Where to go?

1. Introduction

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.

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.

2. Data acquisition and cleaning

For the sake of simplicity, 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 Collection and Processing:

- 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
- 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.

3. 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.