Capstone Project: The Battle of Neighbourhoods (Week 2)

By: Sanya Siddiqui

Introduction

- Toronto, Ontario is known as one of the most multicultural and cosmopolitan cities in the world. As Canada's most populous city, it is recognized as a hub for business, technology, arts, culture, sports and food. With its highly diversified economy, Toronto is attracting more and more people to its urban centre. It is estimated that over the next 50 years, the population of the city will double, and the density will grow at a thrilling rate from 4,000 people per square kilometre today to more than 7,700 in 2066.

https://www.ontario-realhome.com/toronto-properties-1

Business Problem

- With this drastically high population growth rate, contractors and developers have a great opportunity to build more apartment buildings and condos to keep up with growing demands. Thus, the target audience for this project are contractors looking to build new residential buildings in the growing city of Toronto.

https://www.rentseeker.ca/rentals/apartments/ontario/toronto

- My client, Contractor X, has hired my team of data scientists to find the best location in Toronto to build their next apartment building complex. The selection criteria will be based off two main factors: 1) wards with lower apartment building densities and 2) wards with already existing nearby amenities for future tenants (e.g. grocery stores, fitness centres, shopping malls, parks, bus stations, gas stations, etc.).

Datasets

- To solve this business problem for my client, Contractor X, I will use 3 datasets to find the optimal location to build their apartment building complex:

  1. City Wards Data (https://open.toronto.ca/dataset/city-wards/): To find out the latitude and longitude of each of the 25 wards in Toronto). Data includes AREA_NAME, AREA_LONG_CODE, LONGITUDE, LATITUDE.

  2. Apartment Building Registration Data (https://open.toronto.ca/dataset/apartment-building-registration/): To find out the density of registered apartment buildings for each ward in Toronto). Data includes PCODE, SITE_ADDRESS, WARD, YEAR_REGISTERED.

  3. Foursquare Data: To find out which amenities (venue categories) are located in each ward. Data includes: WARD_NAME, WARD_CODE, LONGITUDE, LATITUDE, VENUE_LATITUDE, VENUE_LONGITUDE, VENUE_CATEGORY.

Data Importing, Processing & Cleaning

- The City Wards data was first imported from the City of Toronto open data catalogue using the pandas read_csv function and then processed into a workable data frame using the data features of interest (renamed to WARD, WARD_NAME, LATITUDE and LONGITUDE)

```
df_citywards = df[['AREA_LONG_CODE','AREA_NAME','LATITUDE','LONGITUDE']]
df_citywards = df_citywards.rename(columns={'AREA_LONG_CODE':'WARD','AREA_NAME':'WARD_NAME'})
df_citywards.sort_values(['WARD'], ascending = True, axis = 0, inplace = True)
df_citywards
```

	WARD	WARD_NAME	LATITUDE	LONGITUDE
14	1	Etobicoke North	43.719405	-79.584667
16	2	Etobicoke Centre	43.664431	-79.552534
15	3	Etobicoke-Lakeshore	43.621646	-79.520874
13	4	Parkdale-High Park	43.650121	-79.467340
22	5	York South-Weston	43.694951	-79.493371

- The Apartment Building Registration Data was also imported from the City of Toronto open data catalogue using the pandas read_csv feature. In order to get a count of the total number of apartment buildings in each ward, the value_counts function was used.

```
aptcount = df2['WARD'].value_counts()
aptcount.index.name = 'WARD'
aptcount.sort_index(ascending = True, axis = 0, inplace = True)
aptcount.drop(labels =['YY'], inplace=True)
df_apartments = aptcount.to_frame('APT_BUILDINGS_COUNT')
df_apartments.index = pd.to_numeric(df_apartments.index)
df_apartments.reset_index()
df_apartments
```

	APT_BUILDINGS_COUNT
WARD	
1	81
2	130
3	240
4	197
5	233

Data Importing, Processing & Cleaning (Cont.)

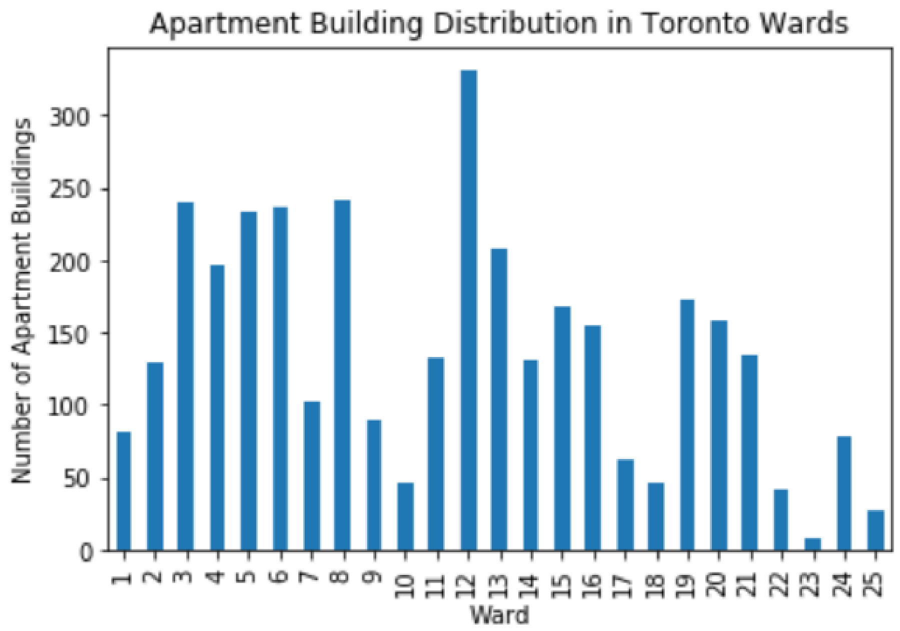
- The final step in the data importing, processing and cleaning process was to merge the two dataframes into one for easy data exploration. The pandas merge function was used, with WARD being the column in common of both dataframes.

```
df_citywards = pd.merge(df_citywards,right = df_apartments,on = 'WARD')
df_citywards
```

	WARD	WARD_NAME	LATITUDE	LONGITUDE	APT_BUILDINGS_COUNT
0	1	Etobicoke North	43.719405	-79.584667	81
1	2	Etobicoke Centre	43.664431	-79.552534	130
2	3	Etobicoke-Lakeshore	43.621646	-79.520874	240
3	4	Parkdale-High Park	43.650121	-79.467340	197
4	5	York South-Weston	43.694951	-79.493371	233
5	6	York Centre	43.752565	-79.467810	236

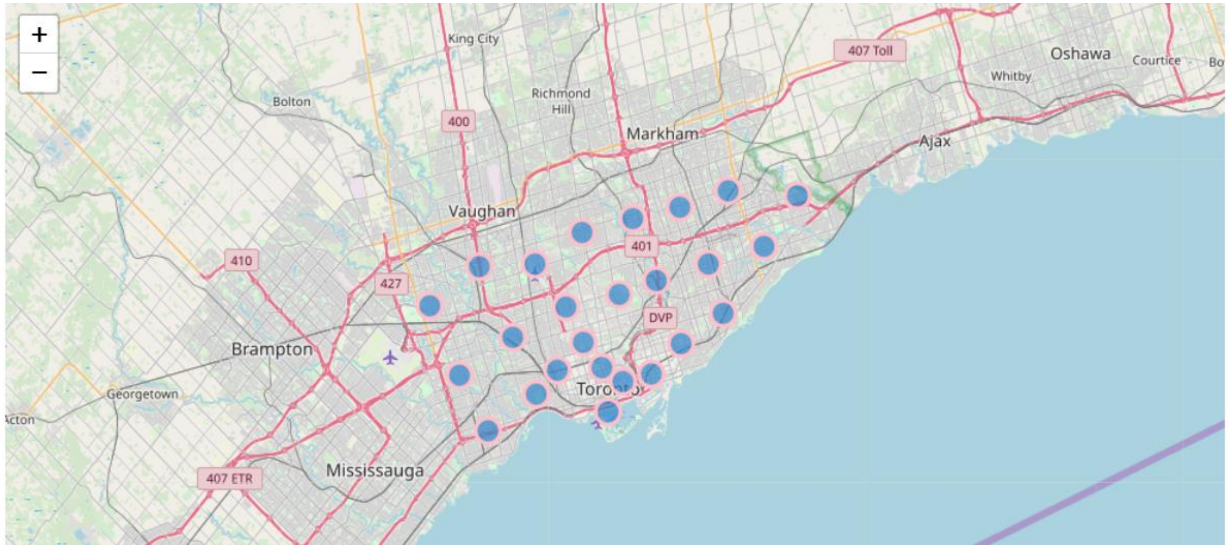
Data Exploration & Visualization (Bar Chart)

- The first visualization tool used on the data was a bar chart. A bar chart is a type of plot where the length of each bar of the independent variable is proportional to the value of the dependent variable. In this case, the independent variable is WARD and the dependent variable is APT_BUILDINGS_COUNT. After importing the matplotlib and its scripting interface, the plot(kind = 'bar') function was used to plot the bar chart. The chart below visualizes the apartment building distribution in the 25 Toronto wards.



Further Data Visualization (folium maps)

- Further visualization was done using folium maps. When using folium maps, the necessary libraries were first imported including json, geopy.geocoders and folium among others. Using the geopy library, the latitude and longitude values of Toronto were obtained. A map of Toronto was created using the folium.Map function. Markers for each of the wards were then superimposed on top using a function. In this function, the ward longitude and latitude values from the df_citywards dataframe was passed through and a circular marker was placed on the map for each ward.



Foursquare Location & Venue Data

- Next, the Foursquare API was used to the explore the wards in the df_citywards dataframe to obtain the top venues in each ward. After obtaining my Foursquare credentials, and passing the ward longitude and latitude values in my function, I sent a get request to extract the category of the venues into a new pandas dataframe, as shown. The get_dummies function was used on the data frame by passing the VENUE CATEGORY data feature for one hot encoding. The rows in the resulting data frame were grouped using the WARD_NAME and by taking the mean of the frequency of occurrence of each venue category, as shown in the table below.

	WARD_NAME	American Restaurant	Amphitheater	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bank	...	Taiwanese Restaurant	Tennis Court	Thai Restaurant	The
0	Beaches-East York	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
1	Davenport	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	Don Valley East	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	...	0.0	0.0	0.0	0.0
3	Don Valley North	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	...	0.1	0.0	0.0	0.0
4	Don Valley West	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

Data Analysis & Interpretation

- In order to address the business problem of our client, the top five most common venue categories in each ward were identified and analyzed. A dataframe was created using several functions to display the venue categories. K-means clustering was performed on this dataframe and it was then merged with the existing df_citywards dataframe. The values in the dataframe were sorted using the sort_values function by apartment building density. Because our wards of interest are those with the lowest apartment building densities a new data frame was created with the five wards with the lowest apartment building counts.

```
#Filter data to the five wards with the lowest apartment building densities
df_wards_low = df_wards_data.iloc[20:,]
df_wards_low
```

	WARD_NAME	LATITUDE	LONGITUDE	APT_BUILDINGS_COUNT	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
WARD									
10	Spadina-Fort York	43.635801	-79.387335	47	Yoga Studio	Garden	Coffee Shop	Brewery	Performing Arts Venue
18	Willowdale	43.776886	-79.416500	46	Pizza Place	Bubble Tea Shop	Korean Restaurant	Seafood Restaurant	Coffee Shop
22	Scarborough-Agincourt	43.796483	-79.308135	42	Chinese Restaurant	Grocery Store	Hotpot Restaurant	Greek Restaurant	Caribbean Restaurant
25	Scarborough-Rouge Park	43.805647	-79.176842	28	Gas Station	Pharmacy	Home Service	Pizza Place	Bus Station
23	Scarborough North	43.809672	-79.254671	8	Fast Food Restaurant	Indian Restaurant	Coffee Shop	Business Service	Salon / Barbershop

Amenities Scoring System

- Next, a scoring system was created to select best candidate ward from the wards of interest. Each amenity was given a score of either 1, 2 or 3 depending on its importance as a basic necessity, with 3 being the most important. The table shows a snippet of the scores for each amenity in the wards of interest.
- Using this scoring system, each one of the top five amenities in each ward was assigned a score. The sum of the scores was calculated for each ward and the ward with the highest score was recommended to the client as an ideal apartment building development site.

Amenity	Score
Bubble Tea Shop	1
Bus Station	3
Business Service	2
Brewery	1
Caribbean Restaurant	1
Chinese Restaurant	1
Coffee Shop	1
Fast Food Restaurant	1
Garden	2
Gas Station	3
Greek Restaurant	1
Grocery Store	3

Results

- After following the rigorous data science methodology outlined above, a final result was obtained as a recommendation to the client. Each of the amenities in the wards of interest (Wards 10, 18, 22, 25 and 23) was assigned with a score. The scores were tallied and as shown in the results table below, the ward with the highest score was Ward 25, Scarborough-Rouge Park. The results table acts as a decision matrix that is often used in engineering design.

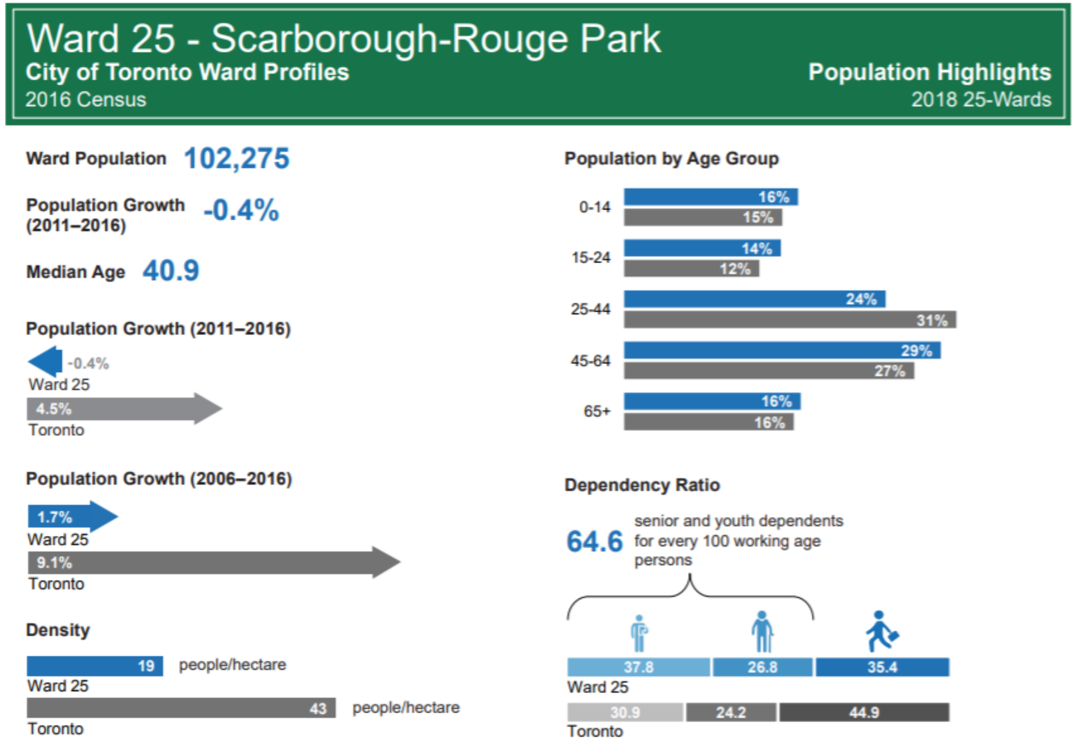
	WARD_NAME	Amenity 1 Score	Amenity 2 Score	Amenity 3 Score	Amenity 4 Score	Amenity 5 Score	Total Score
WARD							
10	Spadina-Fort York	2	2	1	1	2	8
18	Willowdale	1	1	1	1	1	5
22	Scarborough-Agincourt	1	3	1	1	1	7
25	Scarborough-Rouge Park	3	3	2	1	3	12
23	Scarborough North	1	1	1	2	2	7

Discussion

- As seen in the results, a ward was selected as a viable option for the development of an apartment building complex. It is important to discuss the results with respect to real-world information. Using the Ward 25 – Scarborough-Rouge Park profile found on the City of Toronto website (https://www.toronto.ca/wp-content/uploads/2018/09/9708-City_Planning_2016_Census_Profile_2018_25Wards_Ward25.pdf), we can use the 2016 census data to understand the population, dwellings and socioeconomic highlights.

Population Highlights

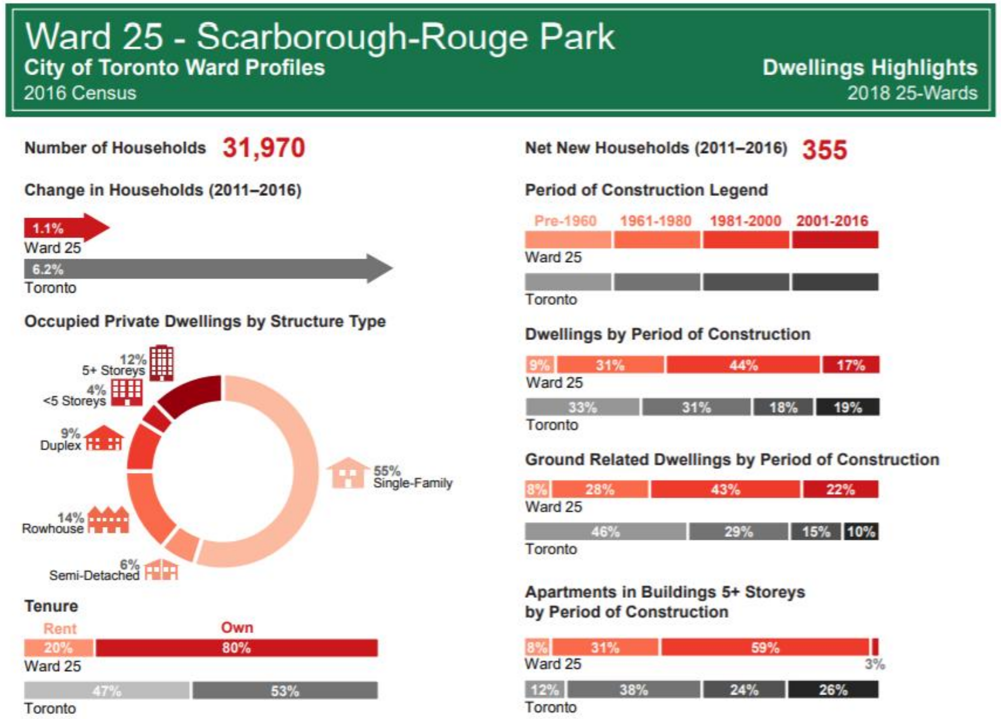
- Starting off, we can see that the population density of Ward 25 is about half of that of Toronto as whole. From this information, we can see that Ward 25 is not very congested with people and that with the predicted growth of the city, it would be an ideal location for new residents to live.



https://www.toronto.ca/wp-content/uploads/2018/09/9708-City_Planning_2016_Census_Profile_2018_25Wards_Ward25.pdf

Dwellings Highlights

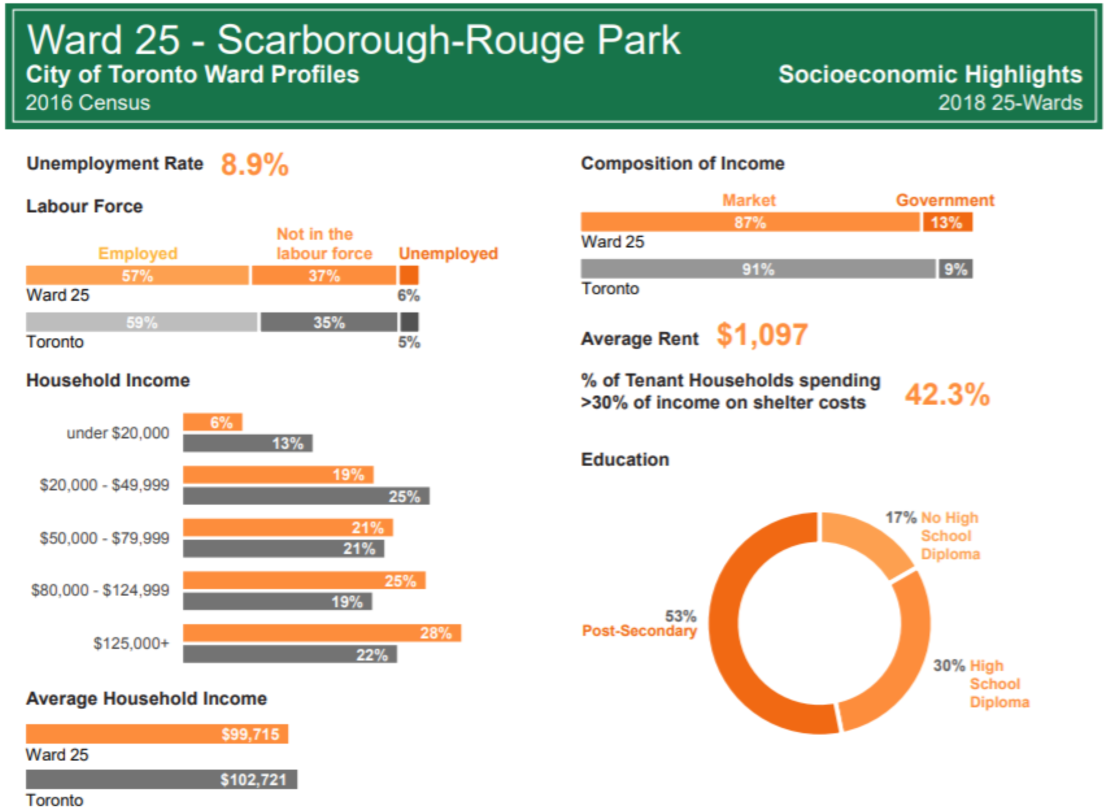
- When looking at the dwellings highlights data, we can see that only 12% of dwellings are apartment buildings. Most of the construction of apartment buildings (59%) occurred between 1981 and 2000. Therefore, there is a clear lack of newer apartment buildings with more modern facilities and would be a good fit for a contractor to begin new development.



https://www.toronto.ca/wp-content/uploads/2018/09/9708-City_Planning_2016_Census_Profile_2018_25Wards_Ward25.pdf

Socioeconomic Highlights

- One challenge for our client would be to adjust rental costs to compete with existing costs in the ward. According to the socioeconomic highlights, the average rent in Ward 25 is $1,097 which is substantially lower than other wards in Toronto. The challenge for our client will be to construct a new apartment building complex in this ward that can accommodate to this cost and still provide attractive facilities for future tenants.



https://www.toronto.ca/wp-content/uploads/2018/09/9708-City_Planning_2016_Census_Profile_2018_25Wards_Ward25.pdf

Recommendations

- Upon the completion of this data science project, I have several recommendations for future work.
- The first recommendation I have is to conduct a market research study to accurately identify the top amenities that attract most tenants. In this project, there were various amenities that fell into different categories, including fitness, arts, cuisine, shopping etc. It would be a good idea to understand which amenities are most important to tenants and how much of a mix of the different categories is necessary to attract a diverse clientele. By implementing this recommendation, the scoring system would hence be chosen by the target clientele, rather than being personally biased to me.
- Another recommendation I have is to include census data as another dataset in this study. Using population data, dwellings data, households & family data, immigration & mobility data and socioeconomic data, we can create a more rigorous selection system for our client with better real-world understanding.

Conclusions

- In conclusion, with Toronto’s population estimated to drastically double in the next 50 years, more apartment buildings will need to be built to keep up with the city’s growing demands.
- Our task was to identify which one of Toronto’s 25 wards would be an ideal location for our client, Contractor X, to develop a new apartment building complex.
- Following the processing and cleaning of our City Wards and Apartment Building Registration data, a FourSquare API was used to map, explore and analyze the nearby amenities in each ward.
- Once a data frame with the top five amenities for each ward was created, the results were filtered for the wards with the lowest apartment building densities.
- Each amenity in the wards was assigned a score, based on a defined scoring system.
- The ward with the highest score, recommended to our client was Ward 25, Scarborough- Rouge Park.
- Further work can be done to use and assess demographics data to understand other factors and challenges a client may face. In addition, a market research study can be conducted to understand which amenities in the business, technology, arts, culture, sports and food categories are most attractive to potential tenants in Toronto.