Purchasing a Rental Property in Toronto, Ontario

Ankit Sharma

March 13, 2021

Introduction

Business Problem

There has been a significant increase in the individuals looking at owning rental properties as a source of income over the past decade. The real estate market in Toronto, Ontario has been very popular for the past decade, especially with Toronto becoming a business hub and tourist destination. Prices in Toronto are quite high, which is why it is extremely important that consumers make an informed decision when purchasing an income generating rental property.

Location is the key when purchasing a property, and two key characteristics that contributes to the success of an income generating property are the amenities nearby and the crime rate.

The purpose of this study will be to look at the most attractive neighborhoods to purchase property in Toronto based on the access to amenities and crime rates.

Target Audience

The target audience of this report would be anyone who wants to buy an income generating property in the city Toronto, Ontario.

Data Acquisition and Sources

The data used in this project has been linked below. Fortunately, Foursquare has lots of venues available in the city of Toronto compared to other Canadian cities and there are numerous public datasets that describe various aspects of Toronto.

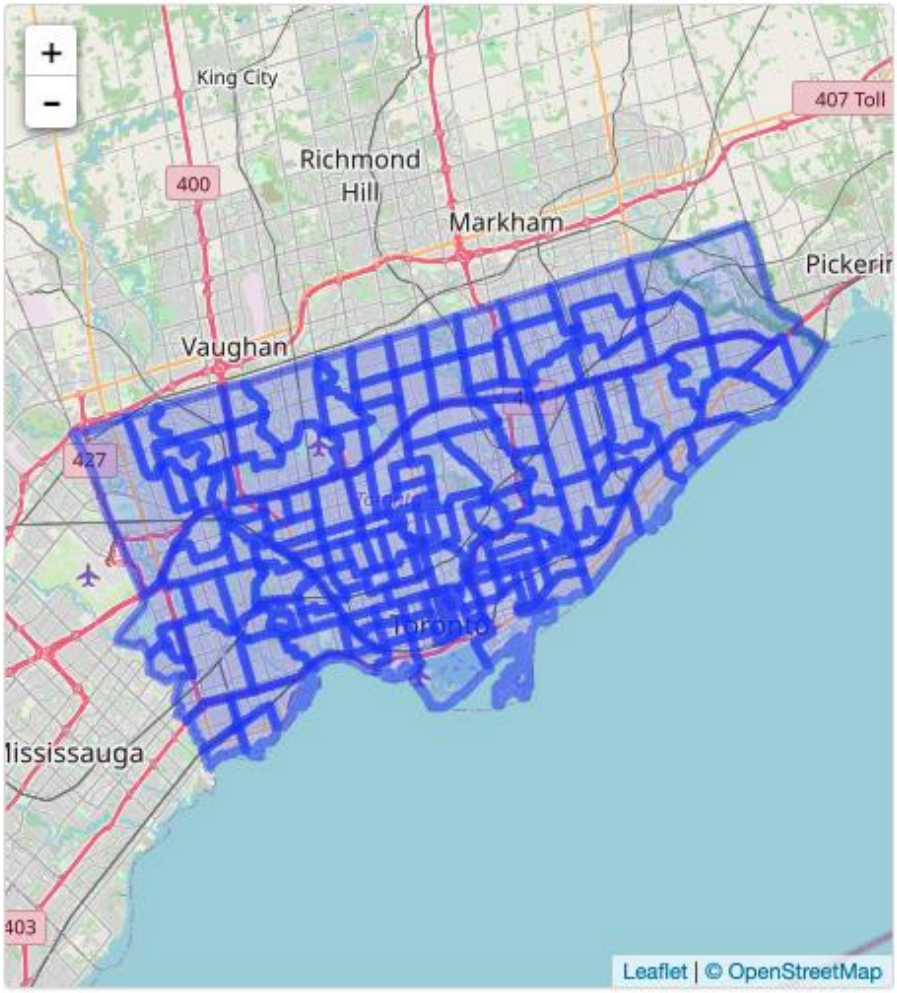
Toronto Crime Data (https://open.toronto.ca/dataset/neighbourhood-crime-rates/)

This dataset includes the 2014-2019 Crime Data by Neighbourhood. Counts are available for Assault, Auto Theft, Break and Enter, Robbery, Theft Over and Homicide. Data also includes five-year averages and crime rates per 100,000 people by neighbourhood based on 2016 Census Population.

About Neighbourhood Crime Rates

This dataset includes the 2014-2019 Crime Data by Neighbourhood. Counts are available for Assault, Auto Theft, Break and Enter, Robbery, Theft Over and Homicide. Data also includes five year averages and crime rates per 100,000 people by neighbourhood based on 2016 Census Population.

DATA PREVIEW



DATA FEATURES

DOWNLOAD DATA

Toronto Postal Code Data (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

This dataset includes data about Postal Codes, Boroughs, and Neighbourhoods in Toronto.

Toronto - 103 FSAs [edit]

Note: There are no rural FSAs in Toronto, hence no postal codes should start with M0. However, the postal code M0R 8T0 is assigned to an Amazon warehouse in Mississauga, and the postal code M0R 2A2 is used for the Gateway postal facility in Mississauga, suggesting that Canada Post may have reserved the M0 FSA for high volume addresses.

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue, Humber Valley Village
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	Not assigned
M3B	North York	Don Mills
M4B	East York	Parkview Hill, Woodbine Gardens
M5B	Downtown Toronto	Garden District, Ryerson

Foursquare API

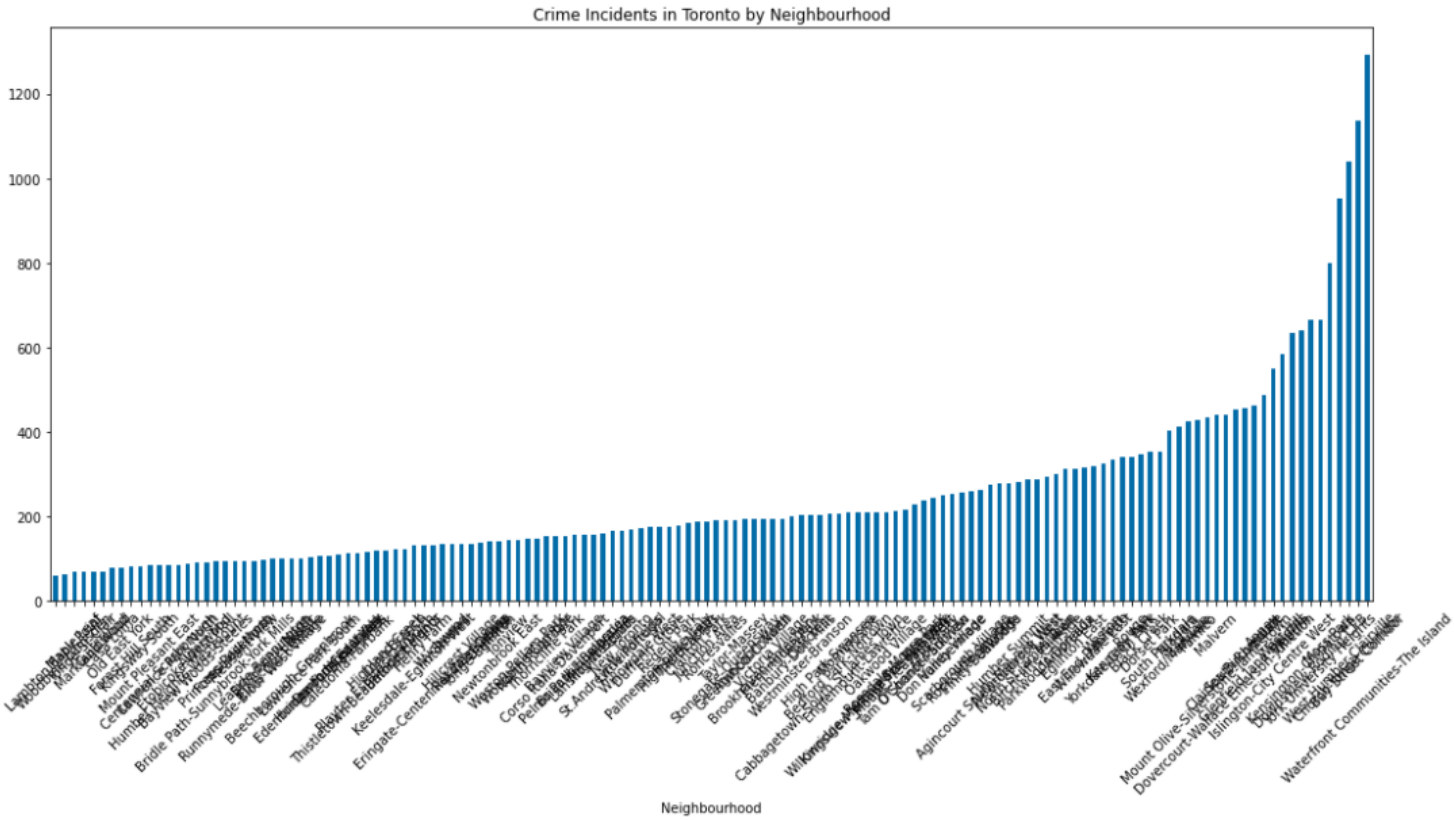
Foursquare is an independent data platform which allows you to retrieve information about the most popular spots in any city. Calling the Foursquare API will allow us to turn location data into a data frame for analysis in python.

Data Analysis Methodology

Crime Data

The crime for every neighbourhood in Toronto was obtained and saved as a .csv file. Once this was loaded on to Jupyter Labs, we used the *pandas* package to manipulate the table to reduce the column into only relevant data.

To simplify the data, we averaged the crimes reported annually from 2014 to 2019 into an *Average Crimes Per Year* for each neighbourhood.



It was visually evident that an average of 200 Total Crimes per year is a below average number for the city of Toronto. I decided to use an Average Total Crime of "less than 200 incidents" from our crime dataset and considered those neighbourhoods as the "safe" neighbourhoods where I would promote buying a rental property.

```
[7]: #get a dataframe with the 10 lowest incident neighborhoods
     cols = ['TotalCrime_AVG']
     crime3[cols] = crime3[crime3[cols] < 200][cols]
     crime4 = crime3.dropna()
```

Postal Code Data

The data was obtained from a *Wikipedia* article utilizing the *BeautifulSoup* method to scrape the table from the web. All postal codes with Boroughs listed as “Not assigned” were removed from this data.

[11]:
```
# Remove all rows that have "Not assigned"
toronto_data = toronto_data[toronto_data.Borough != 'Not assigned']
toronto_data = toronto_data[toronto_data.Neighbourhood != 'Not assigned']

toronto_data.tail()
```

[11]:

	Postal Code	Borough	Neighbourhood
160	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
165	M4Y	Downtown Toronto	Church and Wellesley
168	M7Y	East Toronto	Business reply mail Processing Centre, South C...
169	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

Merging the Data

The *Crime Data* and *Postal Code* data were merged to obtain a final dataset which provides the dataset fundamental to this project.

[14]:
```
#start by merging the datasets and making a new dataset
rental_data = toronto_data1.merge(crime4, on=['Neighbourhood'])

rental_data
```

[14]:

	PostalCode	Neighbourhood	Population	TotalCrime_AVG
0	M4A	Victoria Village	17510	193.7
1	M5A	Regent Park	10803	176.6
2	M6C	Humewood-Cedarvale	14365	107.8
3	M1E	Guildwood	9917	68.8
4	M4E	The Beaches	21567	185.9
5	M2H	Hillcrest Village	16934	135.2
6	M3H	Bathurst Manor	15873	121.1
7	M4H	Thorncliffe Park	21108	152.0
8	M6J	Little Portugal	15559	169.9
9	M2K	Bayview Village	21396	154.7
10	M8V	New Toronto	11463	166.5
11	M8W	Alderwood	12054	91.0

Obtaining Longitude and Latitude

Geopy is utilized to obtain longitude and latitude coordinate for each neighbourhood in our analysis.

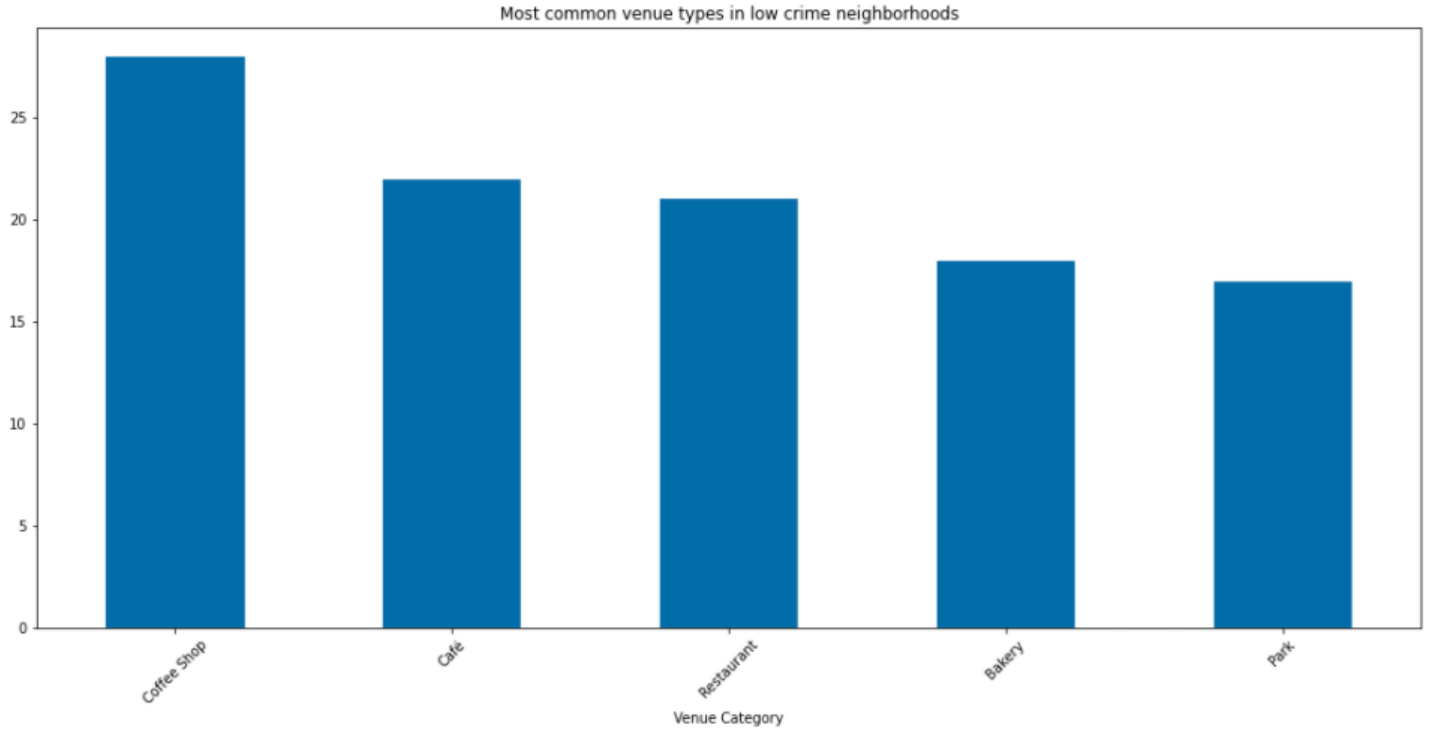
[15]:
```
from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent="Toronto,Ontario")
rental_data['Coordinates'] = rental_data['Neighbourhood'].apply(geolocator.geocode).apply(lambda x: (x.latitude, x.longitude))
rental_data
```

[15]:

	PostalCode	Neighbourhood	Population	TotalCrime_AVG	Coordinates
0	M4A	Victoria Village	17510	193.7	(43.732658, -79.3111892)
1	M5A	Regent Park	10803	176.6	(43.6607056, -79.3604569)
2	M6C	Humewood-Cedarvale	14365	107.8	(43.6883215, -79.4280805)
3	M1E	Guildwood	9917	68.8	(43.7552251, -79.1982293)
4	M4E	The Beaches	21567	185.9	(43.6710244, -79.296712)
5	M2H	Hillcrest Village	16934	135.2	(43.6816953, -79.4257118)
6	M3H	Bathurst Manor	15873	121.1	(43.76389295, -79.45636693710946)
7	M4H	Thorncliffe Park	21108	152.0	(43.704553, -79.3454074)
8	M6J	Little Portugal	15559	169.9	(43.64741325, -79.43111632546047)
9	M2K	Bayview Village	21396	154.7	(43.7691966, -79.3766617)
10	M8V	New Toronto	11463	166.5	(43.6007625, -79.505264)
11	M8W	Alderwood	12054	91.0	(43.6017173, -79.5452325)

Foursquare Data Analysis

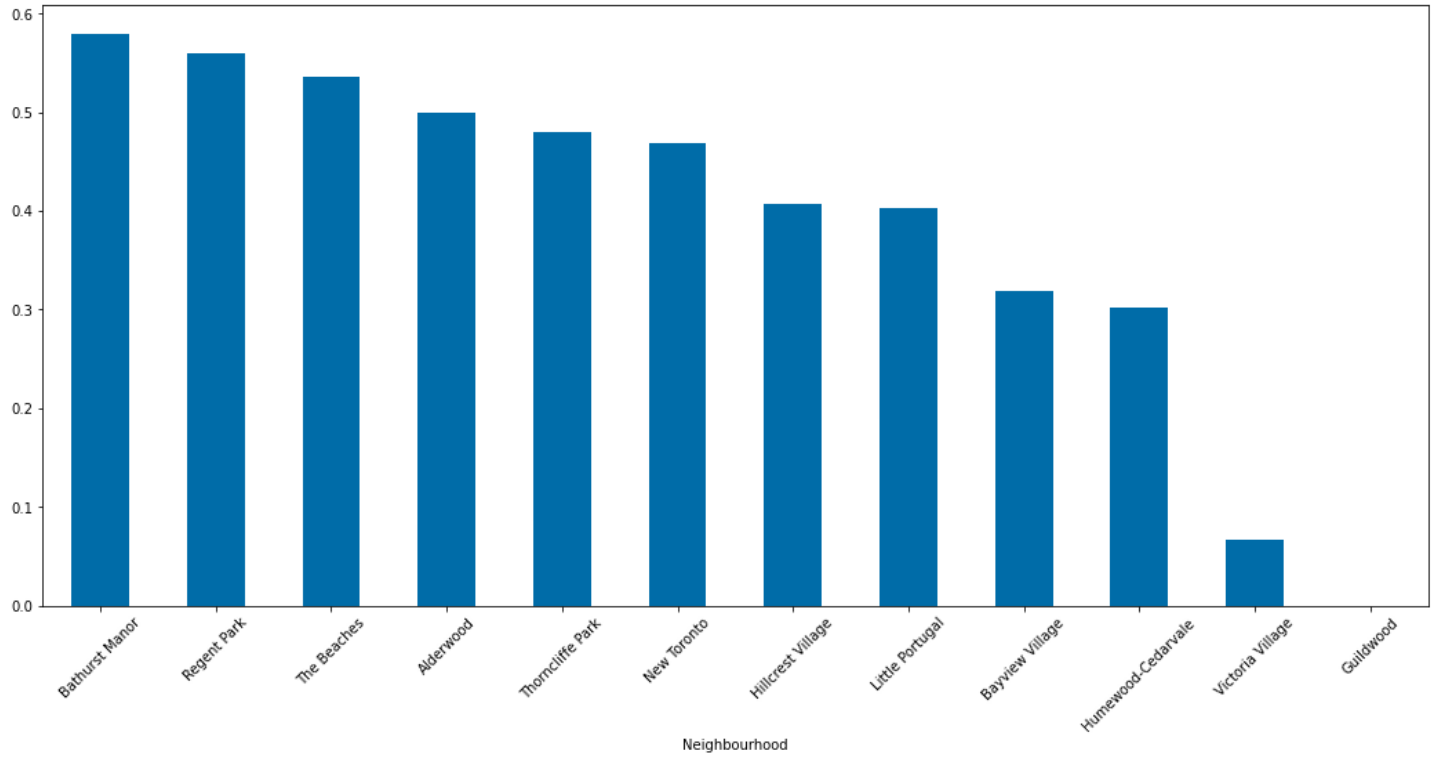
The Foursquare API is leveraged to obtain all venues for each of the neighbourhoods. The most common venue types for these low crime neighbourhoods was obtains.



Doing some research online, it was evident that in order to determine the “best” neighbourhoods to purchase a rental property, the following venues needed to be present **within 800 meters** of the property.

1. Train Station *(Only one neighbourhood [Guildwood] has a Train Station nearby, so we can leave it out of the analysis below and keep it as an FYI variable to influence our analysis later)*
2. Grocery Store
3. Supermarket
4. Restaurants
5. Bars
6. Cafes
7. Coffee Shops

We utilised the Foursquare analysis to find which neighbourhoods had the highest availability of the venues mentioned above.



It is clear the Guildwood should not be considered for this analysis due to its low availability of venues; the only venue that promotes purchasing in Guildwood is its proximity to a Train Station.

Clustering the Data

Utilizing K means clustering on the Foursquare data and then mapping it with *Folium*:

[79]:
```
# set number of clusters
kclusters = 6

rental_grouped_clustering = rental_grouped.drop('Neighbourhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(rental_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

[79]:
```
array([3, 1, 5, 2, 5, 1, 1, 5, 5, 1], dtype=int32)
```

Results and Discussion

We have pulled data on crime rates to filter out neighborhoods in Toronto that are deemed as "unsafe" and leveraged this data with information regarding popular venues that determine the value of a neighbourhood's rental property value. Our analysis has informed us that:

- Coffee shop, Cafes, Bakeries, Restaurants, and Parks are the most common venues in low crime neighborhoods.
- The neighbourhood Guildwood has a transit station, but there are not many amenities nearby, therefore it should not be considered a good candidate to purchase a rental property.
- When Grocery Stores, Supermarkets, Restaurants, Bars, Cafes, and Coffee Shops are used to determine the "best" neighbourhood to buy a rental property, Bathurst Manor comes out on top.
- Clustering the data yielded results that described similar neighbourhoods:
    - The Beaches, Humewood-Cedarvale, Bathurst Manor, and Little Portugal were similar
    - Regent Park, Hillcrest Village, and New Toronto, Bayview Village were similar

The top 3 neighbourhoods in Toronto to buy a valuable property are:

1. Bathurst Manor
2. Regent Park
3. The Beaches

After these three neighbourhoods, there is a downward step change in the number of amenities available. Amenities are what attract tenants, and amenities are what ensure that you will continue to have tenants looking for rentals in that specific neighbourhood.

It is important to note that this project could have been a more in-depth analysis but was limited due to being based on Foursquare' s data for popular venues. There are plenty other ways to assess the value of neighborhoods and venue popularity/crime rate is only one of them.

It will be more helpful to look at average income for each neighbourhood and the ratings for the "best venues" in each neighbourhood.

Conclusion

We made it! We went from learning importing python libraries to figuring out the top three neighbourhoods to purchase a rental property in Toronto. The experience of learning Data Science fundamentals in this course was extremely fun and the journey has taught me so much. The tools that can be developed with Data Science have applications in every industry; I plan on using them in the Oil and Gas industry. Curiosity is king when it comes to applying Data Science and formulating an idea to dive into is the steppingstone.

I wish everyone all the best and happy learning!