

Coursera - Capstone Project – Final Report

Mauricio F. O. Munhoz – March 04th, 2020

Introduction

This project was inspired by the requirements of week 4 of the Course Applied Data Science Capstone. It presents an analysis of the New York region, based on data from the Foursquare API. The main objective is to find nice places for a new brazilian restaurant in New York.

Business problem

A brazilian stakeholder wants to start his own restaurant brand and as such has chosen New York City as his prime location for his first business outside Brazil. The location has its own set of risks. His requirements for the new restaurant is a safe place and it needs to be family friendly, as that is at the core of his success in Brazil. The cuisine he specializes in is a unique mixture of Brazilian classics such as barbecue, fried fish, feijoada etc., infused with recipes derived from local demographics. In other words, he opens his restaurants in areas with a high number of immigrants and chooses to give the local dishes an extra kick with select herbs and spices which represent the fabric of local communities.

The restaurant reflects the culinary diversity of Brazil: it's cuisine, festive culture and hospitality. It needs to be spacious and inviting, featuring a large bar, and two private parties rooms that accommodate up to 100 guests. Also the menu features includes steaks, exotic salads, vegetarian pastas, seafood, brazilian barbecue and more.

Problem formulation

As mentioned before, the location has to be in a safe area. Ideally, it would be on an intersection as that would give his restaurant the most visibility. However, the intersection should be in a borough with the least amount of vehicle collisions. Once the location has been deemed to be safe, an overview of the existing restaurants, and more importantly, the cuisines they cater to will be completed such that a brazilian restaurant can formulate it's own recipes which will make his restaurant a success story.

Data source and usage

The predominant data source comes from the NYC Open Data page and can be found in their official site. Vehicle collisions range from 2012 to 2019 and list the location coordinates (longitude and latitude) as well as the contributing factors from both vehicles in the collision.

- 1) The data will be cleaned, analyzed and presented as a chart which shows the borough with the least amount of collisions;
- 2) The contributing factors will be presented to have an overview of the factors responsible for the collisions;

- 3) Time permitting, an optional analysis would be conducted on the biggest contributing factors to get an in-depth look in the fabric of societal, social and economic issues which were responsible for the contributing factors.

The second source of data is also freely available from NYU. It gives the names of neighborhoods in New York City and also can be found on their site.

Foursquare location data about restaurants in neighborhoods will be essential to determine which kind of business already exist in different boroughs. This will be used by the stakeholder to make new recipes and select the place to set up his new business.

Methodology and idea development

Data source: acquiring data

1) Safety requirements: New York City traffic collision data

The traffic collision data was accessed to determine the safest neighborhood in terms of collisions. It's a way to search for a safe place.

The traffic collision data was accessed from the JSON API endpoint, since the downloaded file was quite large (1.4M+ rows in a csv file). However, it was observed that accessing the data from the API only gave a 1000 entries, which was quite a bit lower than expected. Nevertheless, the dataset was imported and some initial analysis was conducted to find out the type of data present, as shown below.

Table 1 - New York City Collision Data

```
Data columns (total 29 columns)
borough
collision_id
contributing_factor_vehicle_1
contributing_factor_vehicle_2
contributing_factor_vehicle_3
contributing_factor_vehicle_4
contributing_factor_vehicle_5
crash_date
crash_time
cross_street_name
latitude
location
longitude
number_of_cyclist_injured
number_of_cyclist_killed
number_of_motorist_injured
number_of_motorist_killed
number_of_pedestrians_injured
number_of_pedestrians_killed
number_of_persons_injured
number_of_persons_killed
off_street_name
on_street_name
vehicle_type_code1
vehicle_type_code2
vehicle_type_code_3
vehicle_type_code_4
vehicle_type_code_5
zip code
```

After further cleaning (dropping NaN rows, isolating the boroughs and contributing factors, it was discovered that Staten Island had the lowest number of collisions, which was 18, when looking at the first vehicle for accidents. This was substantially lower than the other boroughs, as can be seen in the comparison below.

Table 2 - Number of accidents in Staten Island (contributing factors)

STATEN ISLAND	Backing Unsafely	2
	Driver Inattention/Distracted	3
	Failure to Yield Right-of-Way	1
	Following Too Closely	6
	Other Vehicular	1
	Passing Too Closely	2
	Traffic Control Disregarded	1
	Unsafe Lane Changing	1
	Unspecified	1

Thus, Staten Island was chosen as the safest borough for the Brazilian new restaurant since it had the least number of collisions.

Table 3 - Number of accidents in various boroughs

BROOKLYN	204
QUEENS	177
MANHATTAN	147
BRONX	98
STATEN ISLAND	18
Name: borough, dtype: int64	

Five were unspecified whereas five were from backing the vehicle. All in all, with the data at hand, Staten Island appears to be a safer borough when compared with others.

2) New York City borough data

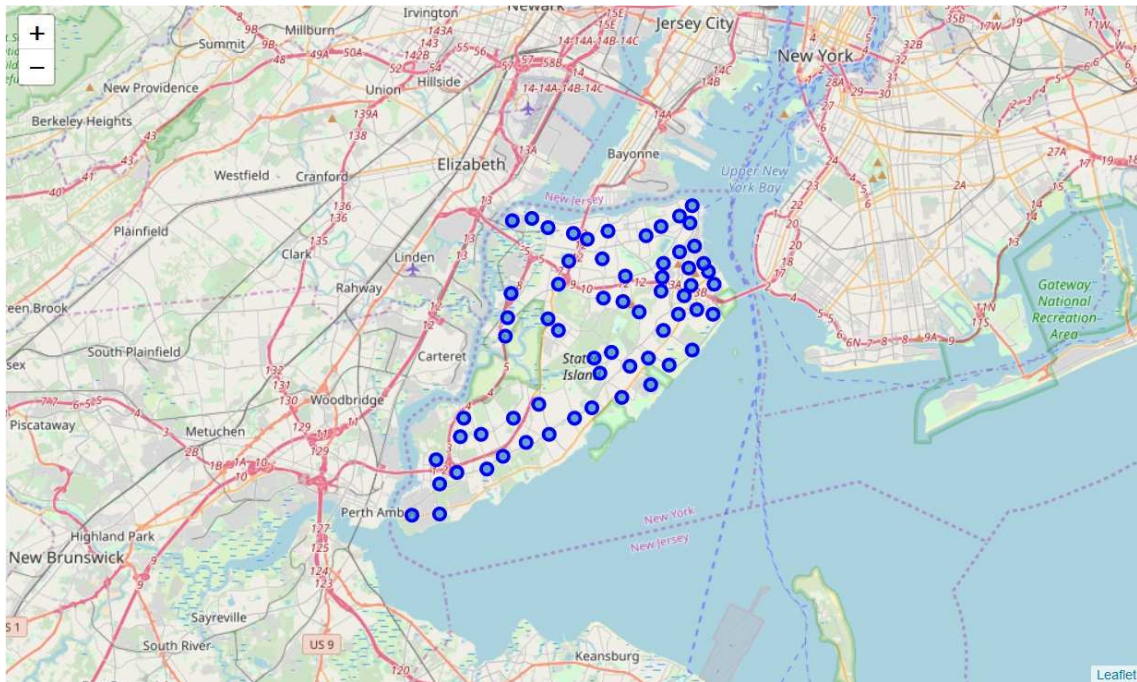
The New York City borough data was also in JSON, and further examination revealed that the various boroughs and neighborhoods were under the “feature” tag in the data, as shown below.

Table 4 - Features of boroughs and neighborhoods (JSON data)

```
{'type': 'FeatureCollection',
'totalFeatures': 306,
'features': [{ 'type': 'Feature',
'id': 'nyu_2451_34572.1',
'geometry': {'type': 'Point',
'coordinates': [-73.84720052054902, 40.89470517661]},
'geometry_name': 'geom',
'properties': {'name': 'Wakefield',
'stacked': 1,
'annoline1': 'Wakefield',
'annoline2': None,
'annoline3': None,
'annoangle': 0.0,
'borough': 'Bronx',
'bbox': [-73.84720052054902,
40.89470517661,
-73.84720052054902,
40.89470517661]}},
{'type': 'Feature',
'id': 'nyu_2451_34572.2',
'geometry': {'type': 'Point',
'coordinates': [-73.82993910812398, 40.87429419303012]},
'geometry_name': 'geom',
'properties': {'name': 'Co-op City',
'stacked': 2,
'annoline1': 'Co-op',
'annoline2': 'City',
'annoline3': None,
'annoangle': 0.0,
```

The dataset was subsequently cleaned and analyzed until all the neighborhoods in Staten Island were isolated, as shown below.

Map 1 - Neighborhoods in Staten Island



Once this was done, we could move on to using Foursquare API and looking at the most popular venues in Staten Island.

Exploring Neighborhoods in Staten Island

Now that the data has been acquired and cleaned, the next step is to explore the venues already present in Staten Island, which will help point us in the right direction for the Brazilian restaurant. The number of venues for some neighborhoods is shown below.

Table 6 - Venues in various neighborhoods around Staten Island

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Annadale	10	10	10	10	10	10
Arden Heights	5	5	5	5	5	5
Arlington	6	6	6	6	6	6
Arrochar	19	19	19	19	19	19
Bay Terrace	10	10	10	10	10	10
Bloomfield	5	5	5	5	5	5
Bulls Head	45	45	45	45	45	45
Butler Manor	5	5	5	5	5	5
Castleton Corners	13	13	13	13	13	13
Charleston	30	30	30	30	30	30
Chelsea	7	7	7	7	7	7
Clifton	21	21	21	21	21	21
Concord	10	10	10	10	10	10
Dongan Hills	24	24	24	24	24	24
Egbertville	5	5	5	5	5	5
Elm Park	8	8	8	8	8	8

If we just look at the unique values, we get about 170 unique categories. Further analysis gives the top ten venues in terms of popularity in each neighborhood, as shown below.

Table 7 - Most popular venues in different Staten Island neighborhoods

	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	Annadale	American Restaurant	Pizza Place	Sports Bar	Liquor Store	Diner	Restaurant	Dance Studio	Train Station	Asian Restaurant	Falafel Restaurant
1	Arden Heights	Deli / Bodega	Home Service	Pizza Place	Coffee Shop	Pharmacy	Dog Run	Event Space	Food	Flower Shop	Fish & C Shop
2	Arlington	Bus Stop	American Restaurant	Coffee Shop	Business Service	Deli / Bodega	Hotel	Event Service	Food	Flower Shop	Fish & C Shop
3	Arrochar	Deli / Bodega	Italian Restaurant	Pizza Place	Bus Stop	Hotel	Mediterranean Restaurant	Cosmetics Shop	Sandwich Place	Outdoors & Recreation	Food Tru
4	Bay Terrace	Supermarket	Insurance Office	Salon / Barbershop	Sushi Restaurant	Donut Shop	Liquor Store	Shipping Store	Plaza	Italian Restaurant	Discount Store
5	Bloomfield	Recreation Center	Discount Store	Bus Stop	Theme Park	Park	Diner	Falafel Restaurant	Food	Dessert Shop	Flower S
6	Bulls Head	Pizza Place	Bus Stop	Chinese Restaurant	Pharmacy	Ice Cream Shop	Deli / Bodega	Food Truck	Japanese Restaurant	Grocery Store	Mobile P Shop
7	Butler Manor	Pool	Baseball Field	Convenience Store	Yoga Studio	Event Space	Food & Drink Shop	Food	Flower Shop	Fish & Chips Shop	Filipino Restaurant
8	Castleton Corners	Pizza Place	Japanese Restaurant	Bagel Shop	Mini Golf	Tattoo Parlor	Grocery Store	Bank	Sandwich Place	Park	Burger J
9	Charleston	Big Box Store	Cosmetics Shop	Japanese Restaurant	Gift Shop	Scenic Lookout	Pizza Place	Pet Store	Music Venue	Liquor Store	Kids Sto
10	Chelsea	Steakhouse	Italian Restaurant	Bus Stop	Spanish Restaurant	Sandwich Place	Park	Theater	Event Service	Flower Shop	Fish & C Shop

Clustering the Venues

Table 8 - Clustered places of interest (POI) in Staten Island

	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
0	Staten Island	St. George	40.644982	-74.079353	3.0	Clothing Store	Sporting Goods Shop	Bar	Deli / Bodega	Italian Restaurant	Pharmacy	Burger Joint	Monu / Landn
1	Staten Island	New Brighton	40.640615	-74.087017	1.0	Bus Stop	Deli / Bodega	Park	Daycare	Bowling Alley	Discount Store	Playground	Falafe Resta
2	Staten Island	Stapleton	40.626928	-74.077902	3.0	Discount Store	Café	Pizza Place	Mexican Restaurant	Sandwich Place	Bank	Restaurant	Motor Shop
3	Staten Island	Rosebank	40.615305	-74.069805	3.0	Pizza Place	Grocery Store	Italian Restaurant	Electronics Store	Filipino Restaurant	Sandwich Place	Breakfast Spot	Storaç Facilit
4	Staten Island	West Brighton	40.631879	-74.107182	3.0	Coffee Shop	Pharmacy	Music Store	Bank	Bar	Breakfast Spot	Italian Restaurant	Cosm Shop

Overview of Clusters

Table 9 – Neighborhood cluster 1

	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
18	Annadale	American Restaurant	Pizza Place	Sports Bar	Liquor Store	Diner	Restaurant	Dance Studio	Train Station	Asian Restaurant	Falafel Restaurant
22	Silver Lake	American Restaurant	Dance Studio	Burger Joint	Dog Run	Golf Course	Falafel Restaurant	Food & Drink Shop	Food	Flower Shop	Fish & Chips Shop

Table 10 – Neighborhood cluster 2

	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
1	New Brighton	Bus Stop	Deli / Bodega	Park	Daycare	Bowling Alley	Discount Store	Playground	Falafel Restaurant	Food	Flower Shop
5	Grymes Hill	Bus Stop	Dog Run	Yoga Studio	Event Space	Food & Drink Shop	Food	Flower Shop	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant
27	Arlington	Bus Stop	American Restaurant	Coffee Shop	Business Service	Deli / Bodega	Hotel	Event Service	Food	Flower Shop	Fish & Chips Shop
52	Randall Manor	Bus Stop	Bagel Shop	Carpet Store	Yoga Studio	Falafel Restaurant	Food & Drink Shop	Food	Flower Shop	Fish & Chips Shop	Filipino Restaurant
56	Willowbrook	Bus Stop	Deli / Bodega	Spa	Pizza Place	Bagel Shop	Event Space	Food	Flower Shop	Fish & Chips Shop	Filipino Restaurant

Table 11 – Neighborhood cluster 3

	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
51	Emerson Hill	Food	Historic Site	Yoga Studio	Event Service	Food & Drink Shop	Flower Shop	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market

Table 12 – Neighborhood cluster 4

	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 Common Venue
0	St. George	Clothing Store	Sporting Goods Shop	Bar	Deli / Bodega	Italian Restaurant	Pharmacy	Burger Joint	Monument / Landmark	Farmers Market	Scenic Lookout
2	Stapleton	Discount Store	Café	Pizza Place	Mexican Restaurant	Sandwich Place	Bank	Restaurant	Motorcycle Shop	Skate Park	Spanish Restaurant
3	Rosebank	Pizza Place	Grocery Store	Italian Restaurant	Electronics Store	Filipino Restaurant	Sandwich Place	Breakfast Spot	Storage Facility	Restaurant	Martial A Dojo
4	West Brighton	Coffee Shop	Pharmacy	Music Store	Bank	Bar	Breakfast Spot	Italian Restaurant	Cosmetics Shop	Fast Food Restaurant	Mexican Restaurant
7	South Beach	Beach	Pier	Deli / Bodega	Athletics & Sports	Falafel Restaurant	Food Truck	Food & Drink Shop	Food	Flower Shop	Fish & C Shop
8	Port Richmond	Rental Car Location	Sports Bar	Playground	Donut Shop	Park	Yoga Studio	Event Service	Flower Shop	Fish & Chips Shop	Filipino Restaurant
9	Mariner's Harbor	Deli / Bodega	Italian Restaurant	Bus Stop	Other Repair Shop	Insurance Office	Event Space	Food	Flower Shop	Fish & Chips Shop	Filipino Restaurant
11	Castleton Corners	Pizza Place	Japanese Restaurant	Bagel Shop	Mini Golf	Tattoo Parlor	Grocery Store	Bank	Sandwich Place	Park	Burger J
12	New Springville	Pizza Place	Deli / Bodega	Health & Beauty Service	Coffee Shop	Mexican Restaurant	Pharmacy	Hookah Bar	Donut Shop	Restaurant	Sandwich Place
13	Travis	Hotel	Deli / Bodega	Bowling Alley	Comedy Club	Spanish Restaurant	Park	Gym	Gym / Fitness Center	Baseball Field	Sports C

Table 13 – Neighborhood cluster 5

	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
6	Todt Hill	Park	Yoga Studio	Event Space	Food & Drink Shop	Food	Flower Shop	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market

Discussion and Conclusion

Looking at the clusters shown in the previous section, it appears that clusters 2 (Table 10) and 4 (Table 12) have more distinct results when compared with other clusters. Cluster 2 has a clear bias towards Eastern European Food whereas cluster 4 has some neighborhoods which prefer Italian food. This is more of a guideline, as existing preference for one cuisine does not automatically rule out the possibility of another cuisine flourishing. If anything, this result could be counter-intuitive in that something radically different to the most common venues could also succeed.

To sum up, at first place, the best location inside NY to open our new brazilian restaurant is Staten Island. There are the lowest car accident level, a very good indicator.

Second place, the best cluster to choose a neighborhood is the cluster 4, because is the largest neighborhood, a true gastronomic hub! Probably, a lot of people go around there looking for lunch or dinner.