Segementation of Parking Locations in Toronto Using Crowdsourced Data and Foursquare Locations API

Otasowie Osifo

October 14, 2019

1 Introduction

iIn this project, I will be using foursquare location data to analyze and characterize the average percentage number of vehicles searching for parking by the hour around a given parking location in different neighbourhoods of Toronto, Canada.

For different neighbourhoods, the activities at different venues in each neighbourhood has an effect on the average time taken for car owners to find parking space in such neighbourhoods. For example for neighbourhoods with many resturants, night clubs and cafe, the average time taken to find parking space could be much longer in the evenings when compared to the average time taken to find parking space in neighbourhoods that do not have such points of interest.

2 Background

The content of this paragraph is from the web site:[1]. Studies reveal that drivers searching for parking in large urban centers not only causes frustration, but contributes to traffic congestion and additional greenhouse gas (GHG) emissions.[2] To put the problem into perspective, on average, American drivers spend 17 hours a year searching for parking, which is estimated to cost $345 per driver in time, fuel, and emissions.

Drivers are not the only ones impacted by this problem - cities are as well. Out of 6,000 U.S. drivers that responded to a survey, sixty three percent reported they avoided driving to a destination due to parking challenges. This can consequently impact local businesses and economic activity.[3] As large metropolitan cities continue to grow year-over-year, traffic congestion continues to rise with no sign of slowing down.[4]

3 Stakeholders

This type of analysis could be interesting for city planners that need clear understanding of how parking location are used with respect to the activities in the different neighbourhoods of a city. The analysis could also be of interest to would-be enterprenours that are planning to invest in new businesses in different neighbourhoods in a city

4 The Data

Two sets of data will be used in this project:

1. Searching for parking data from Geotab.[1]
2. Foursquare location data

4.1 Searching for Parking Data

This data set is a crowd sourced data set that is provided free of charge by Geotab. Full description is give at the web site:[1]. The data set is constructed from a rolling average for the last six months and it is available in a Google Big Query table that is updated on the second of every month. Table 1 gives a summary of some of the fields in the table:

Field	Type	Description
Geohash	STRING	Geohash at the 7 character level (153m x 153m) that identifies the geohash associated with the parking area
Latitude	FLOAT	The average latitude of the parking location
Longitude	FLOAT	The average longitude of the parking location
City	STRING	City (or municipality) within which the geohash resides (U.S., Canada, and Mexico only)
County	STRING	County within which the geohash resides (U.S. and Mexico only)
State	STRING	State within which the geohash resides (U.S., Canada, and Mexico only)
Country	STRING	Country (or territory) within which the geohash resides (English common name)
AvgTimeToPark	FLOAT	The average time taken to search for parking (in minutes)
AvgTimeToParkRatio	FLOAT	The ratio between the average time taken to search for parking and of those not searching for parking in the current geohash
TotalSearching	FLOAT	The number of drivers searching for parking
PercentSearching	FLOAT	The percentage of drivers that were searching for parking
SearchingByHour	STRING	JSON object representing the average percentage of vehicles searching for parking within the hour
PercentCar	FLOAT	Percentage of vehicles with parking issues that were cars.
UpdateDate	DATETIME	Date and time that the record was updated.
Version	STRINGT	Version number of the dataset.

Table 1: Fields in the Big Query Table

4.1.1 Getting the Data from Google Big Query Table

The data for Toronto, Canada was extracted from the table into csv files and stored locally for analysis. The data extracted sets that were extracted were from July, August, September and October. The datasets were extracted into a csv files from Google's BigQuery Table. In this project, the dataset that was extracted on 20190801 was used in the analysis described below and in total, the datataset contained 3308 unique geohash parking locations for Toronto. Figure 1 shows the average percentage number of vehicles searching for parking by the hour around a given parking location.

Figure 1: Average percentage number of vehicles seaching by the hour at a parking location

4.2 Foursquare Venues Categories

Foursquare provides an API that can be used to access information about different categories of venues around a geographical . The list below shows the main categories that are avaliable through the API. Each main category contains one or more sub-categories.

1. Arts & Entertainment (4d4b7104d754a06370d81259)
2. College & University (4d4b7105d754a06372d81259)
3. Event (4d4b7105d754a06373d81259)
4. Food (4d4b7105d754a06374d81259)
5. Nightlife Spot (4d4b7105d754a06376d81259)
6. Outdoors & Recreation (4d4b7105d754a06377d81259)

7. Professional & Other Places (4d4b7105d754a06375d81259)

8. Residence (4e67e38e036454776db1fb3a)

9. Shop & Service (4d4b7105d754a06378d81259)

10. Travel & Transport (4d4b7105d754a06379d81259)

Full details are available from the web site:[5]. In this project, the API will be used to query for information using the main categories.

5 Exploratory Data Analysis

Exploratory analysis of the data was performed in order get an understanding of the data. The first step in this regard, was to create a calculated Search for Parking Index as described in [1]. This index gives a rough measure of how time consuming searching for parking is and how prevalent it is in any given city. The index is defined according to:
Let μ = Average time to park ratio, ϕ = Searching for parking ratio and θ = Search for Parking Index; then
$\theta = a\mu + b\phi$
Where a and b are parameters that can be varied.[1] Figure 2 shows a heat map of the calculated Serach for Parking Index using the 3308 geohashes in the dataset. The Search for Parking Index θ was scaled to take values between 1 and 10.

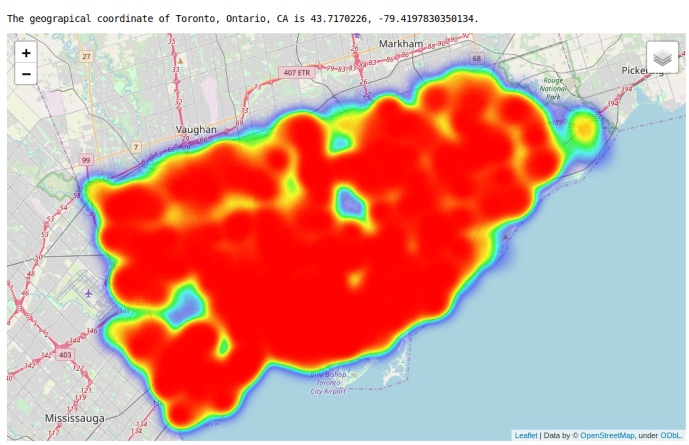

Figure 2: Heat Map of the calculated Search for Parking Index for the dataset in this project

5.1 Statistical properties of the searching for parking data

Below are histograms and Kernel Density Estimation (KDE) plots for: the average time to park and the calculated search for parking index.

5.1.1 Average time to park

The probability distribution of the average time to park seems to be roughly Gaussian like and from the Kernel Density Estimation plot in figure 3, we can see that the most probable (about 25% probability) average time for vehicles to find parking space in Toronto is around 5 minutes.

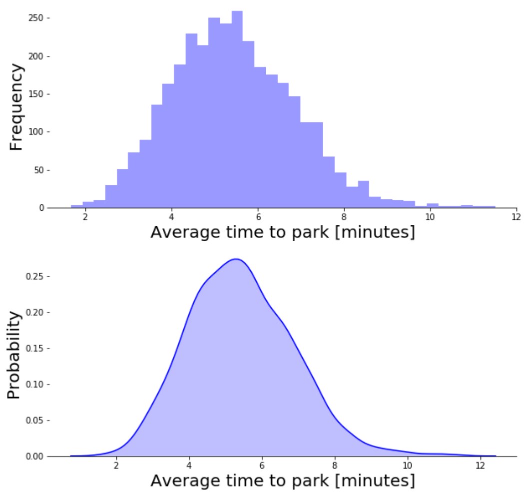

Figure 3: The average time taken for vehicles to find parking using the dataset

5.1.2 Search for Parking Index

The calculated Search for Parking Index has right skewed distribution with the most probable index being around 2 (about 20% probability). This is shown in figure 4 below.

6 Selection of Data set for Segmentation

Given the relatively large dataset of 3308 unique geohashes, it is neccessary to use a subset for analysis due to restrictions on the number of API calls for the free Foursquare subscription account. To this effect, only parking locations with calculated Index around 10 were selected for further analaysis. The selected region is in the far right tail of the Karnel Density Estimation plot in figure 4. The slection criteria resulted in a dataset of 55 parking locations; a heat map of the selected dataset is shown in figure 5 below.

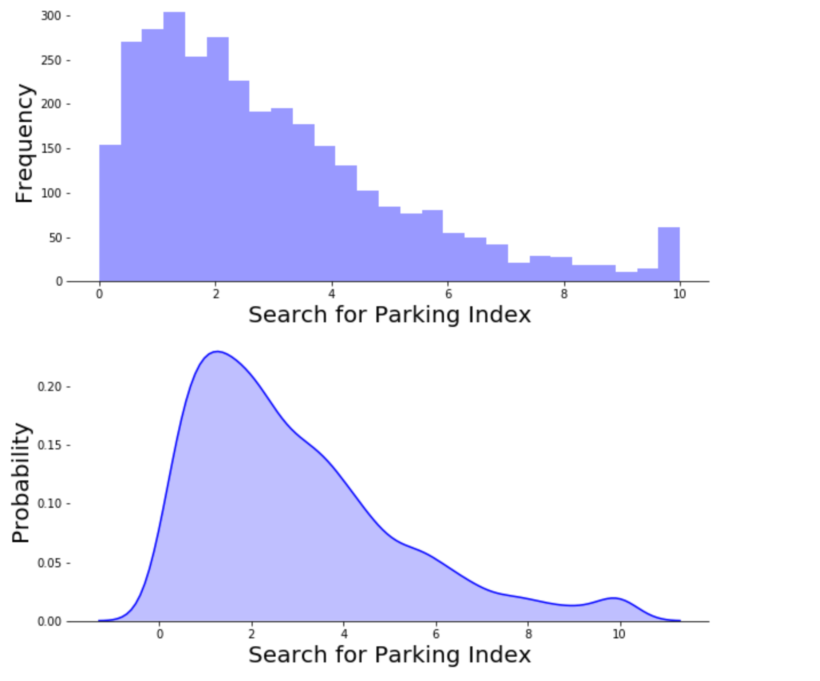

Figure 4: Calculated Search for Parking Index

6.1 Getting the Locations Foursquare Locations Data

For each parking location in the selected dataset, the Foursquare API was used to query for Points of Interests (POI) within 1000 metres of each of the 55 parking locations. For each parking location, the query was executed for each of the 10 main categories defined in the API. This resulted in 550 API calls. Figure 6 shows a table of the counts of the point of Interests across different categories for a parking location (geohash = dpz893q) while figure 7 shows a boxplot of the counts each of the 10 categories across the 55 parking locations.

7 Segementation of Selected parking Locations

In this section, segmentation/clustering of parking locations based on Foursquare Points of Interests will be performed. Figure 8 shows a table of values of the Foursquare Point of Interests for the geohash: dpz893Iq. The dataset contains a total of 6051 rows.

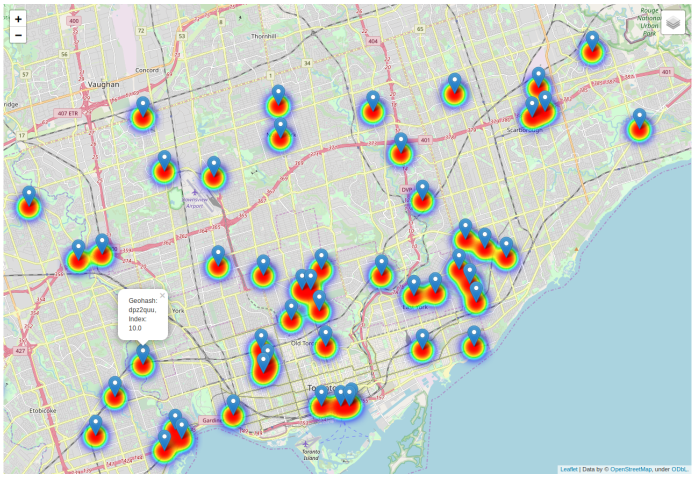

Figure 5: The parking locations that were selected for clustering and segmentation

7.1 Method

The following methods will be used in the analysis and categorization of the searching of parking locations:

1. Use boxplots to show the number of venues in each Foursquare venue category in order to show which activity dominates utilization of the parking location.
2. Use k-means clustering to show the dominant venue categories within 1000 metres of a parking location.

7.2 Clustering/Segmentation of parking locations based on Foursquare locations

Using K-means clustering (unsupervised learning), clusters of the 55 selected parking locations were constructed based on the popular Foursquare locations inside a radius of 1000 metres from each location. The clusters enables the parking locations to be categorized/segmented by popular usage based on the sorrounding Points of Interests. In the procedure, the number of clusters was set to 6 after expmerimenting with different numbers of clusters as input to the for k-means algorithm. Figure 9 shows a map of the clusters, the six clusters are numbered from 0 to 6. In the map, we see clusters of parking locations with similar characteristics as shown by the colors. We have the followings: cluster 0 has color red, cluster 1 has color pink, cluster 2 has color blue, cluster 3 has color 3 light blue, cluster 4 has color light green while cluster 5 has color light brown. Each of these clusters is discussed below. Note that depending on the size of the data returned by the Foursquare API per parking location for each

	Geohash	Category Name	Category Count
0	dpz893q	Arts & Entertainment	18
1	dpz893q	College & University	5
2	dpz893q	Event	1
3	dpz893q	Food	30
4	dpz893q	Nightlife Spot	11
5	dpz893q	Outdoors & Recreation	19
6	dpz893q	Professional & Other Places	30
7	dpz893q	Residence	5
8	dpz893q	Shop & Service	30
9	dpz893q	Travel & Transport	10

Figure 6: Counts of the Points of Interests the main categories accross for a parking locations with geohash: dpz893q

of the 10 main categories, the clusters will look different between runs of the code.

8 Discussion

8.1 Cluster 0

Figure 10 shows a boxplot of the counts of the categories of Foursquare locations within 1000 metres of parking locations that belong to cluster 0. In the boxplot, we see that usage of the parking locations is dominated by Restaurants (Food) and Shop & Service. We also see that the event category of Foursquare locations does not contribute to the usage of parking locations in this cluster.

Figure 11 shows a map of the regions covered by cluster 0 in the different neighbourhoods of Toronto

8.2 Cluster 1

Figure 12 shows a boxplot of the counts of the categories of Foursquare locations within 1000 metres of parking locations that belong to cluster 1. In the boxplot, we see that usage of the parking locations is dominated by Arts & Entertainment, College & University, Event and Residence. We also see that the other six main categories of Foursquare locations do not contribute to the usage of parking locations in this cluster.

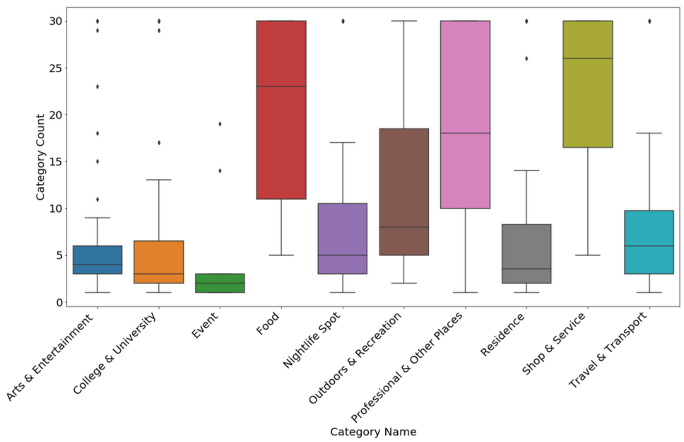

Figure 7: Counts of the Points of Interests in each main category accross the 55 parking locations

Figure 13 shows a map of the regions covered by cluster 1 in the different neighbourhoods of Toronto

8.3 Cluster 2

Figure 14 shows a boxplot of the counts of the categories of Foursquare locations within 1000 metres of parking locations that belong to cluster 2. In the boxplot, we see that usage of the parking locations is dominated by Restaurants, Professional & Other Places, and Shop & Service.

Figure 15 shows a map of the regions covered by cluster 2 in the different neighbourhoods of Toronto

8.4 Clusters 3 and 4

These are single clusters in the map. Further analysis needs to be performed in order to include them in the discussions.

8.5 Cluster 5

Figure 16 shows a boxplot of the counts of the categories of Foursquare locations within 1000 metres of parking locations that belong to cluster 5. In the boxplot, we see that usage of the parking locations is dominated by Restaurants, Outdoors & Recreation, and Shop & Service.

Figure 17 shows a map of the regions covered by cluster 5 in the different neighbourhoods of Toronto

	Venue	Venue Latitude	Venue Longitude	Venue Category	Geohash
0	Cineplex Entertainment - Head Office	43.684314	-79.392285	Movie Theater	dpz893q
1	Andrew Freedom Road	43.690156	-79.390726	Music Venue	dpz893q
2	Skyrims Movie Room	43.687683	-79.391580	Movie Theater	dpz893q
3	tarragon	43.688652	-79.393678	Theater	dpz893q
4	Tunnel of Glam	43.689242	-79.394352	Art Gallery	dpz893q
5	Kaghaz Rangi	43.689002	-79.394745	Art Gallery	dpz893q
6	Your Radio	43.688915	-79.394885	Music Venue	dpz893q
7	Toronto Chamber Choir	43.690963	-79.395103	Music Venue	dpz893q
8	Woodlawn Pottery Studio	43.684445	-79.389843	Art Gallery	dpz893q
9	Cineplex Office	43.684700	-79.392856	Movie Theater	dpz893q
10	Muse Gallery	43.682707	-79.392371	Art Gallery	dpz893q
11	Muse Gallery	43.682450	-79.391729	Art Gallery	dpz893q
12	Learning by Heart Studio	43.682208	-79.390144	Dance Studio	dpz893q
13	Division Gallery	43.687061	-79.399035	Art Gallery	dpz893q
14	the dc beat cathedral	43.696802	-79.391740	Concert Hall	dpz893q
15	Exploration House	43.681235	-79.392431	Art Gallery	dpz893q
16	The Al Green Gallery	43.697037	-79.393850	Art Gallery	dpz893q
17	Compañia Carmen Romero	43.698085	-79.389266	Dance Studio	dpz893q
18	Deer Park Public School	43.689241	-79.391619	College Academic Building	dpz893q
19	English School of Canada	43.688486	-79.391119	College Classroom	dpz893q
20	University of Bahcesehir Campus Toronto	43.697136	-79.392708	College Academic Building	dpz893q
21	Academy Of Design And Technology	43.696640	-79.396000	General College & University	dpz893q
22	Paige Property Inc.	43.694522	-79.399056	College Administrative Building	dpz893q
23	On My Way to Steal Yo Girl	43.692884	-79.379219	Parade	dpz893q
24	9bars	43.688660	-79.391940	Café	dpz893q
25	The Bagel House	43.687374	-79.393696	Bagel Shop	dpz893q
26	Cava Restaurant	43.689809	-79.394932	Tapas Restaurant	dpz893q
27	Capocaccia Café	43.685915	-79.393305	Italian Restaurant	dpz893q
28	Union Social Eatery	43.687895	-79.394916	American Restaurant	dpz893q
29	kibo sushi house	43.682727	-79.392270	Sushi Restaurant	dpz893q

Figure 8: Points of Interests within 1000 metres of the parking geohash: dpz893Iq

9 Conclusion

Segmentation/Charactrisation of parking locations has been performed using Foursquare locations data and crowdsourced gps searching for parking data. By using this type of segementation, parking locations with similar characteristics can be identified by city planners while making decisions regarding improvements and new investments.

References

[1] https://data.geotab.com/urban-infrastructure/searching-for-parking

[2] Sidewalk Talk, "Circling for Parking is Terrible For Cities", September 21, 2016

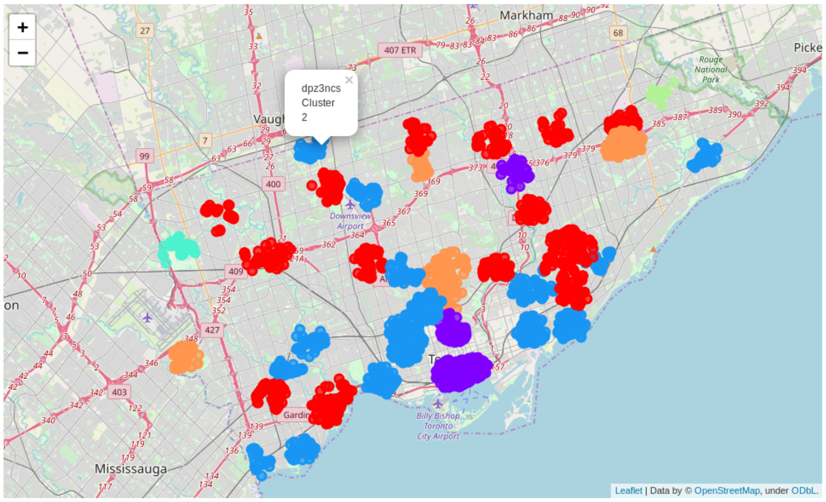

Figure 9: Clusters of parking locations

[3] USA Today, “Drivers spend an average of 17 hours a year searching for parking spots”, July 12, 2017

[4] The US Department of Transportation, Office of Public Affairs, “Driving Topped 262 Billion Miles In March, New Data Show”, May 20, 2015

[5] https://developer.foursquare.com/docs/resources/categories

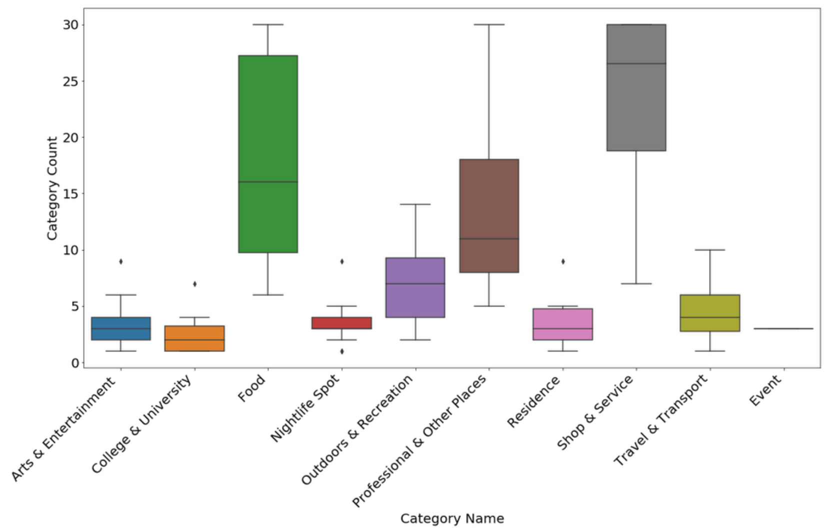

Figure 10: Boxplot of Foursquare locations for the parking geohashes in cluster 0

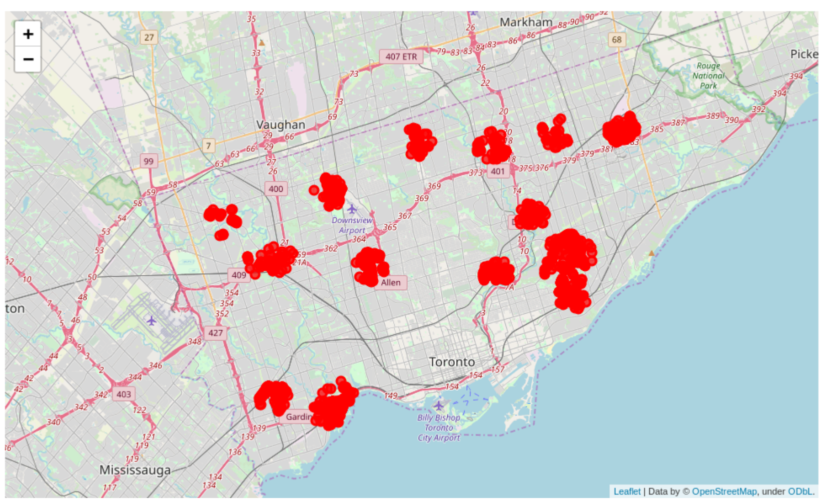

Figure 11: Map of the parking locations in cluster 0

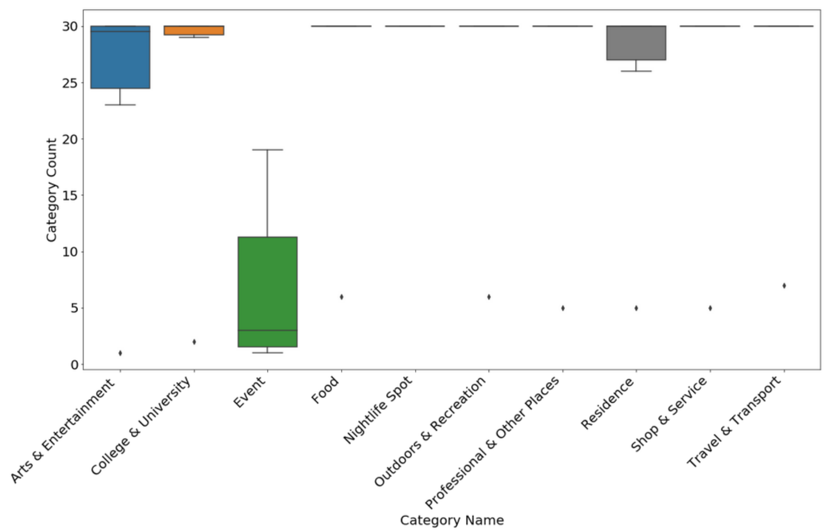

Figure 12: Boxplot of Foursquare locations for the parking geohashes in cluster 1

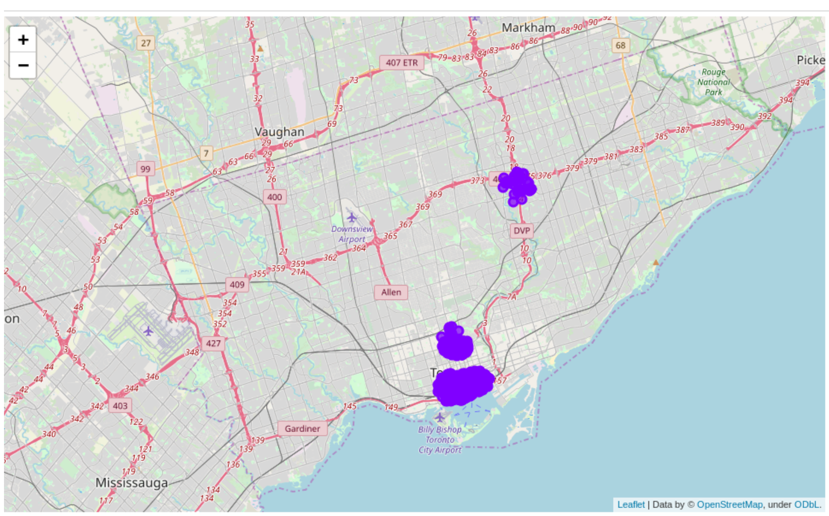

Figure 13: Map of the parking locations in cluster 1

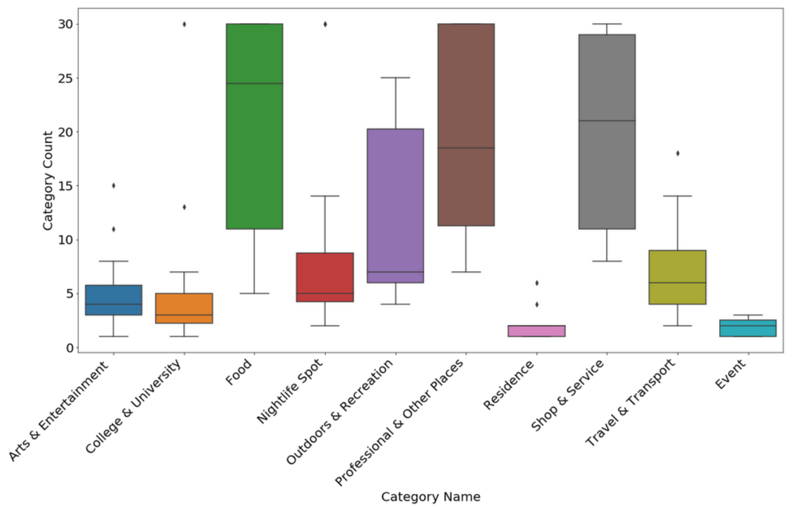

Figure 14: Boxplot of Foursquare locations for the parking geohashes in cluster 2

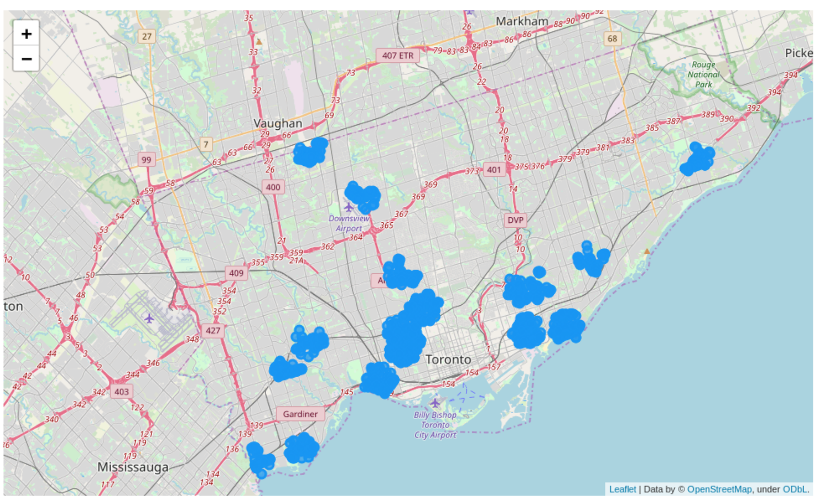

Figure 15: Map of the parking locations with similar characteristics in cluster 2

```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]), <a list of 10 Text xticklabel objects>)
```

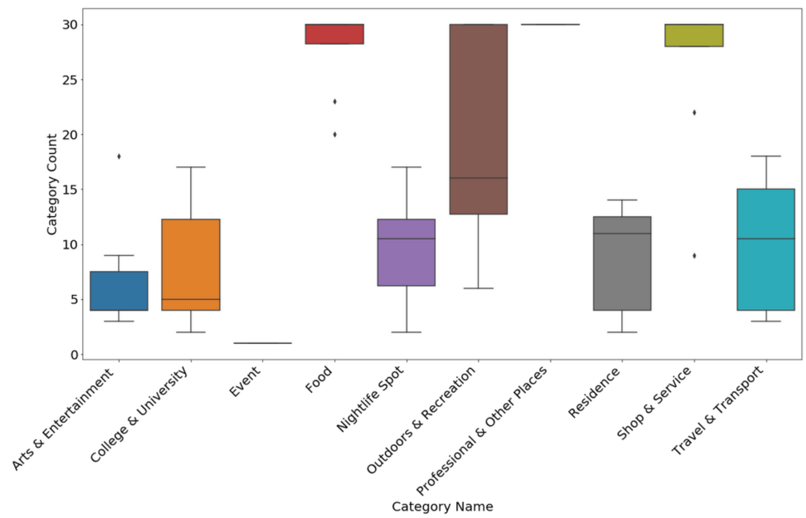

Figure 16: Boxplot of Foursquare locations for the parking geohashes in cluster 5

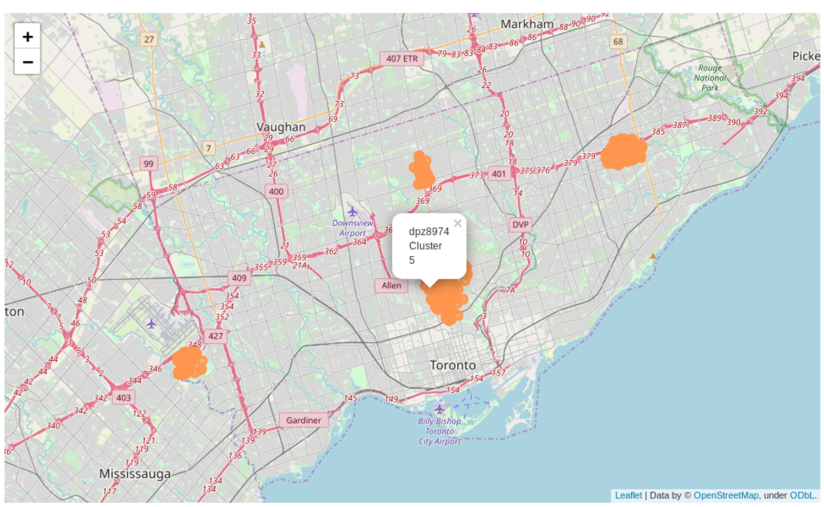

Figure 17: Map of the parking locations with similar characteristics in cluster 5