

BATTLE of NEIGHBORHOODS

Applied Data Science Capstone

IBM Data Science Professional Certificate

Beno Wajsfeld 02-Nov-2020

Battle of Neighborhoods - OBJECTIVE

- ▶ **INTRODUCTION:** Decision to open a business requires a lot research, information in order to build a business plan and try to predict possible risks.
- ▶ **OBJECTIVE:** Open a business in Manhattan or Toronto
 - ▶ 1. What kind o business.
 - ▶ 2. Where and which neighborhood.

Battle of Neighborhoods

DOWNLOAD DATA and EXPLORE DATASET

- ▶ In order to make a more precise decision on which business and where, the following procedure was followed:
- ▶ 1. Build a dataset containing all neighborhoods, their latitude and longitude.
 - ▶ Manhattan: newyork_data.Json
 - ▶ Table in wikipedia with postcode, borough and neighborhood
- ▶ 2. Using api foursquare a dataframe for each city was loaded with maximum 100 venues visited in a radius of 500meters of each neighborhood.
- ▶ 3. Each of those venues, their latitude and longitude were determined and again using api Foursquare all categories of all venues were also saved on a dataframe.

Battle of Neighborhoods

ANALYZING EACH NEIGHBORHOOD

- 4. A dataframe with neighborhood and all venues grouped and categorized will be used to and a mean of the frequency of occurrence of each category was calculated.

	Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports
0	Battery Park City	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.015152
1	Carnegie Hill	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.011236	0.000000	0.011236	0.000000	0.000000	0.000000
2	Central Harlem	0.000000	0.00	0.066667	0.044444	0.00	0.000000	0.000000	0.022222	0.000000	0.000000	0.000000	0.000000
3	Chelsea	0.000000	0.00	0.000000	0.040000	0.00	0.010000	0.000000	0.050000	0.000000	0.000000	0.010000	0.000000

- 5. A new dataframe was created with the 10 biggest values meaning visited categorie.

	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	Hotel	Coffee Shop	Gym	Memorial Site	Shopping Mall	Food Court	Clothing Store	Gourmet Shop	Boat or Ferry
1	Carnegie Hill	Coffee Shop	Café	Yoga Studio	Wine Shop	Italian Restaurant	Gym / Fitness Center	Gym	French Restaurant	Pizza Place	Bookstore

Battle of Neighborhoods

CLUSTERING NEIGHBORHOODS

- 6. Using CLUSTERING operation by applying KMEANS we divided all neighborhoods in clusters (Cluster Labels) based on their 10 most categories visited.



Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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
East Toronto	The Beaches	43.676357	-79.293031	3	Pub	Health Food Store	Trail	Yoga Studio	Distribution Center	Dessert Shop	Dim Sum Restaurant	Diner
East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	1	Greek Restaurant	Coffee Shop	Italian Restaurant	Restaurant	Furniture / Home Store	Bookstore	Ice Cream Shop	Pub
East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572	1	Park	Sushi Restaurant	Liquor Store	Steakhouse	Fish & Chips Shop	Pub	Restaurant	Italian Restaurant

Battle of Neighborhoods

EXAMINE CLUSTERS and RESULTS

- ▶ 7. The clusters were analyzed and for each cluster was observed the categories in common in order to classify them.
- ▶ Results: Showing TORONTO but also applied to Manhattan.

- CLUSTER 0:
 - MOST: Nature: Park, Trail, playground
 - FEW: Stores and restaurants
- CLUSTER 1:
 - MOST: Café, coffee shop, Restaurants (Thai, Italian , Mexican, Vietnamese)
 - FEW: Bar.
- CLUSTER 2:
 - MOST: Sandwich café, café, bakery (Quick meals)
 - FEW but on all neighborhoods: Stores, shops and restaurants
- CLUSTER 3:
 - MOST: Nature and healthy food activities (yoga)



Battle of Neighborhoods

DISCUSSION

- Now is time to choose the business and in which neighborhood of which city, Manhattan or Toronto.
- One thing I know, I would like to open a café/coffee shop or restaurant
- The big question is where. In order to help me find the answer I collected more data and built these summary tables :

TOTONTO				
CLUSTER	Population	Area (mile2)	Density (people/mile2)	AVG income CAD
0	53842	9.87	5455.20	127554.18
1	354784	45.69	7764.96	65372.71
2	84096	8.58	9802.50	63982.77
3	47590	8.86	5371.54	101943.09
4	29249	2.06	14210.00	85991.00
TOTAL:	569561	75.06	7588.35	75160.13
Total of venues collected:	1624	39* neighborhoods	Total of venues considered in clusters:	814

MANHATTAN				
CLUSTER	Population	Area (mile2)	Density (people/mile2)	AVG income US\$
0	749300	8.31	90179.66	118676.79
1	211730	15.79	13408.75	78938.21
2	373280	4.03	92702.50	50010.86
3	16540	0.16	104360.00	96200.00
4	40940	0.60	68240.00	82300.00
5	17110	0.70	24350.00	126800.00
6	118860	1.56	76060.00	37500.00
TOTAL:	1527760	31.15	49045.47	88949.49
Total of venues collected:	3210	40 neighborhoods	Total of venues considered in clusters:	1212

Battle of Neighborhoods

DISCUSSION continuation...

- ▶ Those table can give more information about the cluster and help me to make a better and less risky decision.
- ▶ They can answer:
 - ▶ Which neighborhood has the highest median income?
 - ▶ Which one has the highest area? I could calculate venue per miles² and check which neighborhood could accept another venue without saturating the neighborhood. Which neighborhood is lacking a kind of business?.
 - ▶ Which neighborhood has the highest people density people/miles²? It means more customers.
 - ▶ I can create new indicators and compare businesses and neighborhood that could help to check if the business I have chosen as café/coffee shops (I love coffee) and restaurants bring me the best indicators.

Battle of Neighborhoods

CONCLUSION

- ▶ I know I want a café/coffee shop or restaurant.
- ▶ The next step is to analyze deeper the extra data I collected to build the summary tables and then finally choose city and neighborhood.
- ▶ Clustering was really helpful to classify venues, categories and finally divide all venues in clusters with same characteristics.