The Battle of the Neighborhoods

Capstone Project

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be addressed to interested parties interested in opening an electronics store in Toronto, Canada.

The Battle of the Neighborhoods

Business Problem

Introduction: Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be addressed to interested parties interested in opening an electronics store in Toronto, Canada.

As there are many electronics stores in Toronto, we will try to detect places that are not yet full of stores. We are also especially interested in areas without nearby electronics stores. We also prefer that we have found as close as possible to the city center, assuming that the first two conditions are met.

We will use our data science powers to generate some more promising neighborhoods detected in this criterion. The advantages of each area will be clearly identified so that stakeholders can choose the best possible final location.

Data:

Based on definition of our problem, factors that will influence our decission are:

- number of existing electronic store in the neighborhood
- number of and distance to Electronic store in the neighborhood, if any
- distance of neighborhood from city center
- We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

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

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be addressed to interested parties interested in opening an electronics store in Toronto, Canada.

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Berlin city center

The Battle of the Neighborhoods

Before doing anything, we need to import all the required libraries (such as Pandas, Matplotlib, Numpy, Seaborn, Sklearn, Folium ,Beautifulsoup, Wikipedia) for data collecting, data analysis, modeling and visualization. The required libraries are as follows:

- Pandas: Library for data analysis
- Numpy: Library for handle data in a vectorized manner
- Matplotlib: Library for plotting modules
- Seaborn: Library for plotting modules
- Scikit Learn: Library for building algorithm model
- Folium: Library for visualizing geospatial data
- Beautifulsoup: Library for scraping the HTML data from webpage
- Wikipedia: Library for scraping the data from webpage

Data Preparation

There are two parts for data preparation process. The first part is for boroughs and neighborhood list of Toronto. The second part is for Demographics data of Toronto.

First, the list of districts and neighborhoods in Toronto was imported from Wikipedia using the Beautifulsoup library and turned it into a panda data frame. Since the list imported from Toronto did not include location data, another dataset consisting of location data from neighborhoods and districts was imported and converted to csv format. a panda data frame. After cleaning the data set, the two tables were merged to obtain the final Toronto data set.

	Postcode	Neighborhood	Borough	Latitude	Longitude
0	M1B	Rouge,Malvern,	Scarborough	43.806686	-79.194353
1	M1C	Highland Creek,Rouge Hill,Port Union,	Scarborough	43.784535	-79.160497
2	M1E	Guildwood,Morningside,West Hill,	Scarborough	43.763573	-79.188711
3	M1G	Woburn,	Scarborough	43.770992	-79.216917
4	M1H	Cedarbrae,	Scarborough	43.773136	-79.239476

Data Preparation

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment 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 latitude and longitude coordinates of each neighborhood.

convenience, I downloaded the files and placed it on the server, so you can simply run a wget command and access the data. So, let's go ahead and do that.

The Battle of the Neighborhoods

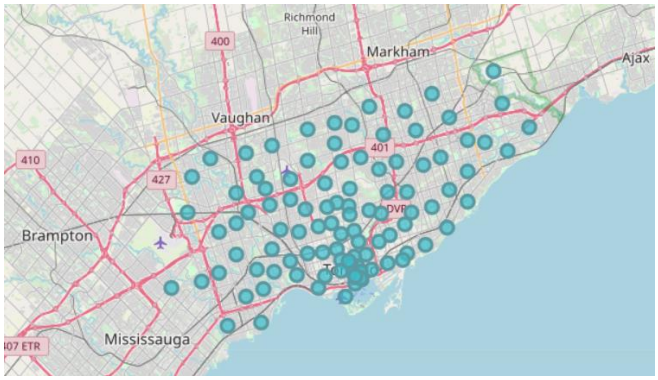
Methodology:

In this project, we will direct our efforts to detect areas of downtown Toronto that have low store density, particularly those with a low number of stores of electronics. We will limit our analysis to the area ~ 6 km around the city center of Toronto.

In the first step, we have collected the required data: location and type (category) of each restaurant less than 6 km from the center of Toronto. We have also identified stores (according to the Foursquare categorization).

The second step in our analysis will be the calculation and exploration of 'store density' in different areas of Toronto: we will use heat maps to identify some promising areas near the center with a low number of stores in general (and there are no stores of electronics nearby) and we focus our attention on those areas.

In the third and final step, we will focus on the most promising areas and within them we will create groups of locations that meet some basic requirements established in the discussion with stakeholders: we will consider locations with no more than two restaurants in a radius of 250 meters, and wants locations without electronic stores within a radius of 500 meters. We will present a map of all those locations, but we will also create groups (using the k-media grouping) of those locations to identify areas / neighborhoods / general addresses that should be a starting point for the final exploration of the 'street level' and look for the optimal location of the place by interested parties

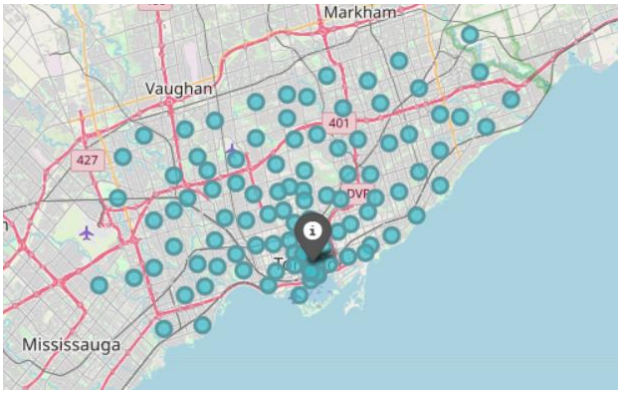


Methodology

In this project, we will direct our efforts to detect areas of downtown Toronto that have low store density, particularly those with a low number of stores of electronics. We will limit our analysis to the area ~ 6 km around the city center of Toronto.

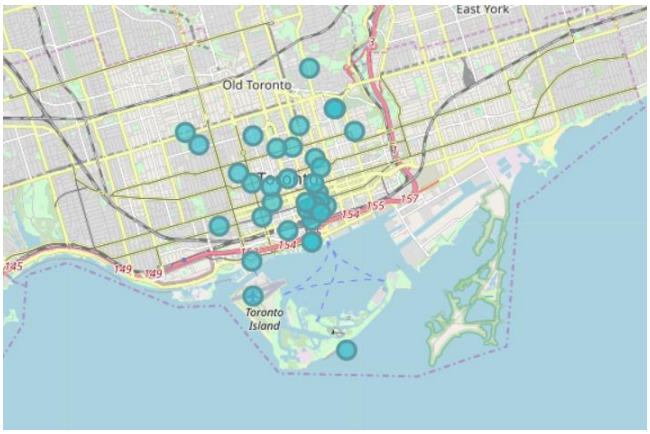
The second step in our analysis will be the calculation and exploration of 'store density' in different areas of Toronto: we will use heat maps to identify some promising areas near the center with a low number of stores in general (and there are no stores of electronics nearby) and we focus our attention on those areas.

The Battle of the Neighborhoods



Analysis

We will create a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Berlin city center (Downtown Toronto).



When all the required data is ready, we will begin to analyze the data based on the category of the place to explore the most popular stores in Toronto.
First, the category of related headquarters in each neighborhood was grouped by "Neighborhood."
After that, the stores in the neighborhood place category was leaked. Then, according to the frequency of occurrence of the restaurant, the data frame was created and the 10 best restaurants were displayed for each neighborhood.

Analysis

In this project, we will direct our efforts to detect areas of downtown Toronto that have low store density, particularly those with a low number of stores of electronics. We will limit our analysis to the area ~ 6 km around the city center of Toronto.

The second step in our analysis will be the calculation and exploration of 'store density' in different areas of Toronto: we will use heat maps to identify some promising areas near the center with a low number of stores in general (and there are no stores of electronics nearby) and we focus our attention on those areas.

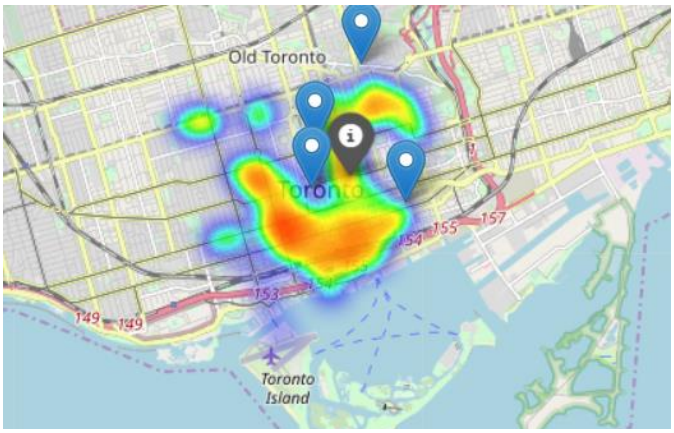
The Battle of the Neighborhoods

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Cabbagetown	43.664473	-79.366986	Fair Trade Jewellery Co.	43.665348	-79.368362	Jewelry Store
1	Cabbagetown	43.664473	-79.366986	Labour Of Love	43.663907	-79.368822	Gift Shop
2	Cabbagetown	43.664473	-79.366986	Butter Chicken Factory	43.667072	-79.369184	Indian Restaurant
3	Cabbagetown	43.664473	-79.366986	No Frills	43.663515	-79.367166	Grocery Store
4	Cabbagetown	43.664473	-79.366986	Spruce	43.663614	-79.367607	Furniture / Home Store

Heat map for stores (General) in downtown Toronto:

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.

There are possible places to locate a store to contribute the heat map.



Heat map

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.

The Battle of the Neighborhoods

Analyze Neighborhood

Analyze Each Neighborhood:

Next, let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category. Let's print each neighborhood along with the top 5 most common venues; First, let's write a function to sort the venues in descending order.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	108-120 Front St W	Coffee Shop (20.0%)	Hotel (15.0%)	Bookstore (10.0%)
1	Adelaide	Clothing Store (40.0%)	Cosmetics Shop (10.0%)	Department Store (10.0%)
2	Bathurst Quay	Grocery Store (50.0%)	Coffee Shop (50.0%)	Dessert Shop (0.0%)
3	Berczy Park	Coffee Shop (15.0%)	Grocery Store (10.0%)	Pet Store (10.0%)
4	CN Tower	Coffee Shop (12.5%)	Shopping Mall (6.25%)	Supermarket (6.25%)

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:

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

20	Kensington Market	Thrift / Vintage Store (10.0%)	Electronics Store (10.0%)	Women's Store (5.0%)
21	King	Clothing Store (15.0%)	Coffee Shop (15.0%)	Department Store (10.0%)
22	King and Spadina	Arts & Crafts Store (20.0%)	Music Store (10.0%)	Coffee Shop (10.0%)
23	Niagara	Boutique (10.0%)	Men's Store (10.0%)	Optical Shop (10.0%)
24	Queen	Clothing Store (40.0%)	Department Store (10.0%)	Sporting Goods Shop (5.0%)
25	Queen's Park	Grocery Store (50.0%)	Discount Store (50.0%)	General Travel (0.0%)

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.

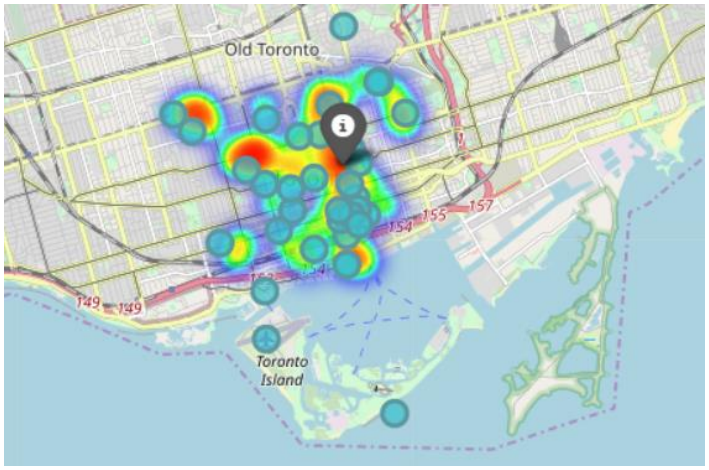
Group rows by neighborhood and by taking the mean of the frequency of occurrence of each category. Let's print each neighborhood along with the top 5 most common venues

The Battle of the Neighborhoods

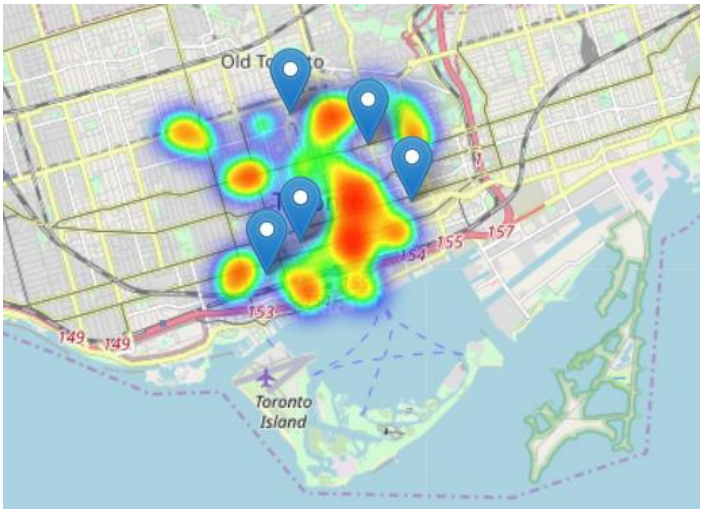
Explore Neighborhood

Explore Neighborhoods in Downtown Toronto (Electronic Store):

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.



There are possible places to locate a store to contribute the heat map.



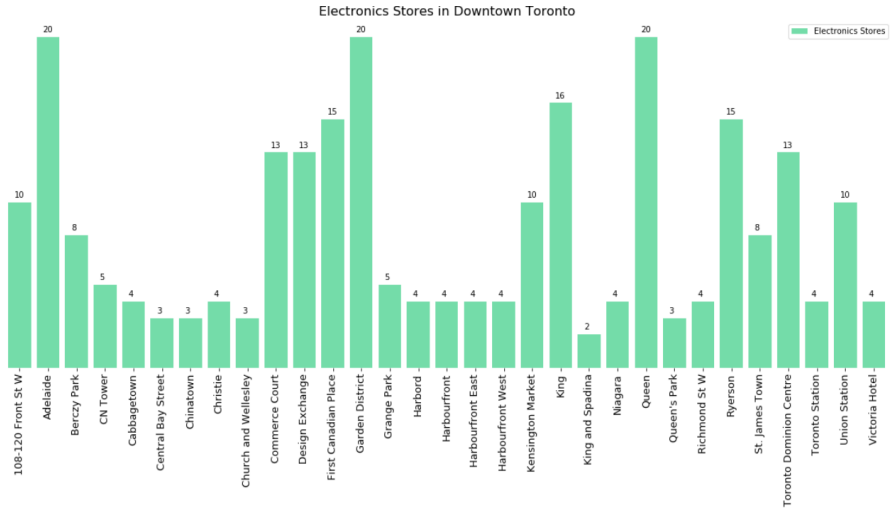
A heat map is a two-dimensional representation of data in which the values are represented by colors. A heat map provides a visual summary of the information we want to represent. Thus heat maps allow to understand quite complex data sets.

The Battle of the Neighborhoods

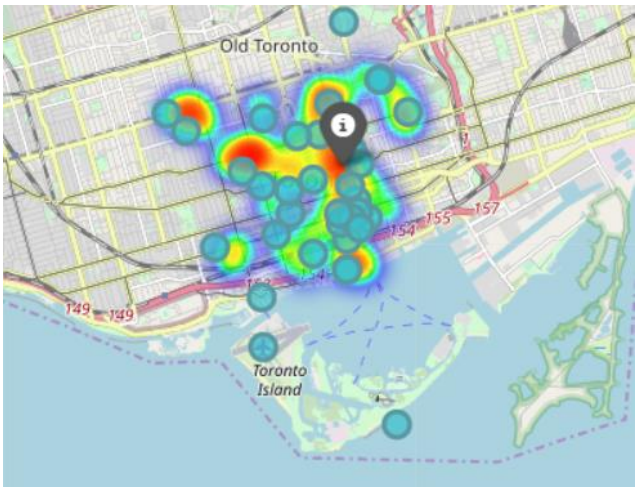
Electronics Stores

Explore Neighborhoods in Downtown Toronto (Electronic Store):

A bar graph is made to identify the neighborhood with the least number of electronic stores and thus be able to rule out candidates.



There are possible places to locate a store to contribute the heat map.



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.

The Battle of the Neighborhoods

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 electronics stores quite close to the city center. . The highest concentration of restaurants was detected in the north, east or 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.

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.

he results 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 electronic stores in particular. This, of course, does not imply that these areas are really optimal locations for a new store! The purpose of this analysis was to provide information only on areas close to the center of Berlin but not full of existing stores (particularly electronic stores).

The Battle of the Neighborhoods

Conclusion :

The objective of this project was to identify the areas of Toronto near the center with a low number of stores (particularly electronics stores) to help interested parties reduce the search for an optimal location for a new electronics store; 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 stores . 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.

We will use our data science powers to generate some more promising neighborhoods detected in this criterion. The advantages of each area will be clearly identified so that stakeholders can choose the best possible final location.