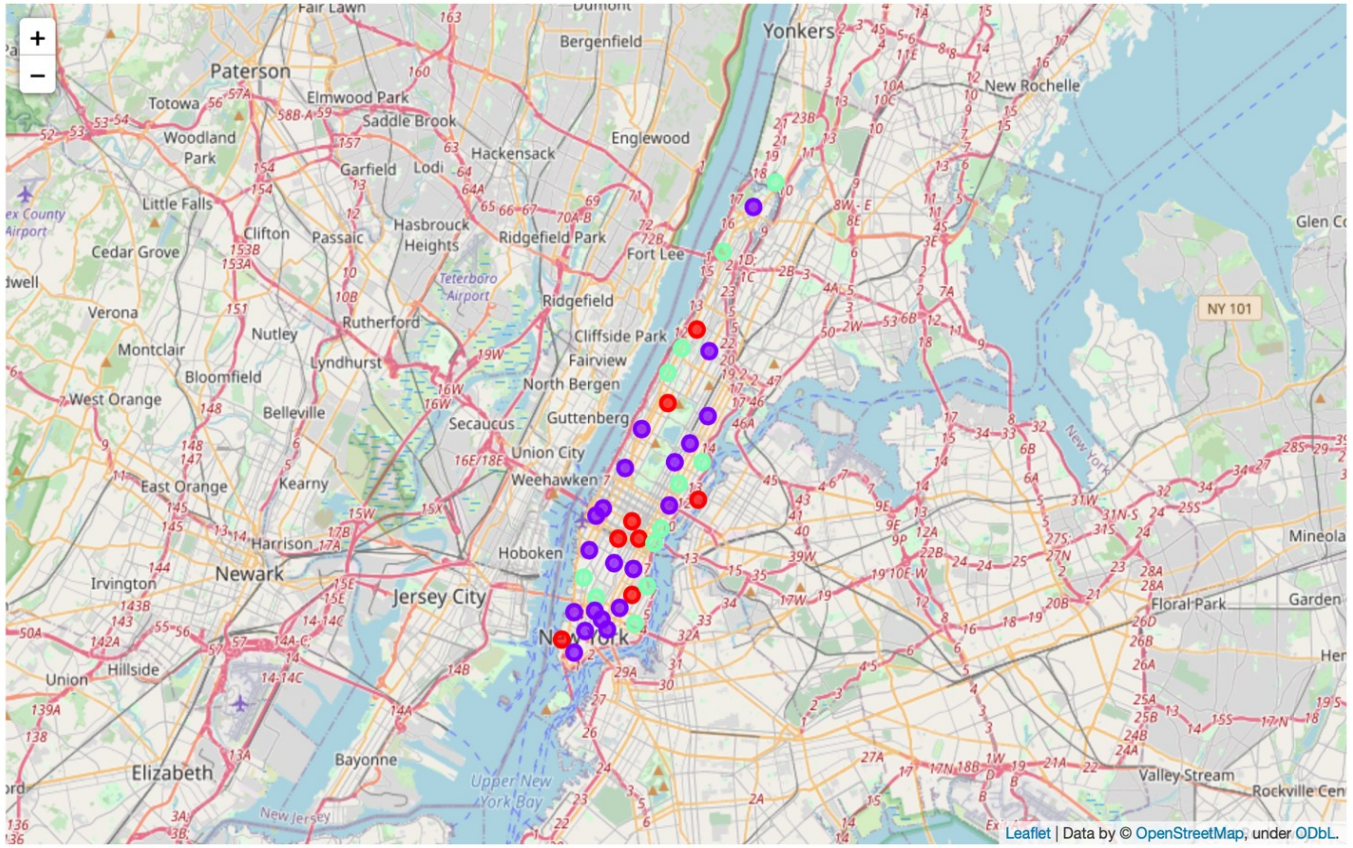
RESULTS



Clustering for tourism venues in Manhattan:
Cluster 1 (red): neighborhoods containing more categories and more tourism venues
Cluster 2 (purple): neighborhoods with least tourism attractions
Cluster 3 (green): have less tourism venues,

Intuitionally, neighborhoods in cluster 1 are popular among tourists.

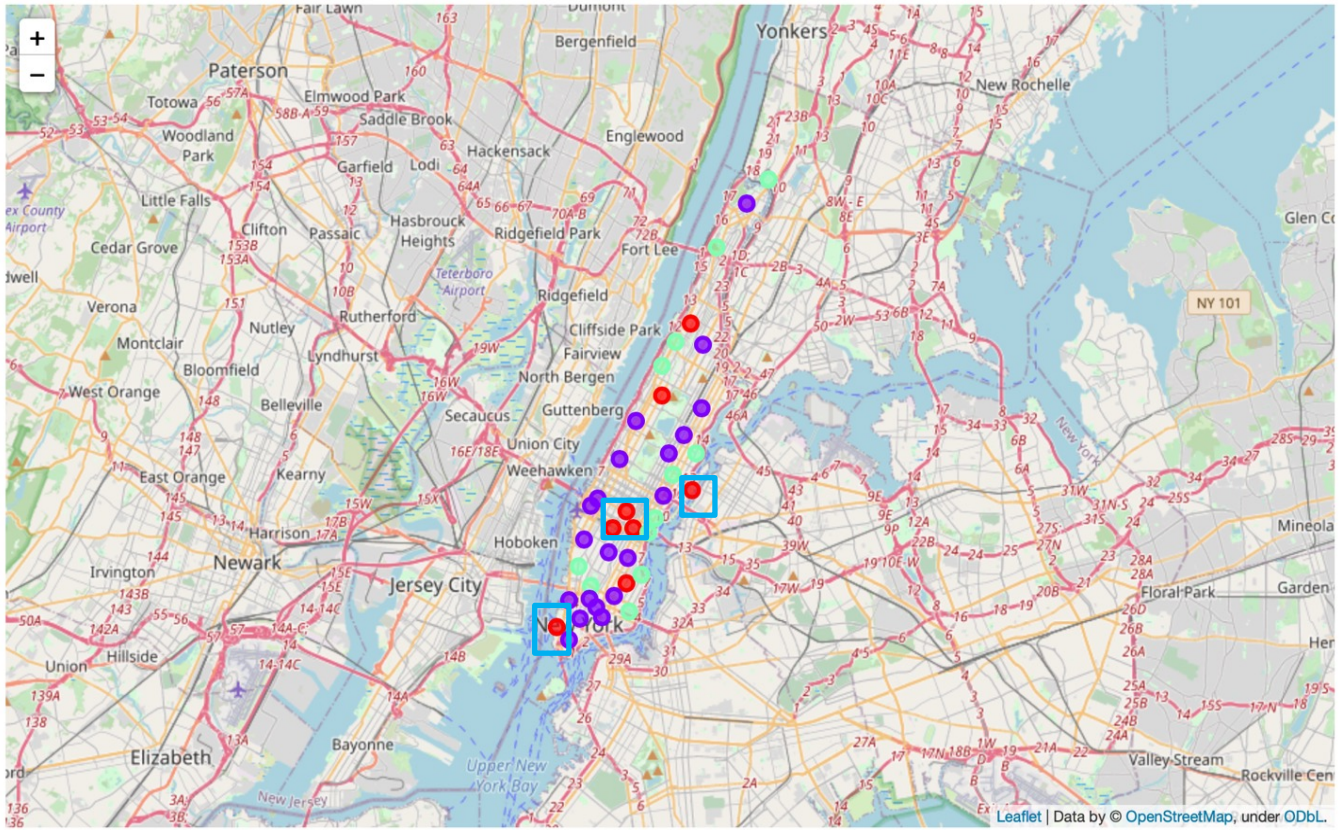
RESULTS



The most popular food types
cluster 1 (red): Mexican food, American food, and Italian food
Cluster 2 (purple): Italian food, American food, French food
Cluster 3 (green): Italian food, fast food, American food, and seafood most.

In general, Italian food and American food are popular in Manhattan region. Neighborhoods in cluster 2 have relatively fierce competition, and cluster 1 has relatively small pressure of competition.

RESULTS



Compared two clustering figures:
Most neighborhoods in (southern part of) cluster 1 of tourist-related business locate at cluster 1 of tourism venues, where there are supposed to be more tourists and less competitors of Italian or American food.

We recommend investors to open an Italian or American restaurant at those blue-squared neighborhoods

CONCLUSION AND FUTURE WORK

- In this project, we use data retrieved from website and Foursquare API to determine applicable location and food type of a restaurant targeting tourists in New York. After data analysis and K-Means Clustering, we find good locations (blue squared) with low competence level near tourism venues.

- In the future, we may further use Foursquare API to get foot-traffic data of restaurants, and determine the popular locations and food types with more statistical data.