BATTLE OF THE NEIGHBORHOODS

NEW YORK CITY AND TORONTO

CONTENTS

1. Introduction

The Umbrella Corporation is considering expanding to North America where there is potentially a large demand for their services. One of their businesses is a boutique targeting the mass affluent market. They want to explore whether the cities of New York or Toronto which could both be potential locations for the new North American business.

Objectives

They have commissioned this project to analyse both cities, based on the following objectives:

- How large is the population, and therefore what is the size of the potential target for the local customer base?
- What are some of the key characteristics of some of the larger neighbourhoods in each city?
- What are the typical salary levels in each city? As we will have to recruit staff and will also be helpful for some of our relocating employees.
- What are the typical housing costs for each city? Will also be useful data to identify the higher and lower income neighbourhoods.
- What are the main types of retailers and other businesses already established? To give an indication of potential competitors and partners.

2. Data

The data is based on a number of open sources. The methodology to obtain the data will vary depending on the source.

Wherever possible the data is collected directly from the webpages, either through scraping or by reading downloadable csv or Excel files.

Businesses data from Foursquare is obtained from their open API for developers.

The intention is to report on the following items:

- Overall population levels in each city and some information on neighbourhood populations where available.
- Demographic data, showing average incomes.
- Housing costs, including average rents.
- The nature and distribution of different types of businesses and amenities in a selection of neighbourhoods in each city.

Main data sources used

	New York City	Toronto
Housing Costs	https://www.zillow.com/research/data/	http://trreb.ca/
Population Data	https://uspopulation2020.com/population-of-new-york-city-2020-largest-city-of-new-york-state.html https://www.census.gov/quickfacts/fact/table/newyorkcountymanhattanboroughnewyork,bronxcountybronxboroughnewyork,queenscountyqueensboroughnewyork,kingscountybrooklynboroughnewyork,richmondcountystatenislandboroughnewyork,newyorkcitynewyork/INC110218	https://populationstat.com/canada/toronto
Income Data	https://www.averagesalarysurvey.com/new-york-city-united-states	https://www.averagesalarysurvey.com/toronto-canada
Geographical Data	US Zip codes with latitude/longitude https://gist.github.com/erichurst/7882666	Toronto neighborhoods: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M Latitude and Longitude: http://cocl.us/Geospatial_data
Local Businesses Data	Foursquare.com data downloaded using developer open API: https://foursquare.com/	

3. Methodology

Housing, population and income data

Data was obtained from open sources online – either scraped directly from the tables on the website or imported using csv or Excel files provided by the sites as downloadable data.

Business data

Latitude and Longitude coordinates mapped to each location were used for requests made to the Foursquare API server.

Foursquare API "search" and "venues" parameters were used to obtain listings of businesses within a selection of neighborhoods in each city. A maximum of 100 venues per neighborhood within a radius of 500metres was specified.

Analysis

Extensive side by side comparisons were carried out based on the cities as well as some more detailed analysis for a selection of neighborhoods.

"K- means clustering" (unsupervised machine learning) was used for the analysis of the business data from Foursquare. Extensive analysis was carried out on a number of neighborhoods which are in the notebooks but are not in this report.

Results

Python visualization libraries were used to output the results such as charts as well as interactive map based visuals.

Data cleaning

Raw data sources were compared against at least two other independent sources for reasonableness before being downloaded.

Data cleaning often resulted in large changes to the original dataset. Validation involved checking the attributes and size of the revised dataframe against previous copies during the cleaning process. Spot checks of the final data were made against the original data to ensure that data wrangling had not corrupted the data.

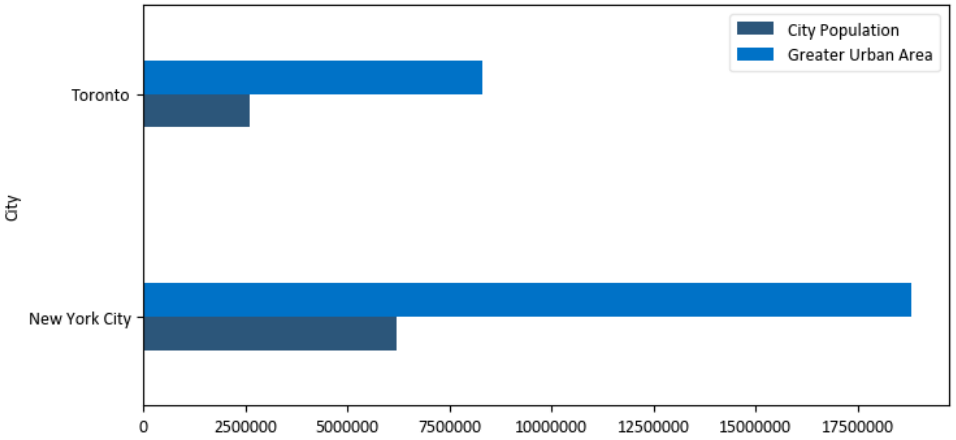
The following Python Libraries were used:

Library	Description/Use
Pandas, Numpy	Data Cleaning, Validation and Analysis
JSON	Working with JSON source files
Geopy	Obtaining location data
Sklearn	Python analytics and machine learning
Matplotlib	Python charts and visualizations
Folium	Mapping library for interactive geospatial visualizations
BeautifulSoup	For scraping data tables directly from webpages

4. Results and Discussion

Population comparison

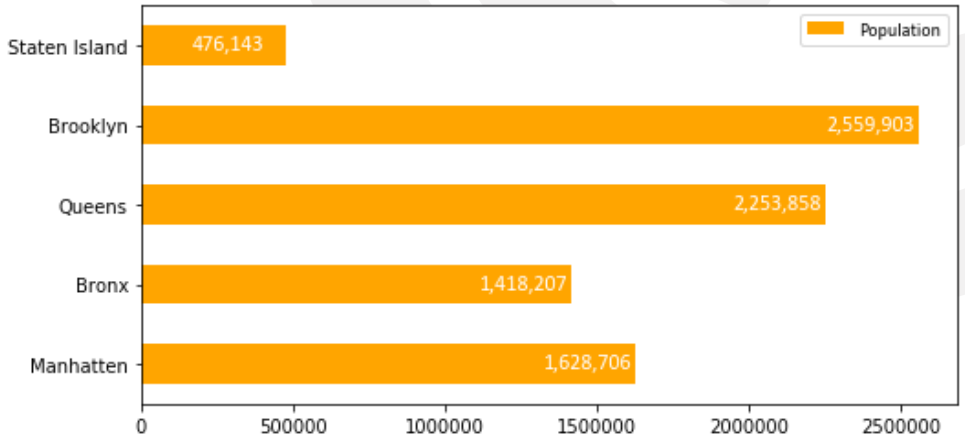
Overall Population



New York City's population (6.2million) is over twice the size of Toronto's city population (2.6million).

Considering the wider surrounding urban areas, over 18million people live in the greater New York area within commuting distance, compared with 8million for Toronto.

New York Population by Borough

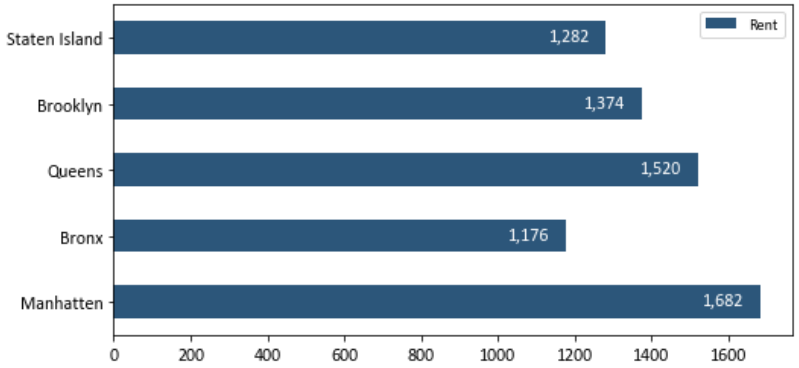


New York City has 5 boroughs. Queens and Brooklyn with 2.5million people are both similar size to the whole population in Toronto's city area.

We have a much larger potential target customer base and potential labour force to hire from in New York.

New York: Housing Costs and Income Levels

Rental Costs for New York City

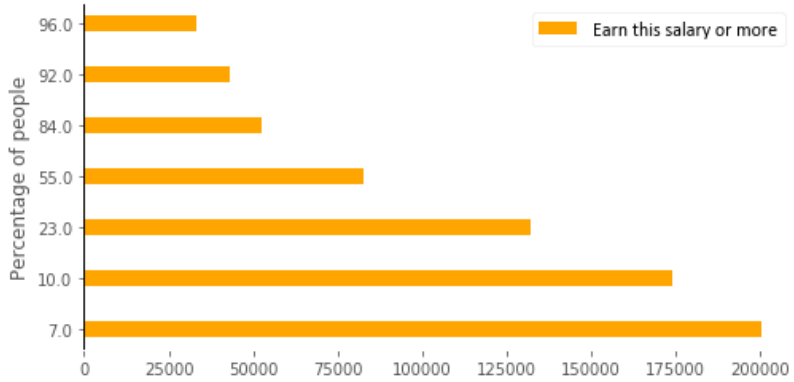


This comparison of housing costs against salary levels provides an indication of disposable income, as rent will usually be one of the largest monthly expenses.

Rental costs are taken from census data and are median values. There is a wide distribution of rents by neighborhood.

Manhatten rents of around $1,700 per month ($20,200 p.a.) are over 40% higher than the Bronx. As the business' target customers are higher end businesses this suggests Manhatten as a potential location to consider in more detail.

Distribution of salaries for New York City



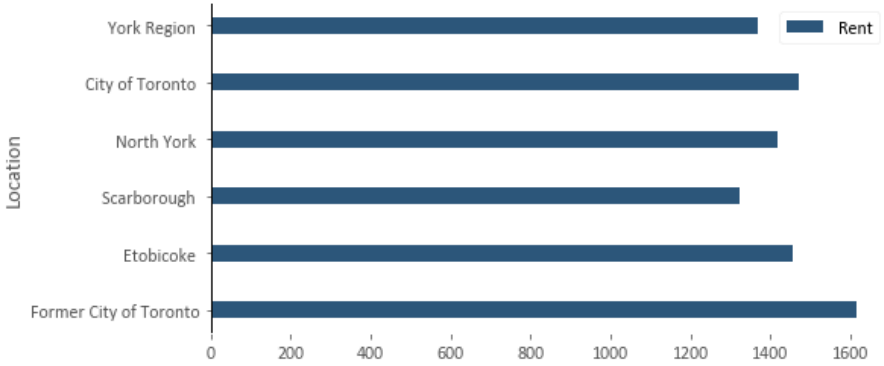
Market salary data may be used as reference for budgeting purposes when hiring staff.

Over 55% of respondents for this market survey earn over $80,000 p.a.

The distribution of salary levels in the survey also suggests that housing costs make up around 25% of income levels.

Toronto: Housing Costs and Income Levels

Rental Costs for Toronto



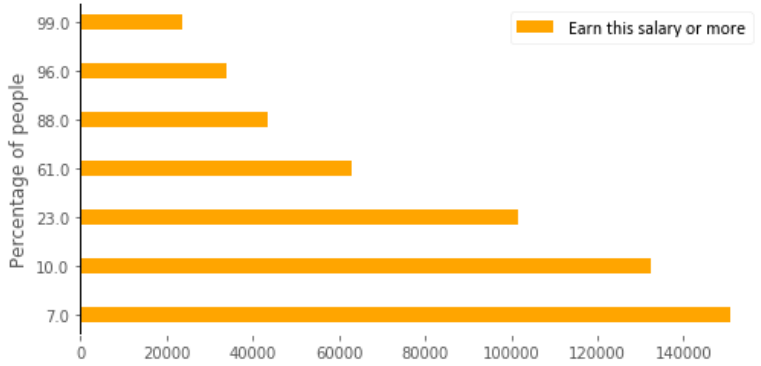
A similar comparison can be made for Toronto.

Rental costs are taken from market survey data and are median values. There is not as much variation in the distribution of rents by neighborhood compared to New York.

Rents of around CAD$1,600 per month (CAD$19,200 p.a.) for the “former city of Toronto” neighborhood are around 20% higher than the lowest rent neighborhood of Scarborough.

Also in absolute terms, Toronto is a much cheaper city to live in as US$1 is around CAD$1.36.

Distribution of salaries for New York City

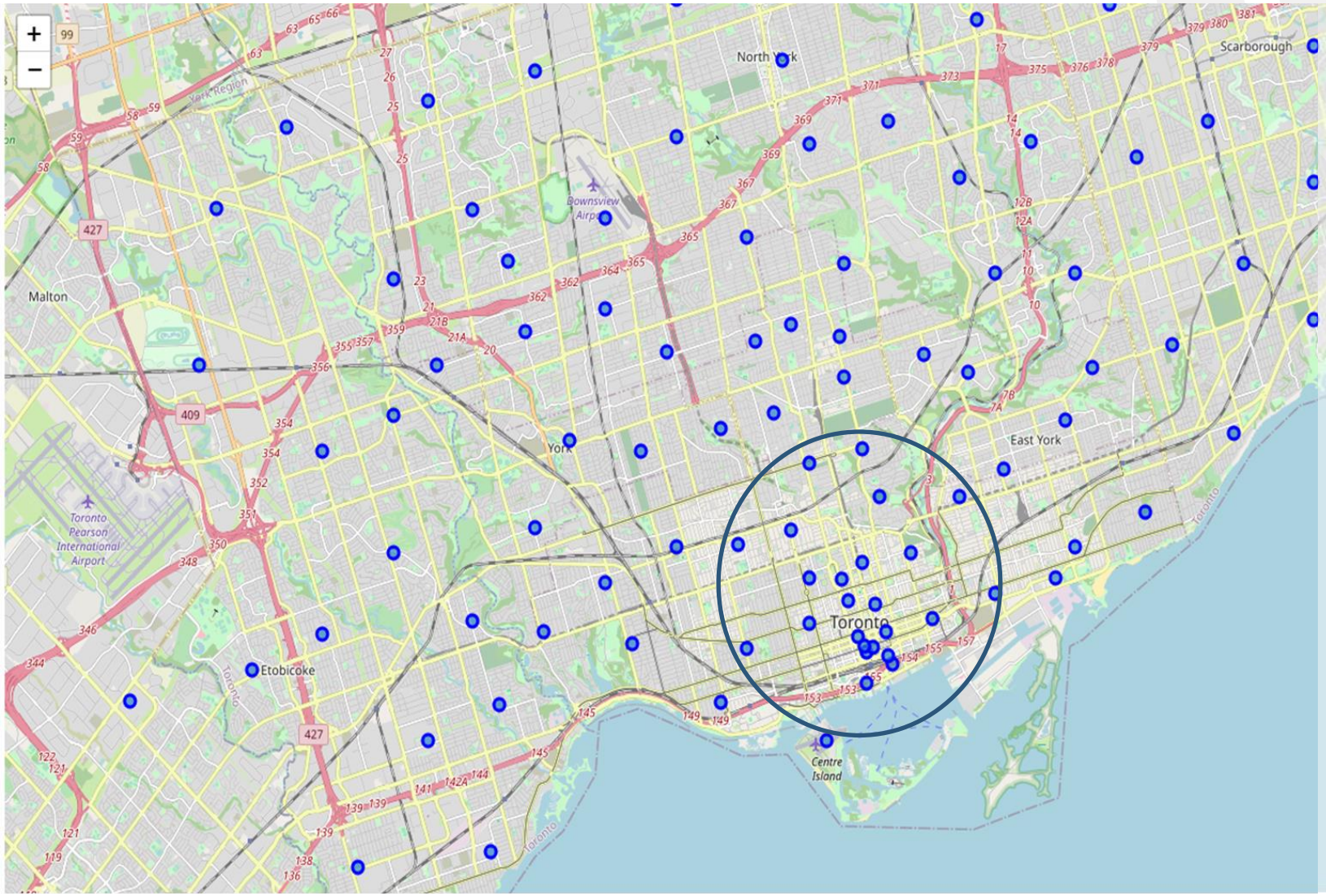


Market salary data may be used as reference for budgeting purposes when hiring staff.

Over 60% of the market survey earn over CAD$60,000 p.a. (around $44,000).

The distribution of salary levels in the survey also suggests that housing costs in Toronto make up around 30% of income levels. Therefore customers may have less disposable income in Toronto than in New York. However this would also suggest lower hiring costs.

Toronto: Downtown

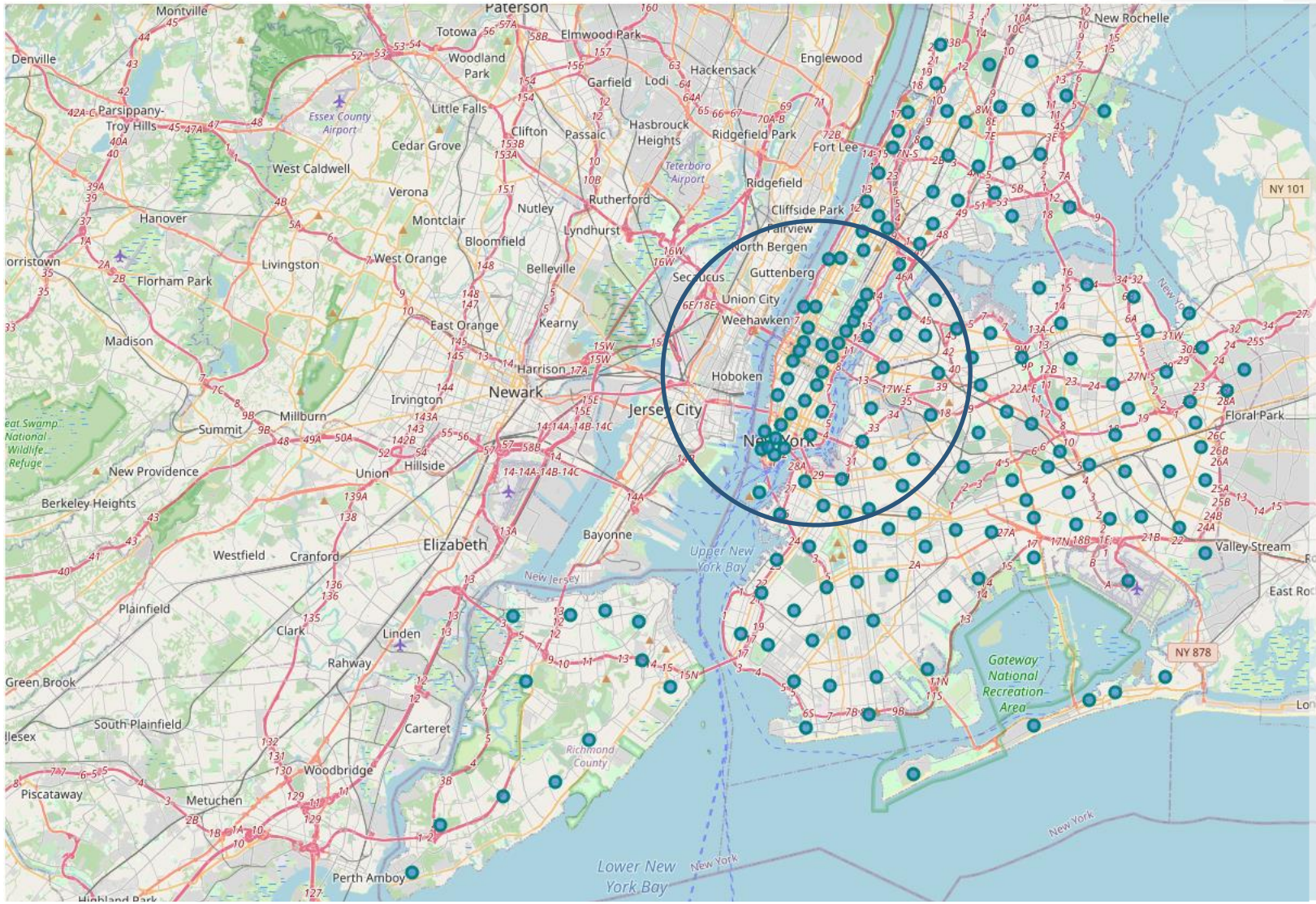


Downtown Toronto will be the focus of our analysis of businesses in Foursquare.

The housing data suggests this is one of the more affluent areas that the business should be targeting as a potential location.

There are **1,224 venues** in this area.

New York: Manhatten



Manhatten will be the focus of our analysis of businesses in Foursquare.

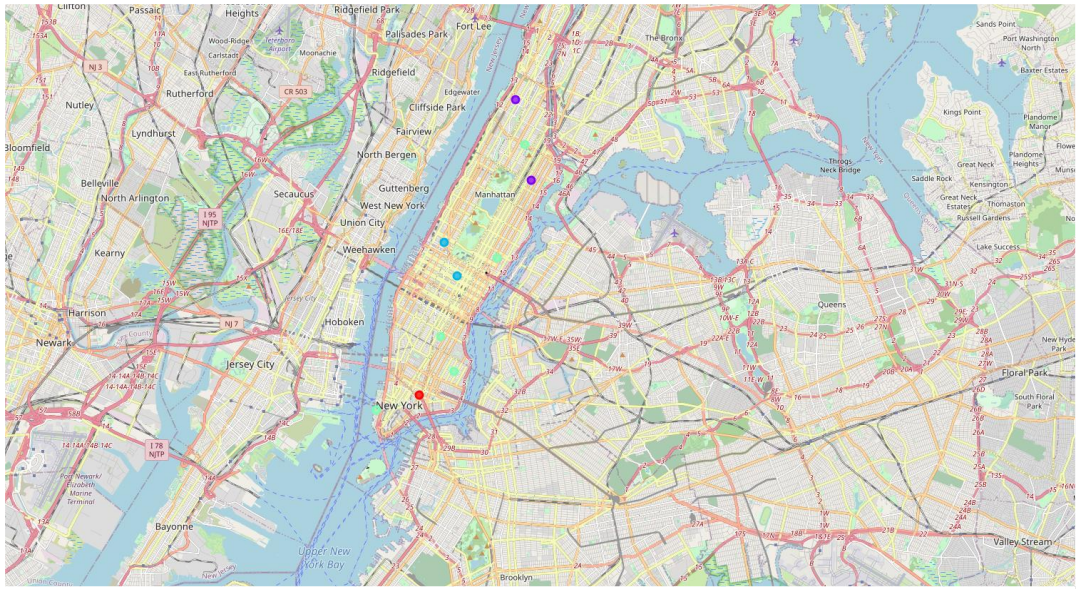
This is the most affluent area that the business should be targeting as a potential location.

There are **3,575 venues** in this area, which **3x** the number of venues in downtown Toronto.

There is also with a much wider variation.

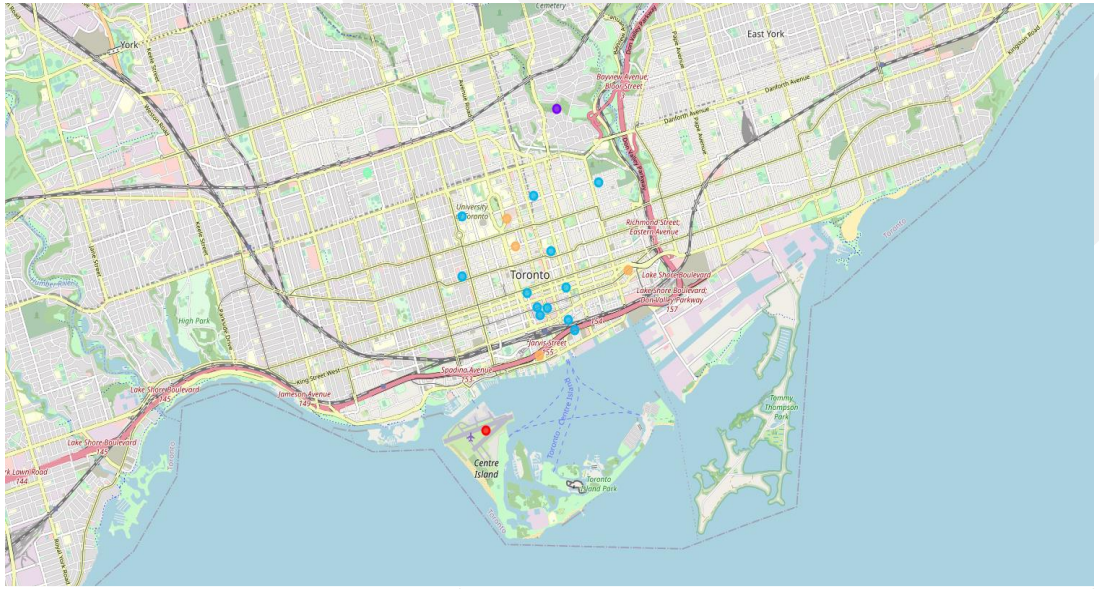
Foursquare analysis – most common venues

K – means clustering results: Manhatten



Overall Number of venues	3575
Cluster 1	Chinese Restaurant
Cluster 2	Pizza Place, Bakery
Cluster 3	Hotel, Italian Restaurant
Cluster 4	Park, Bar, Wine shop, Coffee Shop, Clothing Store
Cluster 5	Bakery

K – means clustering results: Downtown Toronto



Overall Number of venues	1,224
Cluster 1	Airport Services
Cluster 2	Park
Cluster 3	Coffee Shop
Cluster 4	Grocery Store
Cluster 5	Coffee Shop

5. Conclusion

Manhatten, New York

Based on the desktop analysis conducted in this report we would recommend the Manhatten neighborhood in New York based on the following:

- Larger target market of mass affluent customers.
- Higher disposable incomes than Toronto would potentially mean higher $ value of sales and revenue.
- 3x as many venues than downtown Toronto.
- Greater diversity of businesses than Toronto.
- Larger pool of potential hires.

For the next stage in the process, we would suggest further analysis conducted on the costs of setting up businesses, government and state incentives and taxes.

Also other sources of data on businesses should also be analyzed as an independent source, as we have wholly relied on Foursquare.

THANK YOU!

https://github.com/hp-analytica/Coursera_Capstone