Capstone Project - The Battle of the Neighbourhoods (Week 1)

Asad Raja 17 January 2019

Introduction and Business Problem

Background:

A prolific Swedish restaurateur, Torbjorn Magnusson, wants to expand his business outside his local market. Torbjorn wants to expand his own brand and as such has chosen New York City as his prime location for his first restaurant overseas. However, the location within the city poses its own set of risks. Specifically, Torbjorn wants a location within New York which is safe and family friendly, as that is at the core of his success in Sweden. The cuisine he specialises in is a unique mixture of Swedish classics such as meatballs, salmon 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. This makes the recipes extremely popular as they appeal to everyone's taste.

Problem Formulation:

The location of Torbjorn's restaurant 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 Torbjorn can formulate the exact recipes which will make his restaurant a success story.

Data Sources and usage

The predominant data source comes from the NYC Open Data page and can be found below:
https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95
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, analysed and presented as a chart which shows the borough with the least amount of collisions.

2) The contributing factors will be analysed to get an overview of areas with the least amount of 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 can be found below:

https://geo.nyu.edu/catalog/nyu_2451_34572
Foursquare location data about restaurants in neighborhoods will be essential to determine which cuisines already exist in different boroughs. This will be used by Torbjorn to determine the most popular restaurants in neighborhoods which, in turn, will let him know how to target the clientèle.

Methodology

Acquiring Data

1) New York City Traffic Collision Data

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.

```
1 df_ny.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 34 columns):
:@computed_region_92fq_4b7q      926 non-null float64
:@computed_region_efsh_h5xi      920 non-null float64
:@computed_region_f5dn_yrer      926 non-null float64
:@computed_region_sbqj_enih      926 non-null float64
:@computed_region_yeji_bk3q      926 non-null float64
borough                          622 non-null object
contributing_factor_vehicle_1    996 non-null object
contributing_factor_vehicle_2    831 non-null object
contributing_factor_vehicle_3    61 non-null object
contributing_factor_vehicle_4    11 non-null object
contributing_factor_vehicle_5    4 non-null object
cross_street_name                243 non-null object
date                             1000 non-null datetime64[ns]
latitude                         928 non-null float64
location                         928 non-null object
longitude                        928 non-null float64
number_of_cyclist_injured        1000 non-null int64
number_of_cyclist_killed         1000 non-null int64
number_of_motorist_injured       1000 non-null int64
number_of_motorist_killed        1000 non-null int64
number_of_pedestrians_injured    1000 non-null int64
number_of_pedestrians_killed     1000 non-null int64
number_of_persons_injured        1000 non-null int64
number_of_persons_killed         1000 non-null int64
off_street_name                  474 non-null object
on_street_name                   756 non-null object
time                             1000 non-null object
unique_key                       1000 non-null int64
vehicle_type_code1               993 non-null object
vehicle_type_code2               796 non-null object
vehicle_type_code_3              60 non-null object
vehicle_type_code_4              11 non-null object
vehicle_type_code_5              4 non-null object
zip_code                         621 non-null float64
dtypes: datetime64[ns](1), float64(8), int64(9), object(16)
memory usage: 265.7+ KB
```

```
1 df_ny.shape
```

```
(1000, 34)
```

Table 1. New York City Collision Data.

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 21, when looking at the first vehicle for accidents. This was substantially lower than the other boroughs, as can be seen in the comparison below.

```
1 df_ny['borough'].value_counts()
```

```
BROOKLYN         197
QUEENS           182
MANHATTAN        121
BRONX            101
STATEN ISLAND     21
Name: borough, dtype: int64
```

Table 2. Number of collisions in various boroughs

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.

STATEN ISLAND	Alcohol Involvement	1
	Backing Unsafely	5
	Driver Inattention/Distraction	3
	Following Too Closely	2
	Illnes	1
	Passing or Lane Usage Improper	2
	Reaction to Uninvolved Vehicle	1
	Turning Improperly	1
	Unspecified	5

Table 3. Staten Island collision contributing factors

Thus, Staten Island was chosen as the safest borough for opening Torbjorn's new restaurant since it had the least number of collisions by a factor of 5 as compared with the closest second competitior.

2) New York City Borough Data

The New York City Borough Data was also in JSON, and further examination revealed that the various broughs and neughborhoods were under the "feature" tag in the data, as shown below.


```
{'geometry': {'coordinates': [-73.84720052054902, 40.89470517661],
   'type': 'Point'},
  'geometry_name': 'geom',
  'id': 'nyu_2451_34572.1',
  'properties': {'annoangle': 0.0,
   'annoline1': 'Wakefield',
   'annoline2': None,
   'annoline3': None,
   'bbox': [-73.84720052054902,
    40.89470517661,
    -73.84720052054902,
    40.89470517661],
   'borough': 'Bronx',
   'name': 'Wakefield',
   'stacked': 1},
  'type': 'Feature'}
```


Table 4. Features of boroughs and neighborhoods

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

Table 5. Neighborhoods in Staten Island

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

Exploring Neighborhoods

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 Torbjorn's restaurant. The number of venues for some neighborhoods is shown below.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Annadale	13	13	13	13	13	13
Arden Heights	5	5	5	5	5	5
Arlington	7	7	7	7	7	7
Arrochar	18	18	18	18	18	18
Bay Terrace	12	12	12	12	12	12
Bloomfield	4	4	4	4	4	4
Bulls Head	43	43	43	43	43	43
Butler Manor	6	6	6	6	6	6
Castleton Corners	16	16	16	16	16	16
Charleston	31	31	31	31	31	31
Chelsea	4	4	4	4	4	4
Clifton	18	18	18	18	18	18
Concord	10	10	10	10	10	10
Dongan Hills	26	26	26	26	26	26
Egbertville	5	5	5	5	5	5
Elm Park	11	11	11	11	11	11
Eltingville	38	38	38	38	38	38
Emerson Hill	1	1	1	1	1	1

Table 6. Venues in various neighborhoods around Staten Island

If we just look at the unique values, we get 169 unique categories. Further analysis gives the top ten venues in terms of popularity in each neighborhood, as shown below.

	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	Pizza Place	Dance Studio	Cosmetics Shop	Bakery	Park	Diner	Restaurant	Sports Bar	Train Station	American Restaurant
1	Arden Heights	Pool	Coffee Shop	Playground	Pizza Place	Pharmacy	Wings Joint	Donut Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
2	Arlington	Bus Stop	Grocery Store	American Restaurant	Intersection	Food	Boat or Ferry	Wings Joint	Food & Drink Shop	Flower Shop	Filipino Restaurant
3	Arrochar	Bus Stop	Deli / Bodega	Italian Restaurant	Bagel Shop	Cosmetics Shop	Hotel	Food Truck	Pizza Place	Middle Eastern Restaurant	Supermarket
4	Bay Terrace	Italian Restaurant	Supermarket	Donut Shop	Plaza	Insurance Office	Salon / Barbershop	Sushi Restaurant	Liquor Store	Shipping Store	Wings Joint
5	Bloomfield	Recreation Center	Discount Store	Theme Park	Bus Stop	Donut Shop	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Event Space
6	Bulls Head	Bus Stop	Pizza Place	Chinese	Japanese	Pharmacy	Deli / Bodega	Grocery	Baseball Field	Mattress	Fast Food

Table 7. Popular venues in different Staten Island neighborhoods

Clustering the Venues

	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	9th Cor
0	Staten Island	St. George	40.644982	-74.079353	4	Bar	Deli / Bodega	Thai Restaurant	Donut Shop	Bus Stop	Steakhouse	Monument / Landmark	Scenic Lookout	Resta
1	Staten Island	New Brighton	40.640615	-74.087017	1	Bus Stop	Park	Convenience Store	Construction & Landscaping	Bowling Alley	Playground	Discount Store	Chinese Restaurant	B
2	Staten Island	Stapleton	40.626928	-74.077902	4	Pizza Place	Cosmetics Shop	Sandwich Place	Restaurant	Discount Store	Bank	Spanish Restaurant	Fast Food Restaurant	Me Resta
3	Staten Island	Rosebank	40.615305	-74.069805	4	Italian Restaurant	Grocery Store	Pizza Place	Filipino Restaurant	Breakfast Spot	Bus Stop	Mexican Restaurant	Cajun / Creole Restaurant	Eu Resta
4	Staten Island	West Brighton	40.631879	-74.107182	4	Coffee Shop	Ice Cream Shop	Pharmacy	Music Store	Bar	Italian Restaurant	Event Space	Pizza Place	Me Resta

Table 8. Clustered places of interest in Staten Island

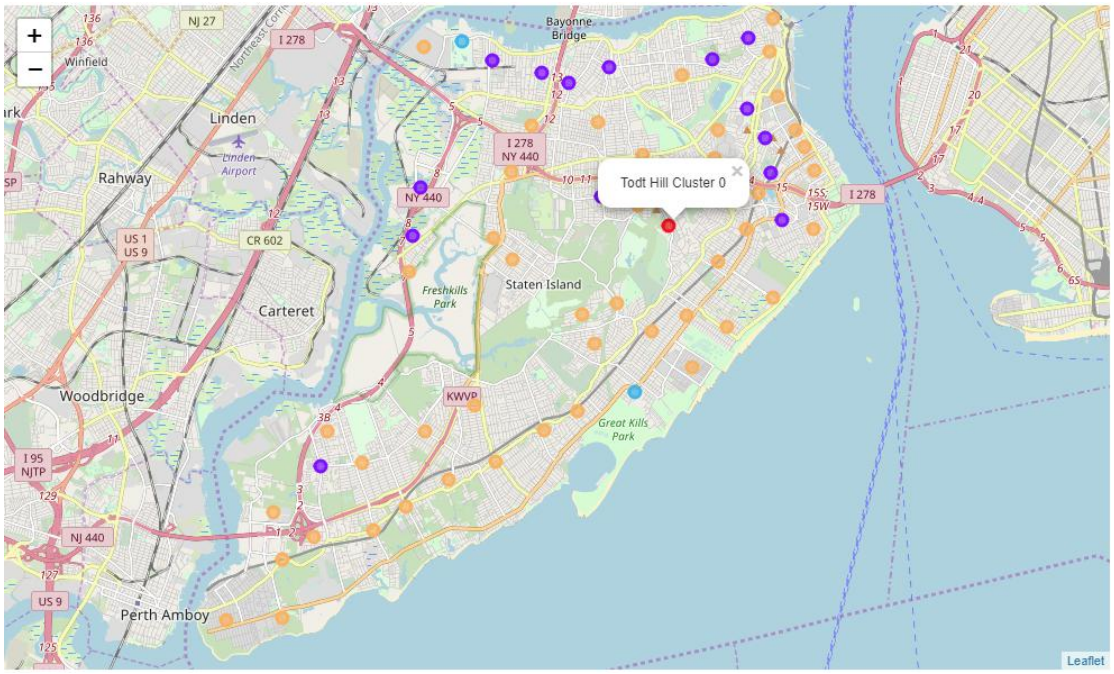

Table 9. Overview of Clusters

Todt Hill, cluster 0, has a Filipino restaurant as 6th most common venue, falafel restaurant as 8th most common and Eastern Europen as the 10th most common venue. Thus, it is hard to come to any logical conclusion from this data.

	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	Wings Joint	Dry Cleaner	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Event Space	Eastern European Restaurant

Table 10. Todt Hill

Cluster 1 has the following venues. Only one neighborhood, Willowbrook, has the Chinese Restaurant as the most common venue. Even looking at the top three most common venues, Marine's Harbor has an Italian restaurant whereas Chelsea has the steakhouse. Like Todt Hill, it is hard to categorically state which type of cuisine is preferred by the locals.

	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	Park	Convenience Store	Construction & Landscaping	Bowling Alley	Playground	Discount Store	Chinese Restaurant	Deli / Bodega	Flower Shop
5	Grymes Hill	Bus Stop	Basketball Court	Dog Run	Wings Joint	Food & Drink Shop	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
8	Port Richmond	Pizza Place	Bus Stop	Rental Car Location	Mexican Restaurant	Donut Shop	Dessert Shop	Eastern European Restaurant	Flower Shop	Filipino Restaurant	Fast Food Restaurant
9	Mariner's Harbor	Bus Stop	Deli / Bodega	Italian Restaurant	Ice Cream Shop	Construction & Landscaping	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
24	Park Hill	Bus Stop	Gym / Fitness Center	Hotel	Coffee Shop	Athletics & Sports	Dry Cleaner	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant
27	Arlington	Bus Stop	Grocery Store	American Restaurant	Intersection	Food	Boat or Ferry	Wings Joint	Food & Drink Shop	Flower Shop	Filipino Restaurant
29	Grasmere	Bus Stop	Food	Bank	Bagel Shop	Japanese Restaurant	Food & Drink Shop	Nail Salon	Restaurant	Juice Bar	Ice Cream Shop
44	Chelsea	Theater	Bus Stop	Steakhouse	Sandwich Place	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Event Space	Eastern European Restaurant
45	Bloomfield	Recreation Center	Discount Store	Theme Park	Bus Stop	Donut Shop	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Event Space
52	Randall Manor	Bus Stop	Deli / Bodega	Wings Joint	Food & Drink Shop	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Event Space
54	Elm Park	Bus Stop	Ice Cream Shop	Toll Plaza	Pizza Place	Chinese Restaurant	Gas Station	Deli / Bodega	Italian Restaurant	American Restaurant	Donut Shop
56	Willowbrook	Chinese Restaurant	Bus Stop	Deli / Bodega	Bagel Shop	Pizza Place	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
57	Sandy Ground	Liquor Store	Market	Art Gallery	Bus Stop	Food & Drink Shop	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
62	Fox Hills	Grocery Store	Deli / Bodega	Sandwich Place	Bus Stop	Donut Shop	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Event Space

Table 11. Cluster 1 venues.

Cluster 2 has a more clear cut preference for venues. Both neighborhoods, Oakwood and Port Ivory, are identical when it comes to preferred venues, as shown below.

	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
10	Port Ivory	Bar	Wings Joint	Eastern European Restaurant	Food & Drink Shop	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Event Space
15	Oakwood	Bar	Wings Joint	Eastern European Restaurant	Food & Drink Shop	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Event Space

Table 12. Identical preferences for venues.

Cluster 3 has only one neighborhood, Emerson Hill and appears to have only a wings joint as the second most common venue. A Filipino restaurant is the fifth most common venue.

	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	Gym	Wings Joint	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Event Space	Eastern European Restaurant	Dry Cleaner

Table 13. Wings joint is a popular venue in Emerson Hill.

The last cluster has a lot of neighborhoods which have restaurants as common venues. The neighborhood New Dorp appears to have a preference for Italian cuisine, as does Great Kills and Annadale.

	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	St. George	Bar	Deli / Bodega	Thai Restaurant	Donut Shop	Bus Stop	Steakhouse	Monument / Landmark	Scenic Lookout	Italian Restaurant	Tapas Restaurant
2	Stapleton	Pizza Place	Cosmetics Shop	Sandwich Place	Restaurant	Discount Store	Bank	Spanish Restaurant	Fast Food Restaurant	Mexican Restaurant	Breakfast Spot
3	Rosebank	Italian Restaurant	Grocery Store	Pizza Place	Filipino Restaurant	Breakfast Spot	Bus Stop	Mexican Restaurant	Cajun / Creole Restaurant	Eastern European Restaurant	Beach
4	West Brighton	Coffee Shop	Ice Cream Shop	Pharmacy	Music Store	Bar	Italian Restaurant	Event Space	Pizza Place	Mexican Restaurant	Liquor Store
7	South Beach	Deli / Bodega	Pier	Beach	Athletics & Sports	Wings Joint	Eastern European Restaurant	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant
11	Castleton Corners	Pizza Place	Ice Cream Shop	Grocery Store	Diner	Sandwich Place	Park	Burger Joint	Bank	Tattoo Parlor	Bagel Shop
12	New Springville	Mobile Phone Shop	Pizza Place	Chinese Restaurant	Bagel Shop	Deli / Bodega	Ice Cream Shop	Soup Place	Sandwich Place	Restaurant	Coffee Shop
13	Travis	Deli / Bodega	Hotel	Bowling Alley	Gym / Fitness Center	Spanish Restaurant	Baseball Field	Park	Gym	Sports Club	Comedy Club
14	New Dorp	Italian Restaurant	Pizza Place	Bank	Chinese Restaurant	Mexican Restaurant	Sandwich Place	Salon / Barbershop	Sushi Restaurant	Taco Place	Bakery
16	Great Kills	Bar	Italian Restaurant	Pizza Place	Cosmetics Shop	Grocery Store	Food & Drink Shop	Falafel Restaurant	Sandwich Place	Chinese Restaurant	Bagel Shop
17	Eltingville	Sushi Restaurant	Pizza Place	Gourmet Shop	Bank	Fast Food Restaurant	Italian Restaurant	Diner	Pharmacy	Restaurant	Chinese Restaurant
18	Annadale	Pizza Place	Dance Studio	Cosmetics Shop	Bakery	Park	Diner	Restaurant	Sports Bar	Train Station	American Restaurant
19	Woodrow	Pharmacy	Cosmetics Shop	Bakery	Pizza Place	Diner	Coffee Shop	Chinese Restaurant	Miscellaneous Shop	Mexican Restaurant	Sushi Restaurant
20	Tottenville	Bus Stop	Deli / Bodega	Cosmetics Shop	Italian Restaurant	Mexican Restaurant	Thrift / Vintage Store	Park	Wings Joint	Flower Shop	Filipino Restaurant
21	Tompkinsville	Brewery	Ice Cream Shop	Rental Car Location	Spanish Restaurant	Bus Stop	Café	Sandwich Place	Rock Club	Caribbean Restaurant	Recording Studio
22	Silver Lake	American Restaurant	Coffee Shop	Golf Course	Burger Joint	Wings Joint	Eastern European Restaurant	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant
23	Sunnyside	Market	Grocery Store	American Restaurant	Spa	Café	Gym	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant

Table 14. Some of the neighborhoods of Cluster 4.

Discussion and Conclusion

Looking at the clusters shown in the previous section, it appears that Clusters 2 and 4 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.

Going back to our case, for Torbjorn's idea of opening a restaurant which should be inspired by local cuisines, he can either go with an Easter European inspired Restaurant in Port Ivory or Oakwood, or he can go with Italian inspired Swedish cuisine in neighborhoods of New Dorp, Great Kills, Eltingville and Annadale. Either of these options will help bring something new to the table.