COURSERA CAPSTONE PROJECT

May 29, 2020

Rongrong Wu
IBM Data Science Professional Certificate

Contents

0.1 Introduction/Business Problem

I live in Chicago, and the best thing I love about Chicago is the diverse culture, and tons of great restaurants it has. The goal of the project is to find the ideal place to open a Chinese Restaurant in Chicago. The project consists of two parts. Part I is the segmentation and clustering of Chicago Restaurants, including types of Italian, French, Japanese and Chinese restaurants. Part II is about choosing the ideal location to open the new Chinese restaurant. Of course, as with any business decision, opening a new restaurant requires serious consideration and is a lot more complicated than it seems. Therefore this project aims to help alleviate the troublesome process of finding the ideal location.

0.1.1 Business Problem

The objective of this capstone project is to find the ideal place to open a Chinese Restaurant in Chicago. Using data science methodology and machine learning techniques such as k-means clustering etc. This project aims to provide in-depth research and thorough analysis.

0.1.2 Target Audience of this project

This project is particularly useful to property develoopers and investors looking to open or invest in new restaurants in Chicago. This project is timely as the city is currently flourishing with good restaurants. Research shows that people are more and more willing to spend money on fine dining restaurants.

0.2 Data

To solve the problem, we will need the following data:

1. List of Community Areas in Chicago. This defines the scope of this project which is confined to the city of Chicago, major city in the Middle West of the United States of America.

2. Latitude and longitude coordinates of those community areas. This is required in order to plot the map and also to get the venue data.

3. Venue data, particularly data related to different types of restaurants. We will use this data to perform clustering on the Community areas.

Sources of data and methods to extract them:

1. The wikipedia page (https://en.wikipedia.org/wiki/Community_areas_in_Chicago) contains a list of community areas in Chicago, with a total of 77 community areas. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates using Python Geocoder package which will give us the latitude and longitude of the community areas.

2. We will then use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the restaurant category in order to help us solve the business problem put forward. The project will make use of many data science skills, from web scraping(Wikipedia), working with API(Foursquare), data cleaning, data wrangling, to machine learning and data visualization(Folium). In the Methodology section, we will discuss the analysis in detail.

0.3 Methodology

0.4 Results

0.5 Discussion

0.6 Conclusion

References