

Applied Data Science Capstone:

Battle of Neighbourhoods

Created by: Jana Chmelova

Introduction/Business problem

I am a very enthusiastic traveller and that's why I am planning to focus on comparison of neighbourhoods in Toronto, particularly which part of the city can offer me and other tourists the best opportunities for sightseeing, culture and fun. I did some research, as I usually do before any journey and I was surprised with this fact: on Jan. 1, 1998, Toronto's amalgamation took effect, merging the six previous municipalities that made up Metro Toronto – Etobicoke, Scarborough, York, East York, North York, and the City of Toronto, into a new singular City of Toronto. It seems Toronto has, like any other big city, a very interesting history.

The aim of this project is to focus on Toronto's neighbourhoods in order to find the best part of the city for tourist activities (with respect to Art & Entertainment). The analysis is built on the information available on Foursquare webpage, so I want to highlight it has some limitations in content.

Data description

The data for this project is extracted from the following sources:

Toronto's Neighbourhoods data

Description: general information about the Toronto's neighbourhoods, such as postal code, name. This data set is used to explore various neighbourhoods of Toronto.

Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Geospace data

Description: geographical coordinates. This source provides data for Toronto's borough boundaries.

Link: http://cocl.us/Geospatial_data

Foursquare API

Description: service where the user can search and discover information about different types of

entertainment, such as coffee shops, restaurants, including the rating and tips. It is needed for detailed neighbourhood exploration, based on different criteria.

Link: <https://api.foursquare.com>

Approach

- gather the Toronto's neighbourhood's data
- clean data
- use Foursquare API to get information about venues for each neighbourhood
- focus on Art and Entertainment category - purposes of this project
- find ratings, tips via Foursquare API
- visualize the neighbourhoods' segments via Folium library

Methodology section

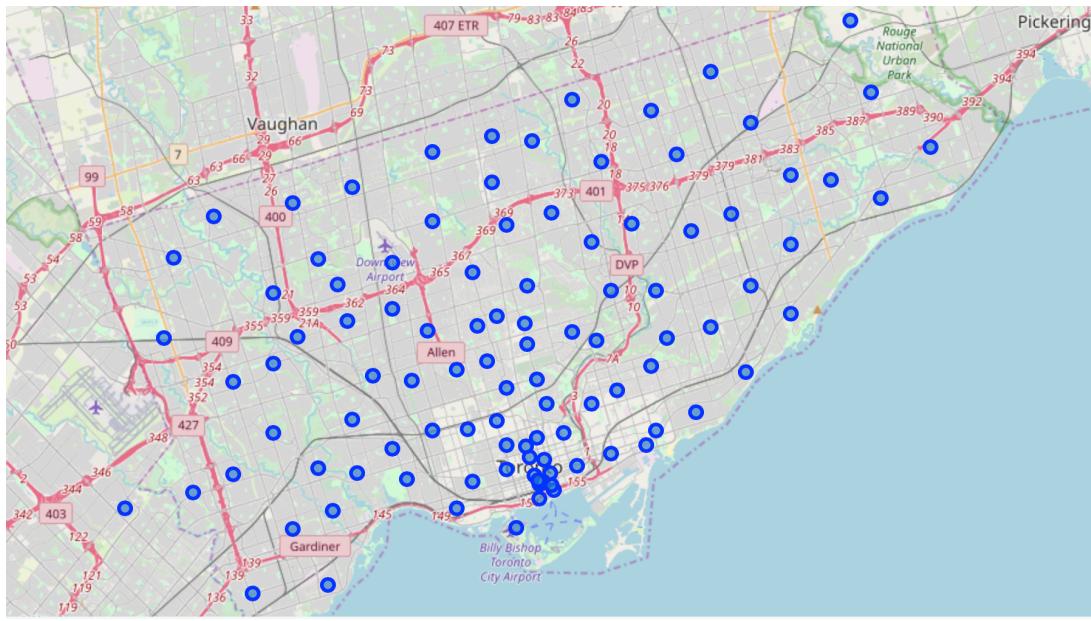
The project should recommend the best neighbourhoods for tourist entertainment in Toronto (Canada). Such neighbourhood should have historic sites and also additional art and entertainment places.

The first stage of the project gathers data from different sources: Toronto's neighbourhoods, geographical coordinates, art and entertainment places from Foursquare API. After cleaning up of the dataset, the analysis focuses on deeper understanding what each neighbourhood can offer. The frequency of occurrence of various categories existing in each neighbourhood is also counted. In the last step, the neighbourhoods are clustered based on the type of category. This cluster part should help the tourist with the starting decision which neighbourhood deserves time.

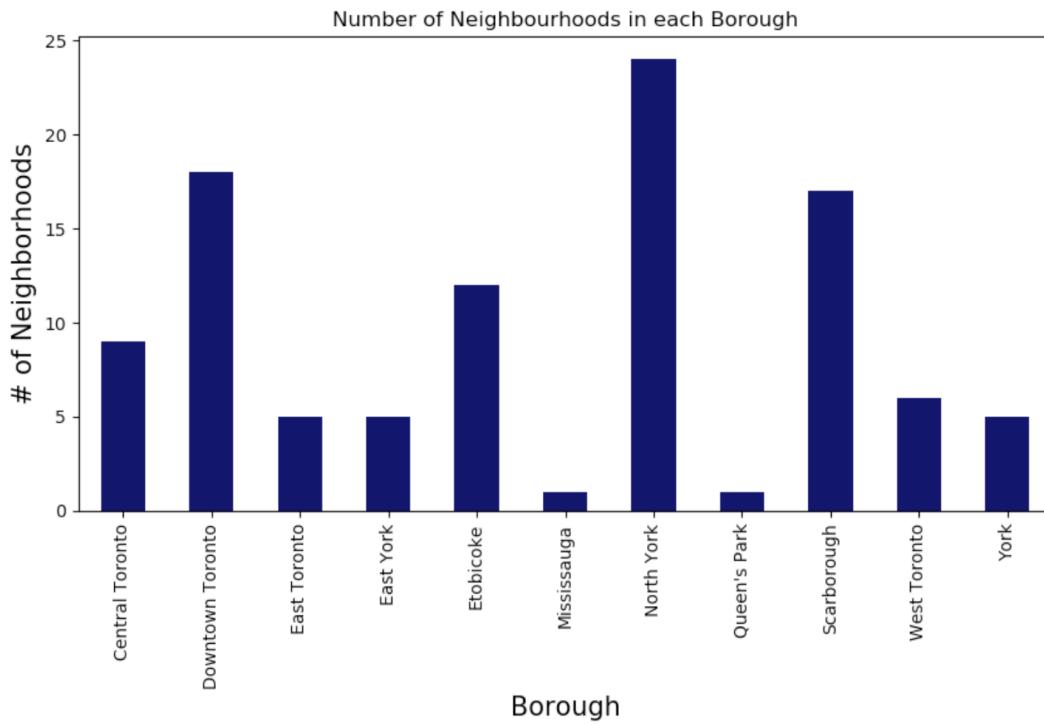
Results section

The project focused into details on art and entertainment category in Toronto's neighbourhoods. I do not want to bother with all steps which had to be done to gather and clean the data.

So let's visualize Toronto's neighbourhoods on a map.

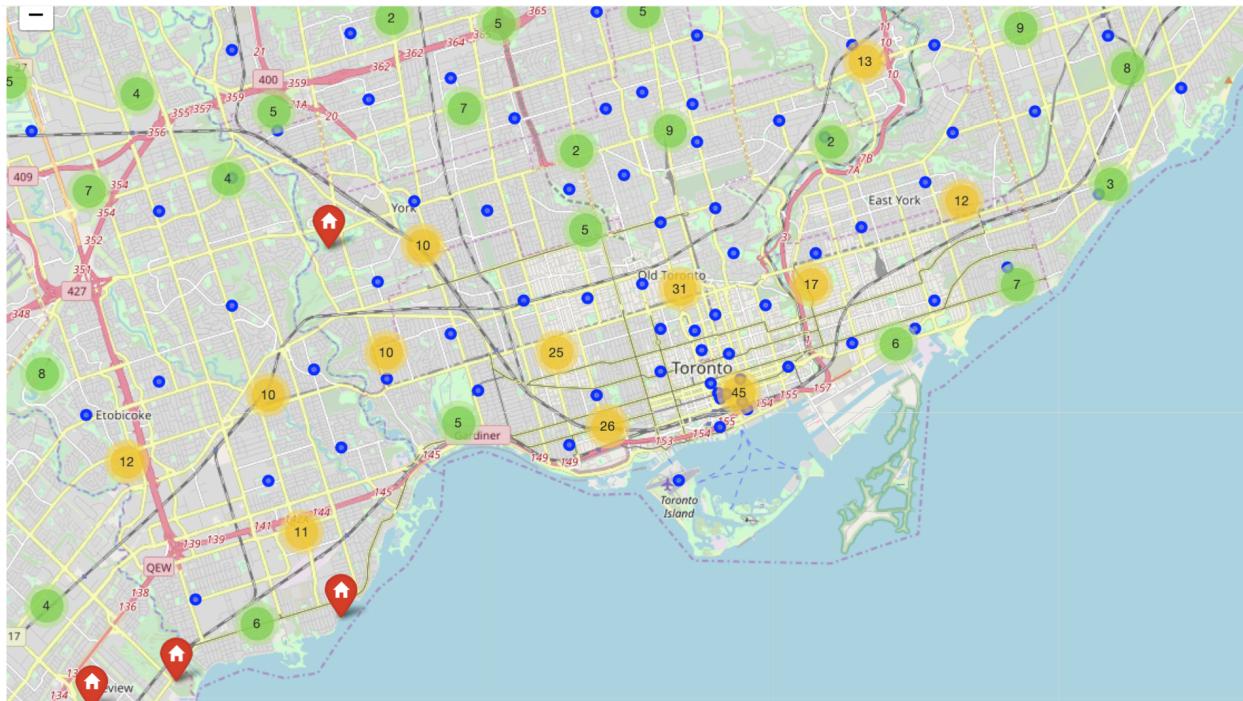


There are 103 neighbourhoods in Toronto city which are being distributed into Boroughs.



It is North York borough with the highest number of neighbourhoods. Next in the row are Downtown Toronto and Scarborough.

When the basis overview about the Toronto's neighbourhoods has been collected, the analysis continued with Art and Entertainment category. Data was extracted from Foursquare. Category "Art and Entertainment" is represented by 90 unique sub-categories in all Toronto's neighbourhood.



Let's look closer on the count of art and entertainment in each neighbourhood.

Top 5 Toronto's neighbourhoods which have the highest number of places for entertainment

	PostalCode	Borough	Neighbourhood	Latitude	Longitude	Entertainment_Count
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	30.0
86	M7R	Mississauga	Canada Post Gateway Processing Centre	43.636966	-79.615819	16.0
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	15.0
14	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	14.0
83	M6R	West Toronto	Parkdale, Roncesvalles	43.648960	-79.456325	14.0

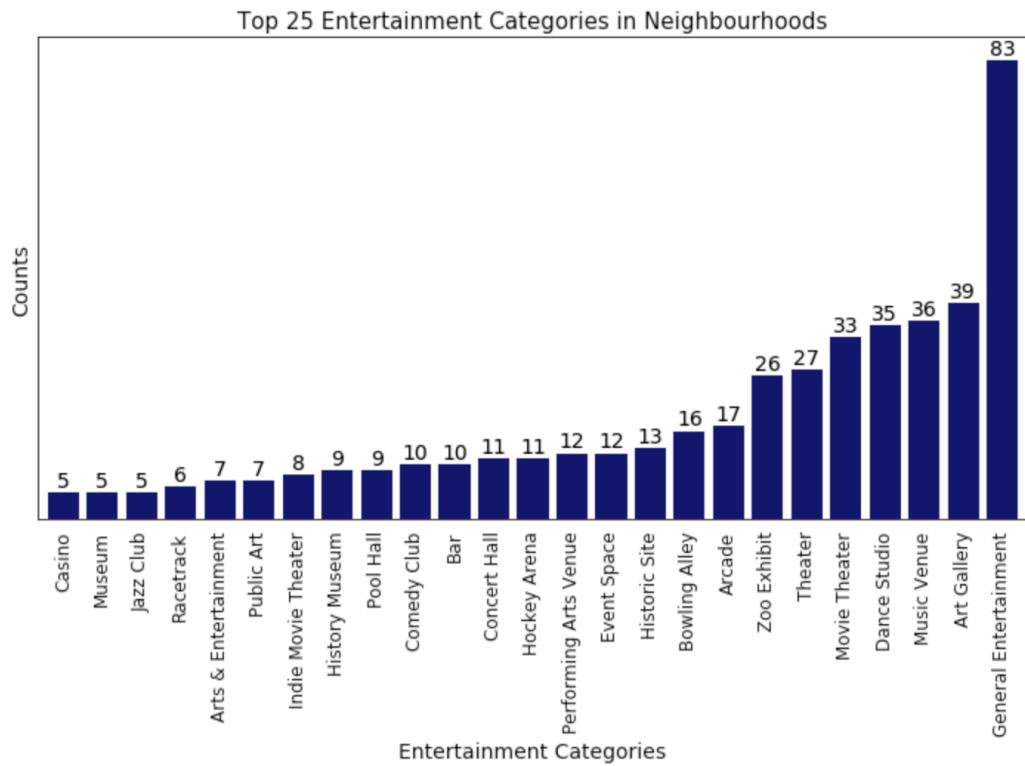
List of Toronto's neighbourhoods which do not have even one entertainment place

	PostalCode	Borough	Neighbourhood	Latitude	Longitude	Entertainment_Count
19	M2K	North York	Bayview Village	43.786947	-79.385975	0.0
16	M1X	Scarborough	Upper Rouge	43.836125	-79.205636	0.0
61	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817	0.0
48	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160	0.0
31	M3L	North York	Downsview West	43.739015	-79.506944	0.0
85	M7A	Queen's Park	Queen's Park	43.662301	-79.389494	0.0
46	M4R	Central Toronto	North Toronto West	43.715383	-79.405678	0.0
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	0.0
77	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750	0.0
75	M6G	Downtown Toronto	Christie	43.669542	-79.422564	0.0
56	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	0.0
57	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	0.0
70	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	0.0
51	M4X	Downtown Toronto	Cabbagetown, St. James Town	43.667967	-79.367675	0.0

It seems like there are 14 neighbourhoods in Toronto which do not have even one place for entertainment. These ones cannot be recommended to the tourists at all.

Now, then we have an idea about the overall picture of art and entertainment, the deeper analysis is needed for each neighbourhood. One hot encoding is used for next analysis (when "1" means that a category is there, "0" in case the category is not there). On the basis of one hot encoding, the mean of the frequency of occurrence of each category will be calculated. As a next step, top 10 Art and Entertainment categories for each neighbourhood is extracted.

Based on the analysis, it seems there are only 13 historic sites in Toronto and 9 history museums. Closer analysis also shows that Zoo Exhibit category is represented only by one place, with a few exhibits. The best category is general entertainment.



Cluster examination

Machine learning technique - clustering - was applied to segment the neighbourhoods into similar objects cluster. This final step could be used by tourists as the starting point for the trip agenda.

Cluster 1 – not recommended; this cluster does not offer very huge range of activities or interesting places for sightseeing.

Cluster 2 – recommended; the second cluster seems to be more tourist friendly, with some venues, galleries and studios.

Cluster 3 – recommended; third segment is oriented a lot on art and free time activities.

Cluster 4 – recommended; this part of the city is defined by historic sites.

Cluster 5 – not recommended; the last cluster is represented mainly by general entertainment category.

More details at: <https://github.com/JanaChmel/bug-free-parakeet/blob/master/Battle%20of%20Neighborhoods.ipynb>

Discussion section

Downton Toronto is recommended for visit in case the tourists prefer historical sites and also some other art and entertainment places. It is difficult to find any dissimilarity in regard to the chosen Art and Entertainment category as the analysis focused on big city, each neighbourhood has a lot of places for relax and fun. The analysis is built on the Foursquare API basis, which means the list of art and entertainment category should not be seen as the final one.

Conclusion section

The above analysis confirmed that Toronto, as any other big city in the world, can offer a lot of interesting places to the tourists. And still it is true that every city is unique in its own way.