

# COMPARING CHICAGO NEIGHBORHOODS TO SAN FRANCISCO NEIGHBORHOODS

David Drees

# Introduction

Chicago is the third most populated city in the United States, and is well known for having a diverse community with a variety of restaurant cuisines, social venues, and tourist attractions. It is also comprised of many unique neighborhoods whose residents are used to being around the venues in the neighborhood. Per U.S. Census data, San Francisco is the most common destination for residents that move out of Chicago (about 6.4% of all migrations out of Chicago are to San Francisco). Many people that move from Chicago to San Francisco hope to find a neighborhood to reside in with familiar surroundings.

# Introduction Cont.

To answer the question of "Which neighborhood should I choose when moving to San Francisco?" I will be exploring the types of venues in each neighborhood of San Francisco and comparing them to the types of venues in each neighborhood in Chicago. This will allow someone to make an informed decision of which new neighborhood to choose when they make the move from Chicago to San Francisco.

# Data

- In order to compare the neighborhoods of Chicago and San Francisco, I needed to gather data from multiple sources. First, I needed to find data on the neighborhoods themselves that would list the neighborhood names as well as their geographical coordinates to be able to explore the types of venues in nearby each one. I was able to download neighborhood data in the form of Geojson files from each city's website.
- I would need to call Foursquare's API for each neighborhood to obtain venue data which would be my last source of collected data.

# Obtaining Neighborhood Information

After loading the geojson data files into my code, I had to pull out the necessary information in order to make them usable. The files contained a lot of extra information, but I was only focused on the geographic coordinates and the name of the neighborhoods that corresponded with the coordinates. I extracted the necessary data by running a loop through each feature in the file to append neighborhood name and coordinate data to a new DataFrame.

The coordinates that a Geojson file provide are boundary coordinates, and there are often times upwards of 100 sets of coordinates just to create the entire boundary of a neighborhood. Knowing that later, when I call the Foursquare API for each neighborhood to obtain venue data, I will need to enter just one coordinate per neighborhood, I needed to find the center of the boundaries provided. To do this, I took an average of the minimum latitude & longitude and the maximum latitude & longitude in each set of boundary coordinates. The same process was completed for both Chicago and San Francisco data.

# Finding Neighborhood Venue Information

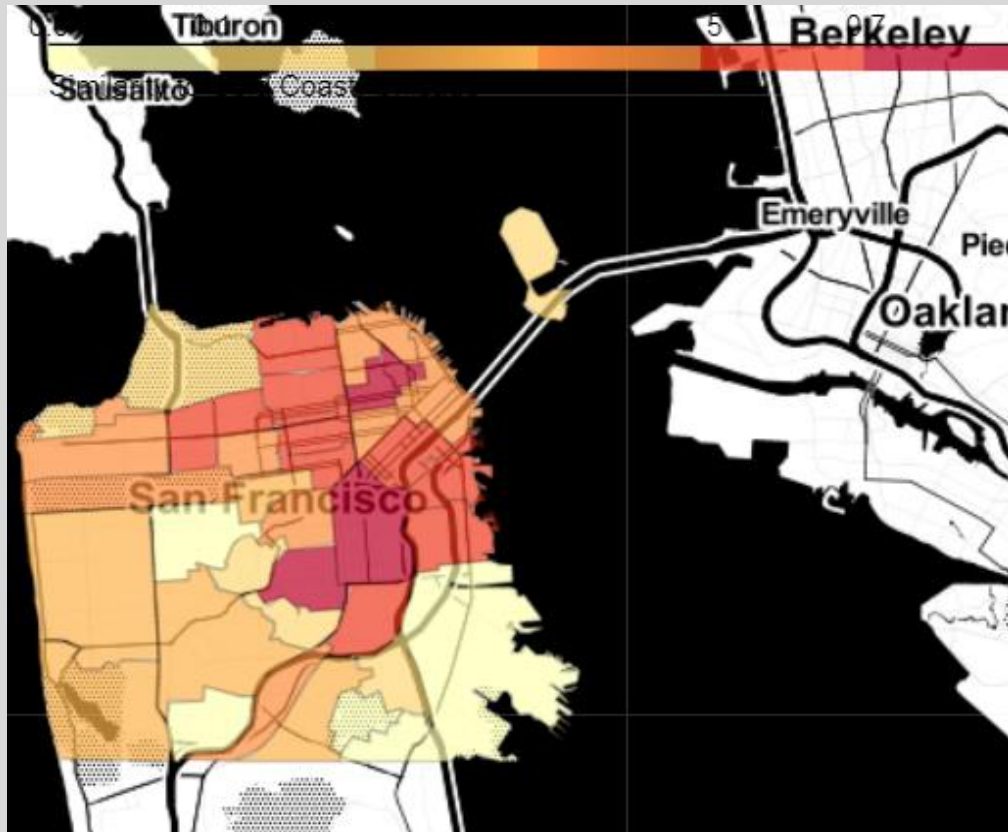
API calls via Foursquare were made in an iteration for every neighborhood in both cities, returning up to the top 100 venues within 500 meters of each neighborhood center. While not all neighborhoods had 100 venues within 500 meters in Foursquare's database, each had enough information to work with. Considering that I am comparing neighborhoods strictly based on the types of venues that they possess, I created a DataFrame for each city listing all of the categories of venues as columns and each neighborhood as a feature, with normalized frequency of the category in the corresponding neighborhood labelling each cell.

# “Scoring” Neighborhoods

Neighborhood	Albany Park	Andersonville	Archer Heights	Armour Square	Ashburn
Bayview Hunters Point	0.00000	0.04082	0.00000	0.00000	0.00000
Bernal Heights	0.47500	0.53316	0.25000	0.64537	0.07500
Castro/Upper Market	0.20000	0.62082	0.34000	0.42000	0.12000
Chinatown	0.40000	0.70408	0.24000	0.65630	0.18000
Excelsior	0.55882	0.51441	0.67647	0.47168	0.14706

- To create a similarity score for each neighborhood, I found the linear distance between each normalized category in different neighborhoods
- I totaled the sum of all distances between two neighborhood's categories
- I repeated the process, then subtracted the total score from 2 so the higher scores reflected the most similar neighborhoods

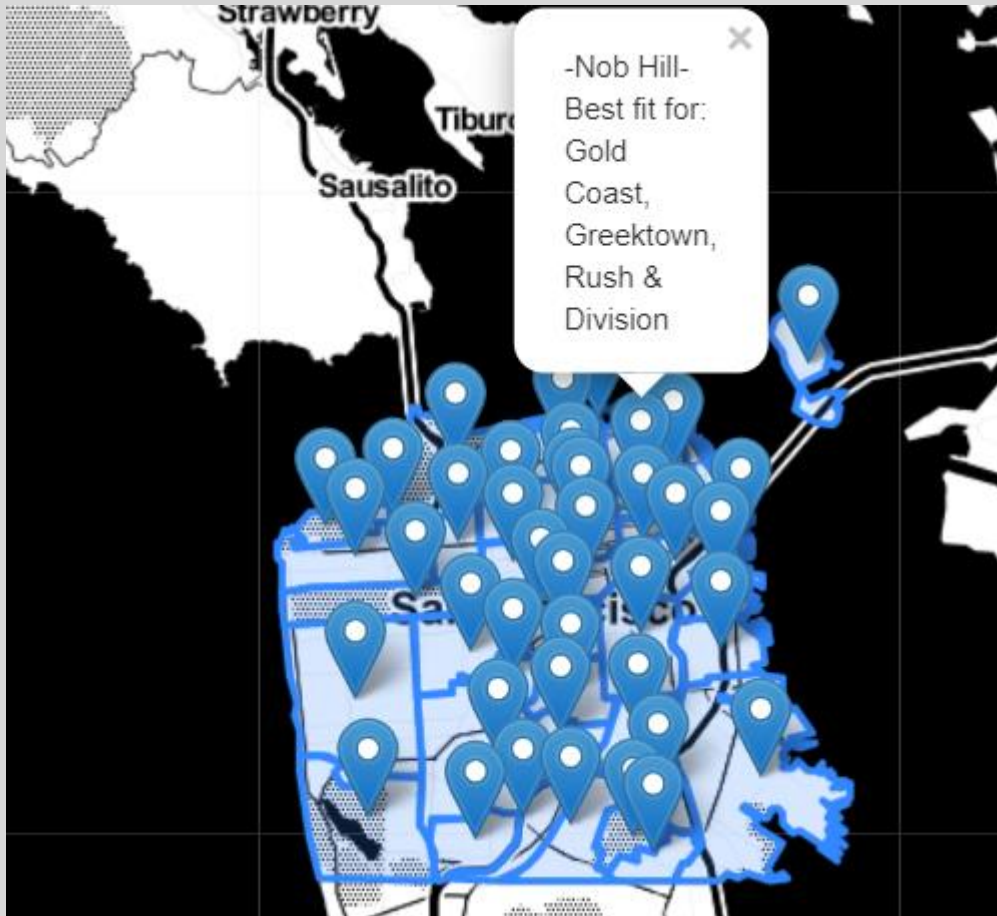
# Results



- I used the scores to create a choropleth map of the neighborhoods in San Francisco that change color based on how similar they are to a neighborhood in Chicago
- The example in this slide shows similarities of San Francisco neighborhoods to Chicago's Gold Coast neighborhood
- The top result was Nob Hill. Both Gold Coast and Nob Hill are known for their fine dining, luxury accommodations, and proximity to downtown



# Results



- To provide a more useful, interactive tool for those interested in similarity of other neighborhoods, I provided a map with an overlay of the boundaries of each San Francisco neighborhood as well as a clickable marker that would describe the name of the San Francisco neighborhood as well as which Chicago neighborhoods it would be the best fit for.

# Observations

- Upon completing this model, I was able to provide “best fit” neighborhoods for anyone in Chicago hoping to find somewhere with similar venues in San Francisco. Recommendations for anyone seeking this information could be found with a simple click in the visualizations in the notebook.

# Future Directions

- While I was able to provide neighborhood similarity strictly based on venue categories, there are certainly a myriad of other factors that play a role in similarity of neighborhoods. Neighborhoods could also be compared based on crime rate, population density, proximity to downtown, etc. that would all contribute to making a more informed decision on finding a similar neighborhood. In addition, Foursquares API only allows to search for venues within a given radius of a coordinate. While this might be an excusable workaround, it might not give the full scope of venue data in a neighborhood. If the comparison were to include all venues within the boundaries of a neighborhood, and possibly even adding weight to the popularity of each venue, it may be possible to achieve more accurate results.

# Conclusion

- In this study, I took venue data from all neighborhoods in both Chicago and San Francisco to answer the question, "Which neighborhood should I choose when moving to San Francisco?" I used the frequency of venue categories in each neighborhood to make a comparison with each other and create a score for each neighborhood. These scores will allow someone to make an informed decision when looking for a neighborhood that brings many of the same characteristics as one they may be used to in Chicago.