CAPSTONE PROJECT

THE BATTLE OF NEIGHBORHOODS

INTRODUCTION

- Introduction where you discuss the business problem and who would be interested in this project.
- Restaurant owners often find out that it is difficult to maintain a business in the multicultural, highly populated city like New York.
- One of the easiest way to find a good location for new business is by finding similar neighborhoods.
- Thus, we will be looking at two ways of finding similar neighborhoods
  - k-means clustering
  - cosine similarity

DATA

- Data where you describe the data that will be used to solve the problem and the source of the data.
- New York City neighborhood dataset
  - https://geo.nyu.edu/catalog/nyu_2451_34572
  - https://cocl.us/new_york_dataset
- Toronto neighborhood dataset
  - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Wikipedia Zip Code data
- Geocode
  - Convert Zip Code data into longitude, latitude pair
- Foursquare API
  - Get relevant information based on the location
- List of Cuisine Types
  - https://en.wikipedia.org/wiki/List_of_cuisines
  - https://developer.foursquare.com/docs/resources/categories

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446

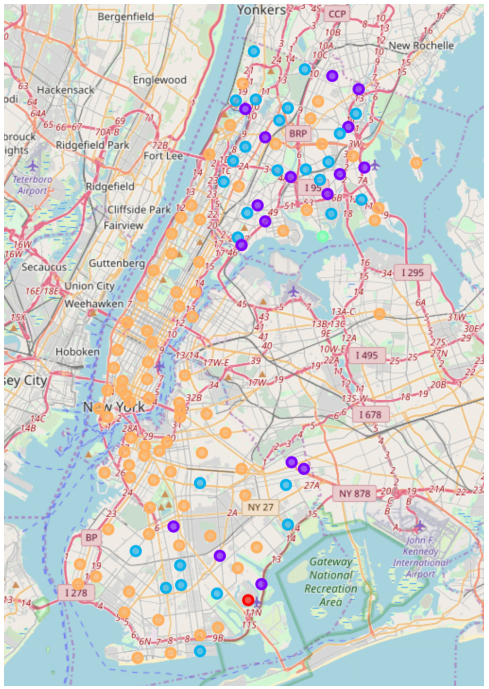
METHODOLOGY

- Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
- K-Nearest Neighbor
- Cosine Similarity

RESULTS (KNN)

- Results section where you discuss the results.
- Cluster 0: Mill Island
- Cluster 1: Wakefield Eastchester Marble Hill Baychester Fordham West Farms Port Morris Longwood Morrisania Country Club Westchester Square Unionport Cypress Hills Flatlands City Line Bergen Beach Prospect Park South
- Cluster 2: Co-op City Kingsbridge Woodlawn Norwood Bedford Park University Heights Morris Heights East Tremont High Bridge Melrose Mott Haven Parkchester Van Nest Morris Park Spuyten Duyvil North Riverdale Schuylerville Castle Hill Pelham Gardens Manhattan Terrace Crown Heights East New York Starrett City Manhattan Beach Borough Park Marine Park Midwood Ocean Parkway
- Cluster 3: Clason Point
- Cluster 4: Riverdale Williamsbridge Pelham Parkway City Island Hunts Point Soundview Throgs Neck Belmont Pelham Bay Edgewater Park Olinville Concourse Bay Ridge Bensonhurst Sunset Park Greenpoint Gravesend Brighton Beach Sheepshead Bay Flatbush East Flatbush Kensington Windsor Terrace Prospect Heights Brownsville Williamsburg Bushwick Bedford Stuyvesant Brooklyn Heights Cobble Hill Carroll Gardens Red Hook Gowanus Fort Greene Park Slope Canarsie Coney Island Bath Beach Dyker Heights Gerritsen Beach Clinton Hill Downtown Boerum Hill Prospect Lefferts Gardens Ocean Hill Georgetown East Williamsburg North Side South Side Fort Hamilton Chinatown Washington Heights Inwood Hamilton Heights Manhattanville Central Harlem East Harlem Upper East Side Yorkville Lenox Hill Roosevelt Island Upper West Side Lincoln Square Clinton Midtown Murray Hill Chelsea Greenwich Village East Village Lower East Side Tribeca Little Italy Soho West Village Manhattan Valley Morningside Heights Gramercy Battery Park City Financial District Murray Hill Chelsea

RESULTS (KNN)


Yonkers
Bergenfield
New Rochelle
Englewood
Hackensack
Ridgefield Park
Fort Lee
Ridgefield
Cliffside Park
Fairview
Secaucus
Guttenberg
Union City
Weehawken
Hoboken
I 295
I 495
I 678
NY 878
NY 27
I 278
Gateway National Recreation Area
John F. Kennedy International Airport

RESULTS (COSINE SIMILARITY)

	Neighborhood	cossim
70	Swansea, Runnymede	0.792118
10	Maryvale, Wexford	0.784633
49	Victoria Hotel, Commerce Court	0.772061
69	Roncesvalles, Parkdale	0.723370
52	Yorkville, North Midtown, The Annex	0.702311
39	St. James Town, Cabbagetown	0.693559

DISCUSSION

- Lack of Data in Foursquare API
- Change K for k-nearest neighbor
- Other similarity measure methods:
  - Euclidean Distance
  - Manhattan Distance

CONCLUSION

- We have looked at the possible location for restaurant owners to go over before they start a new business
- Meaningful results using two different approaches
- Could use same model for other cities, not just Battery Park City