

# The Battle of Neighborhoods

Finding the best place to live

## Introduction

### Background

As a recent graduate, I have obtained a job offer in Irvine, California, and am not familiar with the area. With two dogs and a husband, I am interested in finding a place to live near the new job where we will feel at home and have plenty to do.

### Problem

Since we are unfamiliar with the area, we don't know what to expect when moving in. We will not be able to visit the region before picking a place to live, and will need to find another way to decide where to live.

### Approach

We have a relocation stipend and rent prices in the region are comparable, so the key decision drivers will be nearby venues that we like – namely the number of Coffee Shops, Performing Arts Venues, and Parks. We can use Foursquare API to find this information. We will also cluster the neighborhoods and see if there are similar neighborhoods to where we are moving from – Palms.

## Data Acquisition and Cleaning

### Data Sources

The main data source for this analysis is the Foursquare API, which will provide lists of nearby venues to various neighborhoods in Orange County, CA.

The list of neighborhoods is found at [orangecountyguide.com](http://orangecountyguide.com), and the geolocator Nominatum is used to find the coordinates.

### Data Cleaning

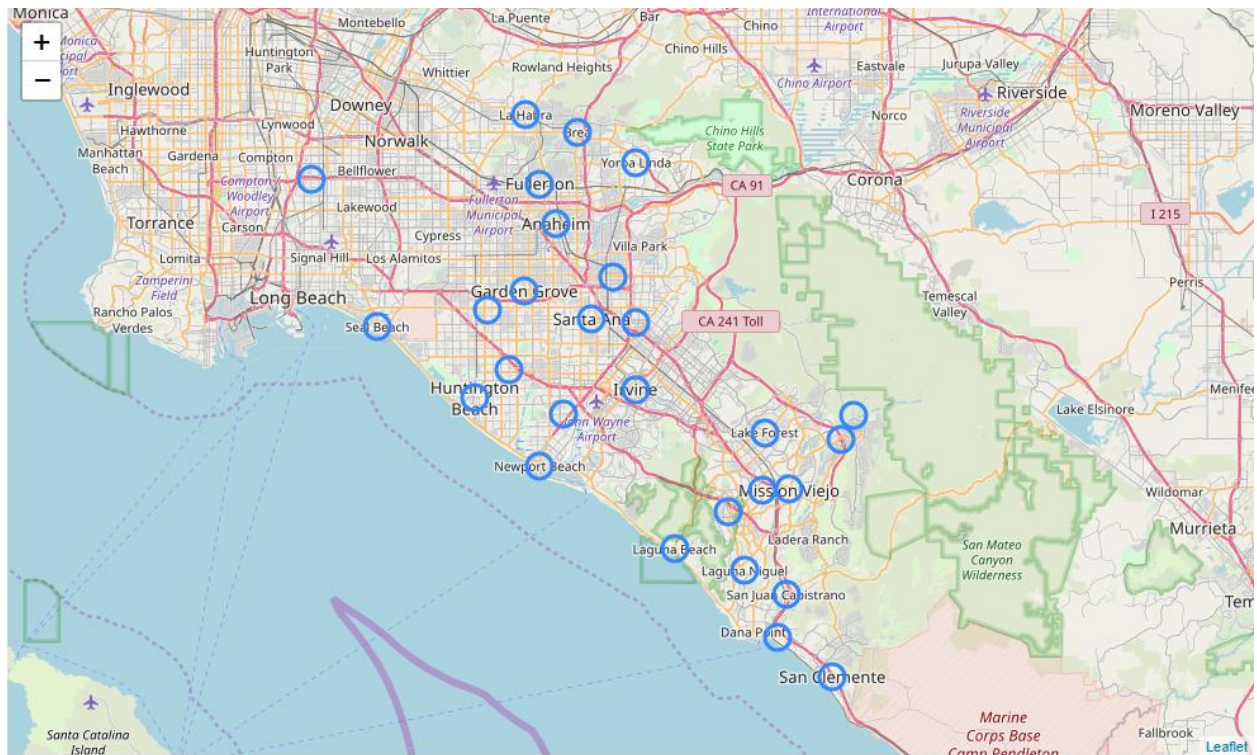
As the list of neighborhoods is short, I manually created a dataframe with the city names. The first pass of the geocoder led to several incorrect coordinates, but this was resolved by specifying each location was in Orange County, California.

The foursquare data came through in a json file, and functions were used to digest this into a more usable format.

## Methodology

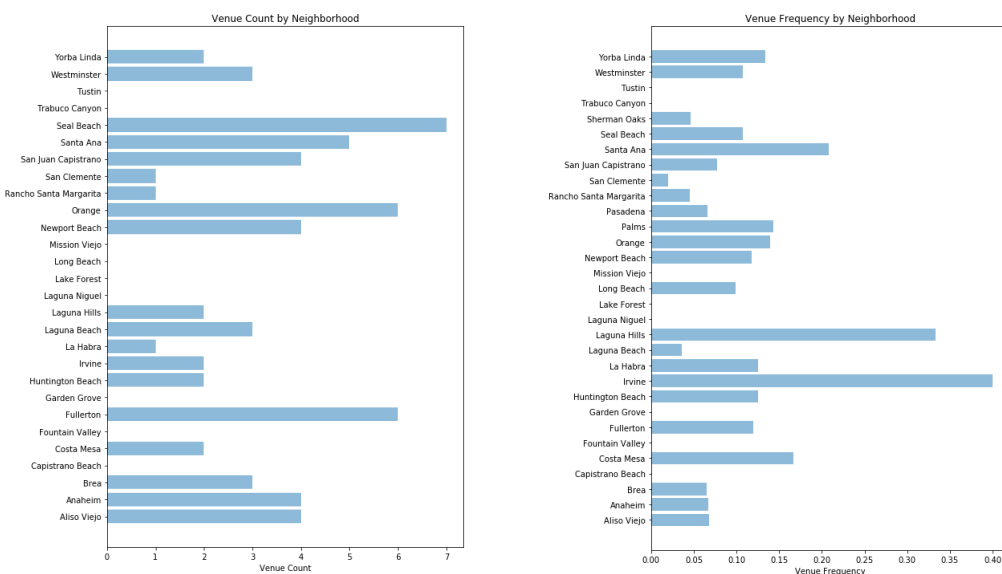
### Analysis

First, the various neighborhoods were plotted on a Folium map to give an idea of the geographic diversity and distribution:



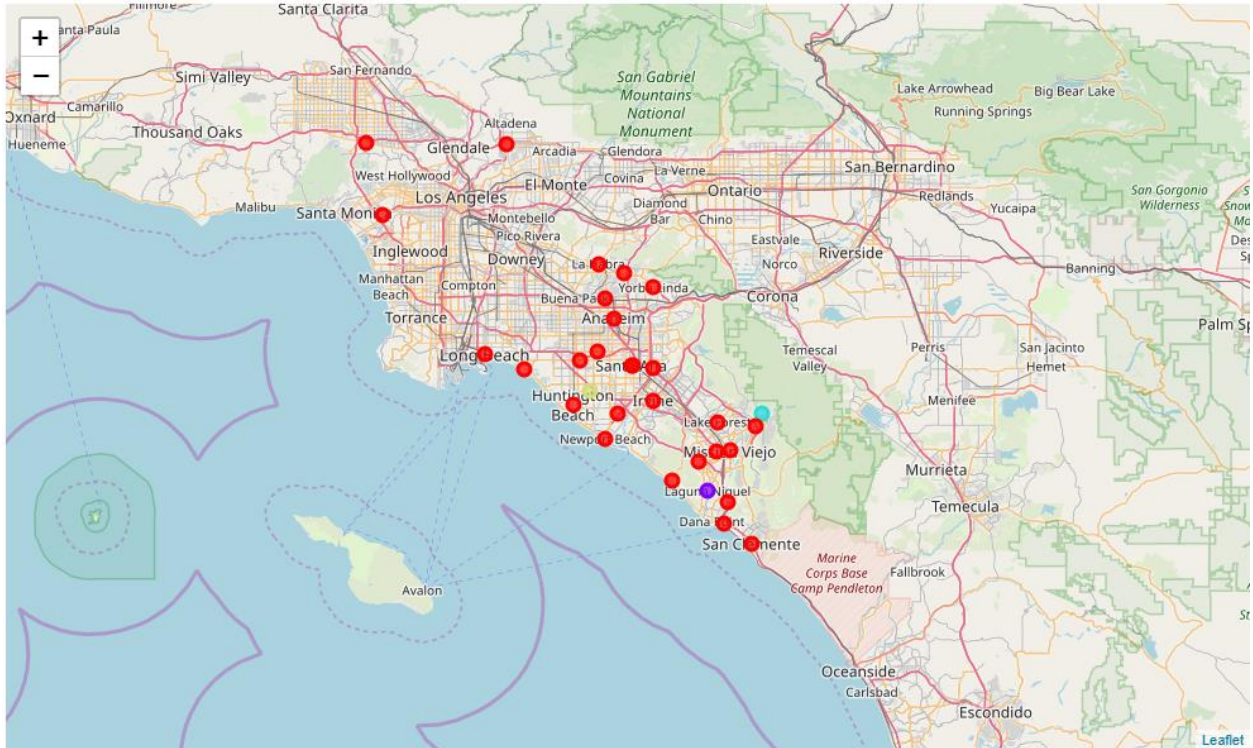
The neighborhoods are spread out but relatively evenly distributed around a center near Irvine, which is the target location where I will be working.

Next, I connected to the Foursquare API to pull lists of venues for each neighborhood. I then filtered this to the venues I was interested in, which were the categories Coffee Shop, Café, Theater, Concert Hall, and Park. I explored and clustered based on both venue type frequency and venue type count.

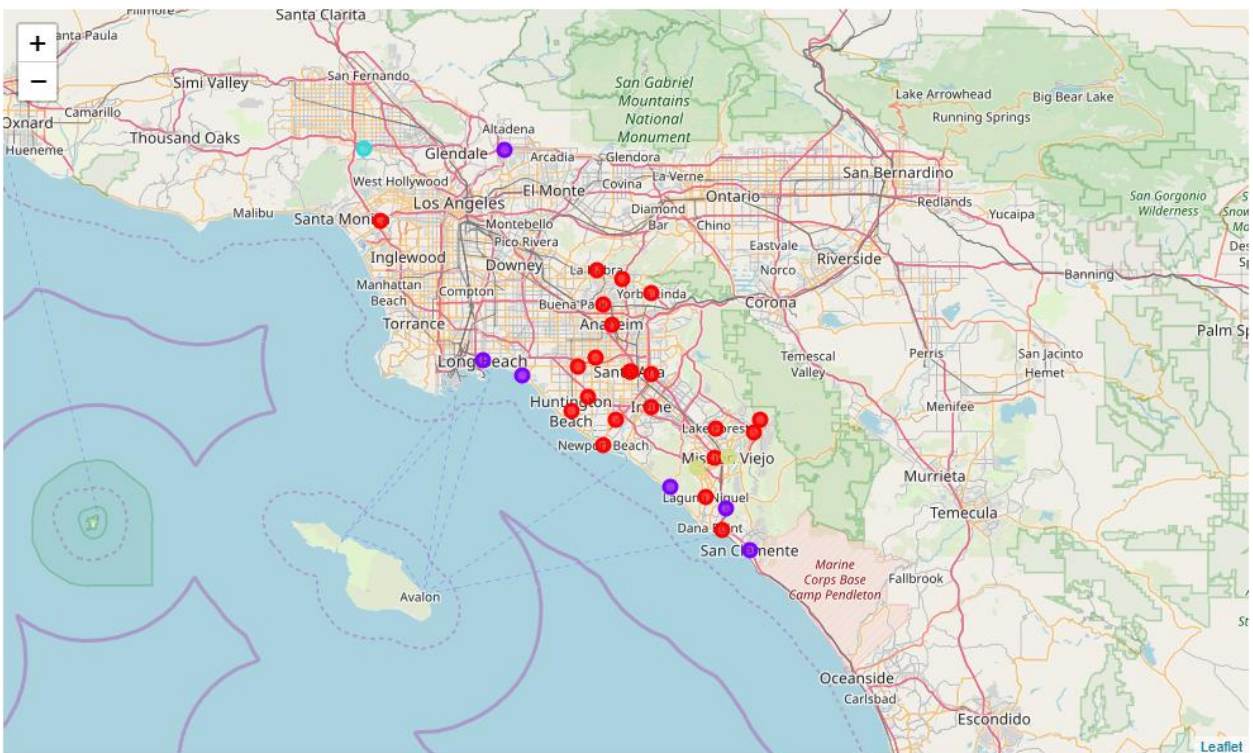


## Results

Using venue frequency, clustering did not create a meaningful way to separate the Cities:



However, when I used venue counts instead of frequency (which actually is a better proxy for density, as I'm looking at a fixed radius for each city center), I found interesting results:





While none of the prospective neighborhoods were like Sherman Oaks, there was a smattering of neighborhoods like Palms (red = cluster 0) and Pasadena (purple = cluster 1). As I had previously lived in Palms, and my husband had lived in Pasadena, this let us make a meaningful decision to narrow down our choices.

## Discussion

From here, we need to move away from the quantitative and apply a more qualitative approach – as this comes to personal choice and preference. We added three locations outside our target area for comparison – Sherman Oaks (where I grew up), Palms (where my husband and I currently live), and Pasadena (where my husband grew up), so we would have known areas to compare in our clusters. We can then decide if we want to live in a place similar to one where we've lived before, and rank the clustered locations by the frequency of our desired features

If we wanted a place like Pasadena (Cluster 1):

- Long Beach (7)
- Seal Beach (7)
- San Juan Capistrano (4)
- Laguna Beach (3)
- San Clemente (1)

If we wanted a place like Palms (Cluster 0), most of Orange County is an option, and we can just rank based on the # of venues on our preferred venue type list. The top 5 would be:

- Fullerton (6)
- Orange (6)
- Santa Ana (5)
- Newport Beach (4)
- Anaheim (4)

From here, we will need to work with real estate agents to determine where we can find an affordable place to live.

## Conclusion

In this report, I have described how I used Foursquare API data to inform my decision on where to live. The output was a ranked list of cities, which I can then take to a real estate agent to prioritize where to look. This is a real issue for me, and I am more confident in my decision to move for a job than I was before completing this exercise.