Opening Colleges in Toronto

This is a hypothetical situation where someone has decided to open up a college in Toronto. In this situation, the management has decided to hire Data Scientists in order to figure out the best place to open up a college in Toronto. The team of Data Scientists has to consider various factors around any neighbourhood to decide the best place to open up a College.

Description of the School

This is a type of college centred around complete self-dependence. This College provides all the essentials that a student needs such as Stationary, General Store, Gymnasium, Sports Facilities, Hostel, Playing Fields, a mess etc.

Methodology

The factors that I would consider while opening up the College are:

1. The Number of Colleges in the Neighbourhood. (The less the better.) - A dominant factor.

2. The Number of Hotels, Restraunts and Bars. (The less the better.) - A dominant factor.

3. The Transport services (Airports, Bus stations, Train Stations etc.) - The more the better.

4. The Number of Active Entertainment Areas (Sports stadiums, Basketball Stadium etc.)

5. The Number of Passive Entertainment Areas (Shopping Malls, Movie Theatres etc.) - The less the better.

6. The Number of Hard Drinks shops. (The less the better.) - A very dominant factor

Data for the Problem

Foursqaure Locations Data

Clearly, this problem can be solved solely by the foursquare data. The Foursquare API can easily provide this data by exploring any given location in Toronto. By classifying all the venues into the categories described above we can clearly determine how many venues of each type are present in any neighbourhood.

The last thing we need to determine is the objective function to minimise in this situation, clearly not all of the venues have the equal amount of weightage or similar impact in determining the type of location. Hence, we would need to define an objective function that properly attaches weights to all of these

parameters. But all of the data we require can be easily obtained from Foursquare.

Discussion

To solve this problem, the first step is obtaining a Data Frame that contains all the neighbourhoods and postal codes of Toronto. We obtained this by using the Wikipedia data on Toronto Neighbourhoods.

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...

After this step was to obtain the latitude and longitude information for each of these neighbourhoods so as to leverage Foursquare API.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724

Next, to use the Foursquare API and get the venues data for each of these locations.

	Postal Code	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M3A	Parkwoods	North York	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	M3A	Parkwoods	North York	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	M3A	Parkwoods	North York	43.753259	-79.329656	Corrosion Service Company Limited	43.752432	-79.334661	Construction & Landscaping
3	M4A	Victoria Village	North York	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	M4A	Victoria Village	North York	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
5	M4A	Victoria Village	North York	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
6	M4A	Victoria Village	North York	43.725882	-79.315572	The Frig	43.727051	-79.317418	French Restaurant
7	M4A	Victoria Village	North York	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place
8	M5A	Regent Park, Harbourfront	Downtown Toronto	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
9	M5A	Regent Park, Harbourfront	Downtown Toronto	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
10	M5A	Regent Park, Harbourfront	Downtown Toronto	43.654260	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
11	M5A	Regent Park, Harbourfront	Downtown Toronto	43.654260	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
12	M5A	Regent Park, Harbourfront	Downtown Toronto	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
13	M5A	Regent Park, Harbourfront	Downtown Toronto	43.654260	-79.360636	Corktown Common	43.655618	-79.356211	Park
14	M5A	Regent Park, Harbourfront	Downtown Toronto	43.654260	-79.360636	Dominion Pub and Kitchen	43.656919	-79.358967	Pub
15	M5A	Regent Park, Harbourfront	Downtown Toronto	43.654260	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot

This data contains a lot of irrelevant information. For example, the presence of a distribution centre would hardly be relevant to someone opening a college. Also, the venue category mentioned in this data are far too specific. For example, it hardly matters to us if a restaurant is Portuguese or a simple Food and Drinks Shop all that does matter is that it is a restaurant.

Hence, we need to clean the data and moreover classify them into 6 broad categories that are relevant to us. After doing that the data frame looked something like this.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Hotels	AE	PE	Bars	Transport	Colleges
0	M3A	North York	Parkwoods	43.753259	-79.329656	1	1	0	0	0	0
1	M4A	North York	Victoria Village	43.725882	-79.315572	4	1	0	0	0	0
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	22	7	2	4	0	0
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	2	0	0	0	0	0
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	21	5	1	2	0	1
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242	1	0	0	0	0	0
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	1	0	0	0	0	0
7	M3B	North York	Don Mills	43.745906	-79.352188	3	1	0	0	0	0
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937	3	2	0	1	0	0
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	51	10	4	4	0	1
10	M6B	North York	Glencairn	43.709577	-79.445073	3	0	0	1	0	0
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724	1	0	0	0	0	0
12	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	0	0	0	1	0	0
13	M3C	North York	Don Mills	43.725900	-79.340923	12	3	1	0	0	0
14	M4C	East York	Woodbine Heights	43.695344	-79.318389	1	3	0	0	1	0
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	47	8	1	9	0	0

Next, we need to come up with an objective function in order to differentiate between these locations.

Objective Function

Remember that in the objective function, the positive and negative weights are to be assigned to each category according to its impact. Moreover, it should be noted that magnitude of weights shall be decided by how dominant any category is in deciding the location for building a college.

Objective Function = 10 x AE + 20 x Transport - 5 x PE - Hotels - 60 x Bars – Colleges

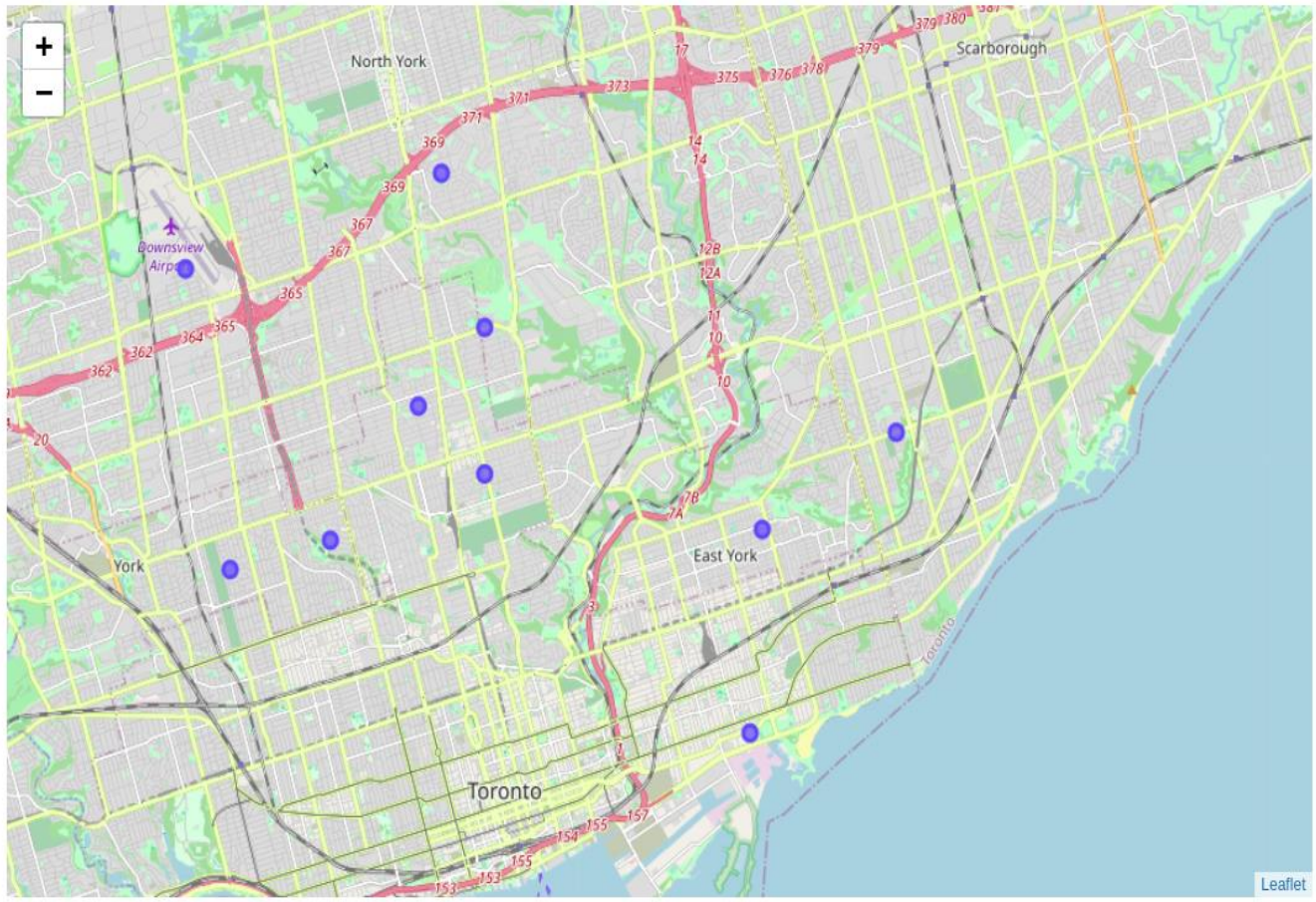
The next step was to just calculate the scores of each of the selected and rank them according to their score. After performing this process, we arrive at this data frame.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Hotels	AE	PE	Bars	Transport	Colleges	Score
44	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577	3	2	0	0	4	0	100
98	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558	4	6	0	0	2	0	100
72	M4R	Central Toronto	North Toronto West, Lawrence Park	43.715383	-79.405678	10	5	0	0	0	0	50
14	M4C	East York	Woodbine Heights	43.695344	-79.318389	1	3	0	0	1	0	50
60	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0	2	0	0	1	0	40

However, we should try to clean up this table so that it contains only the valuable information.

	Postal Code	Neighborhood	Borough	Latitude	Longitude	Score
0	M1L	Golden Mile, Clairlea, Oakridge	Scarborough	43.711112	-79.284577	100
1	M7Y	Business reply mail Processing Centre, South C...	East Toronto	43.662744	-79.321558	100
2	M4R	North Toronto West, Lawrence Park	Central Toronto	43.715383	-79.405678	50
3	M4C	Woodbine Heights	East York	43.695344	-79.318389	50
4	M4N	Lawrence Park	Central Toronto	43.728020	-79.388790	40
5	M4S	Davisville	Central Toronto	43.704324	-79.388790	35
6	M6E	Caledonia-Fairbanks	York	43.689026	-79.453512	30
7	M6C	Humewood-Cedarvale	York	43.693781	-79.428191	30
8	M2P	York Mills West	North York	43.752758	-79.400049	30
9	M3K	Downsview	North York	43.737473	-79.464763	30

Now for the final step let us visualize all of these neighbourhoods on a map.



Conclusion

This result of this project is clear. By determining the venues around a particular Location, we were able to find out the best 10 areas in Toronto that shall be best for setting up a college. Moreover, we determined the hierarchical order in which these areas are better. All of this is possible via the objective function that we have calculated. This objective function helped us determine this result.