Where to Go?

Capstone Project - The Battle of Neighborhoods

Introduction - business problem

- When it comes to exploring a new town, some travelers like to wing it. After all, there's a lot to be said for a day without plans and the freedom to explore.
- But even if you're a Type B traveler, you'll appreciate facing that freedom armed with a little knowledge of how to get around—and maybe a dinner reservation to look forward to. Regardless of your personality type, we believe a little prep work goes a long way.

Introduction – targeted field

- This project aims to help any traveler to find the most interesting places in the city which planning to visit. Furthermore, any company provide tours or any travelling service could use this project.
- For instance, a traveler interested in chines food and cinema so he will get recommendation for neighborhoods contains those.

Data acquisition and cleaning

- Toronto city will be used as a traveler's destination. To achieve that, we will use Foursquare API, which is Location-based services. It seems that everywhere you turn, and other tools to "check in" to their favorite locations, share tips with other users about favorite drinks or dishes, and share all of this activity on other services like Twitter and Facebook.
- The API itself is a RESTful set of addresses to which you can send requests.
- Based on data from foursquare, we will create clusters to recommend suitable neighborhoods for the user.

Data acquisition and cleaning

- Scrapping this webpage :https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario.
- Using beautifulsoup4 library to scrap the previous data.
- Initialize a Dataframe which contains scrapped postal code.
- Ignore not valid and empty neighborhood.
- Get Geolocation data from the file: http://cocl.us/Geospatial_data

Data acquisition and cleaning

- Using Foursquare API to Explore the neighborhoods in Toronto
- After building a dataframe containing neighborhoods and their Venue Category, we build a dummy data using (one hot encoding approach)
- Using K-means Clustering to cluster the neighborhoods in Toronto
- After that, we get several cluster each of them represent a type of area may a person interest in, such as an area contains Asian food or luxury brands. To label each neighborhood a simple study should be done for some neighborhood and its Venue (it's out of scope of this study). So we will only rank a customer by cluster number without label.
- A costumer should provide some favorite venues to match with the most suitable cluster to him.

Data acquisition and cleaning - scraping

	Postcode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M6A	North York	Lawrence Heights, Lawrence Manor
4	M7A	Queen's Park	Queen's Park
5	M9A	Downtown Toronto	Queen's Park
6	M1B	Scarborough	Rouge, Malvern

Data acquisition and cleaning - Geographical coordinates

	postal	lat	lon
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Data acquisition and cleaning - Geographical coordinates

	Postcode	Borough	Neighbourhood	Latitud	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494

Data acquisition and cleaning - Get Venue

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653191	-79.357947	Gym / Fitness Center
3	Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Harbourfront	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot

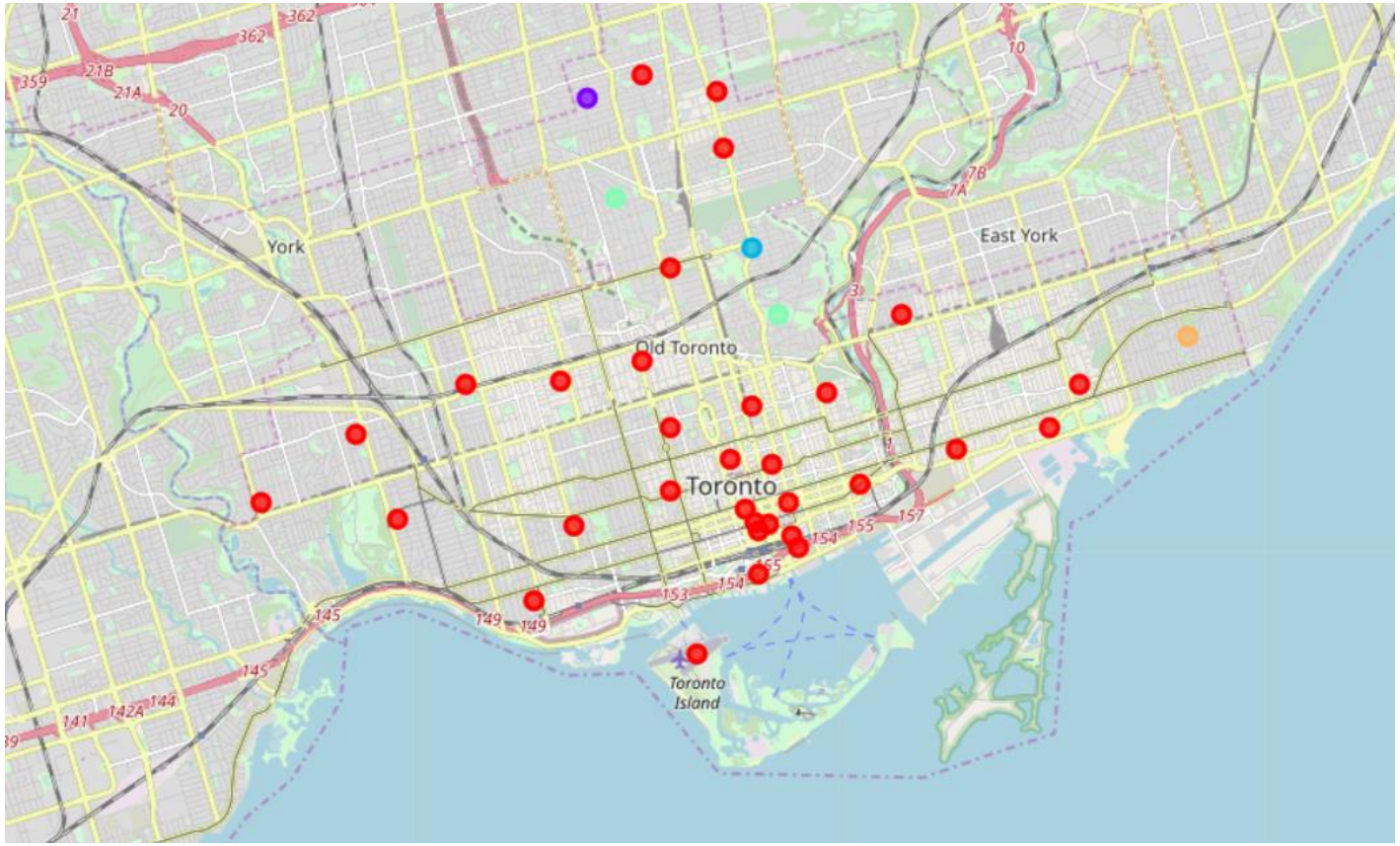
Data acquisition and cleaning - hot encoding

	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theme Restaurant	Thrift / Vintage Store	Toy / Game Store	Trail	Train Station
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

Data acquisition and cleaning – most common venue

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	Thai Restaurant	Burger Joint	Bar
1	Berczy Park	Coffee Shop	Café	Cheese Shop	Bakery	Cocktail Bar	Steakhouse
2	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Café	Coffee Shop	Sandwich Place	Bar	Italian Restaurant
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Yoga Studio	Auto Workshop	Comic Shop	Pizza Place	Recording Studio
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Service	Airport Terminal	Harbor / Marina	Bar	Plane	Rental Car Location

Data acquisition and cleaning – Clustered Data



Result and Conclusion

- As shown in Jupyter Notebook, each neighborhood label with dummy data about its venues, so neighborhood in the same cluster may include different venues but it has the acceptable similarity to each other's. For the traveler who provide his interests to recommend neighborhood with desirable venues to him, we try to recommend the most suitable neighborhood after matching traveler selection with the nearest one to him, so the traveler may get a neighborhoods with venue whom doesn't like.

Thanks !

Bassam