

# San Francisco Restaurant price clustering

## Introduction

Suppose you are looking to open a restaurant somewhere in San Francisco. Being one of the most expensive cities in the world, each neighborhood has its own micro economic climate, which can significantly affect restaurant pricing. A potential restaurant owner may be interested in which neighborhoods are better suited for business needs.

We will explore restaurant pricing data that can potentially serve a future owner on where he or she may want to open. If an owner is looking to open a nice, elegant restaurant, then perhaps he or she would look toward the more expensive areas of town in terms of restaurant pricing. This would allow an owner to maximize their profits and run a successful business. We will assist with this decision by clustering neighborhoods in San Francisco based on pricing.

## Data

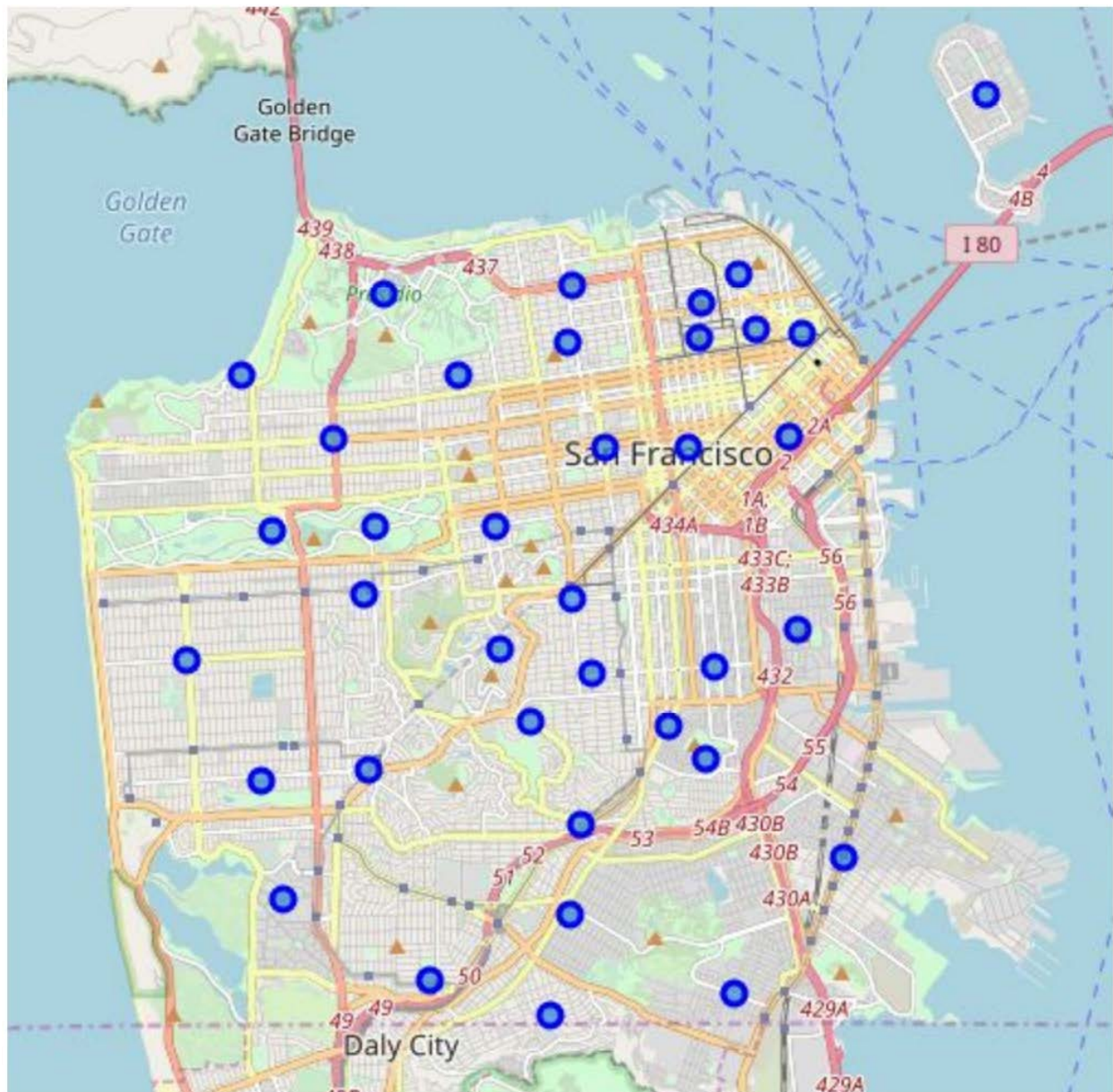
We will begin the data collection process by obtaining a list of all the SF neighborhoods as well as their geographical location. We will then apply Foursquare API, specifically the venue group and explore endpoint, to find all nearby restaurants to each neighborhood point considering their price rating: from \$ to \$\$\$\$ according to FourSquare's pricing bucket. After processing and analyzing the data we will provide a K-means clustering to cluster the neighborhoods based on pricing. This will provide for a visualization in which a future owner will be able to identify similarities in different areas of town and allow he or she to pinpoint locations in which a specific restaurant would be more successful.

## Methodology

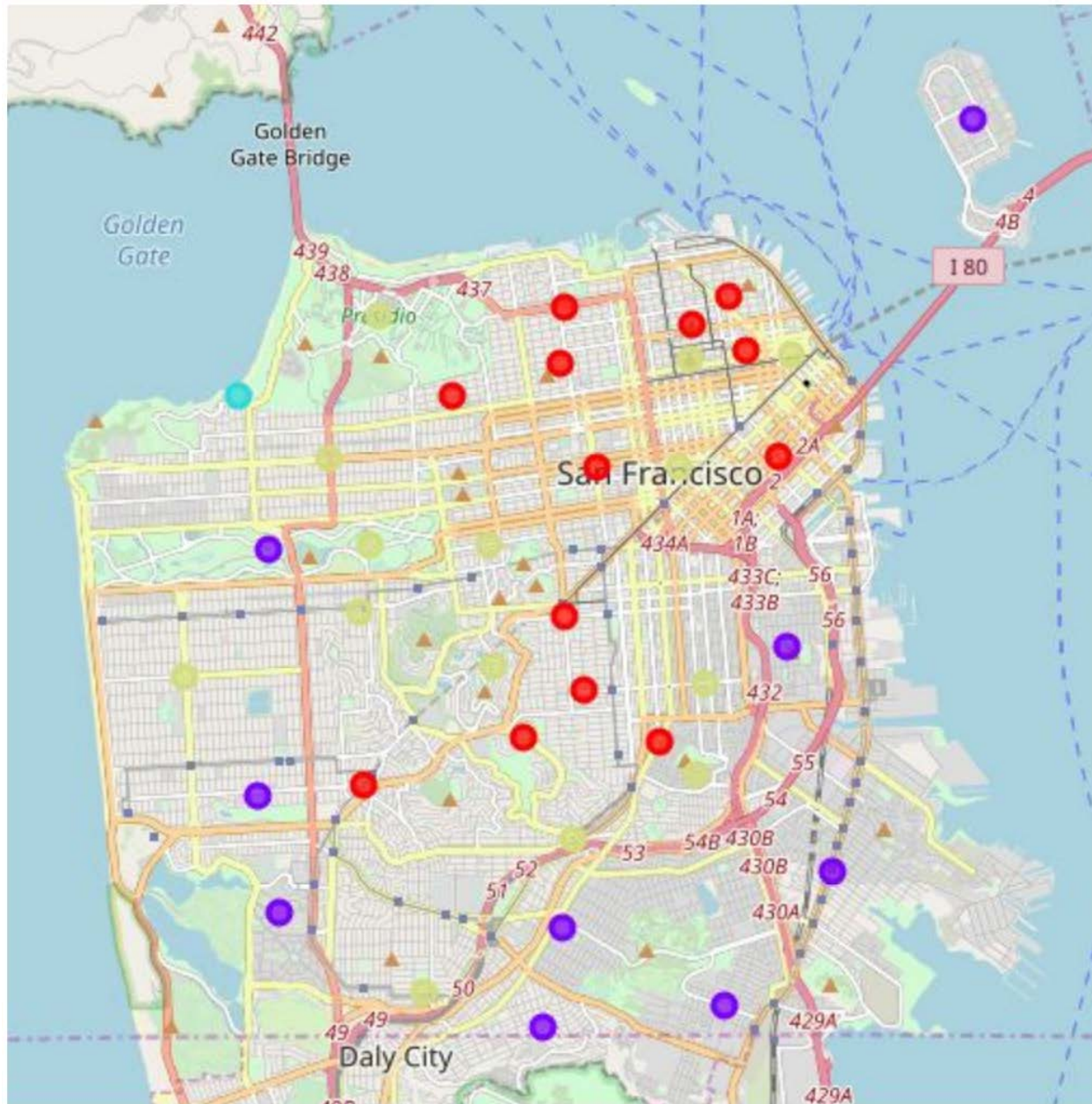
The first step of my project was mining the data containing the neighborhoods and coordinates of San Francisco. For this procedure, I used Nominatim to automate the price but at times had to resort to using Wikipedia to fill in missing data. I then used FourSquare's API to retrieve all the restaurants specified by price bucket within a 500m radius of the coordinates of each neighborhood. I then processed this FourSquare data and used clusters to organize the restaurants based on frequency of price bucket results. I then used Folium to display the clustering in a map setting.

## Results

The following map shows the result of identifying all the neighborhoods in SF with their corresponding coordinates:



Here was my result after clustering the neighborhoods based on restaurant price:

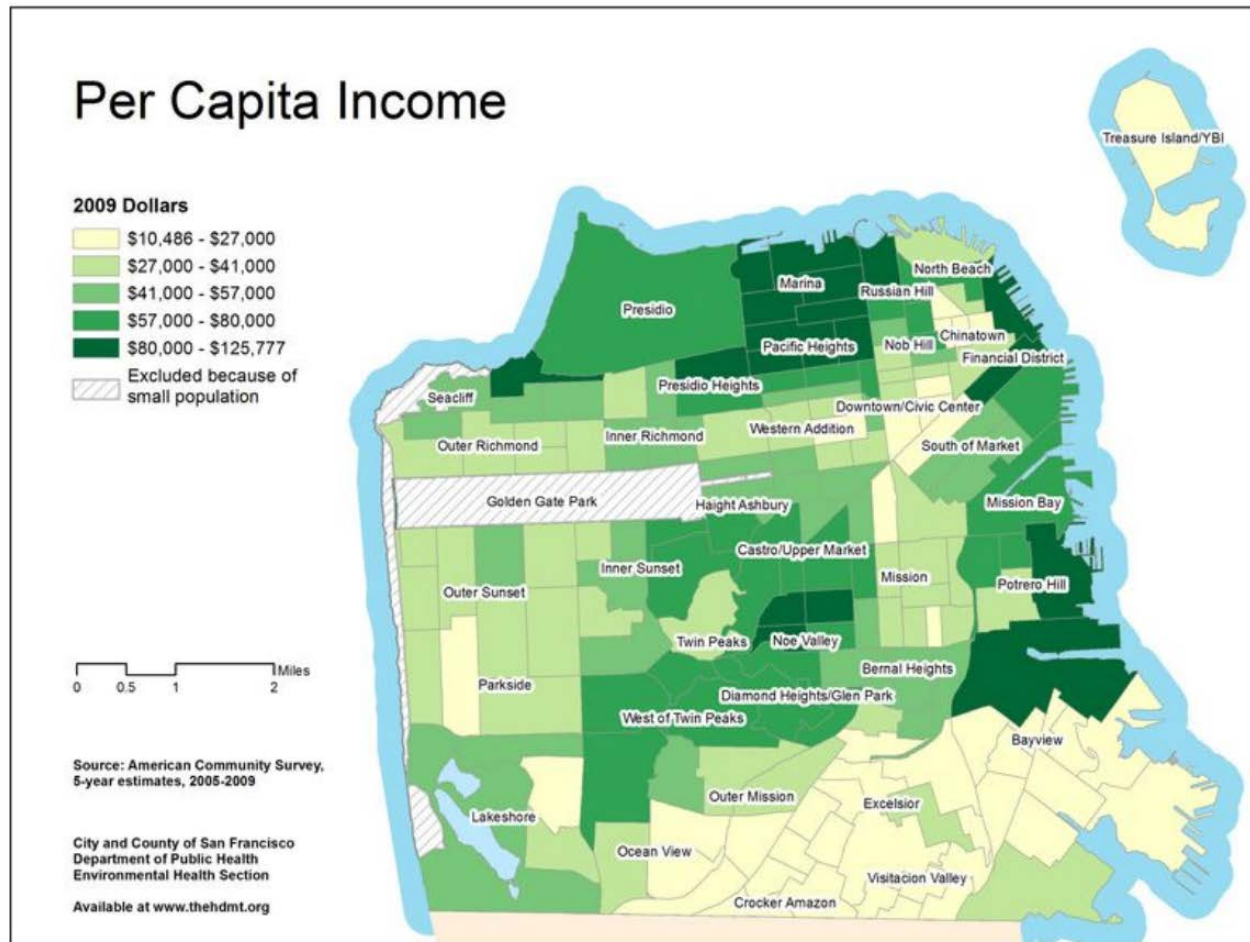


In this cluster analysis, I chose a total of 4 clusters. One cluster, at the top left in light blue, stood by itself as it had no nearby restaurants. The purple, green, and red clusters are somewhat geographically clustered even though I did not factor geographical coordinates into the k-means clustering.

## Discussion

One possible comparison is the restaurant price vs per capita income in SF





Based on the map above, one can conclude that the red cluster might better serve more expensive restaurants, the green cluster might work for medium priced restaurants, the purple might not be suitable for restaurants since it is on the outskirts of the city, and perhaps the blue cluster is not conclusive. It seems that both the green and red clusters are similar, but something more interesting could be the green cluster neighborhoods in the northeast: Chinatown, Nob Hill, and Downtown/Civic Center. Based on the income map, these areas are typically lower-income areas. Perhaps current restauranteurs in red clusters could look to expand to other red cluster areas, depending on business plans.

## Conclusion

For this project, we gathered geo data for San Francisco's neighborhoods, then used FourSquare to find pricing of each restaurant in SF and then clustered them using a k-means approach based on pricing. We then provided maps of the clusters so owners can narrow down where a good restaurant location might be. Comparisons can be drawn from this study such as per capita income vs restaurant pricing that can

lead to a more detailed study. For example, even if a certain area is wealthy, perhaps it is mostly residential and the residents head downtown for more elegant restaurants.