The Battle of Neighborhoods

Capstone Project

In this project we will try to find an optimal location for a park. Specifically, this report will be targeted to stakeholders interested in opening an **Park** in **Toronto**, Canada.

Bussiness problem

The objective of the project is to analyze and select the best solution to open a new park in Toronto, this using data science methodology, allowing obtaining the best place to open the park taking into account, the distance from other parks, the amount of inhabitants of the area and the level of contamination.

Introduction

Toronto is the provincial capital of Ontario. With a recorded population of 2,731,571, and is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.
That is why it is important that it has green areas where people can disperse, play sports and breathe fresh air, in addition to helping to reduce the pollution of the city, Toronto has 30 parks, but leaving areas where there are no green areas, so the following question is asked, what is the most convenient area to open a new park?

Following data sources will be needed to extract/generate the required information:

•Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Page Wikipedia**
•Number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

Data

The data used to solve the problem comes from Foursquare, a geolocation platform, which has 150 million locations. In addition to categorizing this information, we are interested in the category of parks, which will help us solve the problem.

To solve the problem, will need the following data:

- List Toronto neighborhoods.
- Get the coordinates of the neighborhoods
- Venue data, especially data related to parks, clustering neighborhoods

Libraries

- Pandas
- Numpy
- Matplotlib
- Seaborn
- Scikit Learn
- Folium
- Beatifulsoup
- Wikipedia

Data preparation

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood.The first part is for boroughs and neighborhoods lits to Toronto, this was imported from Wikipedia using Beatifulsoup and turned into a dataframe. The second part is for demographics data of Toronto.

	Postal Code	Neighborhood	Borough	Latitude	Longitude
0	M1A	Not assigned,		43.806686	-79.194353
1	M1B	Malvern, Rouge,	Scarborough	43.784535	-79.160497
2	M1C	Rouge Hill, Port Union, Highland Creek,	Scarborough	43.763573	-79.188711
3	M1E	Guildwood, Morningside, West Hill,	Scarborough	43.770992	-79.216917
4	M1G	Woburn,	Scarborough	43.773136	-79.239476
5	M1H	Cedarbrae,	Scarborough	43.744734	-79.239476
6	M1J	Scarborough Village,	Scarborough	43.727929	-79.262029
7	M1K	Kennedy Park, Ionview, East Birchmount Park,	Scarborough	43.711112	-79.284577

Methodology

In this project we will direct our efforts on detecting areas of Toronto that have low parks density, particularly those with low number of parks. We will limit our analysis to area ~6km around city center.

FIRST STEP

Collected the required data: location and type (category) of every park within 6km from Toronto center (Alexanderplatz).
We have also identified parks (according to Foursquare categorization).

SECOND STEP

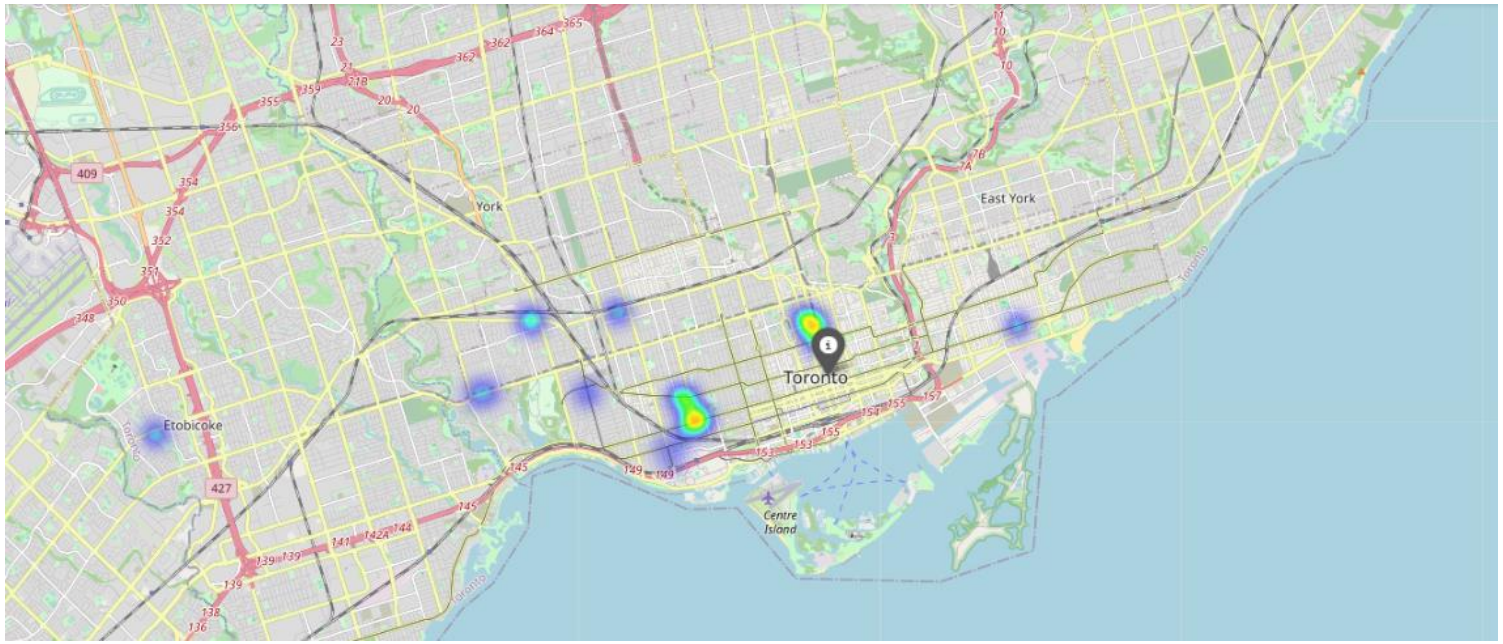
Calculation and exploration of 'park density' across different areas of Toronto - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (*and* no parks in vicinity) and focus our attention on those areas.

FINAL STEP

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 with no more than two parks in radius of 250 meters, and we want locations without parks in radius of 400 meters. 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.

Analisys

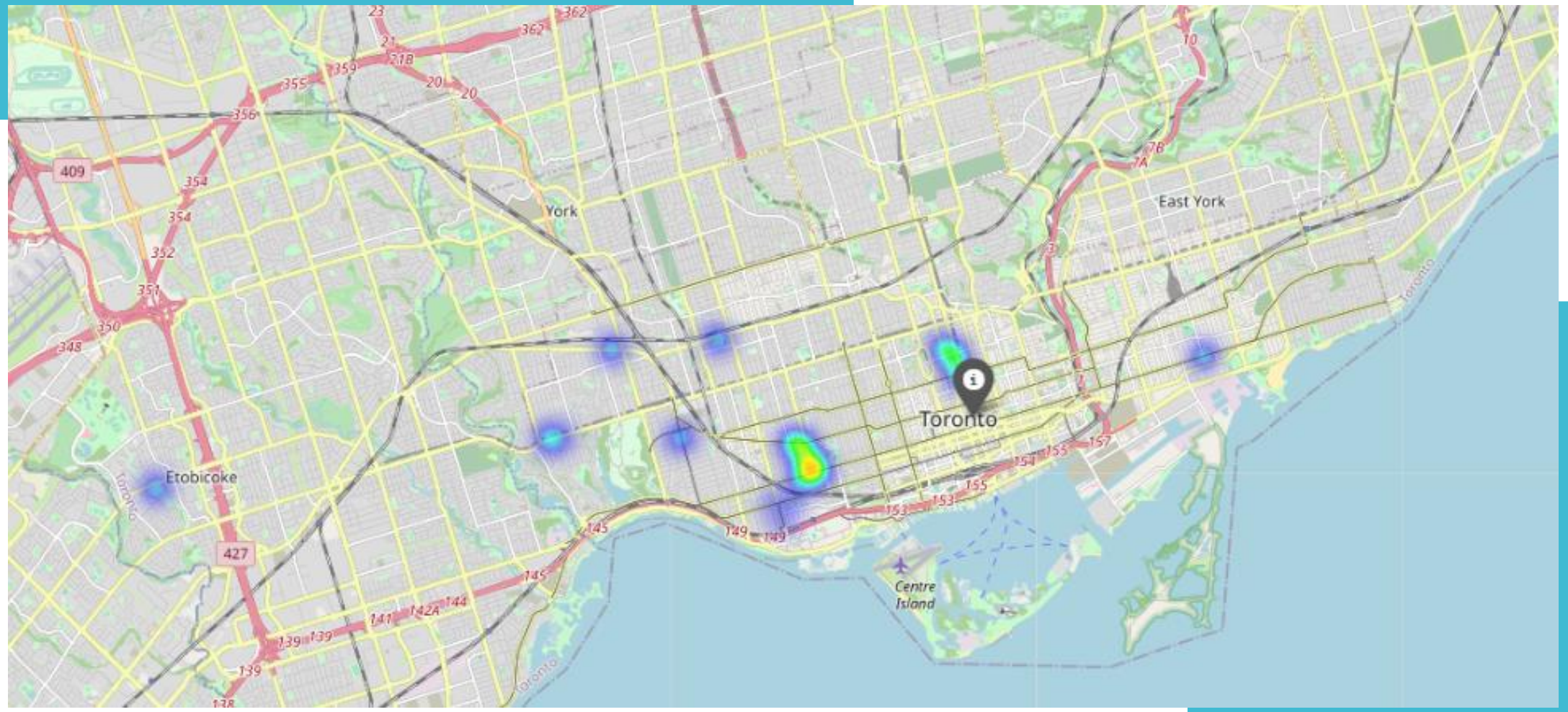
Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the number of parks in every area candidate, and category “Neighborhood”

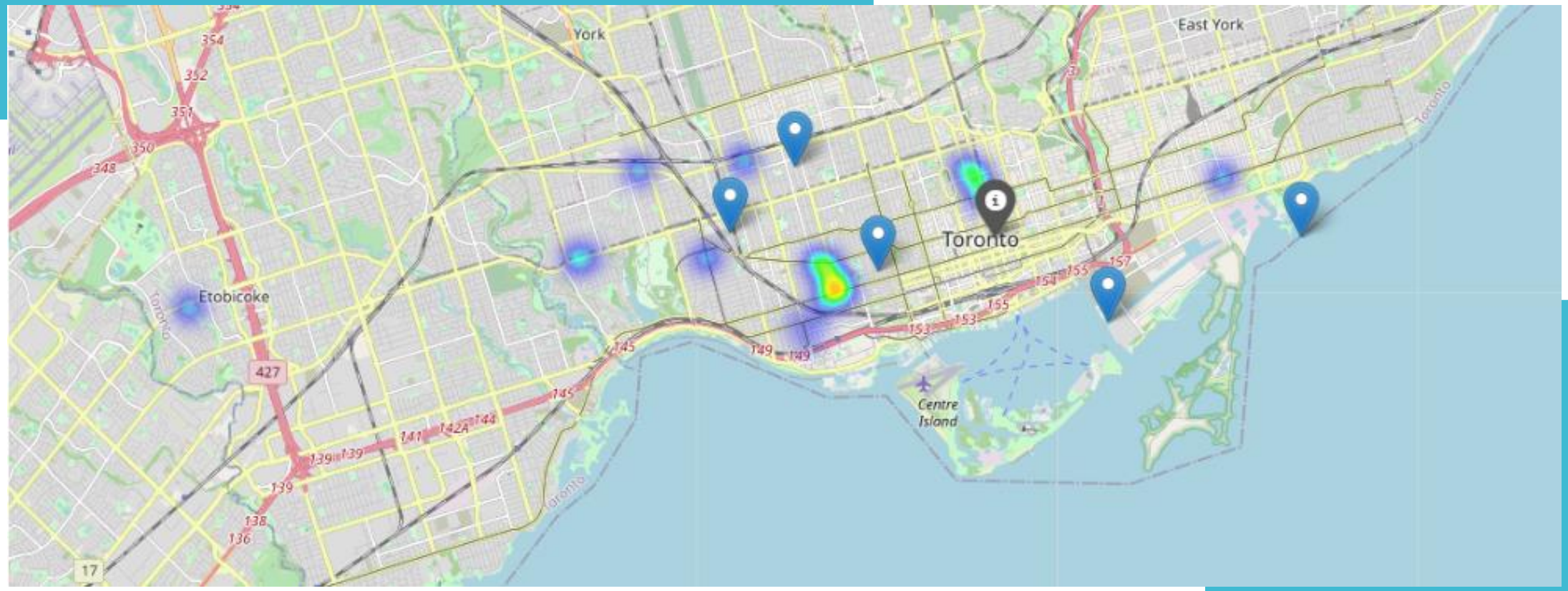


	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rosedale	43.669005	-79.442259	FreshCo	43.667918	-79.440754	Grocery Store
1	Rosedale	43.669005	-79.442259	Rexall	43.667504	-79.442086	Pharmacy
2	Rosedale	43.669005	-79.442259	Galleria Shopping Centre	43.667592	-79.442053	Shopping Mall
3	Rosedale	43.669005	-79.442259	LCBO	43.667002	-79.441747	Liquor Store
4	St. James Town	43.647927	-79.419750	Lost & Found	43.649378	-79.424149	Men's Store

HEAT MAP

A heat map is a data visualization technique that shows the magnitude of a phenomenon as two-dimensional color. The variation in color can be by hue or intensity, giving obvious visual cues to the reader about how the phenomenon is grouped or modified in space.


York
East York
Toronto
Etobicoke
Centre Island
409
427
362
359
355
356
354
352
351
350
348
23
21
21B
20
10
7B
7A
145
149
153
155
157
139
141
142A
144



Possible places to locate the park

Analyze Each Neighborhood

Let's perform some basic exploratory data analysis and derive some additional info from our raw data. First let's count the number of stores in every area candidate.

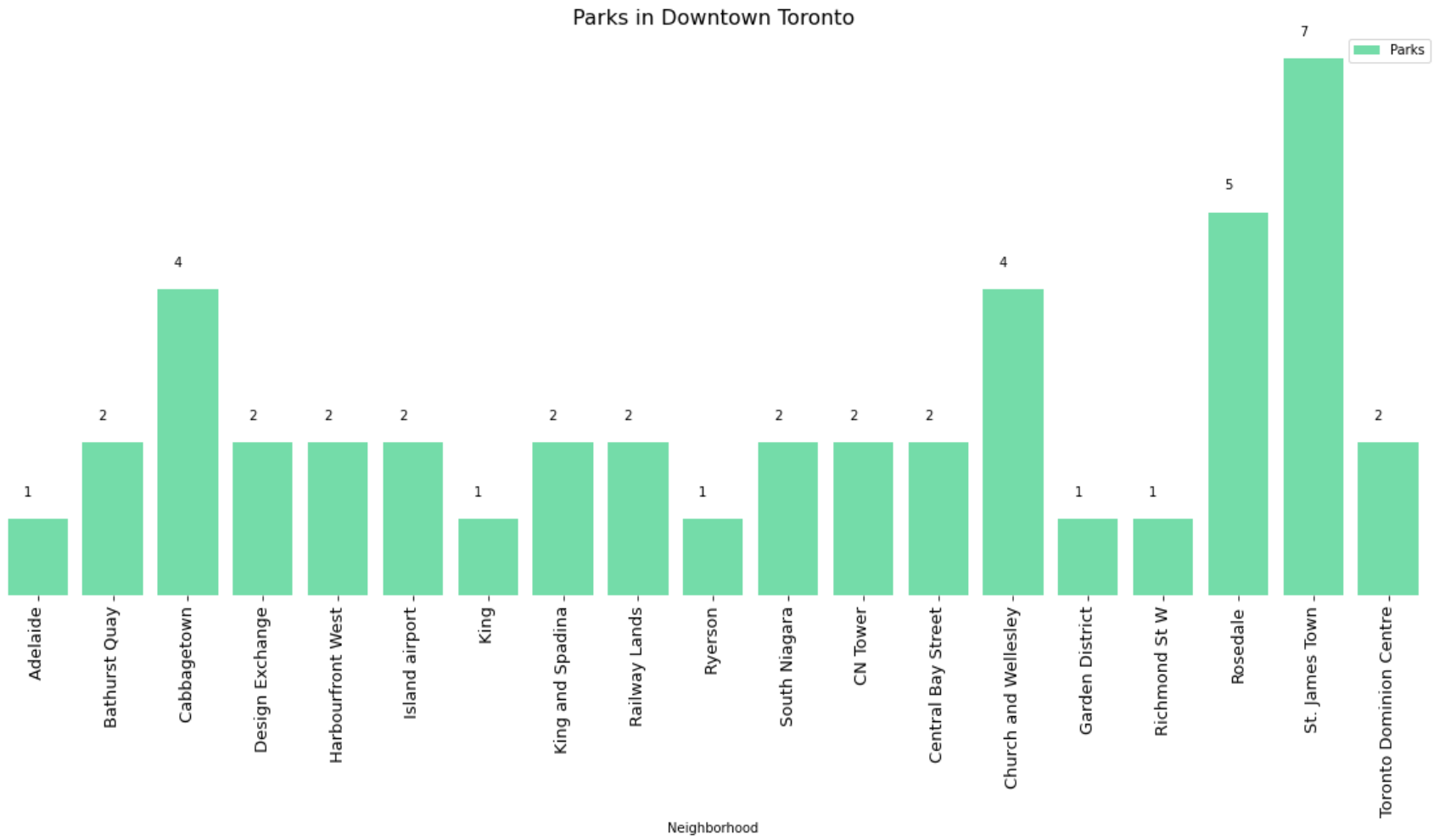
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Adelaide	Supermarket (11.11111111111111%)	Convenience Store (11.11111111111111%)	Shopping Mall (11.11111111111111%)
1	Bathurst Quay	Beer Store (50.0%)	Liquor Store (50.0%)	Supermarket (0.0%)
2	Cabbagetown	Men's Store (30.0%)	Boutique (20.0%)	Bar (10.0%)
3	Design Exchange	Comic Shop (100.0%)	Supermarket (0.0%)	Discount Store (0.0%)
4	Harbourfront West	Beer Store (50.0%)	Liquor Store (50.0%)	Supermarket (0.0%)

Now we only look at the parks, so a calorie map will be plotted to visualize in the areas that there are more parks (Direct Competition) and thus be able to choose the area correctly.

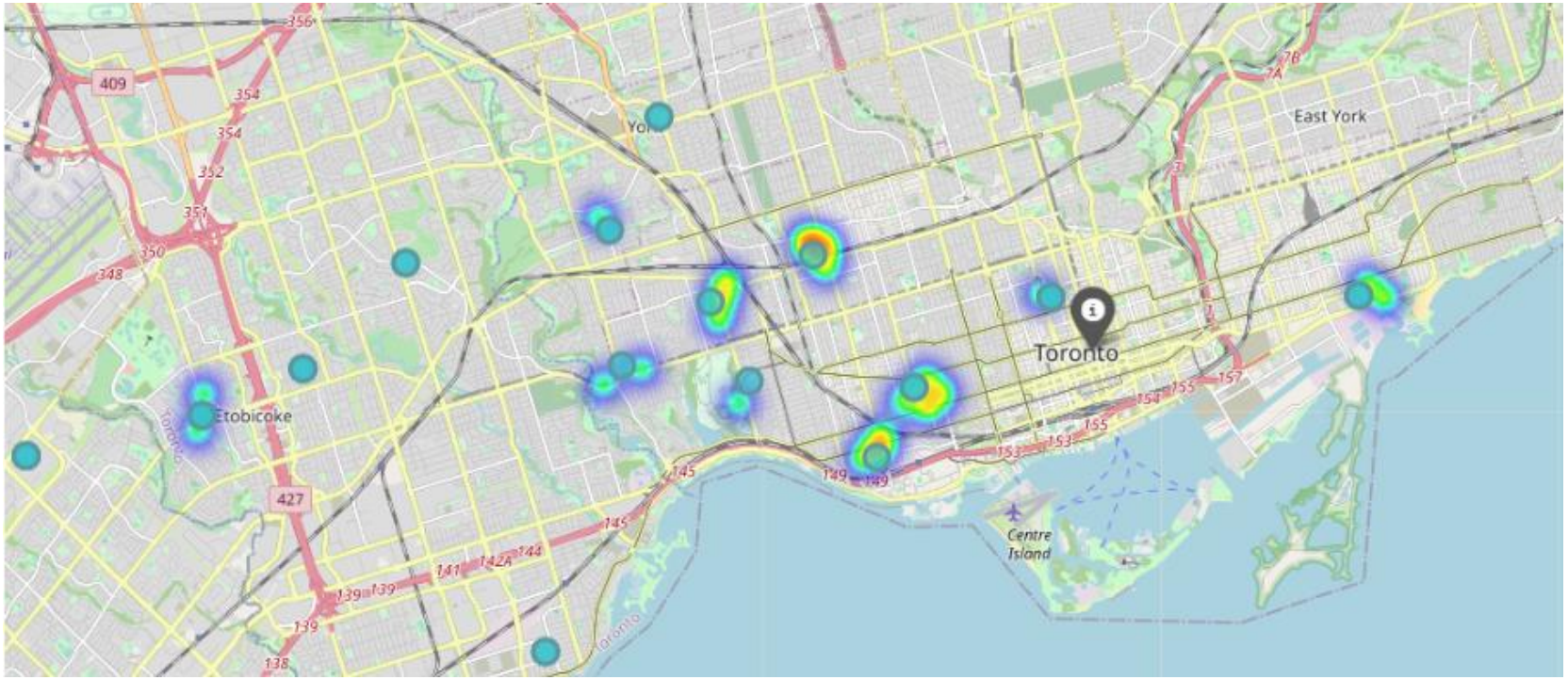
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Parks
0	Rosedale	43.669005	-79.442259	Wallace Emerson Park	43.666933	-79.439449	Park	1
1	Rosedale	43.669005	-79.442259	Chandos Park North	43.670900	-79.442400	Park	1
2	Rosedale	43.669005	-79.442259	Brandon Avenue Parkette	43.670486	-79.440048	Park	1
3	Rosedale	43.669005	-79.442259	Beaver Lightbourn Parkette	43.671400	-79.443500	Park	1
4	Rosedale	43.669005	-79.442259	Bristol Avenue Parkette	43.670700	-79.438498	Park	1
5	St. James Town	43.647927	-79.419750	Trinity Bellwoods Park	43.647072	-79.413756	Park	1
6	St. James Town	43.647927	-79.419750	Trinity Bellwoods Farmers Market	43.649410	-79.417406	Park	1
7	St. James Town	43.647927	-79.419750	Osler Playground	43.646096	-79.422341	Park	1
8	St. James Town	43.647927	-79.419750	Shaw Park	43.644519	-79.416587	Park	1
9	Cabbagetown	43.647927	-79.419750	Trinity Bellwoods Park	43.647072	-79.413756	Park	1
10	Cabbagetown	43.647927	-79.419750	Trinity Bellwoods Farmers Market	43.649410	-79.417406	Park	1

A heat map is a two-dimensional representation of data in which the values are represented by colors. A heat map provides a visual summary of the information we want to represent. Thus heat maps allow to understand quite complex data sets.

In the bar graph, you can see the canidatos neighborhoods for the new park



I can finally see the locations in Toronto and the best candidates



Results and Discussion

Our analysis shows that, although there are many electronics stores in Toronto (in our area of initial interest, which was 12x12 km around downtown Toronto), there are pockets of low-density parks quite close to the city center.

The highest concentration of parks was detected in the south of downtown Toronto, so we focus our attention on the west, southeast, east and southwest areas, which correspond to the Rosedale, Niagara and Toronto Station districts. Another district was identified as potentially interesting (CN Tower, south of the city center), but we focus on Rosedale and Niagara, which offer a combination of popularity among tourists, proximity to the city center, strong socio-economic dynamics and A number of stores low density pockets. After directing our attention to this narrowest area of interest (covering approximately 5x5 km south of the center), restaurants were first obtained near the center; those locations were filtered so that those with more than two stores within a radius of 150 m and those with an electronic store within 500 m were removed.

Results and Discussion

Those placement candidates were grouped together to create areas of interest that contain the largest number of placement candidates. The addresses of the centers in those areas were also generated using reverse geocoding to be used as markers / starting points for a more detailed local analysis based on other factors.
The result of all this is 6 zones that contain the greatest number of potential locations of new stores, depending on the number and distance to existing places, both stores in general and parks in particular. This, of course, does not imply that these areas are really optimal locations for a new park! The purpose of this analysis was to provide information only on areas close to the center of Toronto but not full of existing stores (particularly parks); It is quite possible that there is a very good reason for a small number of stores in any of those areas, reasons that would make them unsuitable for a new store, regardless of the lack of competition in the area. Therefore, the recommended areas should be considered only as a starting point for a more detailed analysis that could eventually result in a location that not only does not have close competition, but also other factors taken into account and all other relevant conditions compliments. It is also seen that much of downtown Toronto is saturated with stores in general.

Conclusion

The objective of this project was to identify the areas of Toronto near the center with a low number of stores and particularly parks to help interested parties reduce the search for an optimal location for a new park; When calculating store density distribution from Foursquare data, we first identify general districts that warrant additional analysis (Rosedale and Niagara), and then generate a wide collection of locations that meet some basic requirements with respect to existing nearby parks . The grouping of these locations was then carried out to create the main areas of interest (containing the greatest number of potential locations) and the addresses of those zone centers were created to be used as starting points for the final exploration by the interested. Interested parties will make the final decision on the optimal location of the store based on the specific characteristics of the neighborhoods and locations in each recommended area, taking into account additional factors such as the attractiveness of each location (proximity to the park or water), noise levels / proximity to main roads, availability of real estate, prices, social and economic dynamics of each neighborhood, etc.