College In Toronto

This is a Data Science Project on determining the best neighborhood in Toronto to build a college in. This project is part of the IBM Applied Data Science Capstone Project on Coursera.

Introduction to the Problem (Opening a College in Toronto)

- **Description of the Situation**: 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.
- The first step in determining the type of college we want to build, since that will determine the type of neighborhood according to our needs.

Description of the college

- The description of the college to be built as provided by the company is:
- It is a completely self – dependent type of college that is that in order to address to day to day needs of a student the college has in-campus facilities.
- To be more clear, the campus contains a barber shop, a general store, an electronics store, a stationery shop, a confectionary shop, some restaurants, a hostel, sports facilities, a mess etc.
- So we want to search for an isolated place so that the students are far away from distractions as they are vulnerable in this age and since the college can provide for their day to day needs there is no need to go out of the campus.

Methodology
(Our Approach to the Problem)

- After due consideration of the description of the college and its needs the following parameters were listed as the most relevant:
- The Number of Colleges in the Neighbourhood. (The less the better.) - A dominant factor.
- The Number of Hotels, Restaurants and Bars. (The less the better.) - A dominant factor.
- The Transport services (Airports, Bus stations, Train Stations etc.) - The more the better.
- The Number of Active Entertainment Areas (Sports stadiums, Basketball Stadium etc.)
- . The Number of Passive Entertainment Areas (Shopping Malls, Movie Theatres etc.) - The less the better.
- The Number of Hard Drinks shops. (The less the better.) - A very dominant factor

Data For the Problem (Foursquare API)

- 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.

Solution to the Problem (Steps involved in Cracking the problem)

- The first step was obtaining the list of neighborhoods and postal codes in the city of Toronto. This was achieved by using the Wikipedia page containing the relevant information.
- The second step was cleaning the data obtained from Wikipedia.
- The third step after obtaining cleaned data was getting the longitudes and latitudes of each neighborhood. This was obtained from the csv file containing this information.
- The fourth step was getting the venues details from foursquare locations API via Longitudes and Latitudes
- The fifth step was cleaning the venues data and obtaining the workable Data Frame containing relevant venues information.
- The sixth step was determining an objective function to maximize and calculating a score of all the neighborhoods.
- The seventh step was selecting the top 10 neighborhoods according to their score and displaying a table with relevant information.
- The final step was visualizing these selected locations on the map.

Cleaned Data Frame (Second Step)

	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...

Data Frame with Coordinates (Third Step)

	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

Venues Table from Foursquare (Fourth Step)

	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

Cleaned Venues Data Frame (Fifth Step)

	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

Objective Function (Sixth Step)

- Remember the following 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

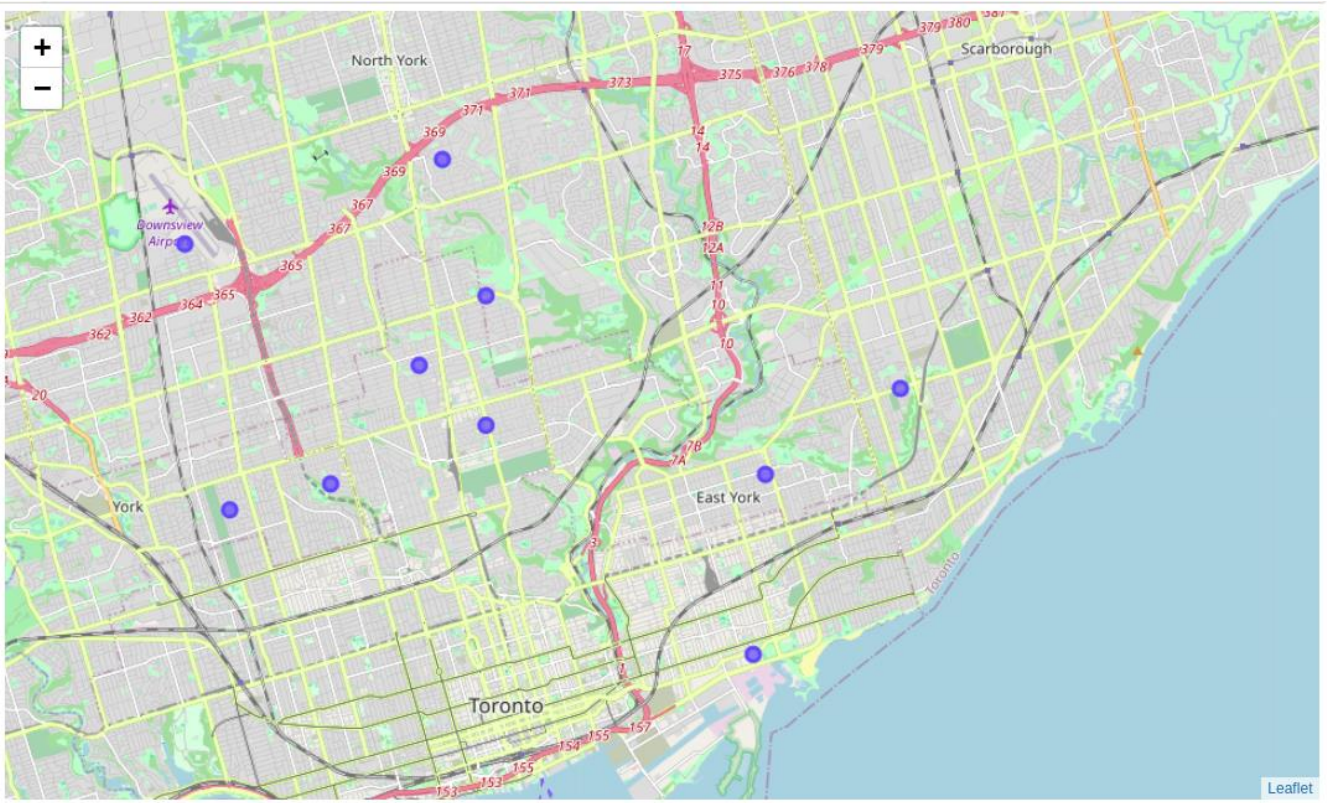
Data Frame with Scores (Evaluated Objective Function)

	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

Top 10 Neighborhoods (Seventh Step)

	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

Map Visualization (Final Step)



Conclusion

- The results of this data science project are pretty much straightforward.
- We determined the top 10 neighborhoods in Toronto to build a college in.
- Lastly, we visualized these top 10 locations on map.