INTRODUCTION/BUSINESS PROBLEM

1. Business. Nanyang Coffee Company (NYC) is a major supplier of the processed coffee beans used in the *Nanyang* style of coffee popularly drank by the people of Southeast Asia. [*Note: Southeast Asian in this context refers to the ten nations that make up the Association of South East Asian Nations i.e. Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand and Vietnam.*] NYC is a dominant player in Southeast Asia, with a market share of the 20% of the roasted coffee beans market for *Nanyang* coffee in the region. Its key products are roasted coffee beans and coffee powders. Its business lines are (1) the wholesale of *Nanyang* style roasted coffee beans and processed products to coffee shops, restaurants, and coffee product retailers; and (2) its own coffee specialty retail outlets that sells its coffee products and take-away coffee beverages.

2. Product Introduction. Unlike the coffee preferred by consumers in America or Europe which uses primarily Arabica beans, the *Nanyang* coffee is made from a mixture of Arabica, Robusta and Liberica coffee beans. In particular, the significant proportion of the Robusta beans in *Nanyang* coffee confers a stronger and thicker flavor on the *Nanyang* coffee, as compared to the smoother and thinner flavor of the coffee drunk by European and American consumers e.g. expressos. The Robusta beans are commonly grown and sourced from Vietnam, Indonesia and Malaysia. Beside the composition of the beans, the roasting of the beans for *Nanyang* coffee also differs substantially, with a longer roasting period at lower heat necessary to mellow the stronger flavor of the Robusta beans, as compared to the shorter period and higher temperature involved in the roasting of the predominantly Arabica coffee beans. In addition, butter, margarine, sugar and/or palm oil are also added to the beans in the roasting process of the Robusta beans to impart a unique and complex flavor to the coffee. Finally, the *Nanyang* coffee is served either black, or with condensed and/or evaporated milk to add a creamy and thick texture to the coffee.

3. Business Problem. NYC is keen to expand beyond its existing market in Southeast Asia to North America. It has identified Canada as a promising country for its maiden entry into the North American market and earmarked the city of Toronto as its first city in Canada to expand it. The latest 2016 Census of Population prepared by Statistic Canada indicated that Toronto had a population of about 2.7 million and is the largest city in Canada. Within this population, Southeast Asians comprise about 7.0% or about 195,000[1]. This represents the largest concentration of Southeast Asian in any Canadian city, significantly more than the next largest such concentration of about 50,000 in Vancouver as similarly indicated in the 2016 Census of Population[2]. NYC views that the large Southeast Asian population in Toronto will constitute a ready market for its product.

4. NYC's expansion plan calls for the setting up of a specialty coffee retail outlet offering a range of roasted coffee products and take-away drinks. Its coffee product will be positioned as a mid-range coffee product that appeals to the middle-income coffee lovers who currently patronize the cafes and coffee shops in Toronto. Within this segment, it seeks to establish its bridgehead customers among the Southeast Asians who are familiar with its products from their home countries, and through their customs and complemented by its marketing campaign, to build awareness amongst the wider coffee lover population in Toronto. Should the above succeed, NYC will consider setting up my outlets in other parts of Toronto, as well as expansion in other Canadian cities. NYC has engaged our consultancy company to advise on the expansion, with a key component being to decide on the location of the first NYC outlet.

5. Goals and Objectives, **Given the above, NYC has determined that the Goal of the expansion into Toronto is to maximize the awareness and penetration of the *Nanyang* coffee among the coffee drinking consumers, measures in terms of sales of this first outlet. The objective of the study is therefore to identify the neighborhood in Toronto to locate this first NYC outlet that best maximize the awareness and sales of the Nanyang coffee among the targeted segment of the**

[1] https://en.wikipedia.org/wiki/Demographics_of_Toronto
[2] https://en.wikipedia.org/wiki/Vancouver#Demographics

coffee drinking consumers in Toronto, with the bridgehead customers being the ethnic Southeast Asians who enjoy drinking coffee.

ANALYTICS APPROACH/DATA

6. **The goal of the study is to identify the neighborhood to locate the first NYC store in Toronto.** The hypothesis is that the location should be in the neighborhood with a relatively high concentration of (1) coffee shops and cafes, as well as (2) restaurants serving ethnic Southeast Asian cuisines. This assumes that a high concentration of coffee shops and cafes, along with a high parallel concentration of ethnic Southeast Asian restaurants would be indicative of a population of ethnic Southeast Asian customers who drinks coffee. This represents the most likely early customers for the *Nanyang* coffee products sold by NYC. This hypothesis is also based on surveys undertaken by NYC in its home market in Southeast Asia that Southeast Asians often patronize coffee outlets to drink coffee and socialized as a post dinner activity. Situating the NYC outlet in this neighborhood is therefore assessed as giving the highest likelihood of building awareness of the *Nanyang* coffee and therefore boosting sales.

7. We will seek to identify the most optimal neighborhood to site the NYC outlet, the analytics approach will be a descriptive one that seeks to identify and rank neighborhood with high concentration of coffee shops and cafes, and ethnic Southeast Asian restaurants. A greater weightage will be accorded to the concentration of ethnic Southeast Asian restaurants as proxy for the Southeast Asians consumers in the selected neighborhood, given the bridgehead customers being ethnic Southeast Asian who like drinking coffee. Besides using the descriptive functionalities of pandas, we will also employ machine learning e.g. clustering to better identify the location for the first NYC outlet. We will also seek to understand the strength of the relationship between the concentration of ethnic Southeast Asian and the concentration of coffee shops/cafes. This seeks to provide data on the consumption behavior of the ethnic Southeast Asians in Canada i.e. whether they enjoy going to coffee shops/cafes as a post-meal activity like the Southeast Asian consumers in NYC's home region. This assumes that there is a

proportional relationship between the two. The analytics approach will be to both use statistical correlation to understand the strength of the relationship which could serve as model for other large Canadian cities.

DATA

8. Geographical Data. This comprises (1) postal codes of boroughs in Toronto, (2) boroughs and associated neighborhoods in Toronto and (3) latitudes and longitudes of the neighborhood. For (1) and (2), these are scrapped using the BeautifulSoup software off the Wikipedia website of the Toronto Postal Codes https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. For the geographical coordinates of each neighborhood in (3), these are obtained from IBM at http://cocl.us/Geospatial_data. The geographical data are then combined into a data frame, to be used as the input via the Foursquare API to extract the venue data of the coffee shops/cafes and Southeast Asian restaurants in each neighborhood for further analysis to address the goals of the study. A sample of the geographical data frame is shown in Figure 1 and the geographical data set (csv format) is attached as Annex A to this document (see github link).

Figure 1: Geographical Data Sample

	PostalCode	Borough	Neighborhood	Latitude_x	Longitude_x
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494

9. Venue Data. The coffee shops/cafes, as well as the various ethnic Southeast Asian restaurants in each Toronto neighborhood will be extracted from Foursquare using the Foursquare API. This will be combined with the geographical data obtained and formatted as outlined in Para 8 as the starting input. The sample of the master data showing all

venues for each neighborhood is as shown in Figure 2 and the venue data set (csv format) is attached as Annex B to this document (see github link). The coffee shops/cafes and the various ethnic Southeast Asian restaurants will be extracted from this master data set subsequently.

Figure 2: Neighborhood Venues Data Sample

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
1	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
2	Parkwoods	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	Café
3	Parkwoods	43.753259	-79.329656	A&W	43.760643	-79.326865	Fast Food Restaurant
4	Parkwoods	43.753259	-79.329656	Bruno's valu-mart	43.746143	-79.324630	Grocery Store
5	Parkwoods	43.753259	-79.329656	High Street Fish & Chips	43.745260	-79.324949	Fish & Chips Shop
6	Parkwoods	43.753259	-79.329656	Food Basics	43.760549	-79.326045	Supermarket
7	Parkwoods	43.753259	-79.329656	Shoppers Drug Mart	43.745315	-79.325800	Pharmacy
8	Parkwoods	43.753259	-79.329656	Shoppers Drug Mart	43.760857	-79.324961	Pharmacy
9	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
10	Parkwoods	43.753259	-79.329656	Pizza Pizza	43.760231	-79.325666	Pizza Place
11	Parkwoods	43.753259	-79.329656	DVP at York Mills	43.758899	-79.334099	Road
12	Parkwoods	43.753259	-79.329656	TTC Stop #09083	43.759655	-79.332223	Bus Stop
13	Parkwoods	43.753259	-79.329656	TTC Stop 9083	43.759251	-79.334000	Bus Stop
14	Parkwoods	43.753259	-79.329656	Sandover Park	43.760277	-79.333305	Park
15	Parkwoods	43.753259	-79.329656	TTC Stop #9075	43.757596	-79.338155	Train Station
16	Parkwoods	43.753259	-79.329656	Dollarama	43.760341	-79.325519	Discount Store
17	Parkwoods	43.753259	-79.329656	Parkwoods Coin Laundry	43.760386	-79.324894	Laundry Service
18	Parkwoods	43.753259	-79.329656	Spicy Chicken House	43.760639	-79.325671	Chinese Restaurant
19	Parkwoods	43.753259	-79.329656	La Notre	43.760704	-79.325396	Coffee Shop
20	Parkwoods	43.753259	-79.329656	Underhill Mini Mart Convenience	43.745836	-79.324835	Convenience Store
21	Parkwoods	43.753259	-79.329656	Parkwoods Village Centre	43.760735	-79.324873	Shopping Mall
22	Parkwoods	43.753259	-79.329656	Family Food Fair Convenience	43.760620	-79.324459	Convenience Store

METHODOLOGY

10. Preliminary Data Wrangling. From the venue data we obtained from Foursquare as outlined in Figure 1, we will first need to understand the variety of restaurants in Toronto. We therefore conduct a filtering of the relevant data through a string search of the relevant data frame column using the parameter ‘restaurant’. This returns 62 different type of restaurants as shown in Figure 3. From inspection of the values above, we identify

the following types of Southeast Asian (SEA) restaurants: (1) Thai, (2) Vietnamese, (3) Indonesian, (4) Malay and (5) Filipino.

Figure 3: Type of Restaurants in Toronto

```
#Identifying the type of restaruants present in Toronto
restaurants=unique_venues[unique_venues[0].str.contains('Restaurant')]
print(restaurants)
```

```
                                0
0             Caribbean Restaurant
3             Fast Food Restaurant
15              Chinese Restaurant
24            Portuguese Restaurant
34                      Restaurant
40         Mediterranean Restaurant
43               Italian Restaurant
44                French Restaurant
49               Mexican Restaurant
50                  Thai Restaurant
55                Asian Restaurant
61               German Restaurant
63        Middle Eastern Restaurant
74                 Sushi Restaurant
77              Pakistani Restaurant
78                Indian Restaurant
80            Vietnamese Restaurant
82                 Greek Restaurant
84               Seafood Restaurant
93                Korean Restaurant
102                Ramen Restaurant
106  Vegetarian / Vegan Restaurant
108                Theme Restaurant
111     Modern European Restaurant
122             Japanese Restaurant
128                Tapas Restaurant
129              Falafel Restaurant
144             American Restaurant
149        New American Restaurant
153       Latin American Restaurant
160               Dim Sum Restaurant
188         Comfort Food Restaurant
194       South American Restaurant
200            Ethiopian Restaurant
201              Jewish Restaurant
202    Eastern European Restaurant
207                Hakka Restaurant
215              Afghan Restaurant
216             Turkish Restaurant
221            Brazilian Restaurant
235                Cuban Restaurant
240            Dumpling Restaurant
247             Tibetan Restaurant
248            Hawaiian Restaurant
257       Indian Chinese Restaurant
263             Empanada Restaurant
265       Cajun / Creole Restaurant
281           Indonesian Restaurant
291              Persian Restaurant
300            Sri Lankan Restaurant
301                Malay Restaurant
302            Cantonese Restaurant
305             Shanghai Restaurant
306            Hong Kong Restaurant
308              Syrian Restaurant
309               Doner Restaurant
310             Belgian Restaurant
312            Taiwanese Restaurant
313             Filipino Restaurant
315                Udon Restaurant
319              Hotpot Restaurant
321             Moroccan Restaurant
```

11. We next proceed to tabulate the (a) the total number of restaurants; (2) and the total number of SEA restaurants in each Toronto neighbourhood. We also sieve out the

coffee outlets based on the strings ‘Coffee shop’ and ‘Café’ in the venue data in Figure 1, before similarly tabulating the total number of coffee outlets in each neighbourhood. The total numbers of restaurants, SEA restaurants and coffee outlets in each neighbourhood are then combined into a single data frame. An example of the data frame is shown in Figure 4. This served as the master data for our subsequent analysis

Figure 4: Data Frame of numbers of restaurants, SEA restaurants and coffee outlets per Toronto Neighbourhood

	index	Neighbourhood	Total Restaurant Count	SEA Count	Coffee Count
0	0	Agincourt	22	2	2
1	1	Alderwood / Long Branch	1	0	1
2	2	Bathurst Manor / Wilson Heights / Downsview North	4	0	2
3	3	Bayview Village	4	0	1
4	4	Bedford Park / Lawrence Manor East	13	1	4
5	5	Berczy Park	20	1	16
6	6	Birch Cliff / Cliffside West	2	1	1
7	7	Brockton / Parkdale Village / Exhibition Place	25	0	14
8	8	Business reply mail Processing CentrE	10	1	4
9	9	Caledonia-Fairbanks	5	0	2
10	10	Canada Post Gateway Processing Centre	19	0	6
11	11	Cedarbrae	8	1	3
12	12	Central Bay Street	23	1	12
13	13	Christie	35	1	13
14	14	Church and Wellesley	22	1	11
15	15	Clarks Corners / Tam O'Shanter / Sullivan	12	2	3
16	16	Commerce Court / Victoria Hotel	21	2	15
17	17	Davisville	37	3	12
18	18	Davisville North	24	1	17
19	19	Del Ray / Mount Dennis / Keelsdale and Silvert...	3	0	1

12. Basic statistics of the data was conducted using the ‘describe’ method in panda. The data was also normalized to minimize distortions of the data. To identify the preliminary location based on the hypothesis where the ideal location has a high concentration of SEA restaurants and coffee outlets (with primary weight to the

concentration of SEA restaurants), the 'nlargest' method in panda was employed to rank the top 5 neighbourhoods with the highest concentration of coffee outlets among the top 5 neighbourhoods with the highest concentration of SEA restaurants. The 'nlargest' method was applied using both the absolute and normalized counts of all restaurants, the SEA restaurants and the coffee outlets. Folium visualization was employed to display the locations identified through the 'nlargest' method to refine the results. Finally, machine learning i.e. clustering using K-means and Folium visualization was applied to the data set comprising the neighbourhoods, coordinates of the neighbourhoods, total number of restaurants, total number of SEA restaurants and total number of coffee outlets in each neighbourhood. This serves to provide a high-level view to identify patterns in the neighbourhood that could allow us to refine the choice of locations.

RESULTS

13. Figure 5 shows the results of using the 'nlargest' method in panda to rank the top 5 neighbourhoods with the highest concentration of coffee outlets among the top 5 neighbourhoods with the highest concentration of SEA restaurants. Figure 6 shows the Folium visualization of the neighborhoods identified by the 'nlargest' method.

Figure 5. 'nlargest' result to rank optimal location for NYC outlet

```
#Top 5 Neighbourhood in terms of number of coffee outlets, among Top 5 neighbourhood in terms of total number of SEA restaurc
df1.nlargest(5,'SEA Count').nlargest(5,'Coffee Count')
```

	Neighbourhood	Total Restaurant Count	SEA Count	Coffee Count
62	High Park / The Junction South	21	5	15
49	Studio District	28	6	12
76	Kensington Market / Chinatown / Grange Park	31	5	12
78	Summerhill West / Rathnelly / South Hill / For...	24	5	10
37	Downsview	17	5	6

Figure 6: Folium visualization of preliminarily identified location

14. From inspection of the results from 'nlargest' in Figure 5, the optimal neighbourhoods to site the NYC outlet could be in High Park/The Junction South, Studio District, Kensington Market/Chinatown/Grange Park and Summerhill West/Rathnelly/South Hill/Forest Hill SE/Deer Park. From Folium visualization in Figure 6, the neighbourhood of High Park/The Junction South, Studio District, Kensington Market/Chinatown/Grange Park and Summerhill West/Rathnelly/South Hill/Forest Hill SE/Deer Park forms a cluster at the southern part of Toronto around downtown, while Downsview is located further away to the north near the airport. **Based on the above, the neighbourhood of Kensington Market/Chinatown/Grange Park is preliminarily selected as the location for the NYC outlet, since it is the closest distance to the other neighbourhood in the cluster**.

15. Figure 7 shows the result of the correlation analysis for the total number of restaurants, the number of SEA restaurants and the number of coffee outlet in Toronto neighbourhoods using the 'corr()' method in panda. From the correlation coefficients, the correlation co-efficient between the number of SEA restaurants and the coffee outlet in neighbourhood is only 0.49199. This implies that there is only a moderate positive

correlation between the number of ethnic Southeast Asian restaurant and number of coffee outlets in a neighbourhood.

Figure 7: Correlation Analysis Results

```
In [83]: df_absolute=df[['Total Restaurant Count','SEA Count','Coffee Count']]
         df_absolute.corr()
```

Out[83]:

	Total Restaurant Count	SEA Count	Coffee Count
Total Restaurant Count	1.00000	0.60722	0.81173
SEA Count	0.60722	1.00000	0.49119
Coffee Count	0.81173	0.49119	1.00000

16. K-means clustering using scikit-learn was performed with cluster=3. Examples of the results in tabular form and sample data of each of the three clusters are as shown in Figure 8. Based on the clustering above, Cluster 1 appears to be cluster with low concentration of restaurants (of all sort) and coffee outlets; Cluster 2 appears to be cluster with high concentrations of restaurant and coffee outlets; Cluster 3 appears to be cluster with moderate concentration of restaurants (include SEA) and coffee outlets. Of note, the neighbourhood of High Park/The Junction South, Studio District, Kensington Market/Chinatown/Grange Park and Summerhill West/Rathnelly/South Hill/Forest Hill SE/Deer Park are all in Cluster 2, while Downsview is located in Cluster 3.

Figure 8: Result of K-means clustering

```
In [119]: DF_cluster.head(10)
```

Out[119]:

	Neighbourhood	Latitude	Longitude	Cluster Labels	Total Restaurant Count	SEA Count	Coffee Count
0	Parkwoods	43.753259	-79.329656	0	3	0	2
1	Victoria Village	43.725882	-79.315572	0	1	0	2
2	Regent Park / Harbourfront	43.654260	-79.360636	1	20	2	19
3	Lawrence Manor / Lawrence Heights	43.718518	-79.464763	2	13	2	3
4	Queen's Park / Ontario Provincial Government	43.662301	-79.389494	1	24	2	10
5	Malvern / Rouge	43.806686	-79.194353	0	5	0	2
6	Don Mills	43.745906	-79.352188	1	24	1	8
8	Parkview Hill / Woodbine Gardens	43.706397	-79.309937	0	2	0	1
9	Garden District, Ryerson	43.657162	-79.378937	1	27	1	12
10	Glencairn	43.709577	-79.445073	0	7	0	2

Cluster 1

```
In [120]: Cluster1.head(20)
```

Out[120]:

	Neighbourhood	Latitude	Longitude	Cluster Labels	Total Restaurant Count	SEA Count	Coffee Count
0	Parkwoods	43.753259	-79.329656	0	3	0	2
1	Victoria Village	43.725882	-79.315572	0	1	0	2
5	Malvern / Rouge	43.806686	-79.194353	0	5	0	2
8	Parkview Hill / Woodbine Gardens	43.706397	-79.309937	0	2	0	1
10	Glencairn	43.709577	-79.445073	0	7	0	2
11	West Deane Park / Princess Gardens / Martin Gr...	43.650943	-79.554724	0	2	0	1
12	Woodbine Heights	43.695344	-79.318389	0	2	2	4
14	Humewood-Cedarvale	43.693781	-79.428191	0	4	0	2
15	Guildwood / Morningside / West Hill	43.763573	-79.188711	0	4	0	2
18	Caledonia-Fairbanks	43.689026	-79.453512	0	5	0	2
19	Woburn	43.770992	-79.216917	0	3	0	2
20	Leaside	43.709060	-79.363452	0	8	0	4
23	Cedarbrae	43.773136	-79.239476	0	8	1	3
24	Hillcrest Village	43.803762	-79.363452	0	4	0	2
25	Bathurst Manor / Wilson Heights / Downsview North	43.754328	-79.442259	0	4	0	2
29	Scarborough Village	43.744734	-79.239476	0	3	0	1
31	Northwood Park / York University	43.767980	-79.487262	0	9	0	3
35	Kennedy Park / Ionview / East Birchmount Park	43.727929	-79.262029	0	7	0	3
36	Bayview Village	43.786947	-79.385975	0	4	0	1
44	Golden Mile / Clairlea / Oakridge	43.711112	-79.284577	0	3	0	2

Cluster 2

```
In [121]: Cluster2.head(20)
```

Out[121]:

	Neighbourhood	Latitude	Longitude	Cluster Labels	Total Restaurant Count	SEA Count	Coffee Count
2	Regent Park / Harbourfront	43.654260	-79.360636	1	20	2	19
4	Queen's Park / Ontario Provincial Government	43.662301	-79.389494	1	24	2	10
6	Don Mills	43.745906	-79.352188	1	24	1	8
9	Garden District, Ryerson	43.657162	-79.378937	1	27	1	12
13	St. James Town	43.651494	-79.375418	1	25	1	16
17	Berczy Park	43.644771	-79.373306	1	20	1	16
21	Central Bay Street	43.657952	-79.387383	1	23	1	12
22	Christie	43.669542	-79.422564	1	35	1	13
27	Richmond / Adelaide / King	43.650571	-79.384568	1	21	1	12
32	East Toronto	43.685347	-79.338106	1	29	3	15
33	Harbourfront East / Union Station / Toronto Is...	43.640816	-79.381752	1	17	1	16
34	Little Portugal / Trinity	43.647927	-79.419750	1	31	3	13
41	The Danforth West / Riverdale	43.679557	-79.352188	1	32	0	13
42	Toronto Dominion Centre / Design Exchange	43.647177	-79.381576	1	26	2	14
43	Brockton / Parkdale Village / Exhibition Place	43.636847	-79.428191	1	25	0	14
46	Commerce Court / Victoria Hotel	43.648198	-79.379817	1	21	2	15
49	Studio District	43.659526	-79.340923	1	28	6	12
53	Willowdale	43.770120	-79.408493	1	40	3	10
60	Davisville North	43.712751	-79.390197	1	24	1	17
62	High Park / The Junction South	43.661608	-79.464763	1	21	5	15

Cluster 3

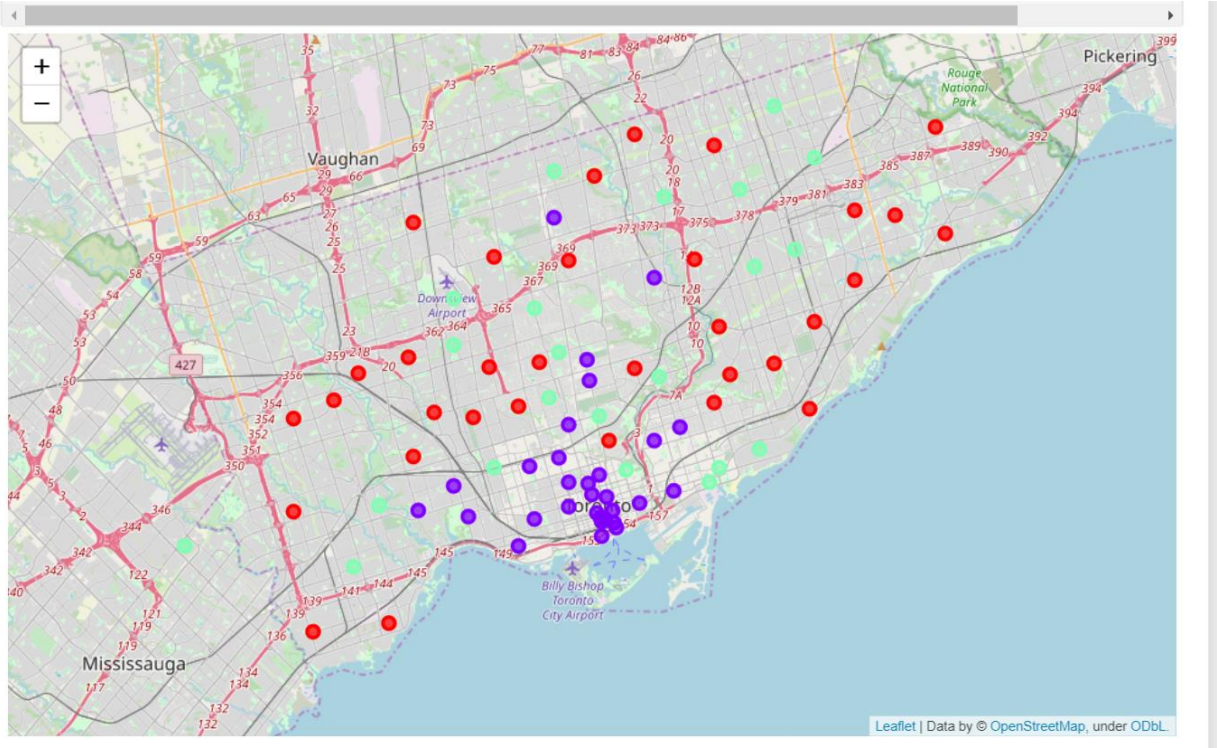
```
In [122]: Cluster3.head(20)
```

Out[122]:

	Neighbourhood	Latitude	Longitude	Cluster Labels	Total Restaurant Count	SEA Count	Coffee Count
3	Lawrence Manor / Lawrence Heights	43.718518	-79.464763	2	13	2	3
16	The Beaches	43.676357	-79.293031	2	16	1	6
26	Thorncliffe Park	43.705369	-79.349372	2	11	0	5
28	Dufferin / Dovercourt Village	43.669005	-79.442259	2	15	2	13
30	Fairview / Henry Farm / Oriole	43.778517	-79.346556	2	9	0	5
37	Downsview	43.737473	-79.464763	2	17	5	6
45	India Bazaar / The Beaches West	43.668999	-79.315572	2	20	0	8
48	Willowdale / Newtonbrook	43.789053	-79.408493	2	11	0	5
50	Bedford Park / Lawrence Manor East	43.733283	-79.419750	2	13	1	4
58	Dorset Park / Wexford Heights / Scarborough To...	43.757410	-79.273304	2	15	1	3
61	Forest Hill North & West	43.696948	-79.411307	2	13	0	6
64	Wexford / Maryvale	43.750072	-79.295849	2	11	1	1
65	North Toronto West	43.715383	-79.405678	2	11	1	5
68	Canada Post Gateway Processing Centre	43.636966	-79.615819	2	19	0	6
70	Agincourt	43.794200	-79.262029	2	22	2	2
74	Clarks Corners / Tam O'Shanter / Sullivan	43.781638	-79.304302	2	12	2	3
75	Moore Park / Summerhill East	43.689574	-79.383160	2	19	4	7
77	Milliken / Agincourt North / Steeles East / L'...	43.815252	-79.284577	2	12	1	1
84	St. James Town / Cabbagetown	43.667967	-79.367675	2	10	2	3
86	The Kingsway / Montgomery Road / Old Mill North	43.653654	-79.506944	2	12	1	5

17. Figure 9 shows the Folium visualization of the clustering results. [*Note: Red as Cluster 1; Purple as Cluster 2; Green as Cluster 3. Of note, we see that Cluster 2 is more concentrated and located at the southern part of Toronto at downtown.*] **Based on this, the clustering further supports the neighbourhood of Kensington Market/Chinatown/Grange Park as the location for the NYC outlet.** It is not only the closest distance to the other three neighbourhood with the highest concentration of coffee outlets among neighbourhood with the highest concentration of SEA restaurants, it is also approximately at the center where most of Cluster 2 is located i.e. downtown Toronto.

Figure 9: Folium visualization of the clusters



DISCUSSION

18. From the results above, the study recommends that the neighbourhood of Kensington Market/Chinatown/Grange Park as the location for the NYC outlet. While in absolute numbers as shown in Figure 5, it has marginally lesser number of SEA restaurants than Studio District (5 versus 6), and marginally lesser number of coffee

outlets than High Park/The Junction South, its geographical location from the visualization places it in the center of a cluster of neighbourhoods with relatively high concentration of restaurants (including SEA restaurants) and coffee outlets. This is assessed as being the most advantageous in facilitating access to and exposure to the NYC outlet by potential customers. The 'outlier' i.e. Downsview in the locations identified by the 'nlargest' method as shown in Figure 5 appears to be an isolated location at the Downsview airport, an aerospace manufacturing hub whose future is uncertain with the troubles encountered by its primary tenant Bombardier.

19. On consumer behavior, NYC should further study the its assumption from its home region that Southeast Asians like to go for coffee after meals and the relevance of this assumption to ethnic Southeast Asians consumers in the Canadian market. This is based on moderate positive correlation between the number of Southeast Asian restaurants and coffee outlets in the neighbourhoods in Toronto. Given the limited number of neighbourhoods in Toronto i.e. 85 neighbourhoods in the study, it is however not conclusive that a stronger correlation does not exist. NYC should therefore seek to undertake a more comprehensive study aggregating similar and other relevant data from other Canadian and even US cities to have a larger data set, and build more robust regression models to understand the consumer behavior of its targeted customer segments.

20. A key learning point in the study is the importance of visualization in complementing other analytical approaches particularly in problems involving geospatial locations. Visualization provides a high-level view and facilitate the sensing of patterns by human analysts that could be missed by statistical methods.

Conclusion

21. In summary, the company recommends NYC to (1) locate its first outlet in the neighbourhood of Kensington Market/Chinatown/Grange Park; and (2) undertake a deeper study on the behaviors and characteristics of its key targeted customer segments

in the Canadian market. Follow on works that the company could undertake on behalf of NYC could include a finer differentiation of its target customer segments in the Canadian market, as well as analysis of the spending and consumption habits of its targeted customer segment, which could allow it to better optimize its product and service offerings in its first NYC outlet in Toronto.

####