

# **Analyzing Neighborhoods to Identify Expansion Opportunities**

December 1, 2020

## **1. Introduction/Business Problem**

### **1.1 Background**

An upscale Italian restaurant chain operating in Milan, Italy is exploring international expansion into North America. The chain's expansion strategy focuses on identifying neighborhoods that are similar to the Milan neighborhoods that the chain and other Italian restaurants currently operate in. Following this strategy offers several benefits to the chain as it seeks to establish a presence in new markets. It minimizes risks of operating in new countries by reducing the unknowns. The country will still be unfamiliar however the neighborhood profiles will be familiar and more normal to operate in. In addition, by identifying neighborhoods that already support Italian restaurants, the chain will be able to benefit from the "second mover" advantage as it is already established that Italian restaurants can be successful there.

### **1.2 Problem**

Identifying the ideal neighborhoods in the new markets involves exploring the opportunity at multiple levels with a focus on the chain's strategy of finding locations that are similar to their current operations. At the city level, since Milan is the financial capital of Italy it makes sense to explore the North American financial capital cities of New York City, USA and Toronto, Canada. At the neighborhood level, an intelligent way to identify and evaluate each neighborhood is necessary. This project will provide this neighborhood analysis based on the commercial venue profile of each neighborhood and their relationship to the commercial venue profiles of the relevant Milan neighborhoods.

### **1.3 Stakeholders**

The neighborhoods that are identified in this project will be utilized to support future expansion decisions by the chain. The stakeholders who will benefit the most from this project are the members of the chain's management team that have been tasked with North America expansion.

## **2. Data**

To successfully complete this project, it will be necessary to identify:

- Neighborhood names and commercial venue profiles of Milan neighborhoods where Italian restaurants are successful

- Neighborhood names and commercial venue profiles of similar Toronto and New York City neighborhoods

## 2.1 Data Sources & Wrangling

The raw data necessary to complete these tasks was identified and scraped from across the internet. Since this will be a comprehensive analysis that covers all neighborhoods of the three cities, it will require a large amount of raw data that will consume the single computer resources available for wrangling and processing. To make the work manageable, it was decided to segment the work for each of the individual cities as much as possible then to concatenate the data for the final analysis.

The Milan borough and neighborhood names were web scraped utilizing Beautiful Soup from the Wikipedia web page: [Municipalities of Milan](#). This data was wrangled into the proper format then geopy was utilized to obtain the geospatial coordinates for each neighborhood. This information was then passed to the Foursquare API which provided the top 100 venues within a 500 meter radius from the geospatial coordinates of each of the 134 neighborhoods of Milan. After further data wrangling, this data was then ready to be analyzed.

Out [197]:

	City	Borough	Neighborhood	Latitude	Longitude
0	Milan	Centro storico	Brera	45.473479	9.188408
1	Milan	Centro storico	Centro Storico	45.447112	9.094054
2	Milan	Centro storico	Conca del Naviglio	45.458560	9.177745
3	Milan	Centro storico	Guastalla	45.458252	9.200023
4	Milan	Centro storico	Porta Sempione	45.477128	9.170598
5	Milan	Centro storico	Porta Tenaglia	45.477821	9.181593
6	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Adriano	45.513572	9.251202
7	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Crescenzago	45.509219	9.247484
8	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Gorla	45.504945	9.224539
9	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Greco	45.502184	9.211233
10	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Loreto	45.484535	9.215276

(3759, 7)

Out [208]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Brera	45.473479	9.188408	Pinacoteca di Brera	45.471979	9.188128	Art Museum
1	Brera	45.473479	9.188408	Palazzo Parigi	45.473378	9.191108	Hotel
2	Brera	45.473479	9.188408	Palazzo di Brera	45.472019	9.188043	College Arts Building
3	Brera	45.473479	9.188408	Hops Beer Shop Milano	45.475240	9.189126	Beer Store
4	Brera	45.473479	9.188408	SUSHI B	45.472153	9.186883	Japanese Restaurant
5	Brera	45.473479	9.188408	Dimore Studio	45.474770	9.186954	Art Gallery
6	Brera	45.473479	9.188408	Osteria di Brera	45.472207	9.187045	Italian Restaurant
7	Brera	45.473479	9.188408	Di Viole Di Liquirizia	45.471460	9.185336	Cupcake Shop
8	Brera	45.473479	9.188408	Orto Botanico di Brera	45.471197	9.189577	Botanical Garden

The Toronto borough and neighborhood names along with their corresponding geospatial coordinates were obtained from a previous project in this course. It is noted that the previous project on Toronto limited the neighborhoods to only those containing "Toronto" in the name whereas this report covers all Toronto neighborhoods thus additional data mining was required.

This data was wrangled into the proper format then passed to the Foursquare API which provided the top 100 venues within a 500 meter radius from the geospatial coordinates of each of the 103 neighborhoods of Toronto. After further data wrangling, this data was then ready to be analyzed.

(103, 5)

Out [204]:

	City	Borough	Neighborhood	Latitude	Longitude
0	Toronto	North York	Parkwoods	43.753259	-79.329656
1	Toronto	North York	Victoria Village	43.725882	-79.315572
2	Toronto	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	Toronto	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	Toronto	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
5	Toronto	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	Toronto	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	Toronto	North York	Don Mills	43.745906	-79.352188
8	Toronto	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	Toronto	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937

(2167, 7)

Out [206]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
5	Victoria Village	43.725882	-79.315572	The Frig	43.727051	-79.317418	French Restaurant
6	Victoria Village	43.725882	-79.315572	Eglinton Ave E & Sloane Ave/Bermondsey Rd	43.726086	-79.313620	Intersection
7	Victoria Village	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place
8	Regent Park, Harbourfront	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery

The New York City borough and neighborhood names along with their corresponding geospatial coordinates were obtained from a previous project in this course. It is noted that the previous project on New York City limited the neighborhoods to only those in the Manhattan borough whereas this report covers all New York City neighborhoods in all five boroughs thus additional data mining was required. This data was wrangled into the proper format then passed to the Foursquare API which provided the top 100 venues within a 500 meter radius from the geospatial coordinates of each of the 306 neighborhoods of New York City. After further data wrangling, this data was then ready to be analyzed.

(306, 5)

Out [210]:

	City	Borough	Neighborhood	Latitude	Longitude
0	New York City	Bronx	Wakefield	40.894705	-73.847201
1	New York City	Bronx	Co-op City	40.874294	-73.829939
2	New York City	Bronx	Eastchester	40.887556	-73.827806
3	New York City	Bronx	Fieldston	40.895437	-73.905643
4	New York City	Bronx	Riverdale	40.890834	-73.912585
5	New York City	Bronx	Kingsbridge	40.881687	-73.902818
6	New York City	Manhattan	Marble Hill	40.876551	-73.910660
7	New York City	Bronx	Woodlawn	40.898273	-73.867315
8	New York City	Bronx	Norwood	40.877224	-73.879391
9	New York City	Bronx	Williamsbridge	40.881039	-73.857446

+

(10478, 7)

Out [43]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
4	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop

### 3. Methodology

The goal of this project is to identify neighborhoods and their corresponding commercial venue profiles for Milan neighborhoods where Italian restaurants are located and then to identify similar neighborhoods in Toronto, Canada and New York City, USA. Once the subset of neighborhoods that meet these general characteristics are identified, it will be further analyzed via data from Foursquare to discover the neighborhoods with the most prevalent data for supporting upscale Italian restaurants.

The Data section detailed the collection and wrangling of data that forms the foundation for the rest of this project. The work described in the Analysis section will combine this data from multiple independent sources to provide meaningful insights and recommendations. As was the case in the Data section, the Analysis section will be segmented for work with each of the individual cities as much as possible then concatenated for the final analysis.

The work in the Analysis section will follow an ordered process:

1. Build a dataframe for each city populated with the commercial venue information for each neighborhood.
2. Concatenate these dataframes into one master neighborhood venue dataframe and to similarly concatenate individual city data to build a master neighborhood location dataframe.
3. Work with the master dataframes and k-Means Clustering to build a new combined master dataframe that lists the ranked top ten commercial venue types for each neighborhood and subdivides the neighborhoods into clusters based on their commercial

venue profiles. Silhouette analysis will be applied to the data in order to identify the best k to utilize during the final k-Means Clustering. in order to provide the most meaningful analysis.

4. Analyze this combined master dataframe to identify the subset of Toronto and New York City neighborhoods that are most promising for future expansion.
5. Final analysis of the subset of neighborhoods to identify the top neighborhoods most appropriate for supporting the upscale-style of Italian restaurant featured by the chain. This will be accomplished by utilizing the Foursquare API to find all Italian restaurants with a price tier of 4 (highest average price for an entree) in the subset neighborhoods.

## 4. Analysis

### 4.1 Commercial Venue Information Dataframes

```
print(milan_neighborhoods_venues_sorted.shape)
milan_neighborhoods_venues_sorted.head(10)
```

(134, 11)

Out[201]:

	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	Acquabella	Fast Food Restaurant	Sporting Goods Shop	Furniture / Home Store	Warehouse Store	Pub	Flower Shop	Hardware Store	Yoga Studio	Event Space	Falafel Restaurant
1	Adriano	Italian Restaurant	Ice Cream Shop	Soccer Field	Performing Arts Venue	Trattoria/Osteria	Toy / Game Store	Cultural Center	Fish & Chips Shop	Event Space	Falafel Restaurant
2	Affori	Café	Italian Restaurant	Pizza Place	Cocktail Bar	Park	Pharmacy	Fried Chicken Joint	Pool Hall	Pool	Health Food Store
3	Arzaga	Pizza Place	Playground	Café	Japanese Restaurant	Cosmetics Shop	Food Court	Supermarket	Pharmacy	Athletics & Sports	Gym / Fitness Center
4	Assiano	Mediterranean Restaurant	Restaurant	Fish & Chips Shop	Empanada Restaurant	Event Space	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Filipino Restaurant
5	Baggio	Café	Convenience Store	Plaza	Dog Run	Gastropub	Supermarket	Bar	Japanese Restaurant	Italian Restaurant	Food
6	Barona	Dog Run	Athletics & Sports	Café	Trattoria/Osteria	Soccer Field	Theater	Yoga Studio	Fast Food Restaurant	Falafel Restaurant	Farm
7	Bicocca	Café	Italian Restaurant	Pizza Place	Restaurant	Plaza	Sushi Restaurant	Pub	Piadineria	College Quad	Sandwich Place
8	Boffalora	Soccer Field	Bakery	Athletics & Sports	Café	Trattoria/Osteria	Dog Run	Theater	Yoga Studio	Fast Food Restaurant	Falafel Restaurant
9	Boldinasco	Tram Station	Hotel	Restaurant	Bakery	Italian Restaurant	Chinese Restaurant	Gun Range	Seafood Restaurant	Café	Bar

```
print(toronto_neighborhoods_venues_sorted.shape)
toronto_neighborhoods_venues_sorted.head(10)
```

< (96, 11) >

Out[70]:

	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	Agincourt	Lounge	Breakfast Spot	Latin American Restaurant	Skating Rink	Clothing Store	Ethiopian Restaurant	Event Space	Escape Room	Electronics Store	Discount Store
1	Alderwood, Long Branch	Pizza Place	Sandwich Place	Coffee Shop	Pub	Pharmacy	Gym	Greek Restaurant	Discount Store	Department Store	Dessert Shop
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Pharmacy	Deli / Bodega	Shopping Mall	Bridal Shop	Sandwich Place	Diner	Restaurant	Middle Eastern Restaurant
3	Bayview Village	Japanese Restaurant	Café	Bank	Chinese Restaurant	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Yoga Studio
4	Bedford Park, Lawrence Manor East	Sandwich Place	Italian Restaurant	Coffee Shop	Greek Restaurant	Thai Restaurant	Liquor Store	Comfort Food Restaurant	Juice Bar	Butcher	Café
5	Berczy Park	Coffee Shop	Restaurant	Bakery	Cocktail Bar	Beer Bar	Farmers Market	Cheese Shop	Seafood Restaurant	Sandwich Place	Beach
6	Birch Cliff, Cliffside West	College Stadium	General Entertainment	Skating Rink	Café	Donut Shop	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant
7	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Nightclub	Coffee Shop	Gym	Climbing Gym	Burrito Place	Bookstore	Stadium	Restaurant
8	Business reply mail Processing Centre, South C...	Pizza Place	Auto Workshop	Garden Center	Garden	Fast Food Restaurant	Light Rail Station	Farmers Market	Comic Shop	Park	Recording Studio
9	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Boutique	Historic Site	Bar	Coffee Shop	Boat or Ferry	Plane	Harbor / Marina	Rental Car Location

```
print(NYC_neighborhoods_venues_sorted.shape)
NYC_neighborhoods_venues_sorted.head(10)
```

< (302, 11) >

Out[73]:

	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	Allerton	Pizza Place	Chinese Restaurant	Deli / Bodega	Spa	Playground	Supermarket	Martial Arts School	Bakery	Electronics Store	Outdoors & Recreation
1	Annadale	Pizza Place	American Restaurant	Train Station	Bagel Shop	Park	Food	Restaurant	Dance Studio	Deli / Bodega	Diner
2	Arden Heights	Pharmacy	Deli / Bodega	Smoke Shop	Coffee Shop	Bus Stop	Pizza Place	Fish Market	Factory	Falafel Restaurant	Farm
3	Arlington	Grocery Store	Gay Bar	Boat or Ferry	Home Service	Bus Stop	Coffee Shop	Deli / Bodega	Fast Food Restaurant	Farmers Market	Food
4	Arrochar	Deli / Bodega	Bus Stop	Italian Restaurant	Athletics & Sports	Food Truck	Supermarket	Outdoors & Recreation	Middle Eastern Restaurant	Mediterranean Restaurant	Bar
5	Arverne	Surf Spot	Metro Station	Sandwich Place	Playground	Burrito Place	Restaurant	Thai Restaurant	Coffee Shop	Board Shop	Pizza Place
6	Astoria	Hookah Bar	Bar	Middle Eastern Restaurant	Seafood Restaurant	Bakery	Indian Restaurant	Pizza Place	Food Truck	Mediterranean Restaurant	Greek Restaurant
7	Astoria Heights	Plaza	Chinese Restaurant	Pizza Place	Playground	Bowling Alley	Bakery	Laundromat	Hostel	Food	Supermarket
8	Auburndale	Italian Restaurant	Mobile Phone Shop	Hookah Bar	Miscellaneous Shop	Toy / Game Store	Athletics & Sports	Comic Shop	Sushi Restaurant	Pharmacy	Bar
9	Bath Beach	Italian Restaurant	Gas Station	Pizza Place	Donut Shop	Pharmacy	Chinese Restaurant	Fast Food Restaurant	Cantonese Restaurant	Bubble Tea Shop	Dessert Shop

## 4.2.A Master Neighborhood Venue Dataframe

```
print(total_grouped.shape)
total_grouped.head(10)
```

(532, 503)

Out[75]:

	Neighborhood	Abruzzo Restaurant	Accessories Store	Adult Education Center	African Restaurant	Airport	Airport Service	American Restaurant	Amphitheater	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Art Museum
0	Acquabella	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	Adriano	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Affori	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Arzaga	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Assiano	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	Baggio	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	Barona	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	Bicocca	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	Boffalora	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	Boldinasco	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

## 4.2.B Master Neighborhood Location Dataframe

```
print(total_coord.shape)
total_coord.head(10)
```

(536, 5)

Out[212]:

	City	Borough	Neighborhood	Latitude	Longitude
0	Milan	Centro storico	Brera	45.473479	9.188408
1	Milan	Centro storico	Centro Storico	45.447112	9.094054
2	Milan	Centro storico	Conca del Naviglio	45.458560	9.177745
3	Milan	Centro storico	Guastalla	45.458252	9.200023
4	Milan	Centro storico	Porta Sempione	45.477128	9.170598
5	Milan	Centro storico	Porta Tenaglia	45.477821	9.181593
6	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Adriano	45.513572	9.251202
7	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Crescenzago	45.509219	9.247484
8	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Gorla	45.504945	9.224539
9	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Greco	45.502184	9.211233

### 4.3.A Silhouette Analysis

```
In [84]: # Silhouette Analysis is performed to determine the optimum value for k

sil = []
kmax = 10

total_grouped_clustering = total_grouped.drop('Neighborhood', 1)

# dissimilarity would not be defined for a single cluster, thus, minimum number of clusters should be 2
for k in range(3, kmax+1):
    kmeans_total = KMeans(n_clusters = k, random_state = 0).fit(total_grouped_clustering)
    labels = kmeans_total.labels_
    sil.append(silhouette_score(total_grouped_clustering, labels, metric = 'euclidean'))

sil

Out[84]: [0.031105065148005373,
0.04306070456186652,
0.04811349246296942,
0.04485577408704314,
0.05036572256509973,
0.05665548349149245,
0.06316349162819809,
0.04227498429296539]
```

### 4.3.B k-Means Clustering

```
In [85]: # Performing the final k-Means Clustering

# set number of clusters
k = 9

# run k-means clustering
kmeans_total = KMeans(n_clusters = k, random_state = 0).fit(total_grouped_clustering)

kmeans_total.labels_[0:10]

Out[85]: array([3, 1, 1, 0, 3, 1, 1, 1, 1, 1])
```



### 4.3.C Combined Master Dataframe

```
print(total_merged.shape)
total_merged.head(10)
```

(538, 16)

Out[93]:

	City	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	
0	Milan	Centro storico	Brera	45.473479	9.188408	1	Italian Restaurant	Ice Cream Shop	Hotel	Japanese Restaurant	Café	Pizza Place	Restaurant	C
1	Milan	Centro storico	Centro Storico	45.447112	9.094054	1	Bar	Pizza Place	Italian Restaurant	Park	Shopping Mall	Cupcake Shop	Fried Chicken Joint	Mo
2	Milan	Centro storico	Conca del Naviglio	45.458560	9.177745	1	Italian Restaurant	Café	Ice Cream Shop	Pizza Place	Cocktail Bar	Seafood Restaurant	Bistro	
3	Milan	Centro storico	Guastalla	45.458252	9.200023	1	Italian Restaurant	Tram Station	Pizza Place	Ice Cream Shop	Restaurant	Kebab Restaurant	Bar	
4	Milan	Centro storico	Porta Sempione	45.477128	9.170598	1	Italian Restaurant	Pizza Place	Cocktail Bar	Japanese Restaurant	Ice Cream Shop	Bakery	Gastropub	Med f
5	Milan	Centro storico	Porta Tenaglia	45.477821	9.181593	1	Italian Restaurant	Japanese Restaurant	Café	Chinese Restaurant	Ice Cream Shop	Pizza Place	Wine Bar	f
6	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Adriano	45.513572	9.251202	1	Italian Restaurant	Ice Cream Shop	Soccer Field	Performing Arts Venue	Trattoria/Osteria	Toy / Game Store	Cultural Center	Fis
7	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Crescenzago	45.509219	9.247484	1	Italian Restaurant	Café	Ice Cream Shop	Bus Stop	Metro Station	Supermarket	Pharmacy	H
8	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Gorla	45.504945	9.224539	1	Italian Restaurant	Pizza Place	Hotel	Plaza	Park	Spa	Brewery	C
9	Milan	Stazione Centrale, Gorla, Turro, Greco, Cresce...	Greco	45.502184	9.211233	3	Supermarket	Seafood Restaurant	Chinese Restaurant	Café	Hotel	Plaza	Pet Store	Ci

## 4.4 Subset of Toronto and New York City Neighborhoods Most Promising for Expansion

Cluster 1

```
In [106]: total_merged.loc[total_merged['Cluster Labels'] == 1, total_merged.columns[[0, 2] + list(range(5, total_merged.shape[1]))]
```

127	Milan	Montalbino	1	Pizza Place	Café	Bakery	Italian Restaurant	Seafood Restaurant	Supermarket	Chinese Restaurant	Pharmacy	Tram Stop
129	Milan	Porta Garibaldi	1	Italian Restaurant	Hotel	Restaurant	Pizza Place	Café	Plaza	Wine Bar	Clothing Store	Coffee Shop
130	Milan	Porta Nuova	1	Italian Restaurant	Restaurant	Café	Pizza Place	Hotel	Ice Cream Shop	Plaza	Sushi Restaurant	Art Gallery
156	Toronto	Christie	1	Grocery Store	Café	Park	Nightclub	Italian Restaurant	Candy Store	Baby Store	Athletics & Sports	Coffee Shop
170	Toronto	Bayview Village	1	Japanese Restaurant	Café	Bank	Chinese Restaurant	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop
187	Toronto	Birch Cliff, Cliffside West	1	College Stadium	General Entertainment	Skating Rink	Café	Donut Shop	Discount Store	Distribution Center	Dive Bar	Dog
209	Toronto	University of Toronto, Harbord	1	Café	Bookstore	Sandwich Place	Bakery	Bar	Japanese Restaurant	Noodle House	Bank	Italian Restaurant
264	New York	Belmont	1	Italian	Pizza Place	Deli /	Bakery	Bank	Dessert Shop	Donut Shop	Coffee Shop	

Cluster 1 is largely Milan-centric. Roughly 68% of the Milan neighborhoods are in this cluster while only around 4% of the neighborhoods for both Toronto and New York City are in this cluster. Most of the neighborhoods have Italian restaurants as the most common type of venue. **This is the cluster that we were hoping to find that is comprised of the Toronto and New York City neighborhoods that are similar to Milan and proven to support Italian restaurants.**

## 4.5 Identifying the Top Neighborhoods for Upscale Italian Restaurant Expansion

```
print(cluster_1_final_price_4_Italian_Restaurants.shape)
cluster_1_final_price_4_Italian_Restaurants
```

(54, 7)

Out[190]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
3	Christie	43.669542	-79.422564	Vinny's Panini	43.670679	-79.426148	Italian Restaurant
27	University of Toronto, Harbord	43.662696	-79.400049	Piano Piano	43.662949	-79.402898	Italian Restaurant
58	Belmont	40.857277	-73.888452	Tino's Delicatessen	40.855882	-73.887166	Italian Restaurant
67	Belmont	40.857277	-73.888452	Arthur Avenue Fiasco	40.854844	-73.887908	Italian Restaurant
72	Belmont	40.857277	-73.888452	Calabria Pork Store	40.854145	-73.888776	Italian Restaurant
73	Belmont	40.857277	-73.888452	Michaelangelo's Coal Fired Brick Oven Pizza Re...	40.857412	-73.886468	Italian Restaurant

```
In [191]: cluster_1_final_price_4_Italian_Restaurants['Neighborhood'].value_counts()
```

```
Out[191]: Belmont                18
Greenwich Village              11
Old Town                      4
Edgewater Park                 4
Shore Acres                    3
Great Kills                    2
Mariner's Harbor               2
New Dorp Beach                 2
Elm Park                      2
Richmond Town                  1
University of Toronto, Harbord 1
Egbertville                    1
Christie                       1
Lighthouse Hill                1
Tottenville                    1
Name: Neighborhood, dtype: int64
```

This analysis clearly shows that the two New York City neighborhoods of Belmont and Greenwich Village have the highest amount of upscale Italian restaurants in Cluster 1 neighborhoods.

## 5. Results and Discussion

Our data collection and wrangling went smoothly. Some minor issues occurred when obtaining the geospatial coordinates from geopy as a few neighborhoods returned coordinates not in the city. These occurrences were rare and handled by eliminating the outlier neighborhoods from further analysis.

Our analysis identified the following Toronto and New York City neighborhoods subset that are similar to Milan neighborhoods where Italian restaurants thrive:

- Toronto Neighborhoods
  1. Christie
  2. Bayview Village
  3. Birch Cliff/Cliffside West
  4. University of Toronto/Harboard
- New York City Neighborhoods
  1. Belmont
  2. Edgewater Park
  3. Greenwich Village
  4. Mariner's Harbor
  5. Great Kills
  6. Tottenville
  7. Old Town
  8. New Dorp Beach
  9. Richmond Town
  10. Shore Acres
  11. Elm Park
  12. Egbertville
  13. Lighthouse Hill

The foursquare API was utilized to further analyze this subset by researching which of the member neighborhoods currently support upscale (price tier = 4) Italian restaurants. This analysis determined that the two New York City neighborhoods of Belmont and Greenwich Village contained the vast majority of the upscale (price tier 4) Italian restaurants in the subset. Thus, **Belmont and Greenwich Village are the most promising candidates by following the chain's strategy for future expansion.**

One footnote for the analysis section is to point out that the silhouette method for identifying the optimum k to use in k-Means Clustering did not produce a strong silhouette value. In the range of 3-10, all of the silhouette values were positive but in the 0.03 – 0.06 range. This is most likely due to the nature of the data being analyzed. All of the neighborhoods had their own commercial venue profiles however on a neighborhood-scale there was a lot of commonality and thus the data did not present the strongly segmented clusters that would have produces a silhouette value closer to 1.

## 6. Conclusion

The goal of this project was to identify neighborhoods in Toronto, Canada and New York City, USA that fit the international expansion strategy for an upscale Italian restaurant chain based in Milan, Italy. The primary focus of this strategy is to identify neighborhoods that have a similar commercial venue profile to the Milan neighborhoods that the chain and other Italian restaurants currently operate in. This type of strategy allows the restaurant chain to take a safer approach to international expansion. Additionally, it also allows the chain to benefit from the "second mover" advantage as it reveals neighborhoods where it is already proven that Italian restaurants can be successful.

By utilizing neighborhood information from the internet and venue information from the Foursquare API, we were able to apply k-Means Clustering to intelligently cluster the neighborhoods of the three cities. This clustering clearly identified a subset of four Toronto neighborhoods and thirteen New York City neighborhoods that are similar to Milan neighborhoods where Italian restaurants thrive. The foursquare API was utilized to further analyze this subset by researching which of the member neighborhoods currently support upscale (price tier = 4) Italian restaurants. This analysis determined that the two New York City neighborhoods of Belmont and Greenwich Village are the best fits for future expansion based on the company's international expansion strategy.

Many factors will ultimately go into expansion decisions. The results of this project are one such factor that provides a data driven direction for stakeholders to consider and further explore.

## 7. Future Developments

The results of this project are useful on their own however future work could expand upon it to include additional data categories/analysis. This could go in many direction but for example one area that would be interesting to explore is to identify neighborhoods that are growing and primed for expanded upscale restaurant opportunities. This could be accomplished by analyzing the neighborhood subset for financial growth projections and also the opening dates of restaurants with a Price tier = 4 in Foursquare.

Another interesting development would be to build a dataframe that is populated with currently available restaurant spaces in the subset of neighborhoods identified as being strong expansion candidates. In fact, this project originally considered including this analysis however no strong commercial property search APIs were found that would be suitable for this work.