Opening Upscale Italian Restaurants in Los Angeles

Introduction/Business Problem

Company A is planning to open several upscale Italian restaurants in Los Angeles. They are now on the stage of deciding the areas in which they are going to open the restaurants. In so doing, the company creates a set of criteria for the areas, sorted by their importance:

First, the areas should have a high median income. The company is targeting high-income customers, and thus prefer to open its restaurants near the areas in which its target customers live. In addition, locating restaurants in such areas will also help the branding of the restaurants.

Second, the areas should not have many Italian restaurants around it. The company wants to catch business opportunities in areas where there are not many competitors. The company hopes to be a pioneer in the areas and thus can attract more customers.

Third, the areas should have a fairly large population size, since the main target customers will be the people who live in the areas. By opening restaurants in areas with a sizeable population, the company hopes to have more customers.

To solve this area selection problem, I decided to help company A by clustering Los Angeles areas based on the criteria mentioned above using the K-means method.

Data

The following is the descriptions of data used in this project and also the sources:

1. Los Angeles median income by ZIP code (2017 data):
   I classify the areas in Los Angeles by their ZIP codes. I then retrieve the data on the median income of these areas (in US$) from the website of Los Angeles Almanac (http://www.laalmanac.com/employment/em12c.php)

2. Los Angeles population by ZIP code:
   I measure the size of the areas by using the number of people living in the area. I retrieve the data from ZIPatlas.com
   (http://zipatlas.com/us/ca/los-angeles/zip-code-comparison/population-density.htm)

3. Italian restaurant by ZIP code:
   I will use the data on the number of Italian restaurants within the radius of 1 kilometer from the centre of the ZIP code area/neighborhood to measure the competitiveness in the areas. I retrieve the data from Foursquare location data.