

# Finding the Optimal Neighborhood in Manhattan for Café Business

---

# Business Problem

---

- There are lots of coffee shops locating in different neighborhoods of Manhattan.
- Their business performances highly depend on various factors including location, residence and other business categories.
- We are interested in café business in Manhattan.

**Goal:**

- **Find the optimal neighborhood for opening a café in Manhattan**

# Data

---

- Manhattan Neighborhood Data
  - source: IBM Skills Network
- Manhattan Population Data
  - source: Wikipedia ([Neighborhoods in New York City](#))
- Foursquare Location Data
  - Foursquare API ([Foursquare venue category documentation](#))

# Factors to measure a neighborhood

---

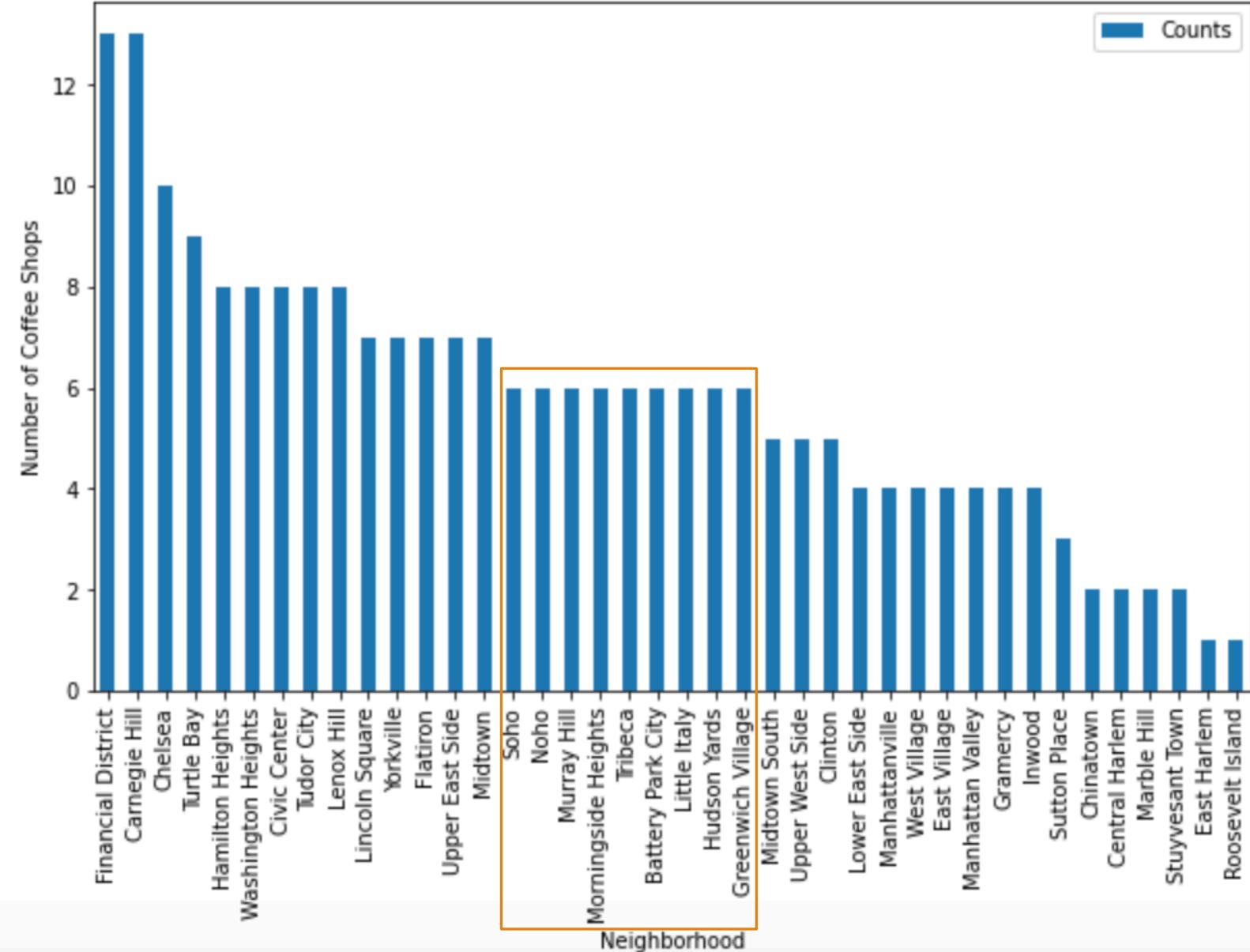
- The number of coffee shops already existing in the neighborhood
- The number of different venue categories in the neighborhood
- Population density of the neighborhood

# Coffee Shop Distributions in Manhattan Neighborhoods

Take away:

1. Financial District and Carnegie Hill are the two neighborhoods with the most coffee shops.
2. East Harlem and Roosevelt Island are the two with the least number of coffee shops.
3. As we want the neighborhoods with not too many and too few coffee shops, we pick the middle 9 neighborhoods as candidates.

Coffee Shop Distribution in Manhattan Neighborhoods



<b>Neighborhood</b>	<b>Neighborhood Latitude</b>	<b>Neighborhood Longitude</b>	<b>Venue</b>	<b>Venue Latitude</b>	<b>Venue Longitude</b>	<b>Category</b>
<b>Soho</b>	100		100	100	100	100
<b>Noho</b>	100		100	100	100	100
<b>Murray Hill</b>	100		100	100	100	100
<b>Morningside Heights</b>	43		43	43	43	43
<b>Tribeca</b>	86		86	86	86	86
<b>Battery Park City</b>	77		77	77	77	77
<b>Little Italy</b>	100		100	100	100	100
<b>Hudson Yards</b>	70		70	70	70	70
<b>Greenwich Village</b>	100		100	100	100	100

Take away:

1. Morningside Heights has the least number of venue categories and is much less than other neighborhoods.
2. We should remove it from consideration because of its limited venue variety.

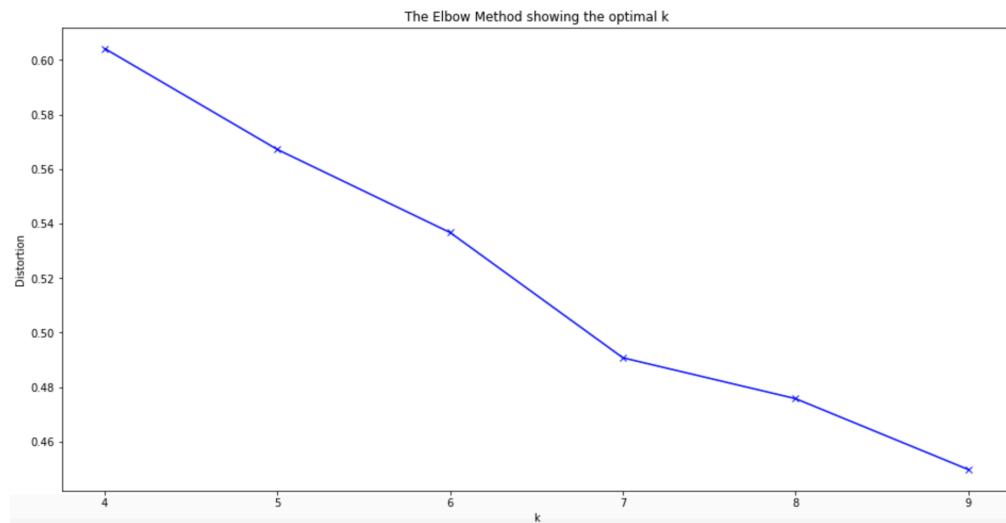
# Top 10 Venue Categories in Our Candidate Neighborhoods

---

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
Battery Park City	Cafe	Park	Hotel	Memorial Site	Clothing Store	Gym	Food Court	Gourmet Shop	Plaza	Playground
Greenwich Village	Italian Restaurant	Cafe	Clothing Store	Sushi Restaurant	American Restaurant	Dessert Shop	Indian Restaurant	Seafood Restaurant	Chinese Restaurant	Gym
Hudson Yards	Cafe	Gym / Fitness Center	American Restaurant	Hotel	Italian Restaurant	Burger Joint	Gym	Dog Run	Nightclub	Park
Little Italy	Cafe	Bakery	Bubble Tea Shop	Italian Restaurant	Chinese Restaurant	Ice Cream Shop	Sandwich Place	Salon / Barbershop	Cocktail Bar	Pizza Place
Murray Hill	Hotel	Cafe	Sandwich Place	Bar	American Restaurant	Italian Restaurant	Burger Joint	Japanese Restaurant	Gym / Fitness Center	Restaurant
Noho	Italian Restaurant	Cafe	Cocktail Bar	Hotel	Yoga Studio	Bookstore	Sandwich Place	Pizza Place	Mexican Restaurant	Grocery Store
Soho	Clothing Store	Cafe	Italian Restaurant	Mediterranean Restaurant	Asian Restaurant	Sporting Goods Shop	Boutique	Salon / Barbershop	Bakery	Pizza Place
Tribeca	Cafe	Park	Italian Restaurant	Spa	American Restaurant	Wine Bar	Men's Store	Greek Restaurant	Bakery	Gym / Fitness Center

# K-Means Clustering – find K

---

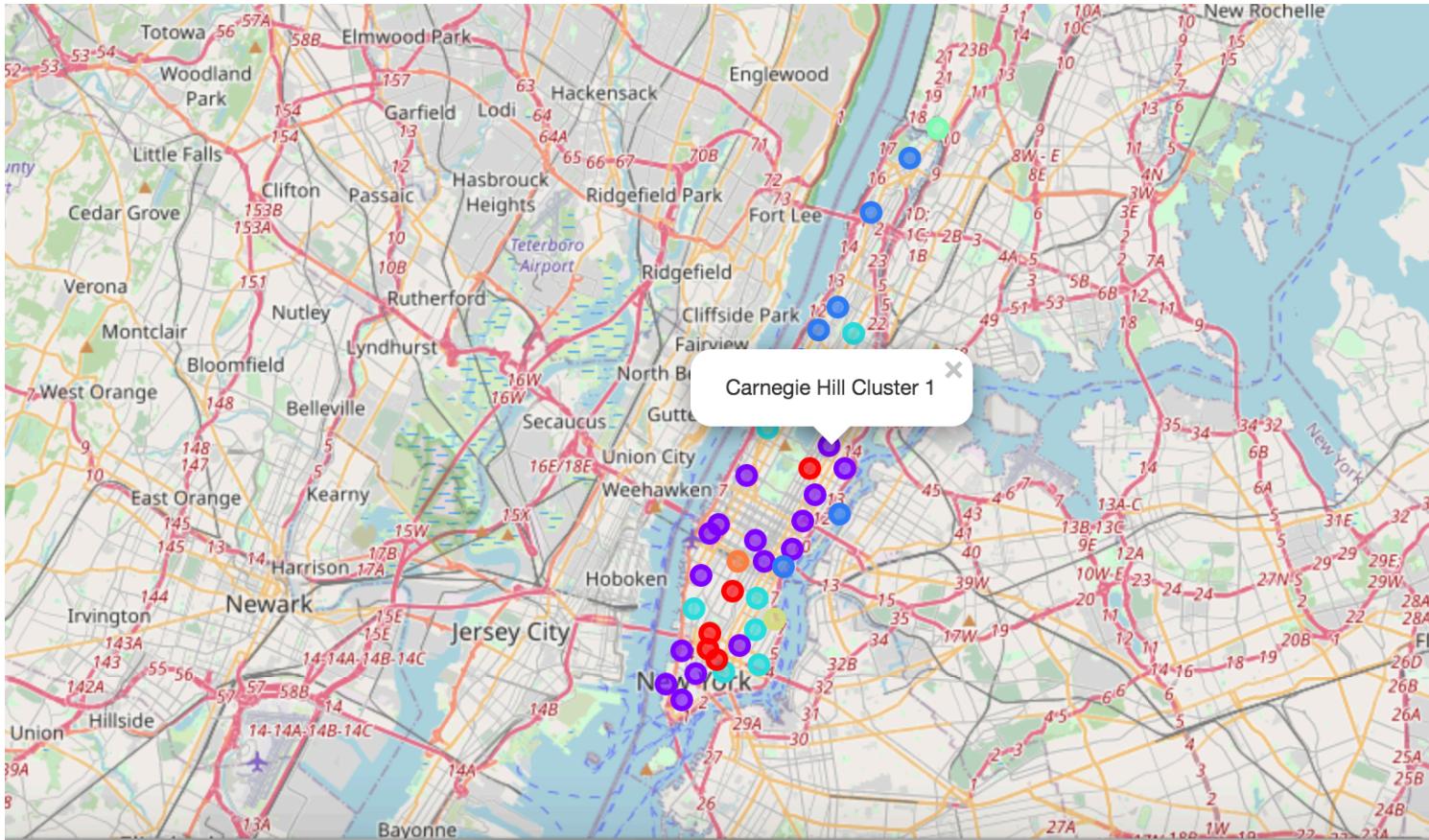


```
For n_clusters = 4 The average silhouette_score is : 0.030734128667779876
For n_clusters = 5 The average silhouette_score is : 0.002508987600141415
For n_clusters = 6 The average silhouette_score is : 0.016984022724632723
For n_clusters = 7 The average silhouette_score is : 0.0131426919175024
For n_clusters = 8 The average silhouette_score is : 0.0026767148998263816
For n_clusters = 9 The average silhouette_score is : 0.006567279598424958
```

1. **The elbow method:** no explicit kink, but a relatively kink around  $K = 7$
2. **Silhouette scores:** not high enough, but relatively high around  $K = 6$  and  $K = 7$
3. **Pick  $K = 7$**

# K-Means Clustering

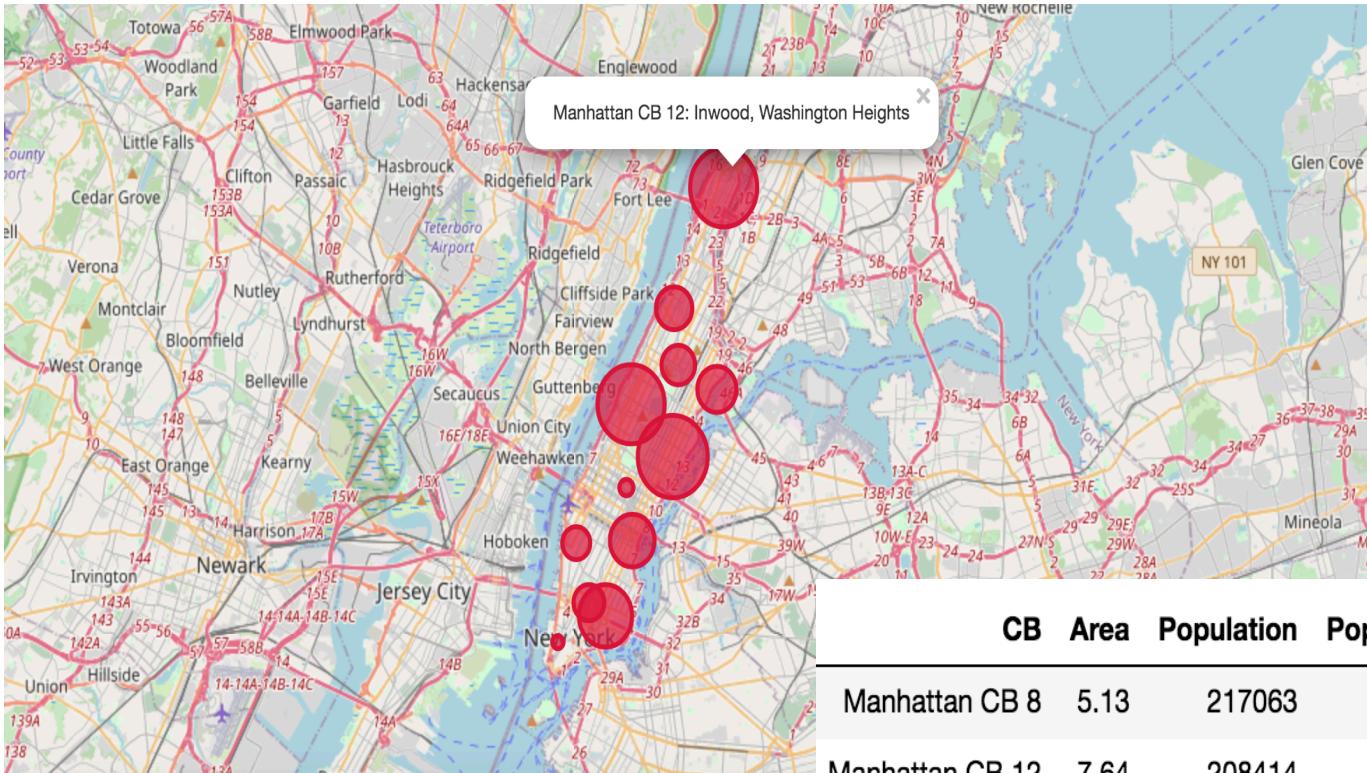
---



# Venue Categories with Cluster Labels

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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Manhattan	Murray Hill	40.748303	-73.978332	1	Hotel	Cafe	Sandwich Place	Bar	American Restaurant	Italian Restaurant	Burger Joint	Japanese Restaurant	Gym / Fitness Center	Restaurant
Manhattan	Greenwich Village	40.726933	-73.999914	0	Italian Restaurant	Cafe	Clothing Store	Sushi Restaurant	American Restaurant	Dessert Shop	Indian Restaurant	Seafood Restaurant	Chinese Restaurant	Gym
Manhattan	Tribeca	40.721522	-74.010683	1	Cafe	Park	Italian Restaurant	Spa	American Restaurant	Wine Bar	Men's Store	Greek Restaurant	Bakery	Gym / Fitness Center
Manhattan	Little Italy	40.719324	-73.997305	0	Cafe	Bakery	Bubble Tea Shop	Italian Restaurant	Chinese Restaurant	Ice Cream Shop	Sandwich Place	Salon / Barbershop	Cocktail Bar	Pizza Place
Manhattan	Soho	40.722184	-74.000657	0	Clothing Store	Cafe	Italian Restaurant	Mediterranean Restaurant	Asian Restaurant	Sporting Goods Shop	Boutique	Salon / Barbershop	Bakery	Pizza Place
Manhattan	Battery Park City	40.711932	-74.016869	1	Cafe	Park	Hotel	Memorial Site	Clothing Store	Gym	Food Court	Gourmet Shop	Plaza	Playground
Manhattan	Noho	40.723259	-73.988434	1	Italian Restaurant	Cafe	Cocktail Bar	Hotel	Yoga Studio	Bookstore	Sandwich Place	Pizza Place	Mexican Restaurant	Grocery Store
Manhattan	Hudson Yards	40.756658	-74.000111	1	Cafe	Gym / Fitness Center	American Restaurant	Hotel	Italian Restaurant	Burger Joint	Gym	Dog Run	Nightclub	Park

# Neighborhood Population Density



- The folium map helps us visualize the population density across all community boards in Manhattan. The higher the density, the bigger the radius of the mark.
- We then compare the 24 neighborhoods with high population density with our candidates.

CB	Area	Population	Population/Area	Neighborhoods	Latitude	Longitude
Manhattan CB 8	5.13	217063	42312	Lenox Hill, Roosevelt Island, Upper East Side,...	40.769891	-73.955392
Manhattan CB 12	7.64	208414	27279	Inwood, Washington Heights	40.854728	-73.930358
Manhattan CB 7	5.46	207699	38040	Lincoln Square, Manhattan Valley, Upper West Side	40.786380	-73.975863
Manhattan CB 3	4.56	164407	36054	Alphabet City, Chinatown, East Village, Lower ...	40.719617	-73.988447
Manhattan CB 6	3.55	136152	38353	Gramercy Park, Kips Bay, Rose Hill, Murray Hil...	40.743383	-73.975280

**Result: Murray Hill**

# Discussion and Conclusion

---

- ❖ Murray Hill – the optimal neighborhood in Manhattan for café business.
  - ❖ It does not have many coffee shops in the neighborhood for now, so we do not need to care about the business competition with other owners.
  - ❖ It has many unique venue categories in it, which indicates business diversity and huge potential for growth.
  - ❖ Relatively larger population density is also another advantage of Murray Hill as it represents popularity and exposure.
- ❖ Discussion and future improvements
  - ❖ k-means clustering needs more feature analysis or metrics to find the optimal K value.
  - ❖ Additional factors need consideration: real estate condition, housing price, business budget, neighborhood center location, etc...