Capstone Project: Finding the perfect place to live in a new city

Introduction

This project will use FourSqaure API to extract location data and we will analyze it to decide where is the perfect location for a specific person to live according to this person's background

Business Problem

Assume that you work at a real estate consulting department. And your client is looking for a perfect location to start their adventure in a new city where they are not familiar with. You will receive client's personal data by communicating with them and use location data to solve this problem.

Here is one client:

Trevor, who is a recent graduate, wants to move to a new place in Toronto. According to his background, he does not own a car and currently looking for a full-time job. He wants to move to a place where it is convenient to commute. Because of budget, he also wants his living expense to be reasonable. Meanwhile, he does not want to travel far for a quick meal in a local restaurant and he enjoys casual shopping and watching movies with friends.

In this project, we will analyze Toronto's location data and figure out a good plan for Trevor's future home location.

Interest

This project will benefit mostly real estate companies who are helping their clients decide where to live. This will list the five most suitable neighborhoods for a specific person to live in. This saves time when the company is facing a large number of clients.

Data

For our project, we have two sets of data. The first one is coming from our clients and mostly sentimental data. The second one is a large data set consist of the nearby venues of every neighborhood in a specific city, in our example: Toronto.

We first extract all the neighborhood data from the Wikipedia page just like week 3. Then we join the data frame with coordinates data to get coordinates for each postal code.

Next, we use foursquare to access each neighborhood's nearby venue information and store them into a new data set 'moving_data.csv'

Here is the link to our moving_data

Methodology

Since our client would like the convenience of having food, shopping, bar, and watching movies, we will first find the areas which have most of these four venues.

We have areas having the most restaurants:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	food_rank
Neighborhood							
First Canadian Place, Underground city	3	3	3	3	3	3	1.0
Little Portugal, Trinity	2	2	2	2	2	2	2.5
Toronto Dominion Centre, Design Exchange	2	2	2	2	2	2	2.5
Commerce Court, Victoria Hotel	1	1	1	1	1	1	7.0
Don Mills South	1	1	1	1	1	1	7.0
Fairview, Henry Farm, Oriole	1	1	1	1	1	1	7.0
Glencairn	1	1	1	1	1	1	7.0
Regent Park, Harbourfront	1	1	1	1	1	1	7.0
Richmond, Adelaide, King	1	1	1	1	1	1	7.0
St. James Town	1	1	1	1	1	1	7.0

Areas having the most shopping malls:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	shop_rank
Neighborhood							
Enclave of M5E	23	23	23	23	23	23	1.5
Garden District, Ryerson	23	23	23	23	23	23	1.5
Harbourfront East, Union Station, Toronto Islands	18	18	18	18	18	18	3.0
Central Bay Street	17	17	17	17	17	17	4.5
Commerce Court, Victoria Hotel	17	17	17	17	17	17	4.5
Toronto Dominion Centre, Design Exchange	15	15	15	15	15	15	6.5
Richmond, Adelaide, King	15	15	15	15	15	15	6.5
Berczy Park	13	13	13	13	13	13	9.0
Regent Park, Harbourfront	13	13	13	13	13	13	9.0
Fairview, Henry Farm, Oriole	13	13	13	13	13	13	9.0

Areas having the most bars:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	bar_rank
Neighborhood							
Little Portugal, Trinity	8	8	8	8	8	8	2.0
Commerce Court, Victoria Hotel	8	8	8	8	8	8	2.0
Enclave of M5E	8	8	8	8	8	8	2.0
First Canadian Place, Underground city	7	7	7	7	7	7	5.5
Berczy Park	7	7	7	7	7	7	5.5
Church and Wellesley	7	7	7	7	7	7	5.5
Harbourfront East, Union Station, Toronto Islands	7	7	7	7	7	7	5.5
Kensington Market, Chinatown, Grange Park	6	6	6	6	6	6	8.5
St. James Town	6	6	6	6	6	6	8.5
Toronto Dominion Centre, Design Exchange	5	5	5	5	5	5	10.0

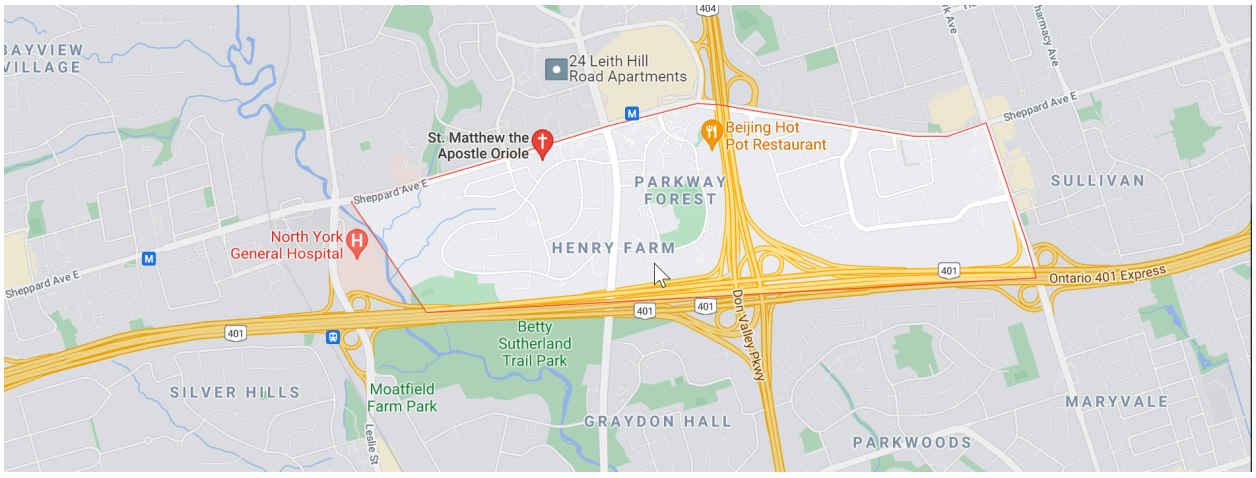
Areas having the most cinemas:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	cin_rank
Neighborhood							
Garden District, Ryerson	2	2	2	2	2	2	1.0
Enclave of M5E	1	1	1	1	1	1	5.0
Fairview, Henry Farm, Oriole	1	1	1	1	1	1	5.0
Harbourfront East, Union Station, Toronto Islands	1	1	1	1	1	1	5.0
India Bazaar, The Beaches West	1	1	1	1	1	1	5.0
Parkdale, Roncesvalles	1	1	1	1	1	1	5.0
Runnymede, Swansea	1	1	1	1	1	1	5.0
Willowdale South	1	1	1	1	1	1	5.0

After we gather these data, we concatenate them and find the area which has four category venues at once. After concatenation, we find that only Neighborhood: **Fairview, Henry Farm, Oriole** satisfy this condition.

Fairview, Henry Farm, Oriole	9.0	5.0	11.0	7.0

However, after digging into this location, we find that it is far from the main area of Toronto and hard to commute to downtown.



Now we consider dropping one category. Since according to our data, not a lot of areas having cinema around. We then only consider three factors and we got the following data.

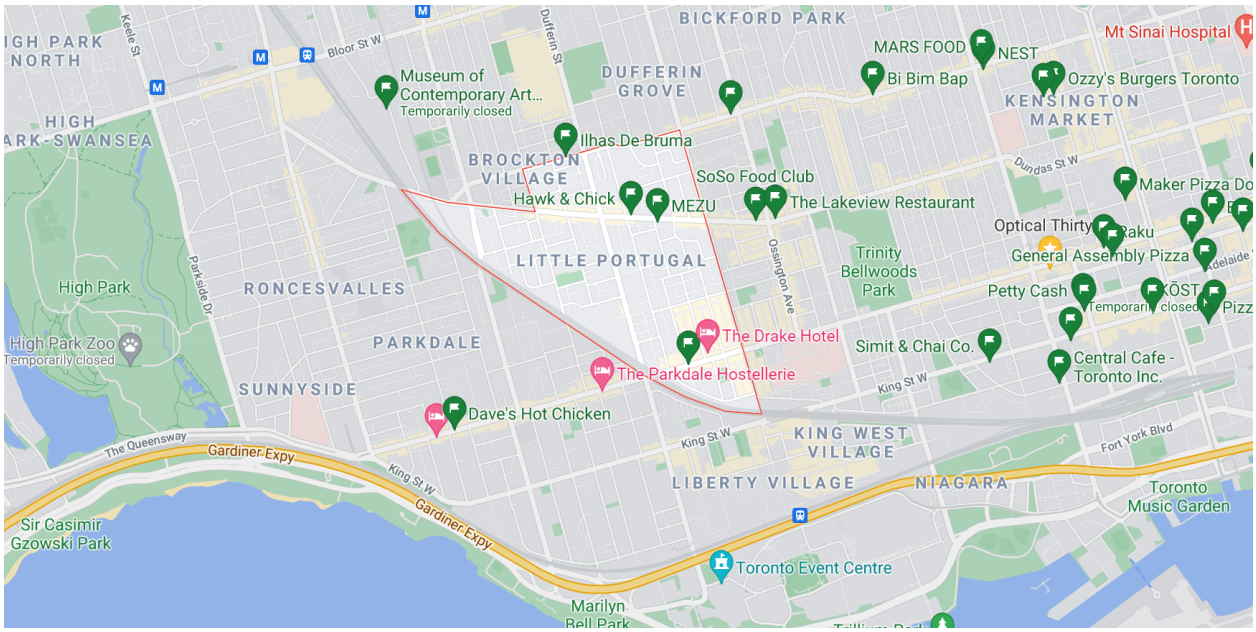
	shop_rank	bar_rank	food_rank	all_rank
Commerce Court, Victoria Hotel	4.5	2.0	7.0	13.5
Toronto Dominion Centre, Design Exchange	6.5	10.0	2.5	19.0
First Canadian Place, Underground city	14.5	5.5	1.0	21.0
Little Portugal, Trinity	17.0	2.0	2.5	21.5
Richmond, Adelaide, King	6.5	12.5	7.0	26.0

We rank them over the sum of the ranks from three categories and yield the top five candidates.

Results

Now after doing research on all five areas, we found that only Little Portugal, Trinity is suitable for living. The rest four areas are located in the financial district and commonly having high living expenses. According to our client's background, he needs his living expenses to be reasonable.

Therefore, we conclude our client will be happy to live at Little Portugal, Trinity.



Discussion and Conclusion

From this report, we can see that given the client's preference, we can determine which area is perfect for the client to live in. However, we might encounter some problems. For example, the movie theater preference here. In the future, I think the data we collect from our client needs to be more specific. For instance, we might introduce ordinal data. So that our client can rank how strong a factor affects this client's opinion of the choice of location. In case that our data do not represent a factor, we can omit it or remedy it in some other way according to the ordinal data.