Manhattan Eating Preferences

Adrian Tarniceriu
20.08.2019

1.Introduction

Goal:

- Analyze the eating preferences for the Manhattan neighborhoods
- Group the eating venues in three categories: Cafe, FastFood, Restaurant

- Help visitors find a location more suitable for their preferences,
- Provide entrepreneurs with an insight about possible business opportunities, such as opening a new restaurant (or shutting down an existing one).

2. Data

Location data:

- Obtain neighborhood location data from *https://cocl.us/new_york_dataset*
- 40 neighborhoods in Manhattan

Venue data from Foursquare:

- 1182 venues for the 40 neighborhoods of Manhattan

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue ID	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	4bf58dd8d48988d1ca941735	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	4bf58dd8d48988d102941735	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	4bf58dd8d48988d147941735	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	4bf58dd8d48988d1e0931735	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	4bf58dd8d48988d148941735	Donut Shop
5	Marble Hill	40.876551	-73.91066	Blink Fitness Riverdale	40.877147	-73.905837	4bf58dd8d48988d176941735	Gym
6	Marble Hill	40.876551	-73.91066	TCR The Club of Riverdale	40.878628	-73.914568	4e39a891bd410d7aed40cbc2	Tennis Stadium
7	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	4bf58dd8d48988d1ce941735	Seafood Restaurant
8	Marble Hill	40.876551	-73.91066	T.J. Maxx	40.877232	-73.905042	4bf58dd8d48988d1f6941735	Department Store
9	Marble Hill	40.876551	-73.91066	Starbucks	40.873755	-73.908613	4bf58dd8d48988d1e0931735	Coffee Shop

2. Data

Data preprocessing:

- Select only eating places
- 455 eating venues for the 40 neighborhoods of Manhattan
- Group different venue types to three categories

Venue Category contains:	Assigned to:
'restaurant', 'bodega', or 'diner'	Restaurant
'cafe' or 'coffee'	Cafe
'joint', 'bagel', 'pizza', 'breakfast', 'burger', 'burrito', 'creperie', 'fast food', 'pastry', 'sandwich', 'snack', or 'taco'	FastFood

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue ID	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	4bf58dd8d48988d1ca941735	FastFood
1	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	4bf58dd8d48988d147941735	Restaurant
2	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	4bf58dd8d48988d1e0931735	Cafe
3	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	4bf58dd8d48988d1ce941735	Restaurant
4	Marble Hill	40.876551	-73.91066	Starbucks	40.873755	-73.908613	4bf58dd8d48988d1e0931735	Cafe
5	Marble Hill	40.876551	-73.91066	Subway Sandwiches	40.874667	-73.909586	4bf58dd8d48988d1c5941735	FastFood
6	Marble Hill	40.876551	-73.91066	Boston Market	40.877430	-73.905412	4bf58dd8d48988d14e941735	Restaurant
7	Marble Hill	40.876551	-73.91066	SUBWAY	40.878493	-73.905385	4bf58dd8d48988d1c5941735	FastFood
8	Marble Hill	40.876551	-73.91066	Subway	40.877720	-73.905380	4bf58dd8d48988d1c5941735	FastFood
9	Marble Hill	40.876551	-73.91066	Terrace View Delicatessen	40.876476	-73.912746	4bf58dd8d48988d146941735	Restaurant

3. Methdology

Exploratory data analysis:

- Count of each eating place type per neighborhood

	Neighborhood	Cafe	FastFood	Restaurant	Venues
0	Battery Park City	2.0	4.0	1.0	7.0
1	Carnegie Hill	2.0	4.0	7.0	13.0
2	Central Harlem	0.0	3.0	11.0	14.0
3	Chelsea	1.0	1.0	11.0	13.0
4	Chinatown	0.0	3.0	10.0	13.0
5	Civic Center	1.0	1.0	11.0	13.0
6	Clinton	0.0	2.0	3.0	5.0
7	East Harlem	1.0	2.0	11.0	14.0
8	East Village	2.0	4.0	9.0	15.0
9	Financial District	2.0	2.0	5.0	9.0
10	Flatiron	0.0	0.0	7.0	7.0
11	Gramercy	2.0	4.0	6.0	12.0
12	Greenwich Village	1.0	3.0	10.0	14.0
13	Hamilton Heights	3.0	1.0	10.0	14.0
14	Hudson Yards	1.0	0.0	8.0	9.0
15	Inwood	0.0	1.0	12.0	13.0
16	Lenox Hill	0.0	4.0	6.0	10.0
17	Little Italy	1.0	4.0	5.0	10.0
18	Lower East Side	2.0	0.0	11.0	13.0
19	Manhattan Valley	2.0	3.0	11.0	16.0

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Venues
0	Battery Park City	FastFood	Cafe	Restaurant	7.0
1	Carnegie Hill	Restaurant	FastFood	Cafe	13.0
2	Central Harlem	Restaurant	FastFood	Cafe	14.0
3	Chelsea	Restaurant	FastFood	Cafe	13.0
4	Chinatown	Restaurant	FastFood	Cafe	13.0
5	Civic Center	Restaurant	FastFood	Cafe	13.0
6	Clinton	Restaurant	FastFood	Cafe	5.0
7	East Harlem	Restaurant	FastFood	Cafe	14.0
8	East Village	Restaurant	FastFood	Cafe	15.0
9	Financial District	Restaurant	FastFood	Cafe	9.0
10	Flatiron	Restaurant	FastFood	Cafe	7.0
11	Gramercy	Restaurant	FastFood	Cafe	12.0
12	Greenwich Village	Restaurant	FastFood	Cafe	14.0
13	Hamilton Heights	Restaurant	Cafe	FastFood	14.0
14	Hudson Yards	Restaurant	Cafe	FastFood	9.0
15	Inwood	Restaurant	FastFood	Cafe	13.0
16	Lenox Hill	Restaurant	FastFood	Cafe	10.0
17	Little Italy	Restaurant	FastFood	Cafe	10.0
18	Lower East Side	Restaurant	Cafe	FastFood	13.0
19	Manhattan Valley	Restaurant	FastFood	Cafe	16.0

3. Methdology

Machine Learning Technique:

- K-Means algorithm to partition the neighborhoods based on the eating venue type.
- As there are three different venue types, we will use three different clusters.

4. Results

Exploratory data analysis:

- 39 of the 40 neighborhoods have "Restaurant" as the most common venue, and one neighborhood has "FastFood" as the most common venue.
- The second most common venue is "FastFood" in 28 cases and "Cafe" in 12 cases.

4. Results

Machine Learning Technique:

- K-Means, 3 clusters
- (Figures in the following two slides)

- 5 neighborhoods have been assigned to cluster 0
- 19 neighborhoods have been assigned to cluster 1
- 15 neighborhoods have been assigned to cluster 2
- One neighborhood does not have any returned venue and was removed from the analysis

- Comments in the Discussion section

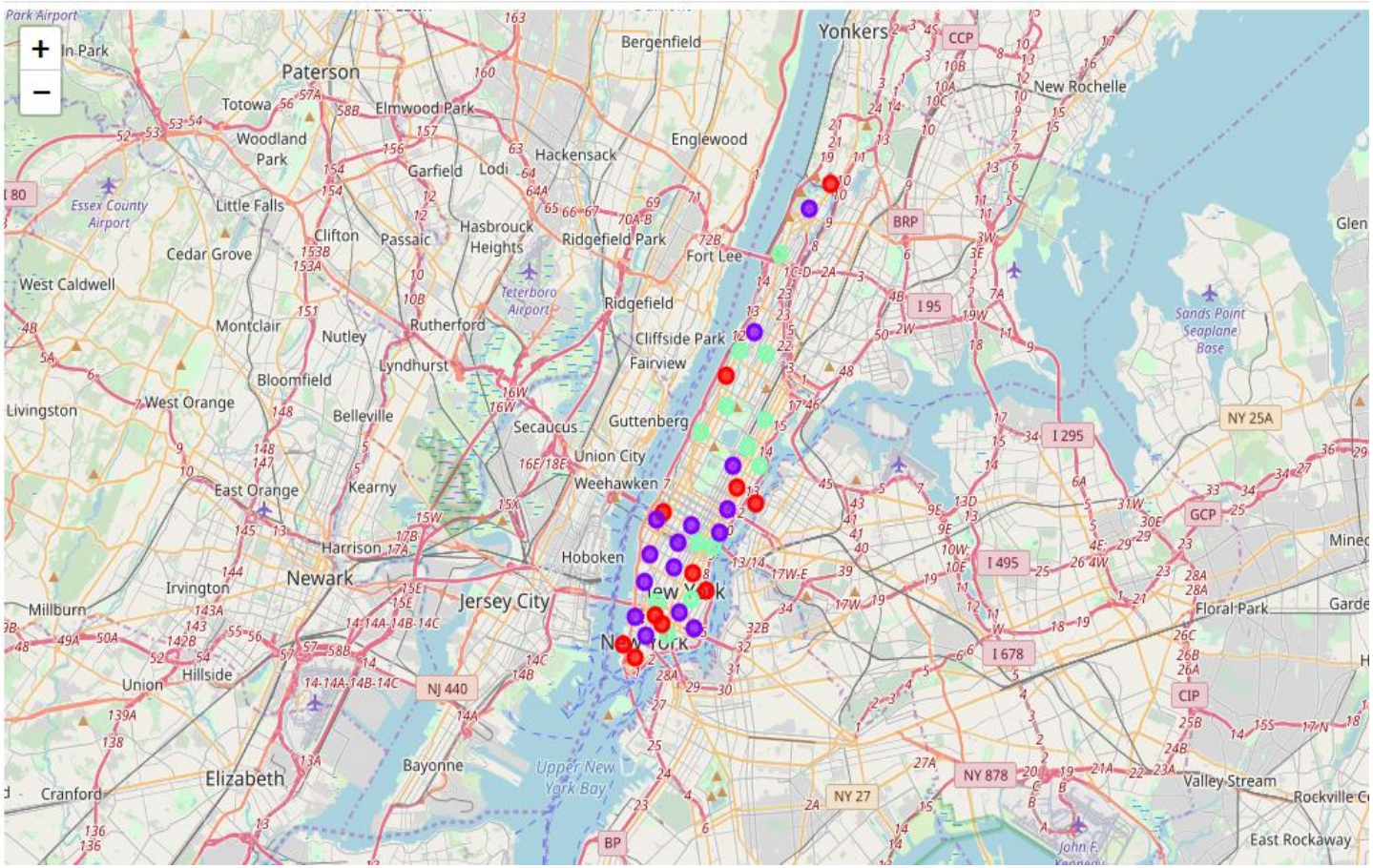
4. Results

Machine Learning Technique:

	index	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Venues
0	28	Manhattan	Battery Park City	40.711932	-74.016869	0.0	FastFood	Cafe	Restaurant	7.0
1	30	Manhattan	Carnegie Hill	40.782683	-73.953256	2.0	Restaurant	FastFood	Cafe	13.0
2	6	Manhattan	Central Harlem	40.815976	-73.943211	2.0	Restaurant	FastFood	Cafe	14.0
3	17	Manhattan	Chelsea	40.744035	-74.003116	1.0	Restaurant	FastFood	Cafe	13.0
4	1	Manhattan	Chinatown	40.715618	-73.994279	2.0	Restaurant	FastFood	Cafe	13.0
5	32	Manhattan	Civic Center	40.715229	-74.005415	1.0	Restaurant	FastFood	Cafe	13.0
6	14	Manhattan	Clinton	40.759101	-73.996119	0.0	Restaurant	FastFood	Cafe	5.0
7	7	Manhattan	East Harlem	40.792249	-73.944182	2.0	Restaurant	FastFood	Cafe	14.0
8	19	Manhattan	East Village	40.727847	-73.982226	2.0	Restaurant	FastFood	Cafe	15.0
9	29	Manhattan	Financial District	40.707107	-74.010665	0.0	Restaurant	FastFood	Cafe	9.0
10	38	Manhattan	Flatiron	40.739673	-73.990947	1.0	Restaurant	FastFood	Cafe	7.0
11	27	Manhattan	Gramercy	40.737210	-73.981376	0.0	Restaurant	FastFood	Cafe	12.0
12	18	Manhattan	Greenwich Village	40.726933	-73.999914	2.0	Restaurant	FastFood	Cafe	14.0
13	4	Manhattan	Hamilton Heights	40.823604	-73.949688	1.0	Restaurant	Cafe	FastFood	14.0
14	39	Manhattan	Hudson Yards	40.756658	-74.000111	1.0	Restaurant	Cafe	FastFood	9.0
15	3	Manhattan	Inwood	40.867684	-73.921210	1.0	Restaurant	FastFood	Cafe	13.0
16	10	Manhattan	Lenox Hill	40.768113	-73.958860	0.0	Restaurant	FastFood	Cafe	10.0
17	22	Manhattan	Little Italy	40.719324	-73.997305	0.0	Restaurant	FastFood	Cafe	10.0
18	20	Manhattan	Lower East Side	40.717807	-73.980890	1.0	Restaurant	Cafe	FastFood	13.0
19	25	Manhattan	Manhattan Valley	40.797307	-73.964286	2.0	Restaurant	FastFood	Cafe	16.0

4. Results

Machine Learning Technique:



5. Discussion

Goal: classify the neighborhoods based on the eating venue type.

- According to the exploratory data analysis, 39/40 neighborhoods have "Restaurant" as the most common venue type.
- Based on this, one might expect three classes ("Restaurant" as most common and "FastFood" as second common, "Restaurant" as most common and "Cafe" as second common, and other).

5. Discussion

- With K-Means clustering, neighborhoods with "Restaurant" as most common and "FastFood" as second common neighborhoods have been assigned to all three clusters (e.g., Carnegie Hill, Chelsea, and Clinton, respectively).
- Automatic K-Means classification also considered the total venue count, and the exact ratio between different venues, not only the popularity order, offering a better segmentation.

6. Conclusion

- The results when using the naïve segmentation based on the most common venue type and the results from K-Means segmentation are not the same
- This a good example for the usability of machine learning techniques, which consider aspects that can be ignored by one's intuition.
- In this case, the examples are relatively basic, but the benefits will increase with more complex data.