**Real Estate & Venues Data Analysis**

**of Toronto and New York City**

Daniel Poliseno

February 6, 2020

**1. Introduction**

**1.1 Background**

Toronto is known as an international center of finance, arts, culture, and business as well as being known as one of the most multicultural cities in the world. Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 as of 2016.2 In comparison, New York City is the most populous city in the United States with an estimated 2018 population of 8,398,748. In addition, New York City is known as a global power city and renowned as the financial, cultural, and media capital of the world.3

For Alphabet Inc., parent company of Google Inc., both New York City and Toronto are important cities. Google’s Toronto Offices serve as a hub-office for many members of the company’s creative team and salespeople. Engineers and computer scientists at the Google office in Toronto work on many of Google’s well known products and services. In addition, in August 2019 Google announced it would be expanding office space in the financial district of Toronto.4 New York is home to Google’s second largest office in the Chelsea neighborhood of Manhattan. “Engineers work on Google Drive, Search, AdWords, and Maps, and the large sales team works with clients that include media companies and ad agencies.”1

**1.2 Business Problem**

For Real Estate Agents, determining the best area or neighborhood and finding a client’s perfect home at the right place can be the most difficult and time consuming aspect. To complete this task, an agent may spend countless hours on researching and showing clients homes or may pay significant amounts of money to hire employees to conduct research and/or pay for expensive subscriptions to companies that aggregate statistical data.

As a Real Estate Agent in New York City, a client has contacted me who has accepted a job offer at Google’s NYC office and is relocating from Toronto. The client is ready to make an offer on their new condo/co-op as quickly as possible. This project will help to understand the diversity of each neighborhood analyzed by leveraging venue data from Foursquare’s Places API and k-means clustering unsupervised machine learning algorithm. The objective of this project will be to propose the 3 best listings I can find for my client who will be arriving next week and is ready to buy their new home. The client would like to buy a condo/co-op in a neighborhood most similar to the one they live in now, as similar as possible in price/size to the one they have now, and closest to Google’s NYC office building.

The apartment in Toronto they are moving from is a 2 bedroom & 2 bathroom condo listed for C$899,000 in the Downtown Toronto (Old Toronto) neighborhood of Toronto, ON, Canada.

The current home address is: 1 A The Esplanade Ave # 2008, Toronto, ON, M5E 0A8.

Google’s NYC office building address is: 75 9th Ave, New York, NY 10011.

**1.3 Interest**

The primary stakeholders, interested in a new way to use quantifiable analysis to understand and profile a neighborhood would be Real Estate Agents and Real Estate Buyers. Previously, neighborhood profiles have always been aggregated and compared based on historic, statistical, and/or demographic information. However, I believe a new approach based on venues and often they’re visited for creating neighborhood profiles in order to compare other neighborhood profiles can provide a basis for a much more accurate area profile.

**2. Data Collection and Cleaning**

**2.1 Data Sources**

The data sources I utilized are; [New York City Dataset](https://cocl.us/new_york_dataset), [Toronto Dataset](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), [Toronto Longitude and Latitude Coordinates Dataset](http://cocl.us/Geospatial_data), [Foursquare API](https://developer.foursquare.com/docs), and the [Zillow API](https://www.zillow.com/howto/api/APIOverview.htm). The New York City Dataset, Toronto Dataset, and Toronto Longitude and Latitude Coordinates Dataset were used to create the Manhattan data frame and Toronto data frame containing neighborhood names and each corresponding latitudes and longitudes. Utilizing the Foursquare API, I collected data on nearby venues and the frequency they are visited in order to create neighborhood profiles which could be compared to find the optimal neighborhood for the client. The Zillow API was used to determine three proposed listings in the optimal new neighborhood for the client moving from Toronto.

**2.1 Data Collection and Cleaning**

In order to address the business problem, it was necessary to gather the neighborhood names and corresponding latitudes and longitudes for Toronto, Ontario and New York, NY. The [New York City Dataset](https://cocl.us/new_york_dataset) contained a complete list of New York City Boroughs, Neighborhoods, and their corresponding latitudes and longitudes. I then formatted the New York City Dataset into a new Manhattan data frame that only contained the neighborhood names and corresponding latitudes and longitudes Manhattan.



To create the data frame Toronto, I scraped two datasets; the [Toronto Dataset](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and the [Toronto Longitude and Latitude Coordinates Dataset](http://cocl.us/Geospatial_data), including only boroughs that contained the word Toronto and each corresponding postal code, borough name, neighborhood name, latitude and longitude.



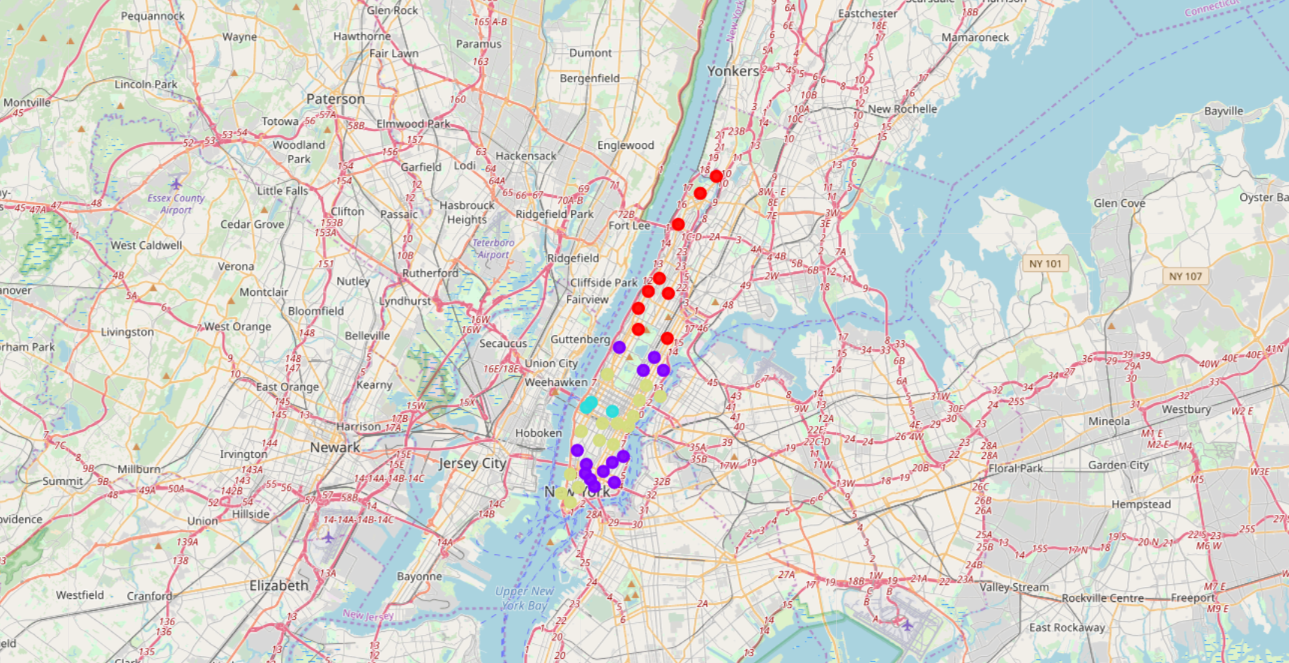
Utilizing the Foursquare API, I analyzed each Manhattan neighborhood creating a new data frame with the top 15 most common venues.



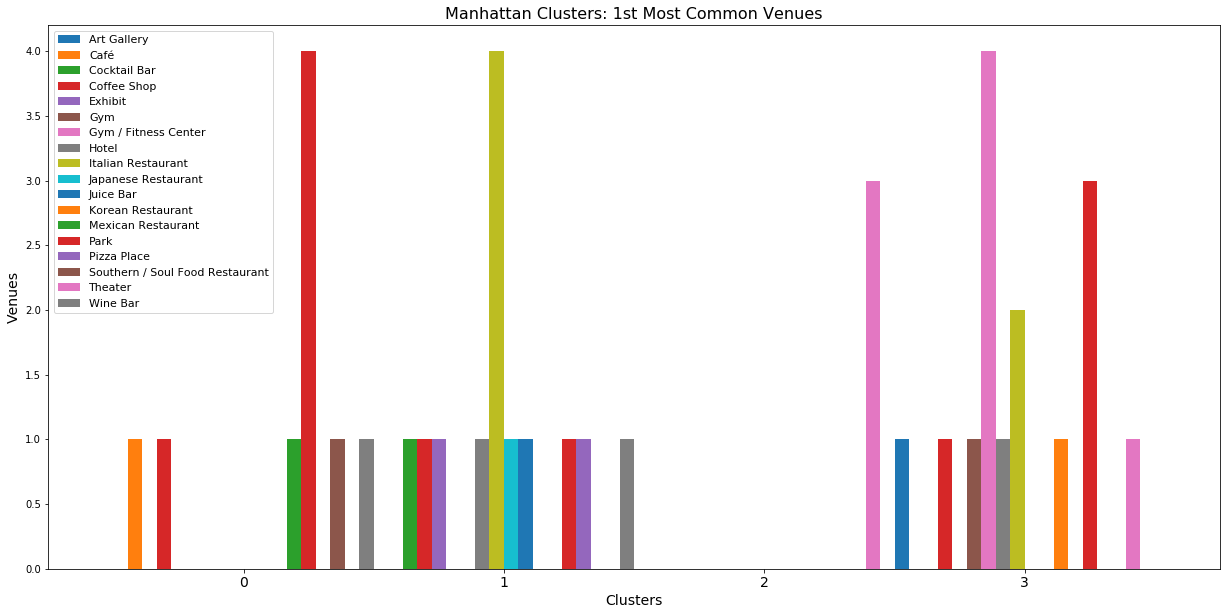
To create the desired neighborhood profiles based on nearby venues and the frequencies visited, I used k-means clustering to create four clusters for the neighborhoods in Manhattan.

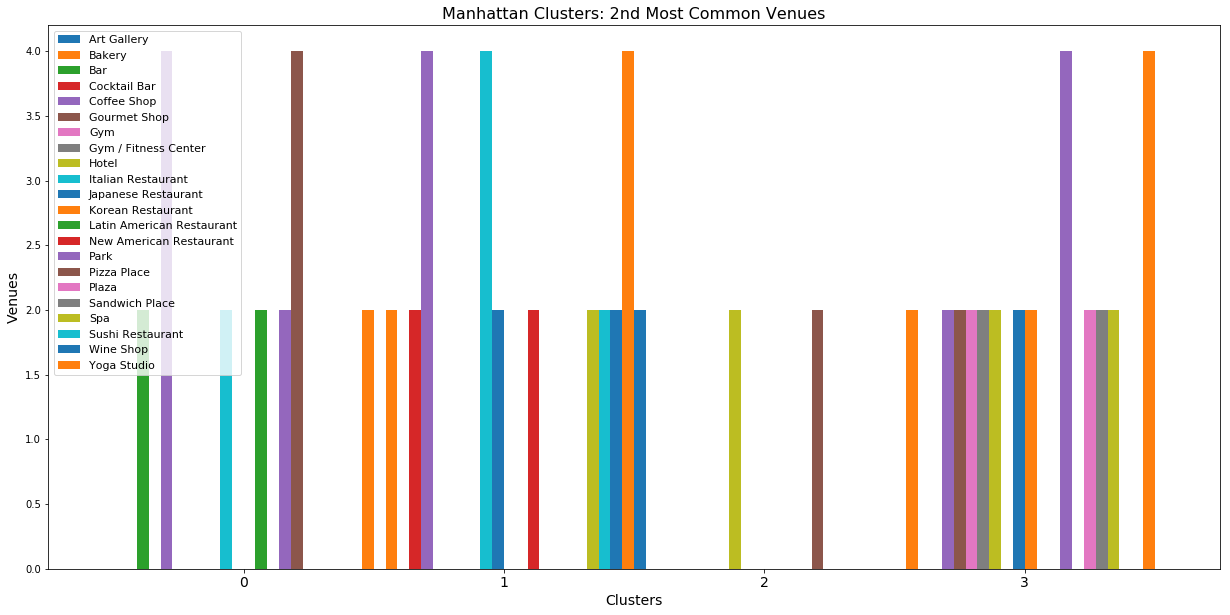


Manhattan Clusters Map:



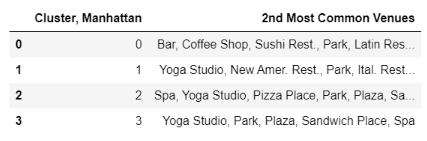
To better visualize the data gathered about the Manhattan neighborhoods, I created two bar plots for the 1st and 2nd most common venues.





Two lists were created for the clustered Manhattan neighborhoods in order to find the neighborhood most similar to the client's current neighborhood in Toronto.





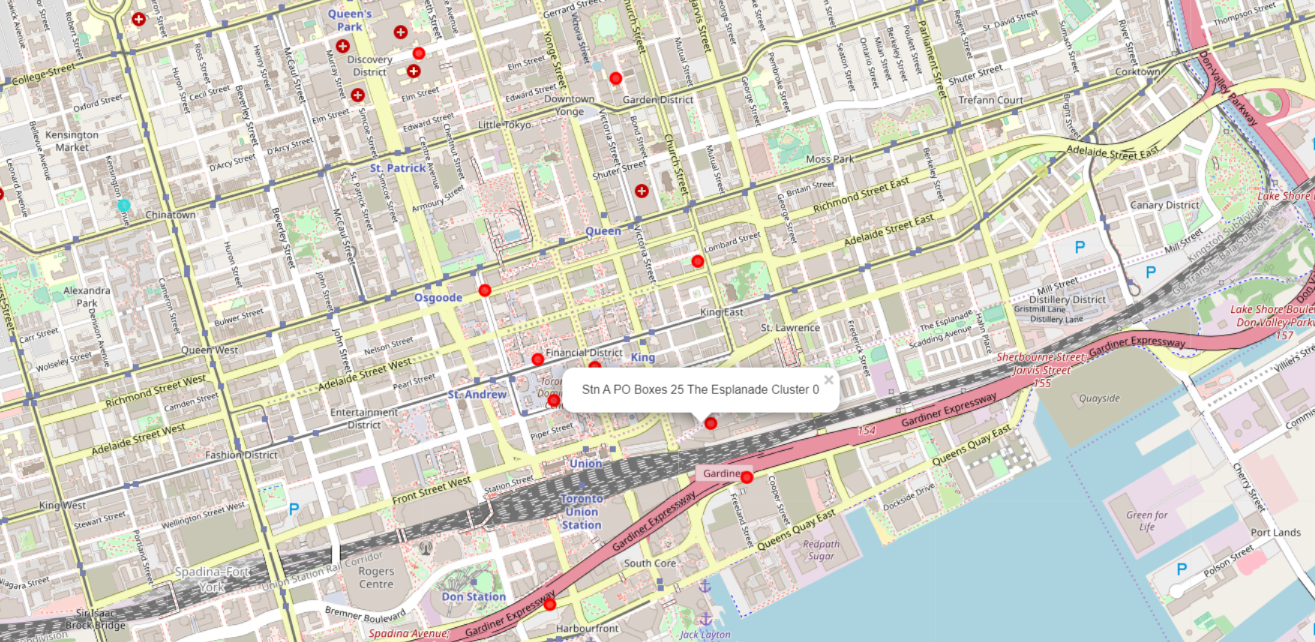
Utilizing the Foursquare API, I analyzed each Toronto neighborhood creating a new data frame with the top 15 most common venues.



To create neighborhood profiles based on nearby venues and the frequencies visited, I used k-means clustering to create four clusters for the neighborhoods in Toronto. Focusing on the client’s current neighborhood, the Esplanade, I created two bar plots that included the most commonly frequented venues in the neighborhood. This resulted in a list of the 1st and 2nd most commonly frequented venues to be compared with the neighborhoods in Manhattan.



Toronto Clusters Map:

****

**3. Methodology & Analysis**

**3.1 Methodology**

Utilizing the data gathered from the Foursquare API, this project aims to assist a Real Estate agent in determining a neighborhood in Manhattan most similar to the client's current neighborhood in Toronto. By leveraging venue data from Foursquare’s Places API and k-means clustering unsupervised machine learning algorithm, we can better understand the diversity of each neighborhood and easily find the most similar Manhattan neighborhood.

First, we collected from the New York City Dataset, a list of Boroughs and Neighborhoods in New York City. We then created a dataframe listing all the neighborhoods within Manhattan and their corresponding latitudes and longitudes, excluding all other boroughs, which we could use to analyze using the Foursquare API. Second, we performed the same procedure to compose a dataframe that listed the neighborhoods in Toronto that only contained the word Toronto and their corresponding latitudes and longitudes. We did this by first creating a dataframe from the Toronto Dataset and joining it with the Toronto Longitude and Latitude Coordinates Dataset, excluding the neighborhoods that did not contain the word Toronto.

Third, using the Foursquare API, we extracted information on the venues in Manhattan. We then created Manhattan neighborhood profiles by using the frequency the venues are visited and k-means clustering machine learning algorithm to compare and contrast the profiles created of each neighborhood. Fourth, using the Foursquare API, we extracted information on the venues in Toronto. We then created Toronto neighborhood profiles by using the frequency the venues are visited and k-means clustering machine learning algorithm to compare and contrast the profiles created of each neighborhood. Focusing on the client's current neighborhood, The Esplanade, we were able to create a neighborhood profile that included the 1st and 2nd most common venues.

Next, we will analyze the Manhattan neighborhood profiles we created based on the most frequented venues in each neighborhood. We will compare the Manhattan neighborhood profiles of each of the four clusters with the Toronto the Esplanade neighborhood profile. We will then identify which Manhattan neighborhood most closely matches the characteristics of The Esplanade neighborhood and propose to the client three current houses within the optimal neighborhood using the Zillow API. Lastly, we will discuss and reasons why the optimal Manhattan neighborhood and three homes to propose to the client were selected.

**3.2 Analysis**

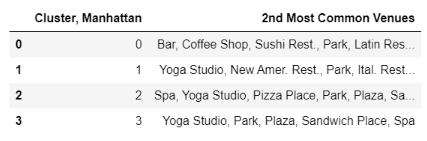
The Neighborhood Profile of the Client's Current Neighborhood in Toronto, the Esplanade:





Manhattan Neighborhood Clusters:

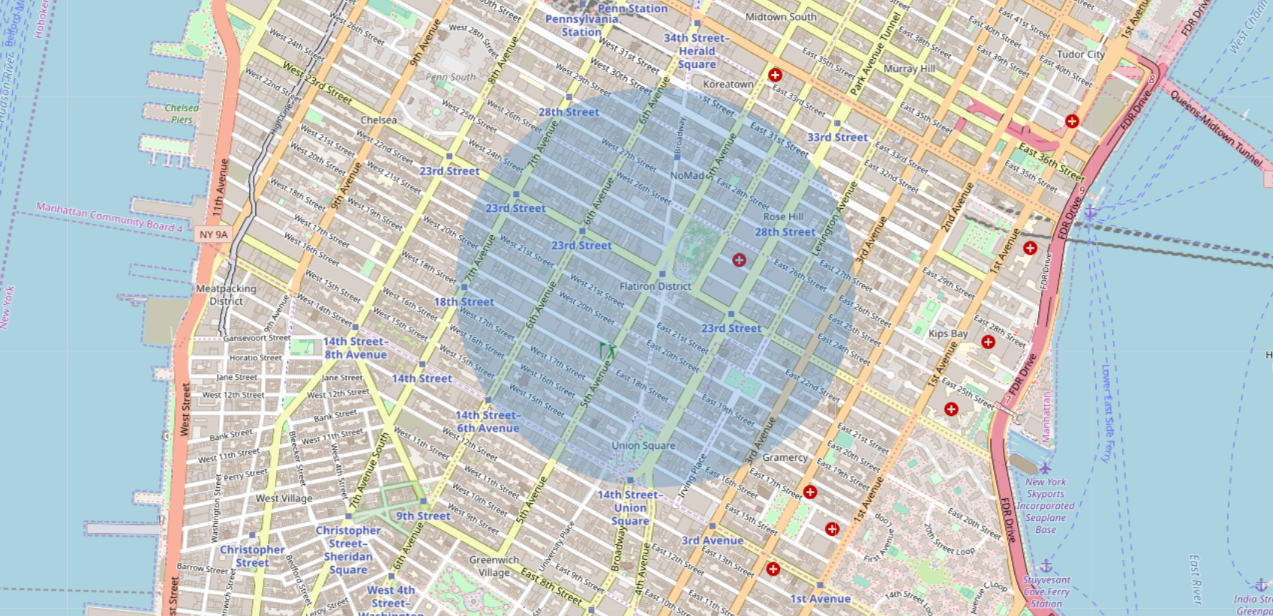




Comparing the Manhattan Neighborhood Clusters to the Esplanade Profile:



It appears the Flatiron District neighborhood would be most similar to the Esplanade and closest to the client's future work at the Google office in Chelsea.



Using the Zillow API, we will determine three current listings for a co-op/condo in the Flatiron District to propose to the client:

Listing 1: 16 West 16th Street, APT 7AS, New York, NY, 10011. Type: Cooperative. 1 Bedroom & 1 Bath. 750 sq. ft. Price: $965,000.

Listing 2: 61 Irving Place, APT 1D, New York, NY, 10003. Type: Cooperative. 1 Bedroom & 1 Bath. 950 sq. ft. Price: $999,000.

Listing 3: 21 E 22nd Street, APT 2A, New York, NY, 10010. Type: Cooperative. 0 Bedroom & 1 Bath. 925 sq. ft. Price: $925,000.

**4. Results and Discussion**

As our analysis has shown that the Esplanade neighborhood profile when compared to Manhattan neighborhood profiles we created, most closely matches the Flatiron District neighborhood. The neighborhood profiles were created based on K-Means Clustering, a form of unsupervised machine learning, along with venue frequency data gathered by using the Foursquare API. The Flatiron District neighborhood appeared to be most comparable to the Esplanade neighborhood in terms of venue frequency and would also be a short commute to client's future employment at the Google office in Chelsea. Based on the results, the Chelsea neighborhood appeared to be a close tie for second. Based on this information we gathered using the Zillow API 3 listings, most similar to the client's current home at 1 A The Esplanade Ave # 2008, Toronto, ON, M5E 0A8. While the client's current home is a 2 bedroom & 2 bathroom condo, listed for C$899,000, almost all 2 bedroom & 2 bathroom condos listed in the Flatiron District were priced significantly higher. The three listings proposed were based first on price and then on sq. ft. and number of bedrooms and bathrooms. Of the three listings, Listing 1 appeared to be the best fit for the client.

The resulting Manhattan neighborhoods listed in cluster 1 (cluster 0) contained the largest number of potential optimal neighborhoods based on the number and frequency of venues. I believe the information gathered, joined with other demographics and statistics not explored in this project, would determine which of the potential neighborhoods would be the very best match. The purpose of this analysis, to create neighborhood profiles based on data gathered via the Foursquare API, was to demonstrate and provide information on areas in Manhattan that might closely match the client's current neighborhood in Toronto. The recommended neighborhood should be considered as a starting point for a more detailed analysis which may result in a location where many other factors have been taken into account.

**5. Conclusion**

As our analysis has shown that the Esplanade neighborhood profile when compared to Manhattan neighborhood profiles we created, most closely matches the Flatiron District neighborhood. The neighborhood profiles were created based on K-Means Clustering, a form of unsupervised machine learning, along with venue frequency data gathered by using the Foursquare API. While each neighborhood profile was based on the venue frequency from Foursquare, there are many other demographics and statistics that could be applied in creating neighborhood profiles.

The primary stakeholders, interested in a new way to use quantifiable analysis to understand and profile a neighborhood would be Real Estate Agents and Real Estate Buyers. Previously neighborhood profiles have always been aggregated and compared based on historic, statistical, and/or demographic information. However, I believe a new approach based on venues and often they’re visited for creating neighborhood profiles in order to compare other neighborhood profiles can provide a basis for a much more accurate area profile.

**The area in Manhattan determined to be most similar to the client's current neighborhood in Toronto, the Esplanade, is the Flatiron District.**

**As a Real Estate Agent in New York City, the three listings I would propose that are most similar to client's current home are:**

**Listing 1: 16 West 16th Street, APT 7AS, New York, NY, 10011. Type: Cooperative. 1 Bedroom & 1 Bath. 750 sq. ft. Price: $965,000.**

**Listing 2: 61 Irving Place, APT 1D, New York, NY, 10003. Type: Cooperative. 1 Bedroom & 1 Bath. 950 sq. ft. Price: $999,000.**

**Listing 3: 21 E 22nd Street, APT 2A, New York, NY, 10010. Type: Cooperative. 0 Bedroom & 1 Bath. 925 sq. ft. Price: $925,000.**

**Sources:**

1. Farfan, Barbara. “Google’s Global Headquarters and Offices Around the World.” The Balance Small Business, The Balance Small Business, 5 Aug. 2019, www.thebalancesmb.com/google-headquarters-offices-2892790.
2. Wikipedia contributors. "Toronto." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 26 Jan. 2020. Web. 26 Jan. 2020.
3. Wikipedia contributors. "New York City." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 19 Jan. 2020. Web. 26 Jan. 2020.
4. Robertson, Becky. “Google Is Opening Another Office in Toronto.” BlogTO, 5 Oct. 2019, www.blogto.com/tech/2019/10/google-opening-another-office-toronto/.