Determining Candidate Neighborhoods For Health Food Store Expansion Using Cluster Analysis

ADAM HORIN

SEPTEMBER 23, 2019

INTRODUCTION

- Background
  - Sunac Natural Food is a health food business that offers organic, health-conscious foods to their customers [1]
- Problem
  - They are looking to expand their locations to another neighborhood in Manhattan
  - They need help identifying factors that would help them select candidate locations for their expansion
- Interest
  - Other businesses in a similar challenge of trying to expand their locations would be interested applying the analysis approach used in this project

DATA

- NYU Repository [2]

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

- Foursquare API [3]

	Neighborhood	Venue	Venue Category
666	Lenox Hill	A Matter of Health	Health Food Store
1011	Clinton	Sunac Natural Food	Health Food Store
3297	Hudson Yards	Sunac Natural Food	Health Food Store
2997	Turtle Bay	The Health Nuts	Health Food Store
2089	Manhattan Valley	The Nutbox	Health Food Store

METHODS

K-means cluster analysis was performed on the Manhattan neighborhoods based on frequency of venue categories.

This is a method of unsupervised machine learning that clusters the data based on similar factors.

EXAMPLE OF VENUE FREQUENCIES IN MANHATTAN NEIGHBORHOODS

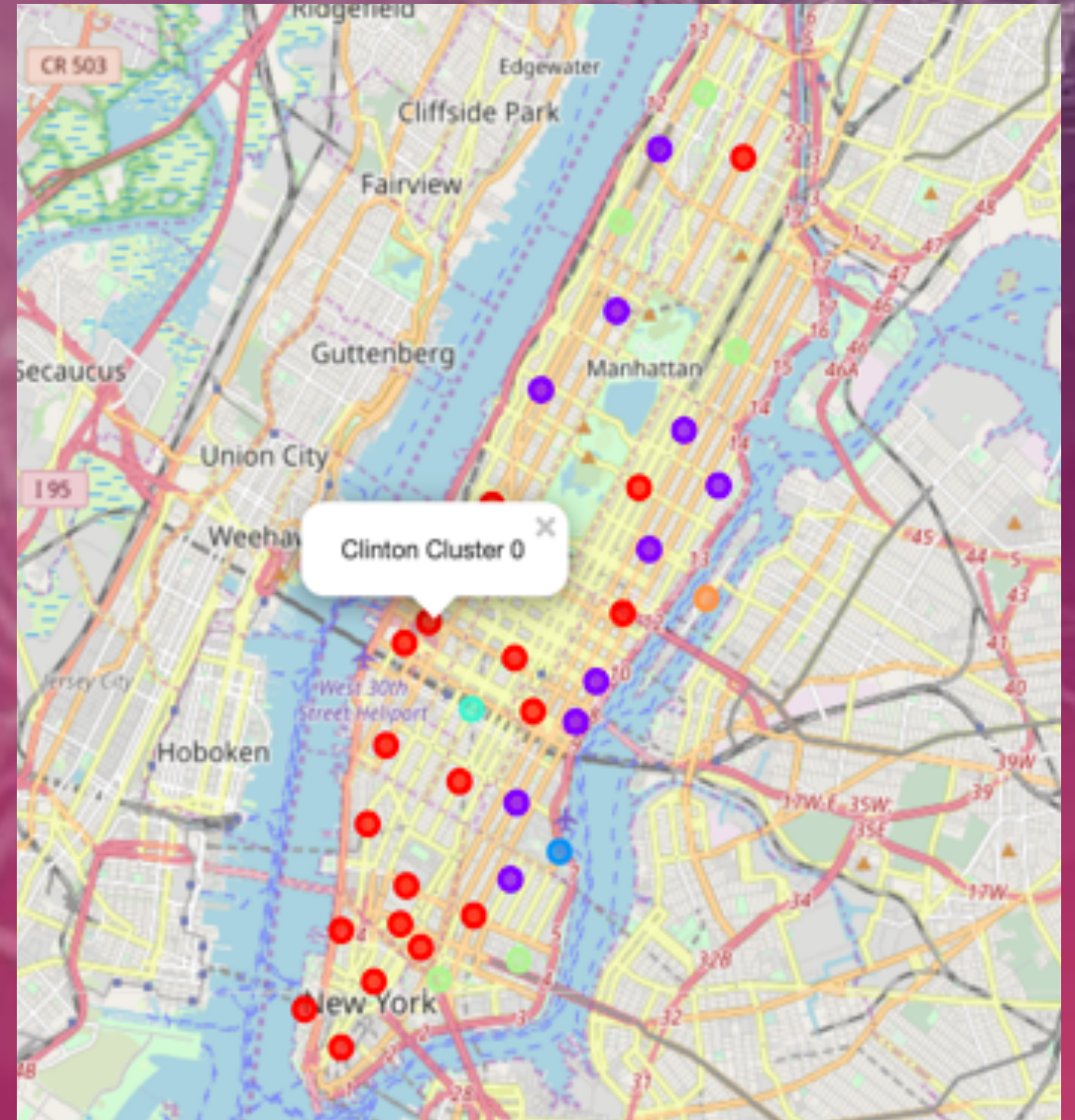|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Memorial Site	Gym	Italian Restaurant	Wine Shop	Clothing Store	Plaza	Playground
1	Carnegie Hill	Coffee Shop	Pizza Place	Cosmetics Shop	Café	Japanese Restaurant	Grocery Store	French Restaurant	Yoga Studio	Bookstore	Bar
2	Central Harlem	African Restaurant	Chinese Restaurant	French Restaurant	Gym / Fitness Center	American Restaurant	Bar	Public Art	Seafood Restaurant	Fried Chicken Joint	Southern / Soul Food Restaurant
3	Chelsea	Coffee Shop	Italian Restaurant	Bakery	Ice Cream Shop	Nightclub	Hotel	Theater	American Restaurant	Seafood Restaurant	Tapas Restaurant
4	Chinatown	Chinese Restaurant	Cocktail Bar	Vietnamese Restaurant	Bakery	American Restaurant	Spa	Salon / Barbershop	Bubble Tea Shop	Korean Restaurant	Asian Restaurant

OPTIMAL NUMBER OF CLUSTERS

For generated models, cluster number vs its distortion was plotted and optimal number of clusters was determined using the elbow method (k=6)

The Elbow Method showing the optimal k
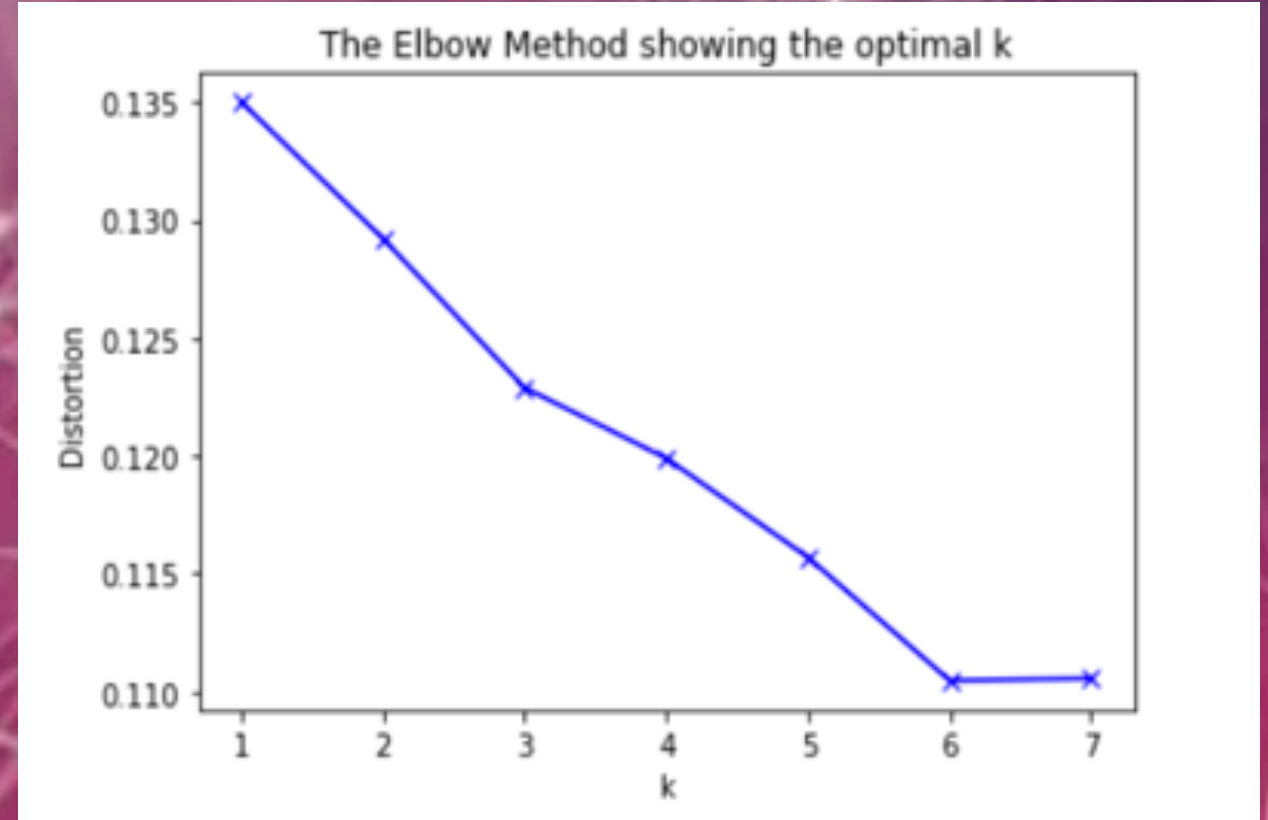
MAP OF CLUSTERED NEIGHBORHOODS

Cluster 0 is represented in red

SUMMARY OF CLUSTER 0 NEIGHBORHOODS

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
6	Central Harlem	African Restaurant	Gym / Fitness Center	French Restaurant
8	Upper East Side	Italian Restaurant	Exhibit	Art Gallery
13	Lincoln Square	Theater	Gym / Fitness Center	Café
14	Clinton	Theater	Gym / Fitness Center	American Restaurant
15	Midtown	Hotel	Coffee Shop	Cocktail Bar
16	Murray Hill	Coffee Shop	Hotel	Japanese Restaurant
17	Chelsea	Coffee Shop	Italian Restaurant	Bakery
18	Greenwich Village	Italian Restaurant	Clothing Store	Sushi Restaurant
21	Tribeca	Italian Restaurant	American Restaurant	Park
22	Little Italy	Bakery	Café	Ice Cream Shop
23	Soho	Clothing Store	Boutique	Art Gallery

CANDIDATE NEIGHBORHOODS

TRIBECA

WEST VILLAGE

FLATIRON

GREENWICH VILLAGE

CHELSEA

MURRAY HILL

MIDTOWN

LINCOLN SQUARE

LITTLE ITALY

SOHO

FINANCIAL DISTRICT

UPPER EAST SIDE

CENTRAL HARLEM

NOHO

CIVIC CENTER

SUTTON PLACE

BATTERY PARK CITY

DISCUSSION

Based on the clustering it could be recommended that Sunac Natural Food opens an expansion location in another neighborhood in Cluster 0. The current sunac natural food locations are in the Clinton and Hudson Yards neighborhoods, which are both in Cluster 0. Many of the neighborhoods in this cluster have a gym/fitness center in their top three most common venues.

- Limitations: limited factors studied

- Future Directions: conduct further analyses using retail market prices and neighborhood demographics

- References

[1] https://www.sunacnaturalmarket.com/

[2] https://cocl.us/new_york_dataset

[3] https://developer.foursquare.com/