Capstone Project: Final Assignment Report

Introduction:

Millions of people travel for other cities and countries every year. Not only for hobby or pleasure, but also for business too. So, for every travel, people has a lot of issues to resolve like the need of hotels, restaurants, recreation, visit the best nearby places and so on. Besides that, some problems can occur like long displacements and the fear of getting lost in unknown places.

This project have the intention to make better a travel experience using the city of Toronto as an example. The intend is to clusterize nearby locations of Toronto using hotels, cafes and other places to do this classification.

Data

As a data source will be used an integration with Foursquare, which is a technology company that uses location intelligence to build meaningful consumer experiences and business solutions. Foursquare API provides a great amount of quality data about locations like restaurants, cafe, gyms, etc. This will allow travelling people to easily decide where the best places to go when in an unknown city.

Besides that it will be necessary to collect postal codes and neighborhoods of Toronto from a Wikipedia page to use with the Foursquare locations.

Methodology

Below are the steps of the developed methodology:

- First, the geojson data for Toronto's neighborhood is obtained from Wikipedia and transformed in a dataframe
- Then, the city data is analyzed using the Foursquare API.
- Lastly, each area in the city and its category is identified using clustering.

Results

Using data manipulation and techniques such as K-means clustering, the results in the next slides obtained common venues in the city nearby.

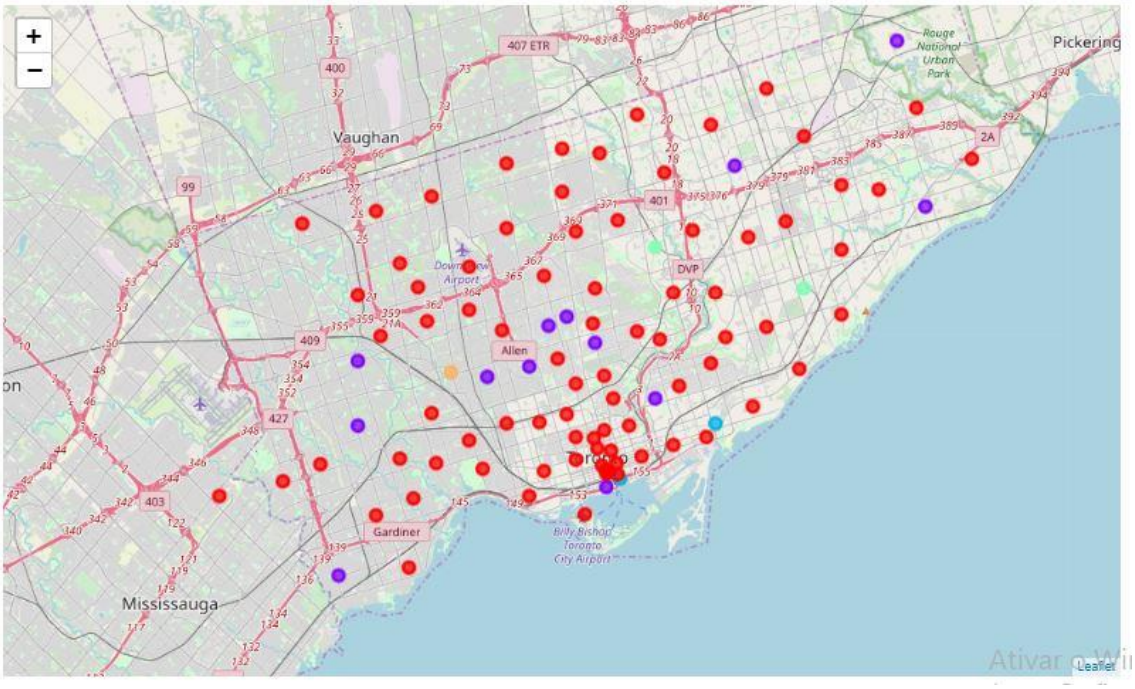
This information can be helpful to tourists since they can focus on what they are trying to experience most during their travelling (food, culture, sports, etc.).

These techniques also provide the clustering visualization of the city of Toronto.

Dataframe of Common Venues in toronto

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Scarborough	Rouge,Malvern	43.806686	-79.194353	0	Fast Food Restaurant	Print Shop	Deli / Bodega	Dessert Shop	Dim Sum Restaurant
1	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	0	Bar	Construction & Landscaping	Women's Store	Diner	Discount Store
2	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	1	Electronics Store	Spa	Pizza Place	Intersection	Tech Startup
3	Scarborough	Woburn	43.770992	-79.216917	0	Coffee Shop	Korean Restaurant	Convenience Store	Dumpling Restaurant	Diner
4	Scarborough	Cedarbrae	43.773136	-79.239476	0	Athletics & Sports	Lounge	Hakka Restaurant	Bakery	Caribbean Restaurant
5	Scarborough	Scarborough Village	43.744734	-79.239476	0	Playground	Women's Store	Donut Shop	Dessert Shop	Dim Sum Restaurant
6	Scarborough	East Birchmount Park,Ionview,Kennedy Park	43.727929	-79.262029	3	Coffee Shop	Playground	Discount Store	Department Store	Women's Store
7	Scarborough	Clairlea,Golden Mile,Oakridge	43.711112	-79.284577	0	Bakery	Bus Line	Soccer Field	Park	Fast Food Restaurant
8	Scarborough	Cliffcrest,Cliffside,Scarborough Village West	43.716316	-79.239476	0	Motel	American Restaurant	Women's Store	Dessert Shop	Dim Sum Restaurant
9	Scarborough	Birch Cliff,Cliffside West	43.692657	-79.264848	0	College Stadium	General Entertainment	Skating Rink	Café	Comic Shop

Clustering venues in Toronto city



Discussion

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining.

The most common algorithm uses an iterative refinement technique.

k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

This results in a partitioning of the data space into Voronoi cells.

Conclusion

Using Foursquare datasets this project successfully created a way to help tourists and business travelers with all the location information handy.

With it travelers can understand a city based in clusters and familiarity places, to avoid problems such as large displacements and getting lost in an unknown city.

For instance, this project showed us that Information such as nearby places, hotels, restaurants and other venues are easily accessible using data science methodology and integrations like de Foursquare API.