



Clustering NYC Neighborhoods to Determine Service Needs

Ryan Brenner ~ September 4, 2018 ~ IBM Applied Data Science Capstone



Project Overview

- Cluster NYC Neighborhoods using Foursquare data
- Cluster NYC Neighborhoods using 311 Request data
- Compare clusters to determine whether Foursquare clusters can predict 311 requests

Foursquare Clusters

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Foursquare Cluster
Allerton	Pizza Place	Deli / Bodega	Chinese Restaurant	0
Annadale	Diner	Bakery	Cosmetics Shop	3
Arden Heights	Pharmacy	Deli / Bodega	Coffee Shop	0
Arlington	Intersection	Coffee Shop	Boat or Ferry	4
Arrochar	Italian Restaurant	Deli / Bodega	Bus Stop	3
Arverne	Surf Spot	Metro Station	Pizza Place	4
Astoria	Middle Eastern Restaurant	Bar	Hookah Bar	4
Astoria Heights	Motel	Bowling Alley	Pizza Place	4
Auburndale	Pharmacy	Supermarket	Pet Store	4
Bath Beach	Pharmacy	Women's Store	Donut Shop	0



311 Request Clusters

NTAName	First	Second	Third	311 Cluster
Airport	Taxi Complaint	For Hire Vehicle Complaint	Water System	0
Allerton-Pelham Gardens	Dead/Dying Tree	Water System	Vacant Lot	0
Annadale-Huguenot-Prince's Bay-Eltingville	Food Establishment	Water System	FLOORING/STAIRS	0
Astoria	Graffiti	Taxi Complaint	For Hire Vehicle Complaint	3
Auburndale	General Construction/Plumbing	Sewer	Water System	0
Baisley Park	Graffiti	Water System	FLOORING/STAIRS	2
Bath Beach	For Hire Vehicle Complaint	Water System	FLOORING/STAIRS	0
Battery Park City-Lower Manhattan	Food Establishment	Taxi Complaint	For Hire Vehicle Complaint	0
Bay Ridge	Special Projects Inspection Team (SPIT)	Food Establishment	Asbestos	0
Bayside-Bayside Hills	For Hire Vehicle Complaint	Street Light Condition	Water System	0



Cluster Matches

Foursquare Group = 311 Group

	A=0	A=1	A=2	A=3	A=4
Matches	18	1	52	3	14
	B=0	B=1	B=2	B=3	B=4
Matches	17	0	72	0	0
	C=0	C=1	C=2	C=3	C=4
Matches	0	0	2	0	0
	D=0	D=1	D=2	D=3	D=4
Matches	0	0	10	0	0
	E=0	E=1	E=2	E=3	E=4
Matches	3	0	14	0	0

Conclusion

- Accuracy of Foursquare clusters predicting 311 cluster: <44% (90/206)
- Accuracy of random guess: 20%