CAPSTONE PROJECT
FINDING THE BEST PLACES TO EAT IN NEW YORK

Introduction and Business Problem

Introduction

The city of New York is a big city and it is packed with restaurants, bars, night life and amazing people. For people that are new to New York, it can be very daunting to figure out what restaurants are worth going to and where they are. For people that used to live in New York or are visiting New York, whether it is for business or pleasure, how do you know what the best places are to get something to eat?

Business Problem

For this capstone project, I am going to go for a cost-friendly project and create a simple guide on where to eat based on what API calls are possible to use with Foursquare having a free membership, mainly speaking: Foursquare likes will be used as a measure of popularity, with the understanding that many likes means a place is good and also cost friendly. Very few likes means the place is not good, and an amount of likes in between is a gray area in which, it may not be a very good place but is not also bad, or it could be a really good place but very costly, therefore not accessible to all people, hence a lower amount of likes. That is why, I will use machine learning to cluster restaurants to provide more insight into making a better choice. Naturally, additional data will be used such as: restaurant category and geographic location data for restaurants in New York.

Data Requirements and Methodology

Data Requirements

For this project, I will be utilizing the Foursquare API to pull the following location data on restaurants in New York:

- Venue Name
- Venue ID
- Venue Location
- Venue Category
- Count of Likes

Data Acquisition Approach

To acquire the data mentioned above, I will need to do the following:

- Get geolocator latitude and longitude coordinates for New York
- Use Foursquare API to get a list of all venues in New York
    - Get venue name, venue ID, location, category, and likes

Methodology

The thought process behind this is that likes are a proxy for quality. The more likes there are, the better the restaurant is. This might be incorrect but API call issues (how many I can use for free) holds me back from getting price / rating data. I will then bin this data into a quality categorical variables so we can cluster appropriately.

I am also going to create new categorical variables for the restaurants to better group them into restaurant, fast food places and healthy food places to eat. This way, if you are a place for a business or formal dinner, or someplace more casual, or just looking to eat healthy, you will know into which category to look for.

I will take the gathered data (see above in Data Acquisition Approach and Data Required sections) and will create a k-means clustering algorithm that groups restaurants into 4-5 clusters so that people looking to eat in New York can easily see which restaurants are the best to eat at, what cuisine is available and where in New York they can look to eat.

Results
Running my clustering algorithm, I was able to generate four clusters of restaurants. These are as follows:

Cluster 1
In this cluster, we can observe most places fall in the other category, with only a few restaurants and fast food places. Therefore, places falling into this cluster should only be considered if we are not able to find the type of place or food we are looking for in other clusters as we will see next.

	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
1	Alba Dry Cleaner & Tailor	4c606c3e1e5cd13ad1a1a1ed	Laundry Service	40.711434	-74.006272	28	below avg	other	0
3	Gibney Dance Center Downtown	53373f26498e940581c90985	Dance Studio	40.713923	-74.005661	56	abv avg	other	0
11	Aahar Indian Cuisine	575dea4c498e2739e43a27e2	Indian Restaurant	40.713307	-74.007994	53	abv avg	restaurants	0
12	Four Seasons Hotel New York Downtown	57c640ad498e74977f98372f	Hotel	40.712612	-74.009380	67	abv avg	other	0
18	Church Street Boxing Gym	4b8dbba4f964a5201e0b33e3	Boxing Gym	40.713354	-74.009067	67	abv avg	other	0
20	Joe's Pizza	5c6f03f30802d4002c16884c	Pizza Place	40.710318	-74.007694	76	abv avg	fast food	0
22	Gran Morsi	5421eb06498e1b6b9c1bfd9c	Italian Restaurant	40.714246	-74.007925	72	abv avg	restaurants	0
30	Potbelly Sandwich Shop	4f4d2593e4b00b42c3f1060f	Sandwich Place	40.714454	-74.005820	50	abv avg	fast food	0
32	Heyday	57ad129c498e05b086594d72	Spa	40.715726	-74.007767	30	below avg	other	0
33	Philip Williams Posters	4b747291f964a52042dd2de3	Antique Shop	40.715284	-74.008781	42	below avg	other	0
37	Chick-fil-A	5ab51b6aa4b51b3dc096ed68	Fast Food Restaurant	40.710419	-74.008550	66	abv avg	fast food	0
39	Konditori	5953e5314382ab0b3b808dc6	Café	40.709474	-74.006630	52	abv avg	other	0
45	Primo's	5ae28a7b15173e002cef3271	Cocktail Bar	40.715501	-74.008977	34	below avg	bars	0
47	Evening Bar	545a8c16498eeafe6d52b176	Hotel Bar	40.715301	-74.009347	39	below avg	bars	0
51	Smyth Hotel	49efcc88f964a52006691fe3	Hotel	40.715144	-74.009183	62	abv avg	other	0
55	Affina Nails & Spa - Fulton Street	535af42f498e0f645e3546e0	Nail Salon	40.709235	-74.005534	40	below avg	other	0
68	Moxy NYC Downtown	5bd4cc646adbf5002c1ced5c	Hotel	40.710783	-74.007865	31	below avg	other	0
76	Kesté	58f54555e2ead178a41de832	Pizza Place	40.709132	-74.004676	68	abv avg	fast food	0
77	Hungry Ghost	5c7d2d843d4791002c881f1c	Coffee Shop	40.715379	-74.007476	28	below avg	other	0
81	La Parisienne	5a47a5e4e679bc7b226989ca	French Restaurant	40.709423	-74.009992	57	abv avg	restaurants	0
83	Sticky's Finger Joint	588251d3ac13691c90477561	Fried Chicken Joint	40.709309	-74.009016	58	abv avg	fast food	0
87	Residence Inn by Marriott New York Downtown Ma...	54aed577498ee5fc196739f7	Hotel	40.709563	-74.009718	54	abv avg	other	0
88	Poke Bowl	58c2e736bf1a6d6b31fa57fa	Poke Place	40.709722	-74.006854	68	abv avg	other	0
93	Hale & Hearty	4eb01d60722e4efd61c49664	Soup Place	40.709903	-74.007052	62	abv avg	healthy food	0

Cluster 2

This is by far the best cluster in which we should look to choose places to eat. All of them are in the 'great' category, meaning they have the most amount of likes. None of them fall in the 'other' category, and we have all the categories of places to eat like restaurants, bars, fast food and healthy food places.

	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
0	The Bar Room at Temple Court	57f0689d498e7d49d9189369	Hotel Bar	40.711448	-74.006802	192	great	bars	1
8	Augustine	58191674ded8f8626ed70af0	French Restaurant	40.711310	-74.006660	238	great	restaurants	1
10	Pisillo Italian Panini	528bf16711d2b7722da6b51c	Sandwich Place	40.710530	-74.007526	292	great	fast food	1
13	Los Tacos No. 1	5d5f24ec09484500079aee00	Taco Place	40.714267	-74.008756	78	great	fast food	1
16	Racines	534c9d7b498e1bdd443a40e1	French Restaurant	40.714754	-74.007581	120	great	restaurants	1
19	Chipotle Mexican Grill	4ee8e058f790a9d738d62d70	Burrito Place	40.714607	-74.006335	126	great	fast food	1
21	Da Claudio	5447e0b2498e49ee7c7b1dc0	Italian Restaurant	40.710826	-74.007639	100	great	restaurants	1
23	Sophie's Cuban Cuisine	4b311e89f964a520890025e3	Cuban Restaurant	40.714803	-74.007656	112	great	restaurants	1
24	Nish Nūsh	50ba9119e4b071a4bae6dc10	Falafel Restaurant	40.715537	-74.007725	403	great	restaurants	1
25	Shake Shack	5787b68e498efcabbebba4f8	Burger Joint	40.710703	-74.009024	505	great	fast food	1
36	FlashDancers Downtown	433b2e80f964a52036281fe3	Strip Club	40.714350	-74.009800	87	great	bars	1
44	Melt Shop	53a7307a498e56d9917d8f32	Sandwich Place	40.709807	-74.006723	175	great	fast food	1
48	Little Park	545c0436498e798e22ce4b2a	American Restaurant	40.715487	-74.009133	466	great	restaurants	1
50	Nish Nush	564cb952498e133963c04186	Falafel Restaurant	40.709418	-74.008056	164	great	restaurants	1
52	Nobu Downtown	4b60c708f964a520d0f829e3	Japanese Restaurant	40.710532	-74.009593	269	great	restaurants	1
54	Atera	4f627061e4b05c1d57815977	Molecular Gastronomy Restaurant	40.716752	-74.005712	183	great	healthy food	1
58	Takahachi	4a8f2f39f964a520471420e3	Sushi Restaurant	40.716526	-74.008101	318	great	restaurants	1
59	Chambers Street Wines	4adcf23cf964a520cc6221e3	Wine Shop	40.715773	-74.009718	90	great	bars	1
61	Juice Press	54148bc6498ea7bb8c05b70a	Vegetarian / Vegan Restaurant	40.714788	-74.011132	86	great	healthy food	1
63	Sophie's Cuban Cuisine	54cad0b2498ea68d80401cbd	Cuban Restaurant	40.709165	-74.005404	84	great	restaurants	1
64	Ward III	4a411ea5f964a520cba41fe3	Cocktail Bar	40.715885	-74.008714	564	great	bars	1
66	GRK Fresh Greek - Financial District	5047c785e4b0bcc0f416cdb3	Greek Restaurant	40.709800	-74.007011	447	great	restaurants	1
71	Hole In The Wall	58fdf5ec6cf01a4f54c7792b	Breakfast Spot	40.708280	-74.005612	327	great	fast food	1
78	Tiny's and the Bar Upstairs	4dadcb124df0522cc5622202	American Restaurant	40.716793	-74.008220	537	great	restaurants	1
79	Khe-Yo	51df85c9498edb5ea3ad2e2a	Asian Restaurant	40.716753	-74.008584	404	great	restaurants	1
80	Weather Up	4cd89eeb6e8b5941660c64d2	Cocktail Bar	40.716741	-74.008666	375	great	bars	1
89	Go! Go! Curry	512a6086e4b023dd9c27e82d	Japanese Curry Restaurant	40.709854	-74.009010	184	great	restaurants	1
92	Gunbae	555b7016498ef57fbcc739d2	Korean Restaurant	40.714529	-74.010242	111	great	restaurants	1
96	Tribeca's Kitchen	54242f3a498e53a21ee14b81	Diner	40.716106	-74.007076	94	great	fast food	1

Cluster 3

This cluster, only grouped places that are not for eating, therefore there is not much we can do with this cluster, unless we were looking for places to visit, since all of these places fall in the 'great' category, based on the high number of likes from Foursquare.

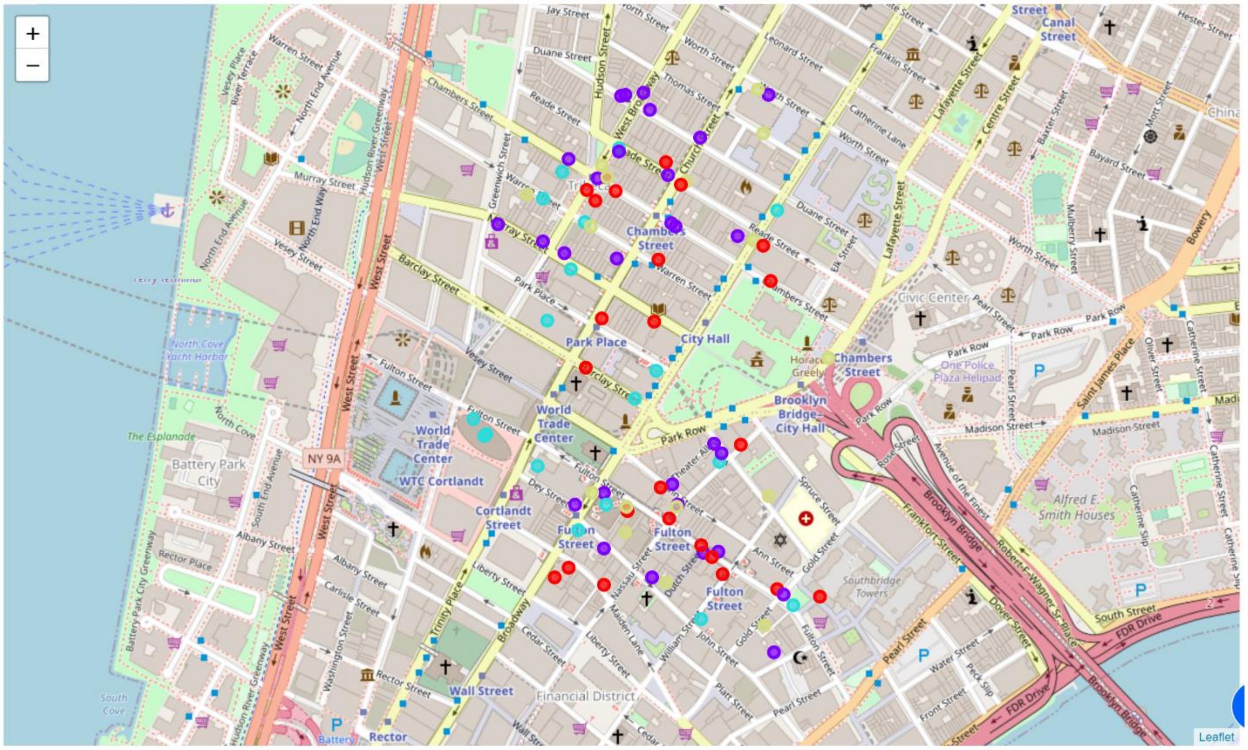
	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
2	The Beekman, A Thompson Hotel	56d8c0f8498edb854f926e6a	Hotel	40.711173	-74.006702	208	great	other	2
5	The Wooly Daily	56093809498e5344ab8835a6	Coffee Shop	40.712137	-74.008395	156	great	other	2
15	African Burial Ground National Monument	4a9442e3f964a520f92020e3	Monument / Landmark	40.714990	-74.005530	93	great	other	2
17	Woolworth Building	4be99814a9900f479a811540	Building	40.712559	-74.007964	116	great	other	2
26	Korin	4af5d65ff964a52091fd21e3	Furniture / Home Store	40.714824	-74.009404	100	great	other	2
28	Equinox Tribeca	4a6e331af964a52031d41fe3	Gym	40.714099	-74.009686	250	great	other	2
56	Ten Over Ten	4ce2a27cd58c60fc0fa1a76f	Nail Salon	40.715941	-74.008721	110	great	other	2
57	Midtown Comics	4cdac798d6656a315f36fc3e	Comic Shop	40.708989	-74.005218	147	great	other	2
60	Oculus Plaza	5984eeb39be522744003dee5	Plaza	40.711822	-74.011632	330	great	other	2
62	Apple World Trade Center	57a240f1cd10ed172db51626	Electronics Store	40.711566	-74.011426	376	great	other	2
69	Voyager Espresso	566de7e3498e30e4798117ae	Coffee Shop	40.708787	-74.007063	173	great	other	2
70	Mulberry & Vine	5171b5cc011cef9833bbb787	Café	40.715177	-74.010227	303	great	other	2
72	Westfield World Trade Center	509fd01de4b070da0a9d3c24	Shopping Mall	40.711602	-74.011368	449	great	other	2
75	Zucker's Bagels & Smoked Fish	4a4baaf8f964a52091ac1fe3	Bagel Shop	40.715580	-74.009850	422	great	other	2
82	Aroma Espresso Bar	4edd43af7bebc29455367992	Café	40.713329	-74.010158	468	great	other	2
95	Fulton Center	540d97b5498e05abd30fd5ca	Shopping Mall	40.710520	-74.008965	763	great	other	2
98	Millennium Hilton	510ad6e2e4b0a2ecec03b322	Hotel	40.711114	-74.010333	143	great	other	2
99	Urban Outfitters	52fbe02a498e1ff915558ae4	Clothing Store	40.710131	-74.009530	113	great	other	2

Cluster 4

This is cluster is basically telling us, all the places that we should avoid at all cost, since all of them fall in the 'poor' category.

	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
7	CrossFit 212 TriBeCa	52001eed498e9ac16ca5e20b	Gym	40.714537	-74.005999	17	poor	other	3
29	Pisillo Italian Cafe	588a4316326c5a4b60559f17	Café	40.710493	-74.007546	12	poor	other	3
31	Modern Martial Arts NYC Tribeca	4e31c3b1814d9a6fbde557c7	Martial Arts Dojo	40.715431	-74.007362	6	poor	other	3
34	Municipal Plaza	5021648ae4b0b9f6b6d566b2	Plaza	40.712755	-74.004498	12	poor	other	3
41	Babesta	4cc368704fcfbfb70a0cbe24	Baby Store	40.714760	-74.009280	18	poor	other	3
49	The Assemblage John Street	5ad8abdec365886896d7a611	Coworking Space	40.710104	-74.008574	14	poor	other	3
53	New York by Gehry Gym	4e9b03e59a52edbd658ca490	Gym	40.710655	-74.005709	14	poor	other	3
65	Gong Cha	5bec5ea23e67417691559498	Bubble Tea Shop	40.710704	-74.009257	8	poor	other	3
67	Lekka Burger	5dc6f6a5ea8dfb00080f6faa	Burger Joint	40.715246	-74.010559	23	poor	fast food	3
73	20 Thomas St. (The Flea Theater)	4ea89e6702d5b8174e85e65e	Theater	40.716161	-74.005826	10	poor	other	3
74	The Frederick Hotel	59e391c3e1f22816d1b09cf5	Hotel	40.715675	-74.009047	10	poor	other	3
85	Maestro Pasta	5d4861420372ce0007e23375	Italian Restaurant	40.709337	-74.007769	5	poor	restaurants	3
86	sweetgreen	5ceeb56dc03635002ce7ab2a	Salad Place	40.715493	-74.008951	4	poor	healthy food	3
90	Water4Dogs Rehabilitaion Center	4a75e7aef964a52098e11fe3	Medical Center	40.716838	-74.005922	8	poor	other	3
91	Pret A Manger	59a4518b3af98819c1b1c80b	Salad Place	40.710484	-74.008563	8	poor	healthy food	3
94	Casa Taqueria	5cd44bec31ac6c003920e96b	Taco Place	40.708694	-74.005791	6	poor	fast food	3

Map of Clusters for Reference



Discussion

It is understandable that the amount of insight we can get from popularity alone, based on the number of likes is very limited, however, as already explained in the Data Section, there is still much we can learn from this indicator alone, and coupled with clustering algorithm, and proper categorization of the types de places to eat, we can narrow down the list of options to only a very small handful, which will be very helpful for any visitor in New York.

Conclusion

From all the insight that was possible to draw, we can conclude that the best options to visit, if we are looking for some place to eat, are the places that fall in Cluster 2. Within these cluster, we have a variety of different places to choose from, like restaurant, bars, fast food and healthy food places.