

Neighborhood Analysis by Venue Ratings

How to find where to get the biggest bang for your buck

Dwayne Thaele

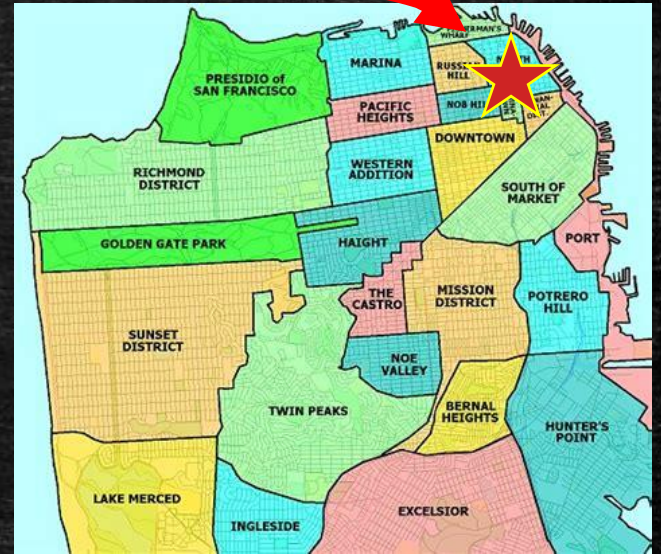
January 20, 2021



You arrive at an unfamiliar city, hungry and uncertain where to go for a highly rated restaurant...

There are many ways to determine your ideal restaurant.

But *Neighborhood Analysis* takes a new approach:
Use user ratings to determine which neighborhoods have the best restaurants



Foursquare – The primary source of locations and venue data


www.foursquare.com

FOURSQUARE

ProductsResourcesPartnersCompanyLog InContact Us

If it tells you where, it's probably built on Foursquare

We believe in the power of location. What people experience in the real world and the places they go are powerful reflections of who they are and what they care about. We help leading global companies tap into this intelligence to create better customer experiences and smarter business outcomes, all based on the world's leading platform for understanding people, places, and the interactions between them.



- ✓ Foursquare collects and stores volumes of data about locations, venues, user ratings, tips, and much more...
- ✓ Best of all, Foursquare offers an API to developers to access this data...
- ✓ The Neighborhood Analysis program utilizes the Foursquare API to acquire data about San Francisco neighborhoods, locations, venues, and ratings to determine the ideal locations to select restaurants

How *Neighborhood Analysis by Venue Ratings* Works

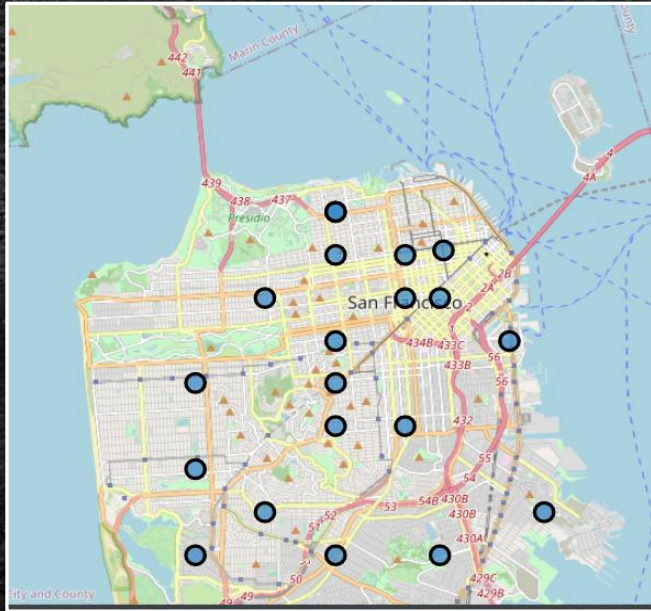
Here is a brief summary of the steps in the program:

- 1) Define neighborhoods in San Francisco
 - a) Zip Codes are used to define neighborhoods
- 2) Obtain geographical coordinates of the neighborhoods
- 3) Create a map of San Francisco
- 4) Plot the neighborhoods coordinates on the map of San Francisco
- 5) Access the Foursquare API and retrieve essential data: venues, coordinates, and ratings
- 6) Color code the venue ratings and plot the venues on the map using their coordinates
- 7) Tabulate the venues and ratings to determine the Average Venue Rating per neighborhood
- 8) **Determine Result of Neighborhood Clustering:**
 - a) **The neighborhood with the highest Average Venue Rating is considered the ideal neighborhood**
- 9) Use the neighborhood and venue information to calculate 5 K-means clusters
- 10) Calculate the centroid for each K-means cluster and coordinates
 - a) Plot the centroids on the San Francisco Map
- 11) Analyze the clusters to determine:
 - a) Cluster with the highest Average Venue Rating
 - b) Cluster with the highest Venue Count
- 12) **Determine Results of K-means clustering:**
 - a) **The cluster with the highest Average Venue Rating is considered to be ideal**
 - b) **The cluster with the highest Venue Count is considered to be secondary**

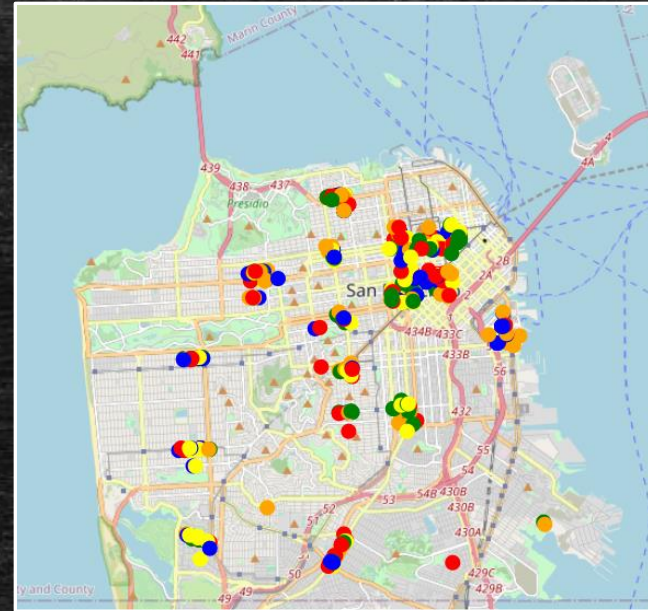
Looking at the Results*:

Good News! The Neighborhood Analysis program displays substantial data to inspect

Plot of neighborhoods
on San Francisco Map



Plot Venue Locations
on San Francisco Map



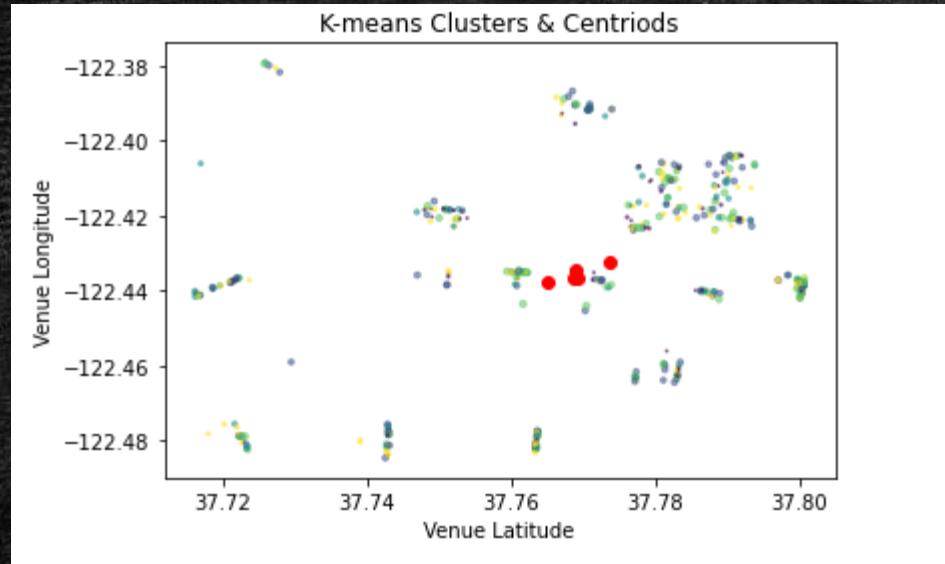
<u>Venue Color</u>	<u>Venue Rating</u>
Green	9 – 10
Blue	7 – 8
Yellow	5 – 6
Orange	3 – 4
Red	1 - 2

*Note – Results will vary over time

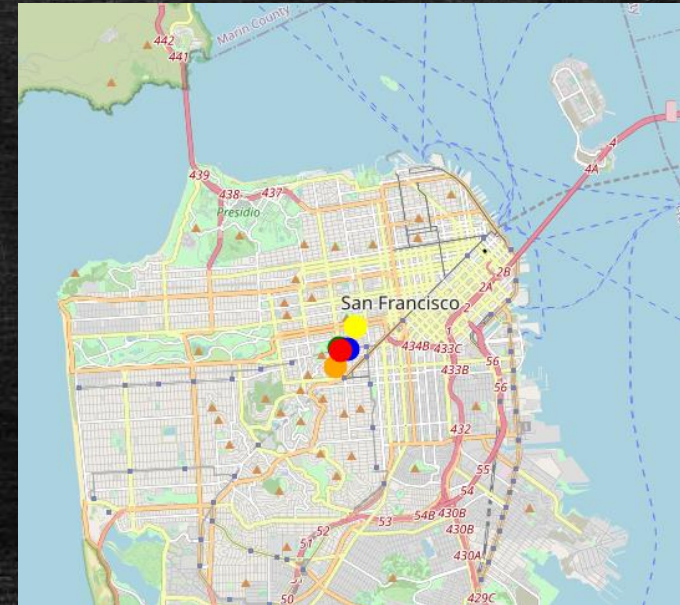
More Results:

Good News! The Neighborhood Analysis program displays substantial data to inspect

Graph of Clusters and Centroids (in red)



Plot of Centroids on San Francisco Map



Output of Neighborhood Clustering

****NEIGHBORHOOD CLUSTERING RESULTS****
The highest Average Venue Rating is 6
The venues with the highest Average Rating are listed below:

	Neighborhood	Venue Count	Avg Rating
16	Bayview-Hunters Point	2	6
4	Chinatown	30	6
11	Haight-Ashbury	12	6
18	Twin Peaks-Glen Park	7	6
14	Sunset	30	6
5	Polk/Russian Hill (Nob Hill)	30	6
15	Marina	30	6

Neighborhoods with the highest Average Venue Rating are listed: Note Venue Counts are also included

	Zip Code	Neighborhood	Latitude	Longitude	Venue Count	Avg Rating	Rating Sum
9	94115	Western Addition/Japantown	37.790	-122.440	15	7	105
11	94117	Haight-Ashbury	37.770	-122.440	12	6	82
15	94123	Marina	37.800	-122.440	30	6	192
16	94124	Bayview-Hunters Point	37.730	-122.380	4	6	24
4	94108	Chinatown	37.791	-122.409	30	6	181
1	94102	Hayes Valley/Tenderloin/North of Market	37.780	-122.420	30	5	155
7	94112	Ingelside-Excelsior/Crocker-Amazon	37.720	-122.440	30	5	160
18	94131	Twin Peaks-Glen Park	37.750	-122.440	7	5	41
14	94122	Sunset	37.760	-122.480	30	5	174

Detailed listed of all Neighborhoods

Output of K-Means Clustering

Output of K-Means Clustering: Part 1 Highest Average Venue Rating

****K-means RESULTS Part 1****

K-Means Cluster 3 has the Highest Average Venue Rating
K-means Clusters Ranked by Highest Average Venue Rating

	Cluster	AVG Rating	Venue Count	Avg Rating	
	3	3	9.55	103.0	9.55
	1	1	7.44	78.0	7.44
	2	2	5.43	89.0	5.43
	0	0	3.45	83.0	3.45
	4	4	1.55	92.0	1.55

Output of K-Means Clustering: Part 2 Highest Customer Count

****K-means RESULTS Part 2****

K-Means Cluster 3 has the Highest Venue Count
K-means Clusters Ranked by Highest Venue Count

	Cluster	AVG Rating	Venue Count	Avg Rating	
	3	3	9.55	103.0	9.55
	4	4	1.55	92.0	1.55
	2	2	5.43	89.0	5.43
	0	0	3.45	83.0	3.45
	1	1	7.44	78.0	7.44

Summary of Venues in cluster 1
Total Venues in Cluster 1 = 88
Average Venue Rating = 9.6

	Cluster Label	Neighborhood	Venue	Venue Category	Venue Rating
389	1	Bayview-Hunters Point	Day Darnet Catering	Food	10
154	1	Inner Mission/Bernal Heights	Al's Place	New American Restaurant	10
246	1	Western Addition/Japantown	My Ivy	Thai Restaurant	10
79	1	Potrero Hill	Poke Delish	Food Truck	10
113	1	Chinatown	Le Colonial	Vietnamese Restaurant	10
121	1	Polk/Russian Hill (Nob Hill)	Bob's Donuts	Donut Shop	10
127	1	Polk/Russian Hill (Nob Hill)	Cordon Bleu	Vietnamese Restaurant	10
94	1	Chinatown	Big 4 Restaurant	American Restaurant	10

Detailed information on all clusters and venues included
(only Cluster 1 shown)

Conclusion: Neighborhood vs K-Means Clustering

General observations:

- ❖ Neighborhood Clustering deliver narrower geographical coverage because the venues are confined to 1 neighborhood, but broader rating ranges
- ❖ K-means clustering provide broader geographical coverage, but weighs venue ratings higher than proximity

Conclusions:

- ❖ If the desired goal is to focus on a central location with acceptance of a mix of venue ratings, then Neighborhood Cluster should be used
- ❖ If the desired goal is to focus on highly rated venues with acceptance of greater distances, then K-Means clustering should be used

Thank You!

Helpful Information

- ❖ GitHub Repository: https://github.com/dthaele-coursera/Final_Project.git
- ❖ GitHub: dthaele-coursera
- ❖ Email: dwayne@comcast.net