

Italian Chef Jimmy De Luca Moves to San Francisco

Garrett George

July 1, 2020

2. Data Acquisition

The Foursquare API will be leveraged to help Jimmy answer who important questions:

- What neighborhood in San Francisco should he move to that is most similar to New York City's Soho neighborhood?
- What is the best spot for his next hit Italian restaurant?

To answer the first question, a dataset of all neighborhoods in San Francisco is needed, which can be found at *data.sfgov.org*. The geocoder library can then be used to get the exact latitude and longitude of each neighborhood. Once compiled, the Foursquare API will help perform a search of the most popular spots around each neighborhood. By also adding in the most popular places in New York City's Soho neighborhood to the dataset, a k-means clustering analysis can be performed to see which neighborhoods in San Francisco share similar characteristics to Soho.

The second question follows a similar methodology. However, determining the optimal location of the restaurant that will set Jimmy up for the greatest success has with it the need to answer the following question: what characteristics of an area will increase the chance of Jimmy's restaurant gaining popularity? After meditating on this question, Jimmy figured that an area with high restaurant density and closeness to his future San Francisco neighborhood would lead both to great success of the restaurant and personal satisfaction knowing he is a short commute away from the establishment. However, when comparing areas with a high density of restaurants, Jimmy wants to find the one that has the *fewest* Italian restaurants to avoid any direct competition with his business.

To determine which areas in San Francisco have the highest density of restaurants, the Foursquare API will be used once again by searching for and classifying the most popular spots in each neighborhood and detecting which ones yielding the highest concentration of restaurants. From this selected list, we can then determine how many Italian restaurants exist within the concentration. The final piece to the puzzle will consist of using a distance formula (using the latitude and longitude coordinates) to determine which neighborhoods are closest in proximity to Jimmy's newfound neighborhood. Here is where we can provide one or more recommendations to Jimmy on optimal locations to open his next restaurant.