18-Jun-20

BEST SPOTS TO OPEN CHINESE RESTAURANT IN TORONTO

1. INTRODUCTION

Toronto is a big and diverse city with people coming from different parts of the world, considering these people from different ethnicities prefer to live in certain areas of the city i.e. areas with mostly Chinese people, another areas with mostly Indians, etc. based on this we can utilize this to recommend where to open a Chinese Restaurant.

A. Business Problem

Finding a method to utilize data that can be used to cluster the city based on ethnicities then recommending best spots to open a Chinese Restaurant within the selected cluster based on the density of already running Chinese restaurants in nearby areas.

B. Target Audience

Chinese people who live in Toronto in areas that has less density of Chinese restaurants.

2. DATA SOURCES

There are three main data sources used in this analysis:

1. Online CSV file that contains data about Toronto neighborhood areas and its demographics and population
2. Geocoder library that provide coordinates as latitude and longitude for given location which is used to determine locations of different neighborhood areas in the city in order to visualize it on map
3. Foursquare API to extract nearby venues, in our example Chinese Restaurants given certain location

3. METHODOLOGY

Our methodology to find the best spots for Chinese restaurant can be defined as follows:

1. Data Preparation

   This stage consists of loading the csv file from the web, loading its contents to Pandas DataFrames, then cleaning and arranging this DataFrame to be ready for visualization and further analysis.

Three DataFrames are constructed on this stage:

- df_neig: That contains areas in Toronto, each area with its latitude and longitude
- df_demo: which contains population of different ethnicities per area
- df_demo_sorted: which is a sorted version of df_demo by top ethnicities

Areas are then visualized on a map, based on each area's latitude and longitude

2. Area Clustering

   DataFrame df_demo is then clustered using K-means clustering algorithm using optimum K as found by elbow method, results are then visualized on the map

3. Exploring Neighborhoods
   In this stage, we'll explore ethnicity groups per each cluster found by the algorithm to determine clusters with most Chinese population

4. Finding the Best Spot
   In this stage we will use FoursquareAPI to determine locations of Chinese restaurants in clusters with most Chinese population, cluster 1 & 3. then we will select areas within these clusters with the lowest number of Chinese restaurants as a candidate spots for our new restaurants

4. RESULTS

Our methodology found top five spots to open Chinese restaurant and visualized the result on the map

5. DISCUSSION

KMeans algorithm provided useful as unsupervised machine learning algorithm that we used to cluster data based on ethnicity population in each city. However, data was clustered based on top ethnicity population per cluster, it’s difficult to be designed to isolate minority distribution across the city, different methodology and algorithms might be utilized for to solve this problem.

Also, foursquare API free edition is very limited in results and number of connections to retrieve data, other options that you may need further exploration is openstreetmaps.org

6. CONCULUSION

Overall KMeans was satisfactory for our problem and successfully clustered neighborhood areas in Toronto to different clusters based on population ethnicity. We then determined areas with less number of Chinese restaurants and assumed that these areas are considered hot spots to open new restaurant in.

Full data and analysis, further discussion and results are in below github link:

https://github.com/moevski/Coursera_Capstone/blob/master/Capstone_project.ipynb