Business Hunting and Intelligent Advisor.

Investment intelligent advisor for investors in Toronto

By: Amr Hafez

Table of Contents

Introduction

Investment in Canada is a very hot topic particularly in Toronto.

Some people find that it is a very tough job to collect a survey of all investments' types in Toronto, therefore I have created this intelligent advisor.

The purpose of this advisor is to dig into all areas of Toronto using FourSquare location and venues information to provide a very nice summary for the investors to help them hunt the most suitable investment to be constructed in order to gain the highest profit.

Audience of this advisor are the business owners who are looking for a great business and profitable investment in Canada and in particular in Toronto.

In the final report, the investor will be able to identify the lack of certain services across different locations, or even the quality of the provided services using the collected tips in order to provide unique or better services.

Data Section

In this section i will collect data from different sources about the Toronto area using the following sources:

1-Wikipedia.

2-Foursquare location data.

These sources contain data about the neighborhoods and venues in Toronto area. Finally i will be using the below steps to collect data from those sources.

Extract data of Toronto neighborhoods from Wikipedia

Out[40]:

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights
7	M6A	North York	Lawrence Manor
8	M7A	Queen's Park	Not assigned

But the returned records contain unprocessed or incomplete data, which needs data cleansing and preparation.

The following steps were taken to beautify the data returned.

Dropping the 'Not assigned' Boroughs

Out[10]:

	Postcode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights
5	M6A	North York	Lawrence Manor
6	M7A	Queen's Park	Not assigned
7	M9A	Etobicoke	Islington Avenue
8	M1B	Scarborough	Rouge

Fill the 'Not assigned' neighborhoods with the name of the Borough

Out[43]:

	Postcode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights
5	M6A	North York	Lawrence Manor
6	M7A	Queen's Park	Queen's Park
7	M9A	Etobicoke	Islington Avenue
8	M1B	Scarborough	Rouge
9	M1B	Scarborough	Malvern

Combine the grouped Boroughs' neighborhoods separated with a comma

Out[11]:

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West

Read the geospatial coordinates from the csv file stored in the capstone project

Out[46]:

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

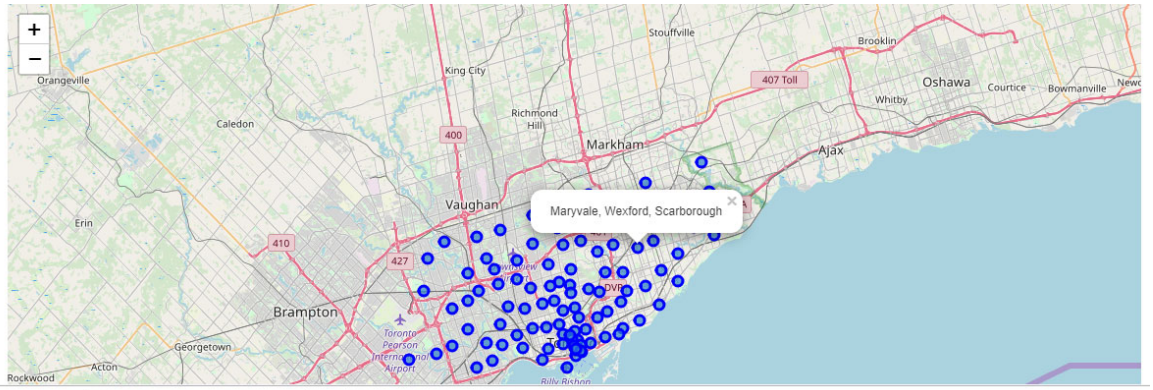
Join the two data frames together

Out[47]:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848

Create the map of Toronto, given the above coordinates

Out[22]:



Create map of the boroughs that contain the word Toronto

Getting the top venues that are in the Beaches within a radius of 500 meters using Foursquare location data.

Out[34]:

	name	categories	lat	lng
0	The Big Carrot Natural Food Market	Health Food Store	43.678879	-79.297734
1	Grover Pub and Grub	Pub	43.679181	-79.297215
2	Starbucks	Coffee Shop	43.678798	-79.298045
3	Glen Stewart Park	Park	43.675278	-79.294647
4	Upper Beaches	Neighborhood	43.680563	-79.292869

Methodology

In this project we will direct our efforts on detecting areas of Toronto that have any, particularly those sort of investments. We will limit our analysis to area around city center.

In first step we have collected the required **data: location and type (category) of every investment within certain radius from Toronto center**. We have also **identified the famous venues** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of '**investment density**' across different areas of Toronto - we will use **maps** to identify a few promising areas close to center with low number of certain investment in general and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations. Also we will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Analysis

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the **types of investments in every area candidate**:

Analyze the neighborhoods

Out[64]:

	Neighborhood	Yoga Studio	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Argentinian Restaurant	Art Gallery	Art Museum	Art Cra St
0	Adelaide, King, Richmond	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.040000	0.000000	0.00	0.000000	0.010000	0.010000	0.000(
1	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.017241	0.000000	0.000(
2	Brockton, Exhibition Place, Parkdale Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000(
3	Business Reply Mail Processing Centre 969 Eastern	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000(

Let's print each neighborhood along with the top 5 most common venues

```
----Cabbagetown, St. James Town----
                venue  freq
0          Restaurant  0.08
1         Coffee Shop  0.08
2                Park  0.06
3  Italian Restaurant  0.04
4              Bakery  0.04


----Central Bay Street----
                venue  freq
0         Coffee Shop  0.15
1  Italian Restaurant  0.05
2               Café  0.03
3     Bubble Tea Shop  0.03
4  Chinese Restaurant  0.03
```

```
neighborhoods_venues_sorted.head()
```

Out[67]:

	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	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	Thai Restaurant	American Restaurant	Burger Joint	Restaurant	Hotel	Asian Restaurant	Sushi Restaurant
1	Berczy Park	Coffee Shop	Cocktail Bar	Seafood Restaurant	Cheese Shop	Bakery	Steakhouse	Restaurant	Farmers Market	Café	Pub
2	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Café	Coffee Shop	Climbing Gym	Italian Restaurant	Bar	Burrito Place	Caribbean Restaurant	Furniture / Home Store	Restaurant
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Pizza Place	Auto Workshop	Comic Shop	Butcher	Burrito Place	Recording Studio	Restaurant	Brewery	Skate Park
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Service	Airport Lounge	Airport Terminal	Plane	Boat or Ferry	Harbor / Marina	Airport	Airport Food Court	Airport Gate	Sculpture Garden

Now, the data has been collected and rendered graphically to the audiences to make comparisons between areas.

Results and Discussion

Our analysis shows that although there is a great number of investments in Toronto, there are pockets of low cloth stores density fairly close to city center. Highest concentration of different investments types were detected north and west from Toronto, so we focused our attention to areas south, south-east and east, corresponding to Cabbage town, St. James Town which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics and a number of pockets of low clothes stores density.

After directing our attention to this more narrow area of interest (covering Cabbage town, St. James Town. We first created a dense grid of location candidates

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is containing largest number of potential new clothes stores locations based on number of and distance to existing venues - both stores in general and non-food investment particularly.

Purpose of this analysis was to enable the investors to choose the right type of investment across Toronto areas in order to get the highest profits.

Conclusion

Purpose of this analysis was to enable the investors to choose the right type of investment across Toronto areas in order to get the highest profits.

By calculating investments density distribution from Foursquare data we have first identified general boroughs that justify further analysis , and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby investments. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal investment type location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.