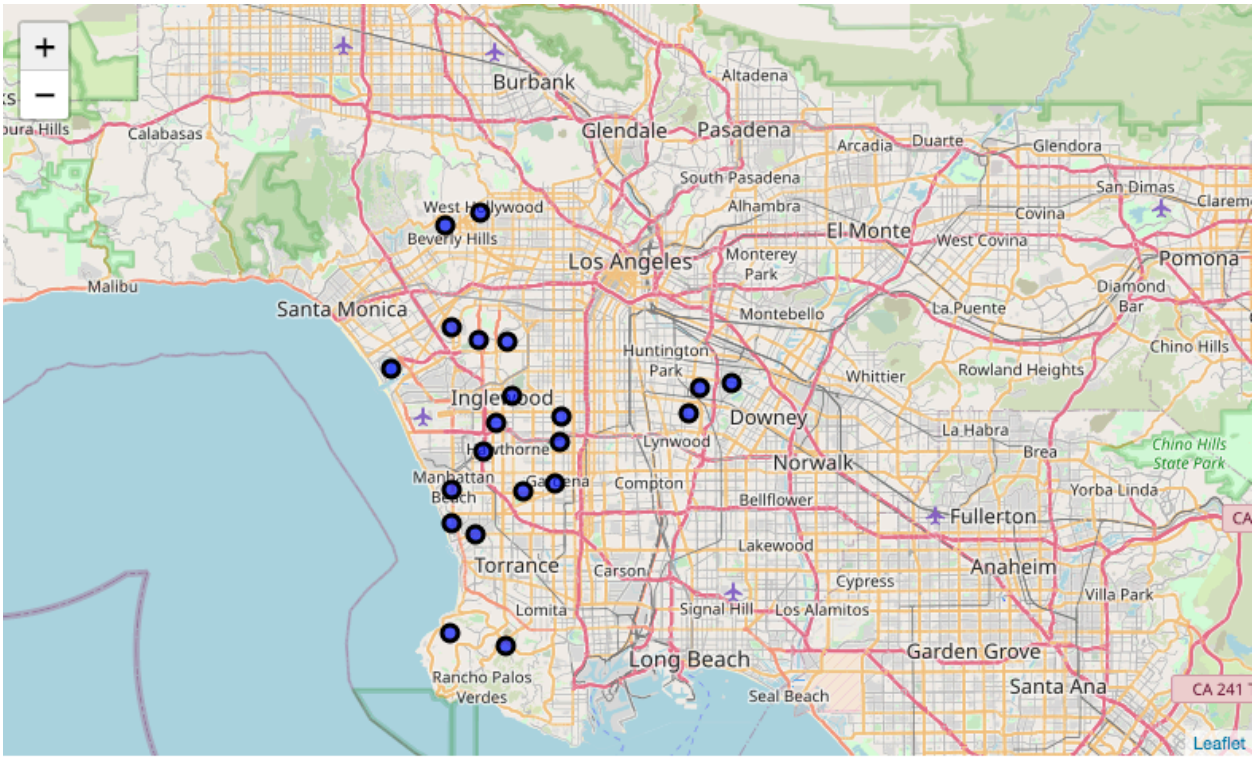
Opening a new Japanese restaurant in Los Angeles county

Sunsoo Chung

6/2/2020

1. Introduction/Business problem
    a. My boss is planning to open a Japanese restaurant. He thinks that there are so many Japanese restaurants in Los Angeles and competitive. So he wants his restaurant to be unique New-Japanese style concept and trying to reach non-Japanese. Since He wants to have one somewhere in Los Angeles county, and have certain distance from existing Japanese restaurants.
2. Data sources and solution
    a. The problem will be addressed using two datasets: The first dataset is a collection of latitude/longitude coordinates for each cities in California (obtained from: https://simplemaps.com/data/us-cities.
        i. It requires us to download the data rather than scraping. So it's imported to the IBM Watson studio.
    b. The second dataset is the Foursquare data. This will be used to obtain information about each location, including restaurants in each area. The data will be used to calculate:
        i. the Euclidean distance from each suburb to the nearest Japanese restaurant,
        ii. the number of restaurants in each district,
        iii. the diversity of restaurants in each area (using the Simpson-Gini coefficient).
    c. Data cleaning
        i. The data above is consisted with 1504 rows.
        ii. To clean the data, I removed the all unnecessary zip codes, and combined with existing ones.
        iii. Narrowed it down to Los Angeles county which starts from 90001 to 90559, and drop the rest.
3. Data Science Methodology
    a. Foursquare API was imported and used to scrape the all Japanese restaurant in Los Angeles county, and got the latitude, longitudes, and zip code to create a map later. In total, there are 541 restaurants in the Los Angeles county.
    b. Calculate the distance to each Japanese restaurant, and also distance from the city using the KMeans Machine Learning algorithm.
    c. Calculate the total number of restaurants within a 3km radius.
        i. Note: there is some overlap; a venue may be counted more than once for adjacent suburbs
    d. Calculates the Gini-Simpson diversity index from a list of categories (cat_list) provided.https://en.wikipedia.org/wiki/Diversity_index#Gini%E2%80%93Simpson_index
        i. Higher values indicate greater diversity.
    e. Map the top 20 recommended location using folium.

4. Results



5. Observation and discussion
    a. There are many fusion restaurants in Beverly Hills and West Hollywood, but there is still opportunities for Japanese restaurant.
    b. Near the beach sounds good place, but it will be difficult to park.
    c. Downey area sounds good idea to open since rent would be affordable and spacious enough to park. Not competitive as well.
6. Conclusion
    a. Top 20 recommended places are plotted on the map.
    b. From the map, opening Japanese restaurant in East side sounds good idea.