BATTLE OF THE NEIGHBORHOODS

Coursera IBM Data Science Capstone Project

By Claire Li

Background & Business Problem

- Client is the owner of a fast-casual Chinese restaurant in Irvine, California and is hoping to expand to Los Angeles
- Client seeks a location that shares similar characteristics with its existing location
- Thus, the specific business problem for this preliminary analysis is:

Which neighborhoods in Los Angeles have a similar demographic profile and competitor landscape to Irvine, California, and are potential locations for the client's proposed new restaurant?

Features for Analysis

A similar demographic profile to the original Irvine location. Factors to be included in the analysis are: total population, age distribution, median income, and race/ethnicity.

Similar consumer preferences. This will be measured by the types of businesses and venues in each area.

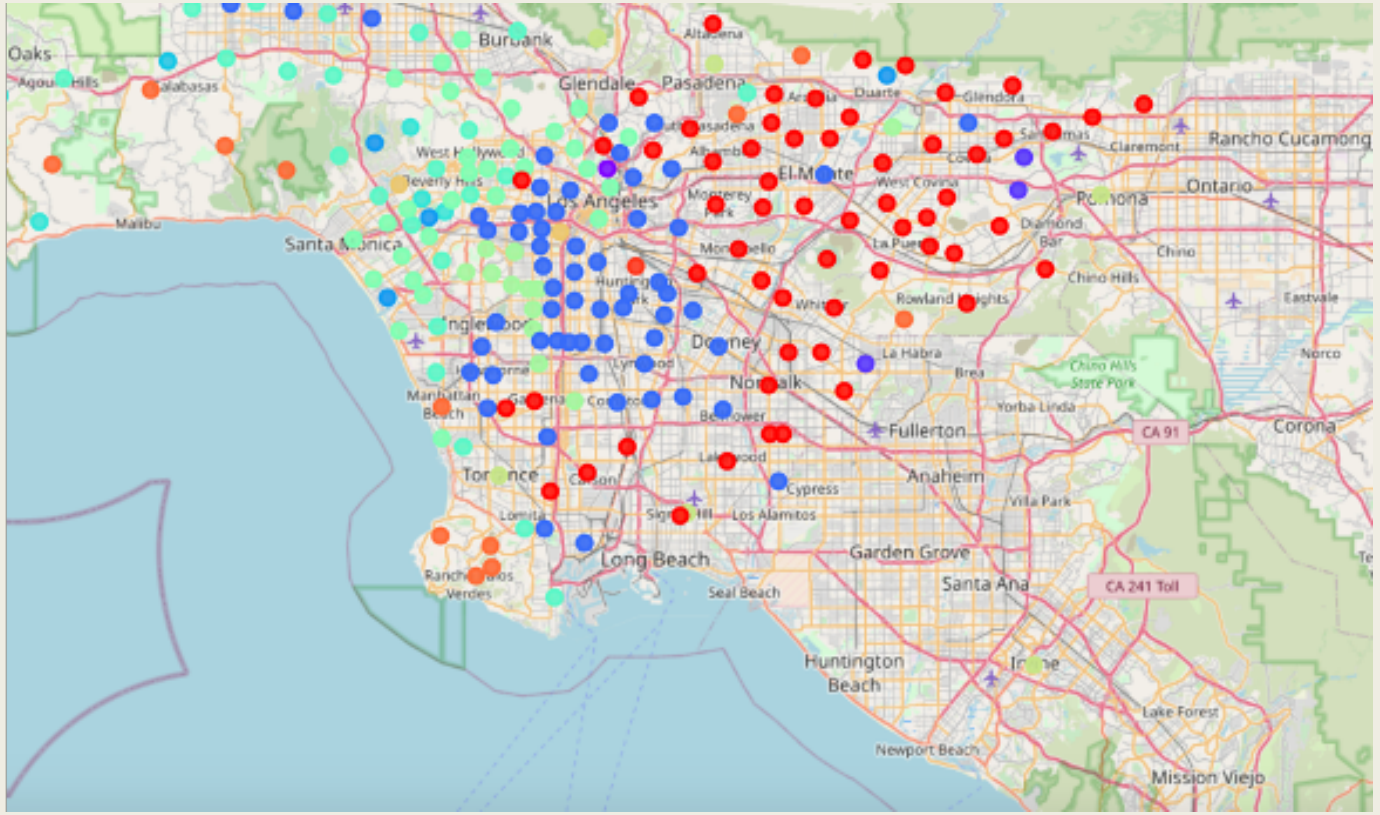
A balance between a proven consumer demand for Chinese food and avoiding oversaturated areas

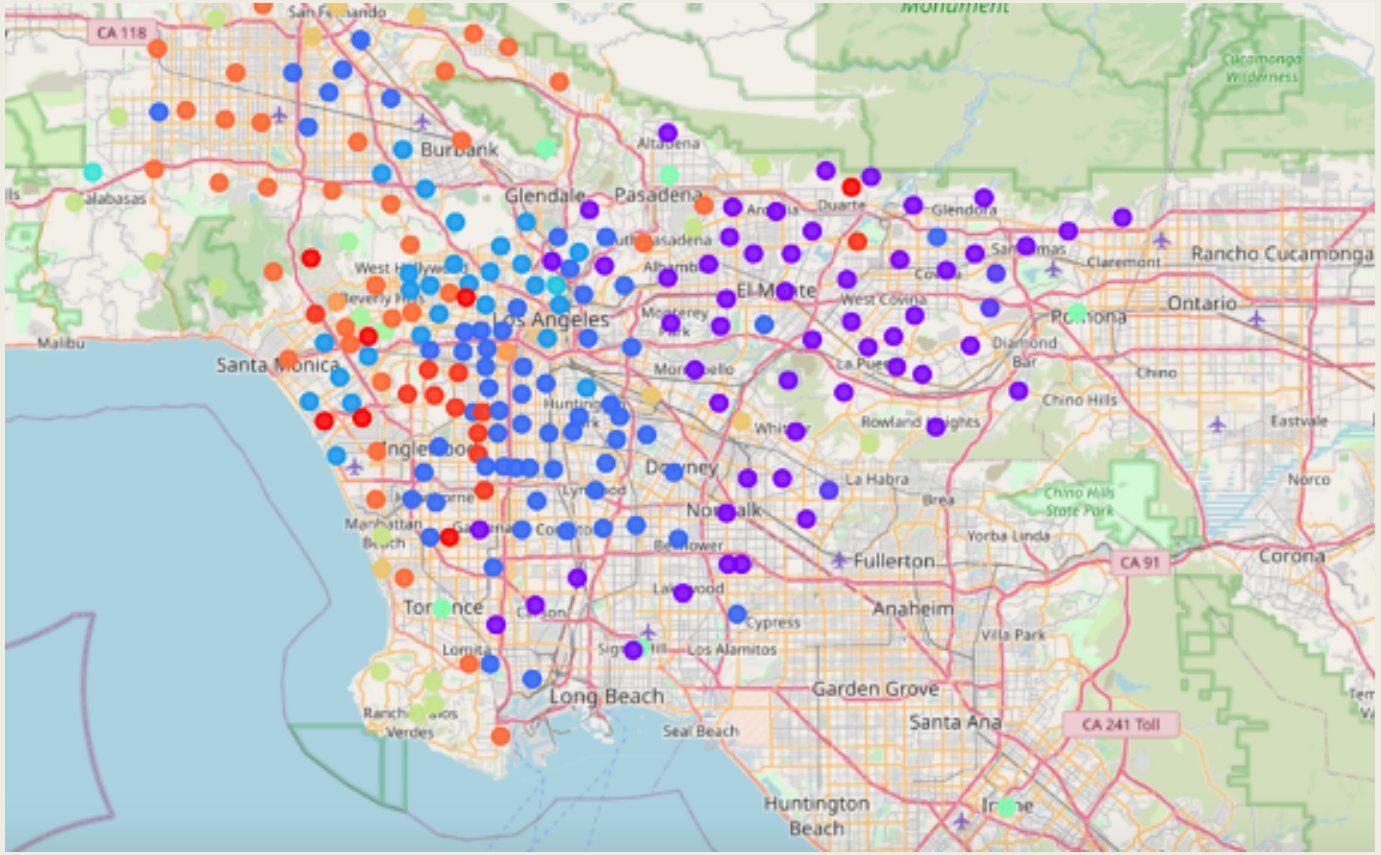
Data Sources

- List of Los Angeles neighborhoods *(Los Angeles Times)*
- Los Angeles demographic data *(US Census Bureau)*
- Orange County (Irvine) demographic data *(US Census Bureau)*
- Foursquare API *(Foursquare Labs Inc.)*

Demographic Data

- 272 Los Angeles neighborhoods in total
- Factors included in analysis: total population, age distribution, median income, and race/ethnicity
- After performing K-means clustering (k = 15), the neighborhoods grouped with Irvine were found to be: **Glendale, Lancaster, Long Beach, Palmdale, Pasadena, Pomona, Santa Clarita, and Torrance.**





Business Landscape

- Retrieved a list of venues and venue categories for each neighborhood (up to 100 venues within 3km of each neighborhood's geographic center)
- Overall: 22381 venues, with 432 unique categories
- Performing K-means clustering with the Foursquare API data itself was not informative. Therefore, we combined the Foursquare API data with the demographic data (k = 15).
- Once again found **Glendale, Lancaster, Long Beach, Palmdale, Pasadena, Pomona, Santa Clarita,** and **Torrance** to be in the same cluster as Irvine

Calculating Relative Frequency of Chinese Restaurants

- Determined relative frequency for list of potential neighborhoods by calculating: **(Number of Chinese Restaurants) / (Total Number of Venues Retrieved)**
- Highest frequency = Palmdale; Lowest frequency = Long Beach
  - To strike a balance between an area with a demonstrated consumer demand and one with an oversaturation of Chinese restaurants, we eliminate these two neighborhoods from our recommendation.

FINAL RECOMMENDATION

Based on this preliminary analysis, we recommend the client explore opening the second location in the following neighborhoods:

- Pasadena
- Torrance
- Glendale
- Lancaster
- Santa Clarita
- Pomona

Further Analysis and Future Steps

- Incorporate more demographic and/or geographic features
  - *(e.g., sex, median household size, proximity to population hubs).*
- Create a ranking system based on the features that the client most prioritizes
  - *(e.g., if the client wishes to shift their focus to college-age students or decides to cater towards office lunch crowds).*