1. Background and Description of the Problem (Introduction Section)

1.1 Background and Problem Description

Toronto is the most populous city in Canada and the fourth most populous city in North America with a population of about 3 million. It is one of the most multicultural cities in the world and the center of many of Canada's industries. It also has 140 neighborhoods.

Living in Toronto can be expensive, and rents are among the highest in Canada. Hence determining where to live in Toronto is one of the most important decision a newcomer must make. There are lot of factors to consider and few resources on the internet to assist in decision making. The purpose of the project is to analyze available datasets for the neighborhoods to extract insights that will be useful in decision-making.

In this project we will go through step by step process to decide which neighborhoods are ideal to live in. We will analyze the neighborhoods in Toronto to identify the most safe, affordable and closest to the venues we want.

1.2 Target Audience

Who will be more interested in this project? What type of clients or a group of people would be benefitted?

1. Immigrants or Toronto residents who want to relocate to better neighborhoods in Toronto.
2. Businesspeople who want to invest in or open businesses in Toronto.
3. Real estate agents who will use the research to advise their clients.

2. Data Acquisition and Cleaning

2.1 Data Sources

Datasets that will be used for the analysis:

1. Toronto geojson and csv files from Toronto's City Government Open Data Portal. They contain the latitudes and longitudes for the neighborhood boundaries and will be used for creating choropleth maps.

2. Toronto Wellbeing data from Toronto's City Government Open Data Portal for 2014. It contains the number and types of crimes committed and average rents in each neighborhood. It will be used for analyzing the average rent and crime rates in different neighborhoods.

3. Foursquare API will be for fetching details about the venues in Toronto and collecting their names, categories and locations (latitude and longitude). It will be used to get the most common venues in Toronto neighborhoods.

2.2 Data Cleaning

a) Downloading Crime and Rent Data from the Toronto Wellbeing website

Selected and downloaded crime and tenant average rent data from Toronto Wellbeing website to the local machine.

Loaded data into Pandas dataframe.

```
In [3]:  1  df.head()
```

Out[3]:

	Neighbourhood	Neighbourhood Id	Drug Arrests	Assaults	Sexual Assaults	Break & Enters	Robberies	Vehicle Thefts	Thefts	Murders	Tenant Average Rent
0	West Humber-Clairville	1	29	259	31	131	82	316	38	3	945
1	Mount Olive-Silverstone-Jamestown	2	24	213	16	34	81	42	3	1	921
2	Thistletown-Beaumond Heights	3	7	35	3	23	12	13	1	0	887
3	Rexdale-Kipling	4	11	57	5	16	15	22	0	0	857
4	Elms-Old Rexdale	5	9	53	2	9	14	16	0	0	966

Then split it into two dataframes, one containing the crime rates and another to contain the average rent.

```
In [5]:   ▶|   1  toronto_avg_rent = df[['Neighbourhood', 'Neighbourhood Id', 'Tenant Average Rent']]
```

```
In [7]:   ▶|   1  toronto_avg_rent.head()
```

Out[7]:

	Neighbourhood	Neighbourhood Id	Tenant Average Rent
0	West Humber-Clairville	1	945
1	Mount Olive-Silverstone-Jamestown	2	921
2	Thistletown-Beaumond Heights	3	887
3	Rexdale-Kipling	4	857
4	Elms-Old Rexdale	5	966

```
In [9]:   ▶|   1  toronto_crime.head()
```

Out[9]:

	Neighbourhood	Drug Arrests	Assaults	Sexual Assaults	Break & Enters	Robberies	Vehicle Thefts	Thefts	Murders
0	West Humber-Clairville	29	259	31	131	82	316	38	3
1	Mount Olive-Silverstone-Jamestown	24	213	16	34	81	42	3	1
2	Thistletown-Beaumond Heights	7	35	3	23	12	13	1	0
3	Rexdale-Kipling	11	57	5	16	15	22	0	0
4	Elms-Old Rexdale	9	53	2	9	14	16	0	0

Added a total column in the toronto_crime dataframe.

```
In [19]:   ▶|   1  toronto_crime.head()
```

Out[19]:

	Neighbourhood	Drug Arrests	Assaults	Sexual Assaults	Break & Enters	Robberies	Vehicle Thefts	Thefts	Murders	Total
0	West Humber-Clairville	29	259	31	131	82	316	38	3	4445
1	Mount Olive-Silverstone-Jamestown	24	213	16	34	81	42	3	1	2070
2	Thistletown-Beaumond Heights	7	35	3	23	12	13	1	0	470
3	Rexdale-Kipling	11	57	5	16	15	22	0	0	630
4	Elms-Old Rexdale	9	53	2	9	14	16	0	0	515

b) Download Toronto Neighborhood Geojson Data

Loaded geojson file into a python variable named toronto_geo to be used for creating the choropleth maps

c) Download Toronto Neighborhood CSV Data

Loaded Toronto neighborhood data into neighborhoods_coords dataframe with four columns; area_short_code, area_name, latitude and longitude.

```
In [26]: ▶  1  neighborhoods_coords.head()
```

Out[26]:

	AREA_SHORT_CODE	AREA_NAME	LATITUDE	LONGITUDE
0	94	Wychwood (94)	43.676919	-79.425515
1	100	Yonge-Eglinton (100)	43.704689	-79.403590
2	97	Yonge-St.Clair (97)	43.687859	-79.397871
3	27	York University Heights (27)	43.765736	-79.488883
4	31	Yorkdale-Glen Park (31)	43.714672	-79.457108

3. Methodology

3.1 Descriptive Statistics

We looked at the descriptive statistics for Toronto crime and rent dataframes.

```
In [19]: ▶  1  toronto_crime.describe()
```

Out[19]:

	Drug Arrests	Assaults	Sexual Assaults	Break & Enters	Robberies	Vehicle Thefts	Thefts	Murders	Total
count	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000
mean	20.757143	108.421429	11.035714	45.221429	20.935714	23.135714	6.957143	0.521429	236.985714
std	26.473516	102.187870	12.221817	31.438891	20.128083	31.197724	8.673843	1.035123	200.854541
min	0.000000	9.000000	0.000000	8.000000	0.000000	1.000000	0.000000	0.000000	40.000000
25%	7.000000	47.750000	4.750000	24.000000	8.000000	9.000000	2.000000	0.000000	116.750000
50%	12.500000	78.000000	8.000000	37.500000	15.000000	16.000000	4.500000	0.000000	176.500000
75%	24.000000	133.750000	12.000000	57.000000	25.000000	26.250000	8.000000	1.000000	277.750000
max	174.000000	712.000000	88.000000	213.000000	112.000000	316.000000	56.000000	6.000000	1292.000000

Crime rates are low on average on most of the neighborhoods in Toronto.

```
In [21]: ▶  1  toronto_avg_rent['Tenant Average Rent'].describe()
```

```
Out[21]: count     140.000000
         mean     1019.792857
         std       219.621994
         min       631.000000
         25%       878.500000
         50%       972.500000
         75%      1124.750000
         max      2388.000000
         Name: Tenant Average Rent, dtype: float64
```
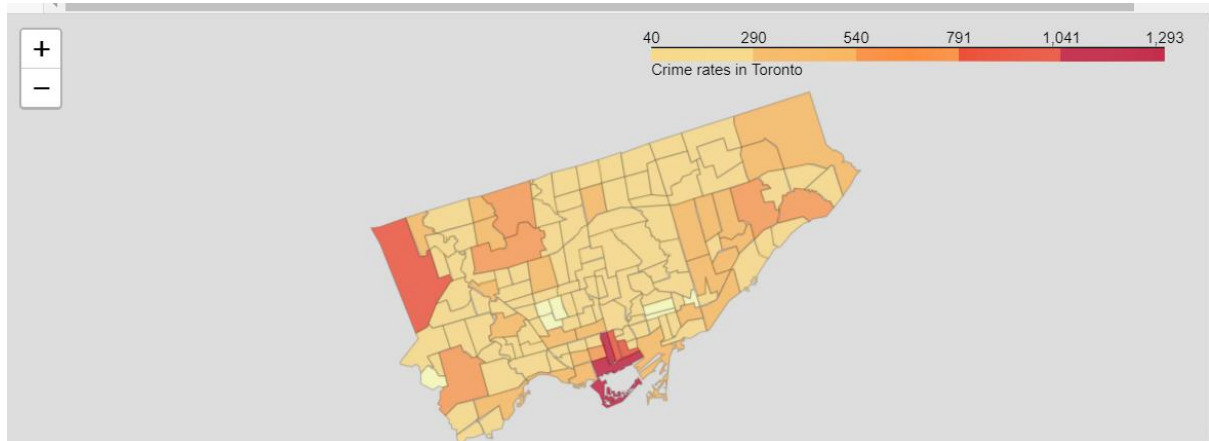
More than 50% of the neighborhoods ask less rent than the average rate of whole of Toronto combined.
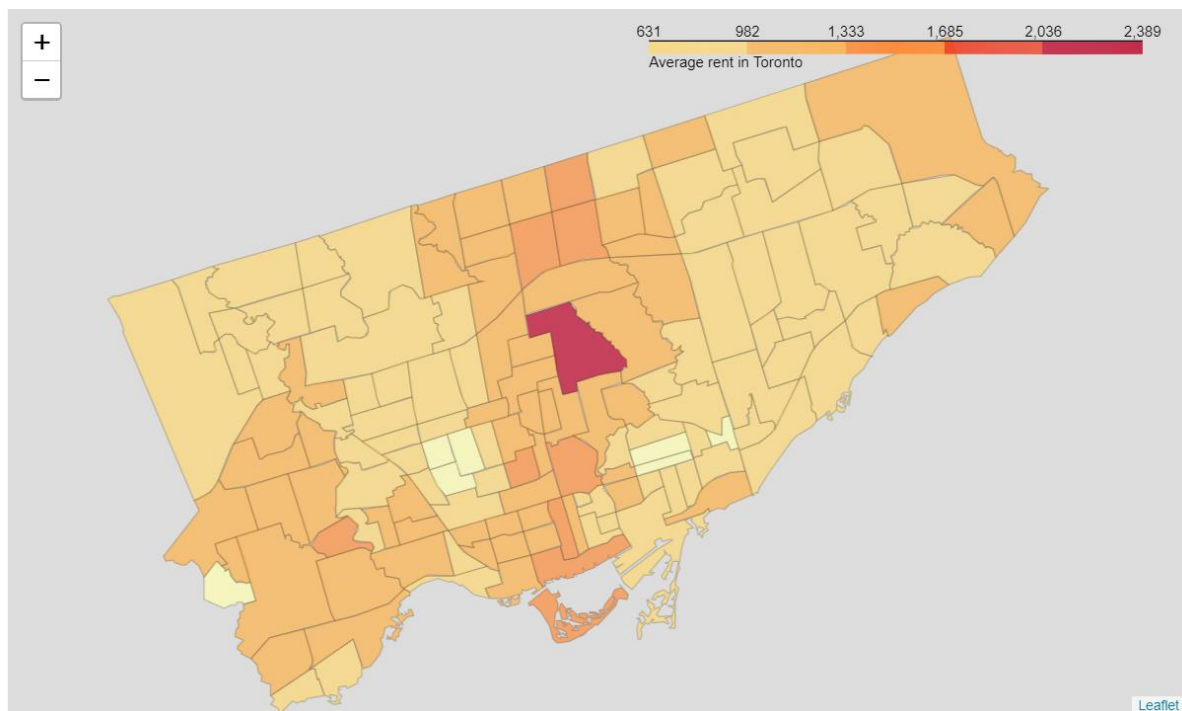
3.2 Choropleth Maps

We use visualization to see crime rates and average rents in the neighborhoods.

1. Toronto Neighborhoods Crime Rates Map

2. Toronto Neighborhoods Average Rents Map

3.3 Clustering Neighborhoods of Toronto

First, we selected 45 safest neighborhoods i.e. neighborhoods will very low crime rates. Then we selected 10 neighborhoods that were more affordable from the safe neighborhoods.

Then we merged the resultant dataframe with the dataframe that contained neighborhood coordinates to create a dataframe that will be used predictive analysis.

```
In [32]:    1  selected_neighourhoods
```
Out[32]:

	Neighbourhood Id	Neighbourhood	Total	Tenant Average Rent	AREA_SHORT_CODE	AREA_NAME	LATITUDE	LONGITUDE
0	48	Hillcrest Village	112	678	48	Hillcrest Village (48)	43.802988	-79.354804
1	28	Rustic	93	740	28	Rustic (28)	43.711609	-79.498091
2	112	Beechborough-Greenbrook	117	749	112	Beechborough-Greenbrook (112)	43.693216	-79.479473
3	69	Blake-Jones	125	778	69	Blake-Jones (69)	43.676173	-79.337394
4	110	Keelesdale-Eglinton West	117	796	110	Keelesdale-Eglinton West (110)	43.685727	-79.471437
5	114	Lambton Baby Point	40	814	114	Lambton Baby Point (114)	43.657420	-79.496045
6	60	Woodbine-Lumsden	70	823	60	Woodbine-Lumsden (60)	43.694107	-79.311164
7	4	Rexdale-Kipling	126	857	4	Rexdale-Kipling (4)	43.723725	-79.566228
8	3	Thistletown-Beaumond Heights	94	887	3	Thistletown-Beaumond Heights (3)	43.737988	-79.563491
9	134	Highland Creek	118	909	134	Highland Creek (134)	43.790775	-79.177472

We will use Foursquare API to get the top 100 venues that are near selected Toronto neighborhoods within a radius of 1000 meters.

```
In [43]:    1  print(selected_neighborhood_venues.shape)
            2  selected_neighborhood_venues.head()
```
```
(62, 7)
```
Out[43]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hillcrest Village	43.802988	-79.354804	Woodbrooke Estate	43.802067	-79.354347	Residential Building (Apartment / Condo)
1	Hillcrest Village	43.802988	-79.354804	McNicoll Park	43.798994	-79.352840	Park
2	Beechborough-Greenbrook	43.693216	-79.479473	McDonald's	43.692926	-79.479982	Fast Food Restaurant
3	Beechborough-Greenbrook	43.693216	-79.479473	York Museum	43.689737	-79.476943	Museum
4	Beechborough-Greenbrook	43.693216	-79.479473	2 Brothers shawarma	43.690118	-79.475150	Turkish Restaurant

We will then get the count of venues by grouping the dataframe by neighborhood column.

```
In [44]:  ▶|  1  selected_neighborhood_venues.groupby('Neighborhood').count()
```

Out[44]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Beechborough-Greenbrook	3	3	3	3	3	3
Blake-Jones	16	16	16	16	16	16
Highland Creek	6	6	6	6	6	6
Hillcrest Village	2	2	2	2	2	2
Keelesdale-Eglinton West	8	8	8	8	8	8
Lambton Baby Point	3	3	3	3	3	3
Rexdale-Kipling	1	1	1	1	1	1
Thistletown-Beaumond Heights	14	14	14	14	14	14
Woodbine-Lumsden	9	9	9	9	9	9

```
In [ ]:  ▶|  1  nnint('Thora ara () uniquos catogonios ' format(lon(soloctod noighhonhood vonuos['Vonuo Catogony'] uniquo()\\\\
```

Encode the data, group again by neighborhood, get the mean of all rows and reset the indexes.

```
▶|  1  selected_neighourhoods_grouped = selected_neighourhoods_onehot.groupby('Neighborhood').mean().reset_index()
    2  selected_neighourhoods_grouped
```

]:

|	Neighborhood	Wine Shop	Asian Restaurant	Bakery	Bank	Beer Bar	Beer Store	Bookstore	Burger Joint	Bus Station	...	River	Sandwich Place	Skating Rink	Spa	Super
0	Beechborough-Greenbrook	0.000	0.000000	0.000	0.000000	0.0000	0.000000	0.0000	0.0000	0.000000	...	0.000000	0.000	0.000000	0.000000	0.
1	Blake-Jones	0.000	0.062500	0.000	0.000000	0.0625	0.000000	0.0625	0.0625	0.000000	...	0.000000	0.000	0.000000	0.000000	0.
2	Highland Creek	0.000	0.000000	0.000	0.000000	0.0000	0.000000	0.0000	0.0000	0.166667	...	0.000000	0.000	0.000000	0.000000	0.
3	Hillcrest Village	0.000	0.000000	0.000	0.000000	0.0000	0.000000	0.0000	0.0000	0.000000	...	0.000000	0.000	0.000000	0.000000	0.
4	Keelesdale-Eglinton West	0.125	0.000000	0.125	0.000000	0.0000	0.000000	0.0000	0.0000	0.000000	...	0.000000	0.125	0.000000	0.000000	0.
5	Lambton Baby Point	0.000	0.000000	0.000	0.000000	0.0000	0.000000	0.0000	0.0000	0.000000	...	0.333333	0.000	0.000000	0.000000	0.
6	Rexdale-Kipling	0.000	0.000000	0.000	0.000000	0.0000	0.000000	0.0000	0.0000	0.000000	...	0.000000	0.000	0.000000	0.000000	0.
7	Thistletown-Beaumond Heights	0.000	0.071429	0.000	0.071429	0.0000	0.000000	0.0000	0.0000	0.000000	...	0.000000	0.000	0.000000	0.000000	0.
8	Woodbine-Lumsden	0.000	0.000000	0.000	0.000000	0.0000	0.111111	0.0000	0.0000	0.000000	...	0.000000	0.000	0.222222	0.111111	0.

Get 10 top venues for each neighborhood.

```
19
20  neighborhoods_venues_sorted.head()
```

t[53]:

	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	Beechborough-Greenbrook	Turkish Restaurant	Fast Food Restaurant	Museum	Gastropub	Garden	Diner	Dim Sum Restaurant	Curling Ice	Convenience Store	Construction & Landscaping
1	Blake-Jones	Café	Coffee Shop	Hostel	Burger Joint	Diner	Dim Sum Restaurant	Nail Salon	Gastropub	Pizza Place	Bookstore
2	Highland Creek	Home Service	IT Services	Bus Station	Construction & Landscaping	Park	Coffee Shop	Garden	Fast Food Restaurant	Diner	Dim Sum Restaurant
3	Hillcrest Village	Park	Residential Building (Apartment / Condo)	Video Store	Caribbean Restaurant	Fast Food Restaurant	Diner	Dim Sum Restaurant	Curling Ice	Convenience Store	Construction & Landscaping
4	Keelesdale-Eglinton West	Pizza Place	Bakery	Latin American Restaurant	Sandwich Place	Wine Shop	Fast Food Restaurant	Thrift / Vintage Store	Coffee Shop	Diner	Dim Sum Restaurant

K-Means to split the selected neighborhoods into six clusters.
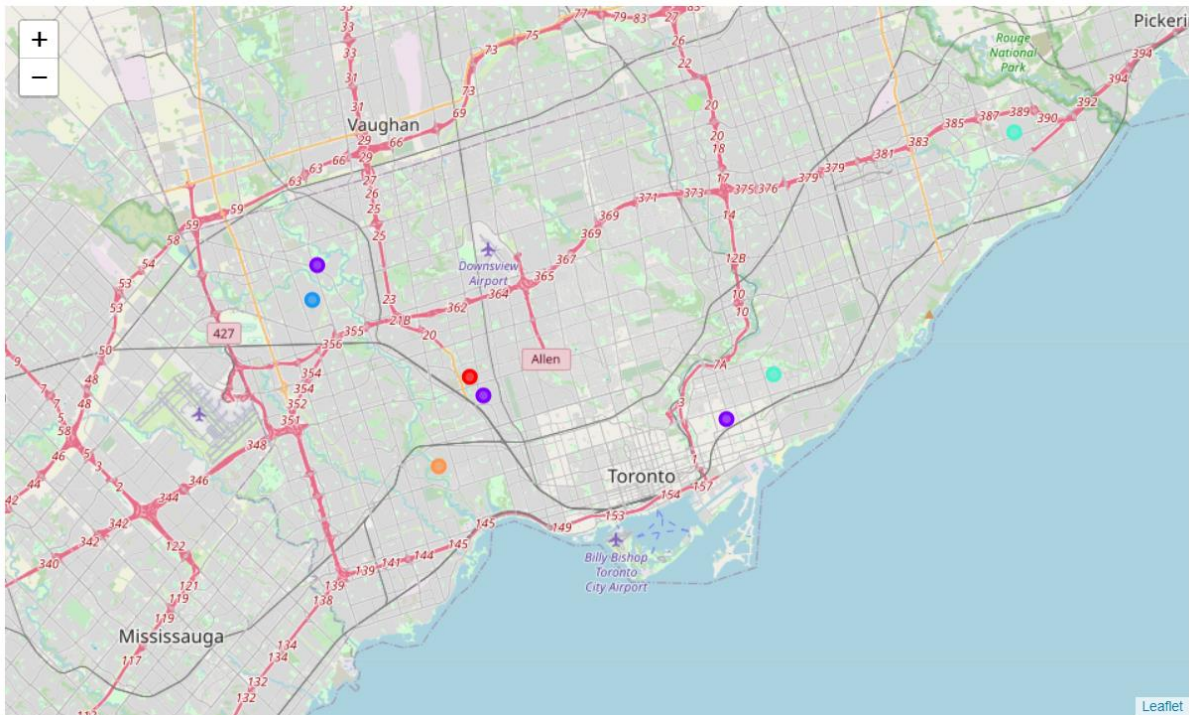
4. Results

Turkish restaurant and museum are unique to cluster 1 neighborhood. This is a touristic neighborhood.

Cluster 1

```
1 selected_neighourhoods_merged.loc[selected_neighourhoods_merged['Cluster_labels'] == 0, selected_neighourhoods_merged.col
```
):

	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	Beechborough-Greenbrook	Turkish Restaurant	Fast Food Restaurant	Museum	Gastropub	Garden	Diner	Dim Sum Restaurant	Curling Ice	Convenience Store	Construction & Landscaping

Cluster 2 has lots unique venues namely: café, hostel, burger joint, nail salon, pizza place, bookstore, bakery, Latin American restaurant, sandwich place, wine shop, thrift/vintage store, Indian restaurant, pharmacy, Asian restaurant, Thai restaurant, supermarket, bank and ice cream shop. This is a foodie neighborhood.

Cluster 2

```
1  selected_neighourhoods_merged.loc[selected_neighourhoods_merged['Cluster_labels'] == 1, selected_neighourhoods_merged.co
```

	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
1	Blake-Jones	Café	Coffee Shop	Hostel	Burger Joint	Diner	Dim Sum Restaurant	Nail Salon	Gastropub	Pizza Place	Bookstore
4	Keelesdale-Eglinton West	Pizza Place	Bakery	Latin American Restaurant	Sandwich Place	Wine Shop	Fast Food Restaurant	Thrift / Vintage Store	Coffee Shop	Diner	Dim Sum Restaurant
7	Thistletown-Beaumond Heights	Indian Restaurant	Caribbean Restaurant	Pharmacy	Asian Restaurant	Thai Restaurant	Supermarket	Bank	Ice Cream Shop	Pizza Place	Coffee Shop

Cluster 3 has the jewelry store as the unique venue. This is a shopping neighborhood.

Cluster 3

```
1  selected_neighourhoods_merged.loc[selected_neighourhoods_merged['Cluster_labels'] == 2, selected_neighourhoods_merged.co
```

	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
6	Rexdale-Kipling	Jewelry Store	Video Store	Home Service	Garden	Fast Food Restaurant	Diner	Dim Sum Restaurant	Curling Ice	Convenience Store	Construction & Landscaping

Cluster 4 has IT services, bus station, beer store, spa and skating rink. This is a commercial neighborhood.

Cluster 4

```
1  selected_neighourhoods_merged.loc[selected_neighourhoods_merged['Cluster_labels'] == 3, selected_neighourhoods_merged.co
```

	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
2	Highland Creek	Home Service	IT Services	Bus Station	Construction & Landscaping	Park	Coffee Shop	Garden	Fast Food Restaurant	Diner	Dim Sum Restaurant
8	Woodbine-Lumsden	Park	Skating Rink	Video Store	Curling Ice	Spa	Beer Store	Convenience Store	Fast Food Restaurant	Diner	Dim Sum Restaurant

Cluster 5 has residential building as the only unique venue. This seems to be purely a residential neighborhood.

Cluster 5

```
1  selected_neighourhoods_merged.loc[selected_neighourhoods_merged['Cluster_labels'] == 4, selected_neighourhoods_merged.co
```

	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
3	Hillcrest Village	Park	Residential Building (Apartment / Condo)	Video Store	Caribbean Restaurant	Fast Food Restaurant	Diner	Dim Sum Restaurant	Curling Ice	Convenience Store	Construction & Landscaping

Cluster 6 has a river as the only unique venue. This is a neighborhood with a lot of natural resources.

Cluster 6

```
1 er_labels'] == 5, selected_neighourhoods_merged.columns[[1] + list(range(2, selected_neighourhoods_merged.shape[1] - 7))]
```

5]:

	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
5	Lambton Baby Point	Garden	River	Park	Caribbean Restaurant	Fast Food Restaurant	Diner	Dim Sum Restaurant	Curling Ice	Convenience Store	Construction & Landscaping

5. Discussion

Deciding where to live in Toronto is not an easy task. Mainly, due to the factors to take into consideration and the number of neighborhoods to consider.

In our study, we decided to use only three factors to make the decision. These factors are crime rates in the neighborhoods, the average rent and venues close to the neighborhoods.

We selected 45 neighborhoods with the least crime rates and from that we selected 10 that were more affordable. Then we searched for venues that were close the selected neighborhoods.

The neighborhoods were split into 6 clusters and analyzed based on their unique venues. Now one has only to choose one of the six clusters and choose a neighborhood if the cluster has more than one neighborhood.

6. Conclusion

Data science and machine learning techniques are very useful for analyzing data and extracting insights that can be used to make informed decisions. In our case, to choose the neighborhood to reside.

The analysis can be improved by using other factors such proximity to schools, hospitals, average income, population density, etc.

REFERENCES

1. https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/a083c865-6d60-4d1d-b6c6-b0c8a85f9c15?format=geojson&projection=4326

2. https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/a083c865-6d60-4d1d-b6c6-b0c8a85f9c15?format=csv&projection=4326

3. http://map.toronto.ca/wellbeing/#eyJ0b3Itd2lkZ2V0LWNsYXNzYnJlYWsiOsSAcGVyY2VudE9wYWNpdHElzcwfSwiY3VzxIJtYcSTYcSXxIBuZWlnaGJvdXJob29kc8S2fcSrxIHEg8SFxIfEicSLdGFixYXEmCLEo3RpdmVUxZBJZMSXxYnEhMWPYi1pbmRpY2HEgnLFhcWIYWdzTWFwxLYiesWCbcSXMTPErHjEly04ODM3NzYzLjXGhDcyN8SsxKc6NTQxMjkzMS4yNMaDMjg1xYjFpMWmxajFqsWSxIDFmMWraW9uxJcyxKxzxaRnbGXFtMSucsSTxJ9UaW1lxZzEqMSsxZbGucajIjfFtHNlxrBjdGVkScWlxafFqcSDxrs6x4EzIsSsxp%2FHjcWqxZxzQcWlV8S5xLt0xZJbxIDEh8WeN8eTxKx3x55odMW5xKzEk8alx4RQb8SOcsSlxo5mYWzHhH1dxYfFiMa%2BZceEx7XIgcWGxKzGs8a1dMeLxqDHjnJNxYPGsceRxq1uxq%2FGscWH

4. https://foursquare.com/developers/apps