Distribution of Chinese Immigrants and Non-immigrants

Ke Guan

June 28, 2019

1.Introduction

1.1 Background

Back to the 1840s, after the discovery of gold in California, countless white gold rushers swarmed, and this craze quickly spread to China. At that time, China was amid two opium wars and the Taiping Rebellion. In order to survive, many Guangdong and Fujian people went to the United States for gold. According to US population record data, there were only three Chinese residents in California in 1848. This number reached 300 in the second year, and in 1852 it surged to 30,000. From 1849 to 1882, a total of 300,000 Chinese entered the United States. By 1882, Congress passed the Chinese Exclusion Act, virtually banning all Chinese immigration into the United State until December 1943, when China became America's ally in World War II. In 1965, all restrictions were lifted and Chinese Start to arrive American again in huge numbers. On June 18, 2012, the United States officially apologized to the practice of discrimination against the Chinese in the form of legislation. Today, the largest Chinatown is in New York City, and the other Chinatowns have emerged in other cities, where tons of Chinese immigrants and non-immigrants live and work. Hence, when one would like to open a store which will have numerous Chinese customers, he must know the best location for opening the store. While the best location would be within a Chinatown or near to where Chinese congregate.

1.2 Problem

Data that might contribute to determining whether the area full of Chinese include the existing of Chinese Supermarket, the number of Chinese Restaurants, and some other stores with Chinese elements. This project aims to show the regions full of Chinese in Manhattan.

1.3 Interest

People who want to open a store or have services for Chinese customers would be interested in the clustering of Chinese. Others who would be tourists who are curious about Chinese culture and custom.

2.Data acquisition and cleaning

2.1 Data sources

Data is requested from the Foursquare, then leverage the location data to compare neighborhoods with Chinese elements. And locate the regions onto the map.

2.2 Data Cleaning

Data downloaded from the Foursquare were combined into one table. There were 5 boroughs and 306 neighbors in New York. As I mainly focused on Manhattan, I retrieved all neighbors for it. Totally, there were 40 Neighbors involved, including Marble Hill, Chinatown, Washington Heights, etc. Around 500 miles from the center location of each neighbor, I got 30 items containing keyword

"Chinese". Reframed these items into a dataset with selected features. Some items have names containing "Chinese" label, but the defined categories are not, so for restaurants, no matter what kind of restaurant. I redefined its category as "Restaurant", as well as the educational institutions. I discovered their distributions of different categories.