Report

Exploring Toronto Neighborhoods and Identifying Suitable Location for Opening a New Chinease Restaurant

1. ## Information Problem

   A need of location for opening a Chinease restaurant in Toronto

2. ## Problem Description

   Toronto is the capital city of the Canadian province of Ontario. It has a recorded population of over 2.5 millions in 2016. It is one of the most multicultural and multiracial cities in the world. In 2016, 51.5% of the residents of the city proper belonged to a visible minority group. According to the ethnic breakdown of 2016 census Chinese ethnic group is on top (12.5% of total population) among the most prevalent ethnic origins in the City of Toronto.

   Keep in mind the population of *Chinese ethnic group*, we aim at identifying suitable location for *opening a new Chinese restaurant*. For this purpose, we use the spatial and non-spatial data of Toronto neighborhoods and analyse the dataset. The data analysis includes *exploring the population distributions in top visible minority groups in neighborhoods, exploring relationship between venues in the neighborhoods, identifying the relationship between neighborhoods and existing Chinese restaurants, identifying relationship between Chinese population and Chinese restaurants, and clustering analysis* for identifying suitable location for opening new Chinese restaurant.

   The findings give an overview for investors interesting in opening new Chinese restaurant by taking into account Chinese population and number of existing restaurants in the vicinity. The analysis also provides information to customers about neighborhoods with lots of option for Chinese restaurants.

3. ## Data Sources, Pre-processing and Data Cleaning

   I have used the Wikipedia data "List of Postal code of Canada: M" (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) to get all the information about the neighborhoods including their postal code, borough & the name of all the neighborhoods present in Toronto.
   The population information is extracted from Wikipedia page https://en.wikipedia.org/wiki/Demographics_of_Toronto. We used demographic data to get population information about Chinese ethnic group in Toronto neighborhoods. The population information together with number of existing restaurants allow us to explore the relationship between population and existing restaurants.
   Spatial information such as latitude and longitude of Toronto neighborhoods is extracted from "https://cocl.us/Geospatial_data"
   Information about various venues in Toronto and their spatial information are extracted using Foursquare's API. Detail on API can be found on https://developer.foursquare.com/docs. We are interested in existing Chinese restaurants. Therefore, we extract the number of existing

restaurants in each neighborhoods. This allows us to explore the relationship between neighborhoods and Chinese Restaurant.

The pre-processing and data cleaning involve scraping Toronto neighborhood data, extracting latitude and longitude of neighborhoods, scraping venue items, extracting Chinese population in neighborhoods and number of existing Chinese restaurants. All the extracted data will be convert in to data frames for further analysis. Each pre-processing and data cleaning steps are discussed, and documented as markdowns together with code in the jupyter notebook.

4. Results

This section concludes the finding of the overall data analysis. In this project, we aim at identifying suitable neighborhood location in Toronto to establish a new Chinese restaurant. For spatial and non-spatial neighborhood data we used the web resources like Wikipedia and Geospatial data. For the extracting and scraping Wikipedia data we used Foursquare API and different python packages such as Wikipedia and beautifulsoup4.We used Toronto neighborhood data and demographic data such as Chinese population in each neighborhood and percentage of existing Chinese restaurants in those neighborhoods. The k-mean clustering technique is used to find the suitable neighborhood for establishing new Chinese restaurant.For the analysis we considered boroughs: Central Toronto, Downtown Toronto, East Toronto, East York, Etobicoke, Mississauga, North York, Scarborough, and York with 103 neighborhoods.Using Foursquare API we have extracted top 274 unique venues in Toronto neighborhoods. We are interested in Chinese restaurant in the neighborhoods. Therefore, we group neighborhoods and percentage of Chinese restaurants.In those 10 boroughs we identified that the neighborhoods of Downtown Tronto, Etobicoke, and North York boroughs have high amount of Chinese restaurants with the help of Violin plots between Number of Chinese restaurants in Borough of Toronto.The relationship between neighborhoods and Chinese restaurants indicates that Scarborough-Agincourt has highest population and Scarborough-Guildwood has lowest population.With the help of k-means clustering and plot visualization looks like Downtown Toronto, Central Toronto, East Toronto, East York, Mississauga, Scarborough, York have least number of Chinese restaurants.Etobicoke and North York are densely populated with Chinese restaurants.

After predictive analysis we came to conclusion that Central Toronto, East Toronto, East York, Mississauga, and Scarborough are ideal locations for opening new Chinese restaurant. However, the population analysis further clarify that Scarborough is most populated borough. Therefor, it has higher number of customers possibility. Thus we can conclude that Scarborough is most suitable location for opening a new Chinese restaurant.

5. Conclusion

This section concludes the predictive analysis for finding suitable location to open new Chinese restaurant. We have use different python packages to extract and scrap the useful spatial and non-spatial data from Wikipedia. We have used Foursquare API to explore the venues in neighborhoods of Toronto. For the visual analysis, we have used seaborn & matplotlib, libraries and neighborhoods and clusters of locations are rendered using Folium leaflet library. Finally, we have used K-means clustering to find clusters of locations suitable for opening new Chinese restaurant.