Helping small business to define where to start in a city

Jorge Cupi
February 2020

Business Problem

Small business such as restaurants, coffee shops, among others might want to know how desired they can be in a new neighborhood if they decide to open a new place.
So, the questions are:
- If someone is looking to open a restaurant, where would you recommend that they open it based on similar venues around the neighborhood?
- Can we use Foursquare to map a list of top venues within a city's different neighborhoods?

If that is possible, then, for sure a restaurant will take a decision to open a new spot at a specific list of neighborhoods based on the number of competitors.

Data

We'll use Foursquare's API to get a list of top venues of each borough (localidad) in Bogota, Colombia.
A list of Bogota's boroughs extracted from Wikipedia that also includes population and population density that can help to improve a decision.

Nº	Localidad	Códigos Postales	Superficie km²[2]	Población[3]	Densidad hab/km²
01	Usaquén	110111-110151	65.31	501 999	7 686.4
02	Chapinero	110211-110231	38.15	139 701	3 661.88
03	Santa Fe	110311-110321	45.17	110 048	2 436.3
04	San Cristóbal	110411-110441	49.09	404 697	8 243.98
05	Usme	110511-110571	215.06	457 302	2 126.39
06	Tunjuelito	110611-110621	9.91	199 430	20 124.11
07	Bosa	110711-110741	23.93	673 077	28 126.91
08	Kennedy	110811-110881	38.59	1 088 443	28 205.31
09	Fontibón	110911-110931	33.28	394 648	11 858.41
10	Engativá	111011-111071	35.88	887 080	24 723.52
11	Suba	111111-111176	100.56	1 218 513	12 117.27
12	Barrios Unidos	111211-111221	11.9	243 465	20 459.24
13	Teusaquillo	111311-111321	14.19	153 025	10 784
14	Los Mártires	111411	6.51	99 119	15 225.65
15	Antonio Nariño	111511	4.88	109 176	22 372.12
16	Puente Aranda	111611-111631	17.31	258 287	14 921.25
17	La Candelaria	111711	2.06	24 088	11 693.2
18	Rafael Uribe Uribe	111811-111841	13.83	374 246	27 060.44
19	Ciudad Bolívar	111911-111981	130	707 569	5 442.83
20	Sumapaz	112011-112041	780.96	6 531	8.36

Data cleaning

We'll proceed to clean the data with the next steps:

- Change of column names and boroughs
- Change of column types
- Splitting the postal codes to later have a list of postal codes per borough

	Borough	SurfaceInKm2	Population	PopulationDensityPerKm2	FirstPostalCode	LastPostalCode
0	Usaquen	65.31	501999	7686.40	110111	110151.0
1	Chapinero	38.15	139701	3661.88	110211	110231.0
2	SantaFe	45.17	110048	2436.30	110311	110321.0
3	SanCristobal	49.09	404697	8243.98	110411	110441.0
4	Usme	215.06	457302	2126.39	110511	110571.0
5	Tunjuelito	9.91	199430	20124.11	110611	110621.0
6	Bosa	23.93	673077	28126.91	110711	110741.0
7	Kennedy	38.59	1088443	28205.31	110811	110881.0
8	Fontibon	33.28	394648	11858.41	110911	110931.0
9	Engativa	35.88	887080	24723.52	111011	111071.0
10	Suba	100.56	1218513	12117.27	111111	111176.0
11	BarriosUnidos	11.90	243465	20459.24	111211	111221.0
12	Teusaquillo	14.19	153025	10784.00	111311	111321.0
13	LosMartires	6.51	99119	15225.65	111411	NaN
14	AntonioNarino	4.88	109176	22372.12	111511	NaN
15	PuenteAranda	17.31	258287	14921.25	111611	111631.0
16	LaCandelaria	2.06	24088	11693.20	111711	NaN
17	RafaelUribeUribe	13.83	374246	27060.44	111811	111841.0
18	CiudadBolivar	130.00	707569	5442.83	111911	111981.0
19	Sumapaz	780.96	6531	8.36	112011	112041.0

Data analysis

Based on the population density, we chose to work only with the Kennedy borough because it has more than 1 million people in the least squared kilometers space across all our different boroughs.

It's important to note that we are taking this approach solely based on population. On a real-world scenario, a data scientist would also talk with the restaurant owners to also take a look at other data per borough such as:

- Average income level
- Housing prices

- Education level
- Population overall
- Dimension of borough in Km2
-

Methodology

- After choosing Kennedy borough, we used the Geocode python library to download the coordinates of each PostalCode from Kennedy.

	PostalCode	Latitude	Longitude
0	110811	4.6544	-74.1540
1	110821	4.6397	-74.1426
2	110831	4.6292	-74.1301
3	110841	4.6075	-74.1458
4	110851	4.6260	-74.1570
5	110861	4.6144	-74.1718
6	110871	4.6455	-74.1675
7	110881	4.6388	-74.1757

- Then, we used Foursquare API to download a list of venues per PostalCode

	PostalCode	PostalCode Latitude	PostalCode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	110811	4.6544	-74.154	Homecenter y Constructor Tintal	4.651588	-74.146031	Hardware Store
1	110811	4.6544	-74.154	Pista De Patinaje El Tintal	4.652716	-74.160928	Skating Rink
2	110811	4.6544	-74.154	Fitness Sports GYM	4.648465	-74.153374	Gym
3	110811	4.6544	-74.154	Centro Comercial Ciudad Tintal	4.652983	-74.159557	Shopping Mall
4	110811	4.6544	-74.154	Merkacol	4.648410	-74.147728	Fruit & Vegetable Store

- Then, to analyze the frequency we applied one-hot encoding to the list of venue types:

	PostalCode	Airport	Arcade	Arepa Restaurant	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Basketball Court	Breakfast Spot	...	Skating Rink	Soccer Field	Soccer Stadium	South American Restaurant	Spa	Spor Go S
0	110811	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
1	110811	0	0	0	0	0	0	0	0	0	...	1	0	0	0	0	
2	110811	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
3	110811	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
4	110811	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	

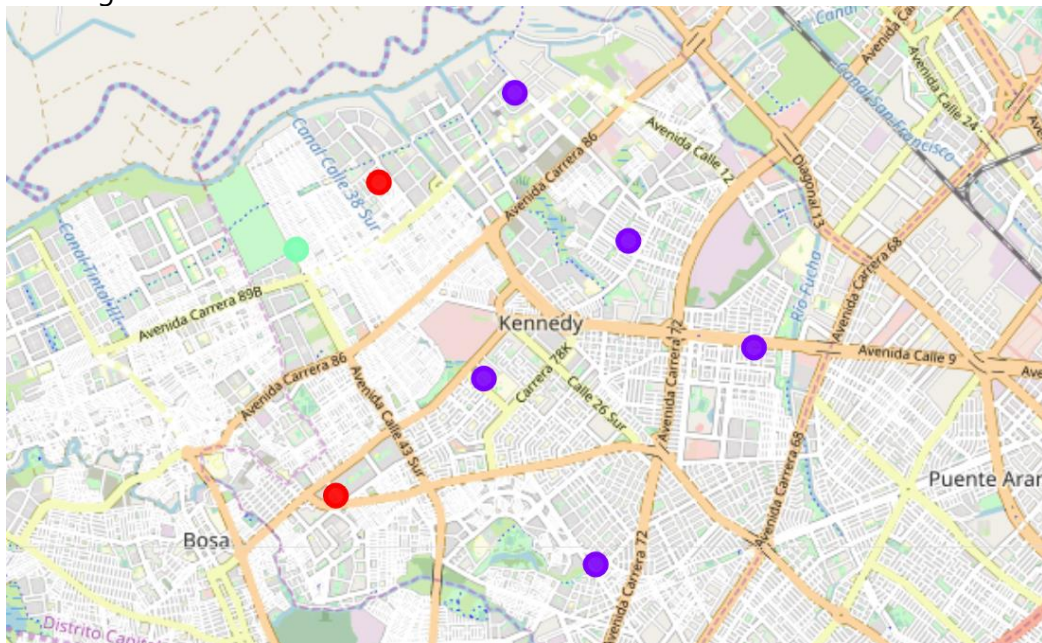|	PostalCode	Airport	Arcade	Arepa Restaurant	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Basketball Court	Breakfast Spot	...	Skating Rink	Soccer Field	Soccer Stadium	South American Restaurant
0	110811	0.000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.166667	0.000000	0.000000	0.000000	0.
1	110821	0.000	0.00	0.000000	0.066667	0.000000	0.133333	0.066667	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.
2	110831	0.000	0.00	0.000000	0.000000	0.000000	0.052632	0.000000	0.052632	0.052632	...	0.000000	0.052632	0.052632	0.052632	0.
3	110841	0.000	0.00	0.000000	0.000000	0.166667	0.166667	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.
4	110851	0.000	0.00	0.052632	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.052632	0.000000	0.000000	0.
5	110861	0.125	0.00	0.000000	0.000000	0.000000	0.125000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.
6	110871	0.000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.
7	110881	0.000	0.25	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.

- After getting the frequency per venue type we were able to take a look at the top 10 venue types per PostalCode:

	PostalCode	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	110811	Hardware Store	Skating Rink	Market	Gym	Shopping Mall	Fruit & Vegetable Store	BBQ Joint	Bakery	Fried Chicken Joint	Fast Food Restaurant
1	110821	Pizza Place	Fast Food Restaurant	BBQ Joint	Gym	Gymnastics Gym	Asian Restaurant	Bakery	Brewery	Seafood Restaurant	Pub
2	110831	Fast Food Restaurant	Pizza Place	Burger Joint	Breakfast Spot	Basketball Court	Wings Joint	Park	Coffee Shop	BBQ Joint	Restaurant
3	110841	Restaurant	Sporting Goods Shop	Athletics & Sports	BBQ Joint	Construction & Landscaping	Park	Coffee Shop	Fried Chicken Joint	Fast Food Restaurant	Farmers Market
4	110851	Restaurant	Farmers Market	Park	Pizza Place	Burger Joint	Gym	Discount Store	Department Store	Historic Site	Ice Cream Shop
5	110861	Park	Airport	Supermarket	BBQ Joint	Convenience Store	Mobile Phone Shop	Shopping Mall	Coffee Shop	Fast Food Restaurant	Farmers Market
6	110871	Supermarket	Shopping Mall	Park	Women's Store	Burger Joint	Fried Chicken Joint	Fast Food Restaurant	Farmers Market	Discount Store	Department Store
7	110881	Women's Store	Arcade	Movie Theater	Park	Coffee Shop	Fried Chicken Joint	Fast Food Restaurant	Farmers Market	Discount Store	Department Store

- And finally applied K-Means clustering to group our PostalCodes by affinity:

	PostalCode	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	9th Most Common Venue	10th Most Common Venue
0	110811	4.6544	-74.1540	1	Hardware Store	Skating Rink	Market	Gym	Shopping Mall	Fruit & Vegetable Store	BBQ Joint	Bakery	Fried Chicken Joint	Fast Food Restaurant
1	110821	4.6397	-74.1426	1	Pizza Place	Fast Food Restaurant	BBQ Joint	Gym	Gymnastics Gym	Asian Restaurant	Bakery	Brewery	Seafood Restaurant	Pub
2	110831	4.6292	-74.1301	1	Fast Food Restaurant	Pizza Place	Burger Joint	Breakfast Spot	Basketball Court	Wings Joint	Park	Coffee Shop	BBQ Joint	Restaurant
3	110841	4.6075	-74.1458	1	Restaurant	Sporting Goods Shop	Athletics & Sports	BBQ Joint	Construction & Landscaping	Park	Coffee Shop	Fried Chicken Joint	Fast Food Restaurant	Farmers Market
4	110851	4.6260	-74.1570	1	Restaurant	Farmers Market	Park	Pizza Place	Burger Joint	Gym	Discount Store	Department Store	Historic Site	Ice Cream Shop

- An image of the clusters formed:

Results

We got some results for each postal code!

- 110811 does look like has a mall, a market, a gym but no restaurants, whatsoever!
- 110821 has pizza places, BBQ joints and overall fast food restaurants
- 110831 has a stadium and a park, and while it has some fast food restaurants. It does look like it could use some variety, specially considering the ammount of people nearby parks or stadiums on any weekend
- 110841 Looks a place in process of rebuildings and some sporting stores could be a hint that some healthy food restaurants could atract some customers
- 110851 with a parks, farmers markets and a soccer field is also giving us hints that some healthy food restaurants are missing in the neighbourhood. While there are general restaurants and some pizza places. There's always room for some healthy competition
- 110861 with an airport at the place and mobile shops(?) being in the top 5 venues gives the appearance to be a more industrial neighbourhood. When there are workers around, some fast food restaurants are always welcomed!
- 110871 has no restaurants! While only 3 top venues are identified in the form of supermarkets, shopping malls and parks. This is THE hint a restaurant owner is looking for.
- 110881 looks like an interesting place to open a restaurant or a fast food restaurant, since the top venues are women's stores, movie theaters, arcade zones and parks.

While we tried to get 5 clusters, it seems reasonable that, for 8 neighborhoods, a total of 3 clusters were formed:

- Cluster One: 110861 and 110871 PostalCodes
- Cluster Two: 110881 PostalCode
- Cluster Three: 110811, 110821, 110831, 110841 and 110851 PostalCodes.

Cluster One

Looks like an all industrial and go to work cluster, not a lot of restaurants but plenty of supermarkets, mobile shops, parks and mobile shops. And with an airport at your neighbourhood it might be an option to open some fancy places that, at the same time deliver food in fast time.

Cluster Two

Looks like it could be merge into Cluster One since it has more stores, movie theathers and arcade zones.

From a restaurant owner perspective, it makes sense to be a different cluster because this looks more like a residential zone rather than an industrial one. So, family-owned type of restaurants with places for the kids to play look like a potential idea. Also, since there seems to be little working activities, chances are the customers of the restaurants will have more time to spare between the arcades, parks and stadiums around the area.

Cluster Three

Had stadiums, sports, parks, gyms but no good enough amount restaurants apart from PostalCode 110821. Interesting enough, 110821 is the 'heart' of this cluster. So this can be a message to our restaurant owners:

Bring competition to 110821 neighborhood or start creating smaller restaurants around the cluster

Recall this looks like a 'healthy' cluster? Consider that when you start building your menus. Also, note that there are some places like stadiums and parks, food-to-go options are a must around those places. So, convenience and fast service will be key to win customers around this cluster

Conclusion

From our findings, we were able to help with a guidance to local restaurant owners in Bogota, Colombia. We only looked at Kennedy borough, but we can easily analyze the rest of the boroughs in Bogota to help our customers to expand their restaurants.

It's worth nothing again that we made the decision to analyze Kennedy solely thinking on density population, there are a lot of extra factors that a restaurant owner should take a look at.