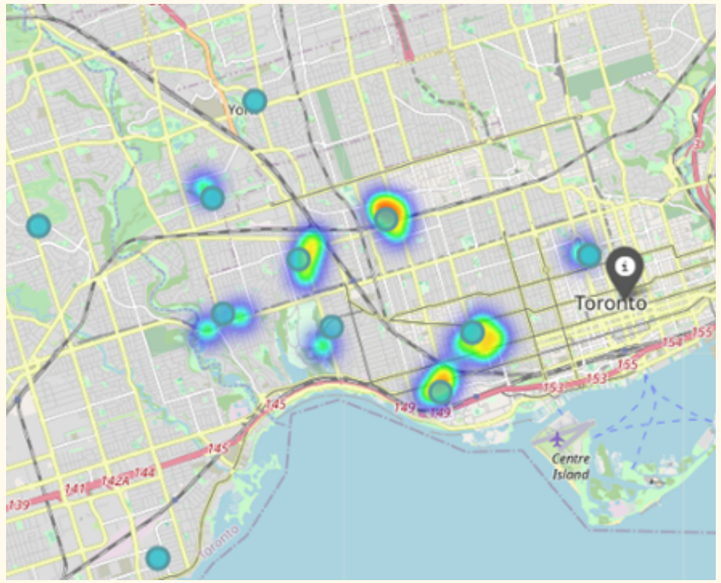
THE BATTLE OF NEIGHBORHOODS

Capstone Project



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

METHODOLOGY

FIRST STEP

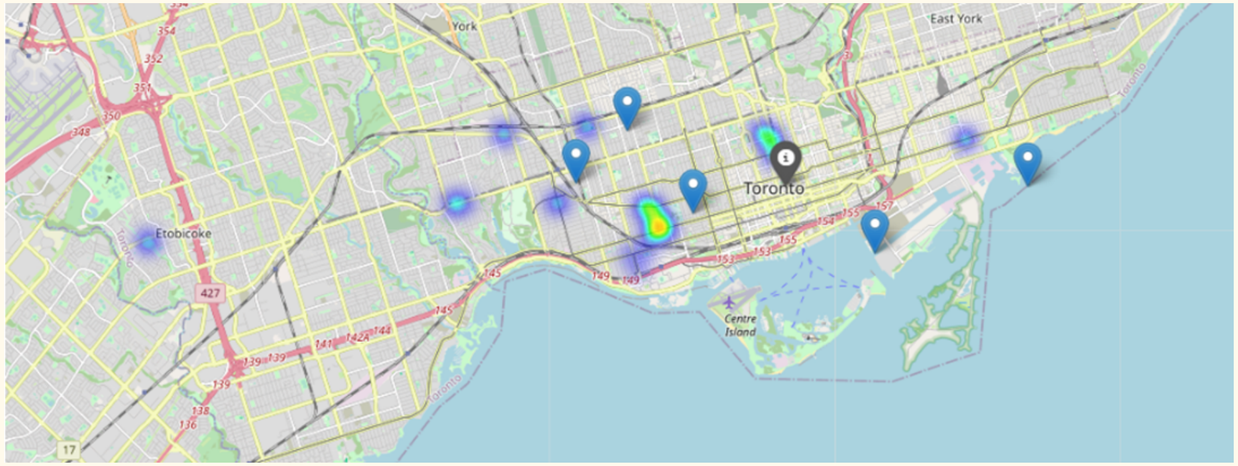
Collected the required data: location and type (category) of every park within 6km from Toronto center (Alexanderplatz).
We have
also identified parks (according to Foursquare categorization).

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.

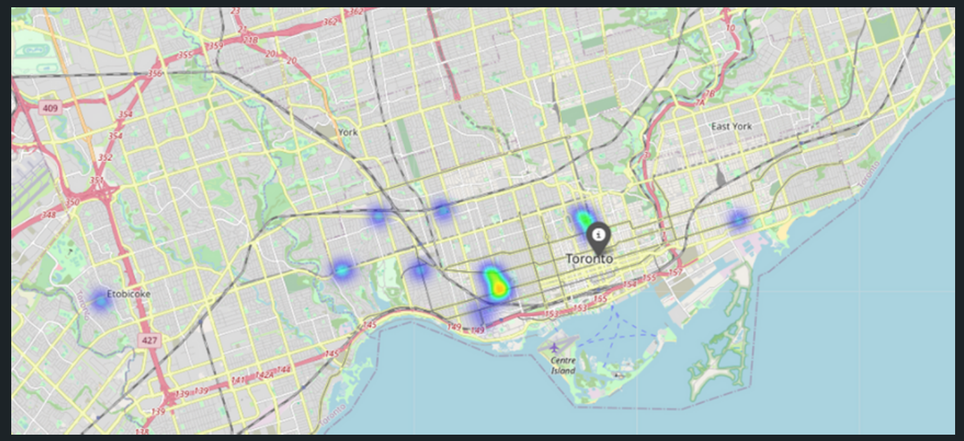
FINAL STEP

Focus on most promising areas and within those create clusters of locations that meet some basic requirements established i 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.



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”

ANALISYS

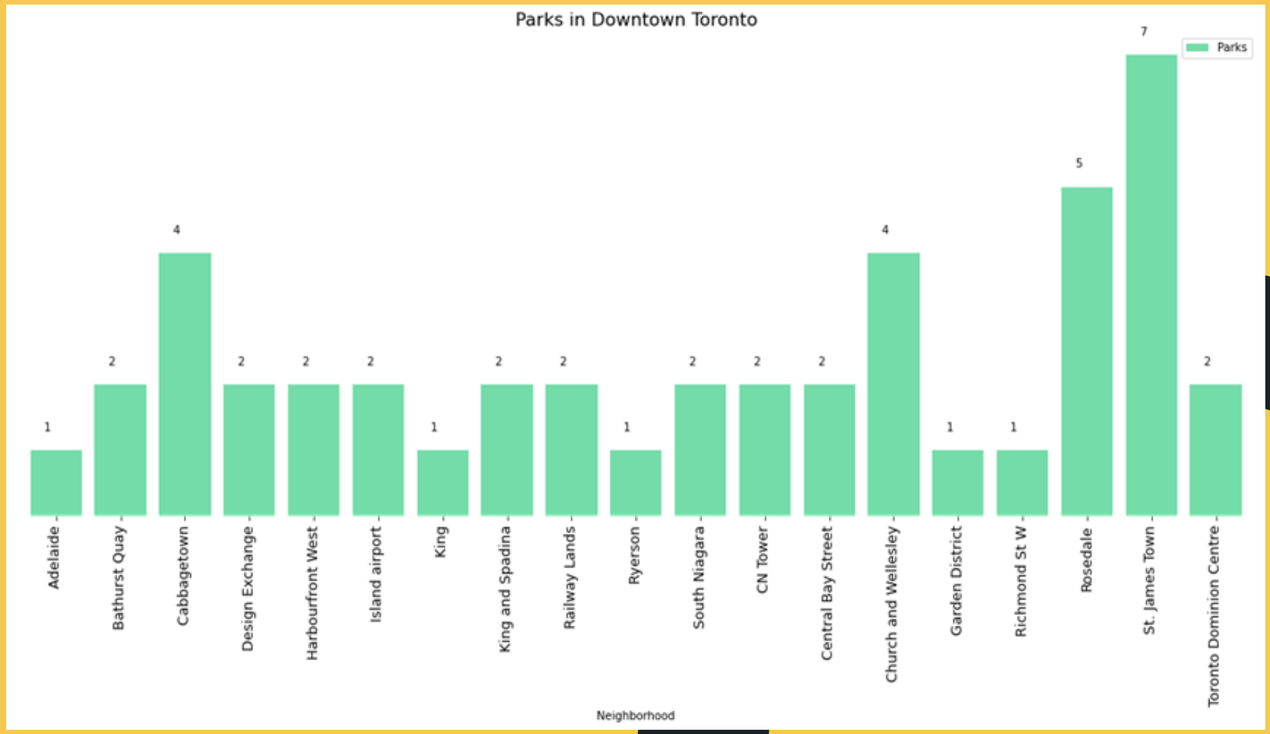


A heat map is a data visualization technique that shows the magnitude of a phenomenon as two-dimensional color. The variation in color can be by hue o intensity, giving obvious visual cues to the reader about how the phenomenon is grouped or modified in space.

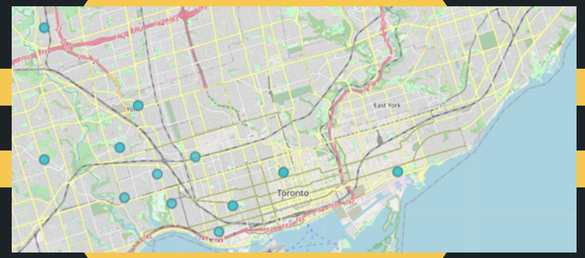
HEAT MAP

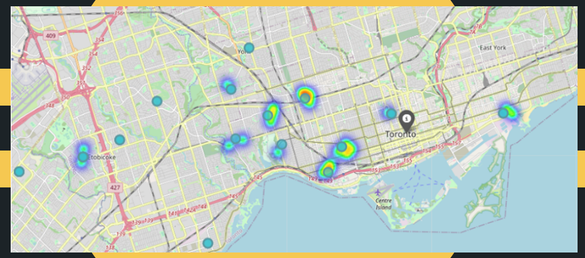
ANALYZE EACH NEIGHBORHOOD

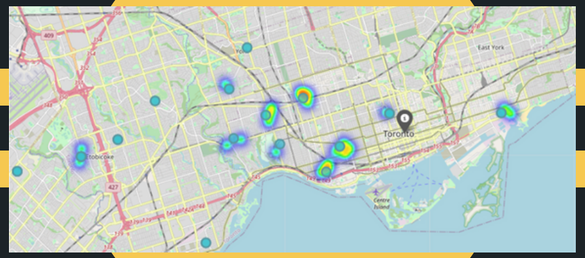
Let's perform some basic exploratory data analysis and derive some additional info from our raw data. First let's count the number of stores in every area candidate.



Explore Neighborhoods







RESULTS AND DISCUSSION

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.
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 existin 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 DYNAMICSOF EACH NEIGHBORHOOD, ETC