Italian Chef Jimmy De Luca Moves to San Francisco

Garrett George

July 19, 2020

1. Introduction

1.1. Background

Ever since he was a child, Jimmy De Luca has been fascinated with the idea of one day moving to San Francisco. A New York City native, Jimmy has grown up to enjoy a successful career as owner and head chef of the hit restaurant, *Jaime Bolognese*. Though the reservations continue to sell out on a nightly basis, Chef JDL (as his staff are accustomed to calling him) is ready for a new challenge, one that involves finally realizing his dreams of moving to the Golden Coast and opening up a second Italian restaurant - one, he believes, can be even more successful than *Jamie Bolognese*.

1.2. Problem

Moving across the country and starting up a new restaurant carries with it the need to plan effectively. The head chef thoroughly loves his quaint three-story townhouse tucked away in the beautiful SoHo neighborhood. His first priority is to find an area in San Francisco with similar characteristics to his current environment. Once found, Jimmy wants to hunt out the closest district to him that will likely have the most demand for Italian food. Right now, *Jamie Bolognese* is in the East Village area, a short train ride away from Jimmy's SoHo complex. He likes the comfort of being able to easily reach his restaurant if need be, especially as he enters the venerable age of sixty-seven this coming November.

1.3. Interest

Jimmy is just one of a population of chefs and restaurant owners alike who are contemplating taking the next step and opening a second establishment. Some, like Jimmy, might even want to move to the restaurant's new location to better ensure its success. For those individuals who might find themselves in the aforementioned situation, this analytics tool could prove invaluable.

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 algorithm 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. In addition, when comparing areas with a high density of restaurants, Jimmy wants to find the one that has the *highest* Italian restaurants to prove there is a substantial demand for authentic Italian food.

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 other 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.

3. Methodology

To kick off the process, we import a dataset of all San Francisco neighborhoods, which can be extracted from data.sfgov.org. From here, the geocoder library can be used to fetch the latitude and longitude coordinates for each respective neighborhood. Before delving into the Foursquare API to gather information on most popular spots within each neighborhood, however, we first add in New York City's SoHo neighborhood to the dataset, which will be referenced later when performing a K-Means clustering analysis. Once added in, we can begin extracting the most popular venues within each neighborhood by leveraging the Foursquare API. The results of the first five rows of this dataset can be found in *Figure 1*.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bayview Hunters Point	37.741286	-122.377633	Heron's Head Park	37.739663	-122.375904	Park
1	Bayview Hunters Point	37.741286	-122.377633	Bay Natives Nursery	37.740532	-122.376845	Garden Center
2	Bayview Hunters Point	37.741286	-122.377633	Speakeasy Ales & Lagers	37.738468	-122.380874	Brewery
3	Bayview Hunters Point	37.741286	-122.377633	Hunter's Point Shoreline	37.738240	-122.376753	Waterfront
4	Bayview Hunters Point	37.741286	-122.377633	USPS Cafeteria	37.740744	-122.382309	Café

Figure 1. San Francisco neighborhoods dataset with respective venues imported using the Foursquare API

With the location data on all neighborhoods in San Francisco now gathered and the most popular venues in each respective area found, we can use a clustering algorithm to group similar neighborhoods together and see which ones share the most characteristics with New York City's SoHo neighborhood.

To do this, we will leverage the K-Means clustering algorithm due to its simplicity and because the data we are working with is unlabeled. Once the clusters are created, we can observe the neighborhoods that were grouped with New York City's SoHo neighborhood and determine which area would be the best for Jimmy to move to based on venue similarity.

In the final step, we must determine where Jimmy De Luca should open his next Italian restaurants. This can be accomplished by looking at neighborhoods nearby the area selected in the previous step and determining which of these areas has a high number of restaurants and a high concentration of Italian restaurants. By taking this approach, we will ensure that the neighborhood is a common area for restaurants to be located while also showing there is a substantial demand for authentic Italian food.

To begin our data analysis, we first transform the current dataset into a usable form for the K-Means clustering algorithm by utilizing one hot encoding, as depicted in *Figure 2*. After this, the K-Means clustering algorithm can be performed on the data and plotted to a Folium map, the results of which are shown in *Figure 3*.

	Neighborhood	Yoga Studio	ATM	Accessories Store	Adult Boutique	Alternative Healer	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	...
0	Bayview Hunters Point	0.000000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...
1	Bernal Heights	0.028571	0.0	0.0	0.000000	0.0	0.028571	0.0	0.0	0.0	...
2	Castro/Upper Market	0.020202	0.0	0.0	0.010101	0.0	0.010101	0.0	0.0	0.0	...
3	Chinatown	0.000000	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...
4	Excelsior	0.021739	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...

Figure 2. One hot encoding is used to transform the dataset into a usable form for K-Means clustering

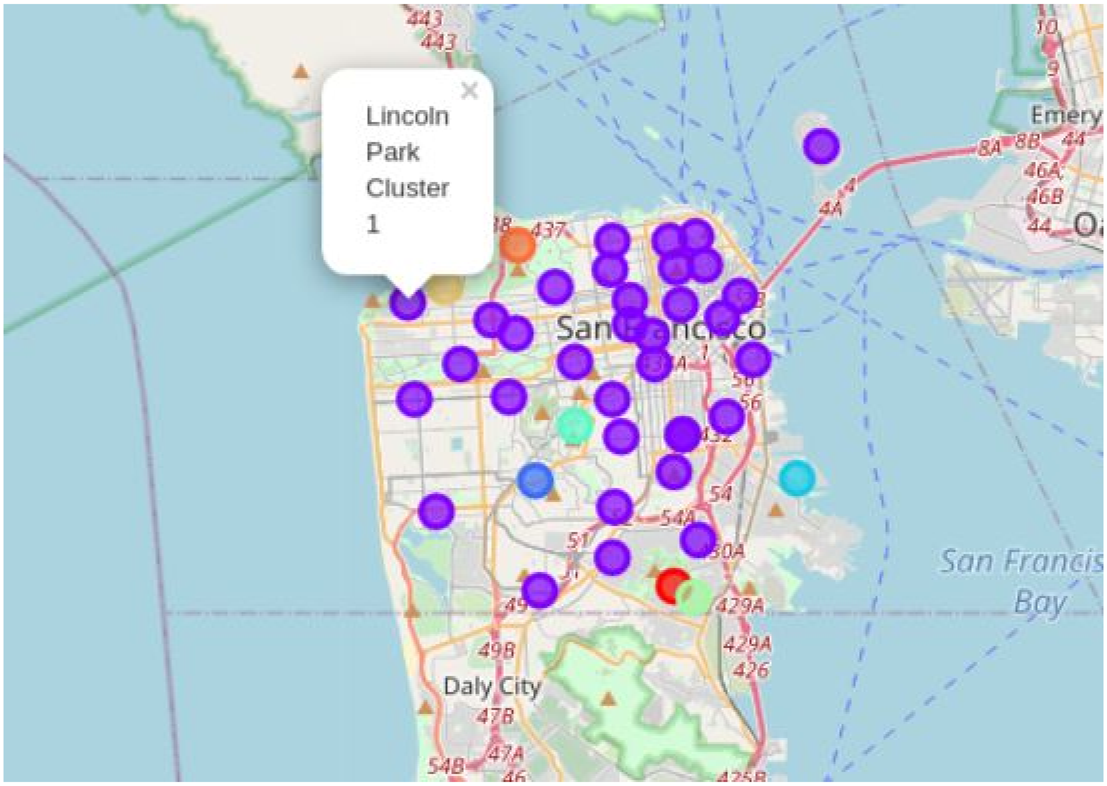

Figure 3. K-Means clustering performed on the San Francisco neighborhoods dataset

At a glance of the map, we can see that the first cluster is the largest cluster of the eight. That being said, we will try to determine from this group which has the most common top venues in its areas as it compares to New York City's SoHo neighborhood. It should be noted that different cluster sizes were tested out to try to break down the first cluster into more granular groupings. However, even after bumping the number of clusters up to eight (and even as high as 12), the grouping stayed about the same size.

To further understand the characteristics of the first cluster, we can show the top ten most popular venues within each neighborhood. This will help us determine the neighborhood in San Francisco most similar to New York City's SoHo neighborhood. *Figure 4* shows the first five rows of what that dataset looks like.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Mo Commo Venu
1	Bernal Heights	Coffee Shop	Trail	Playground	Mexican Restaurant	Italian Restaurant	Bakery	Yoga Studio	Park	Pizza Place	Cocktail B
2	Castro/Upper Market	Gay Bar	Coffee Shop	New American Restaurant	Cosmetics Shop	Thai Restaurant	Arts & Crafts Store	Seafood Restaurant	Playground	Convenience Store	Pet Sto
3	Chinatown	Coffee Shop	Cocktail Bar	Chinese Restaurant	Bakery	Italian Restaurant	New American Restaurant	Hotel	Restaurant	Men's Store	Bubble Te Sho
4	Excelsior	Mexican Restaurant	Bakery	Pharmacy	Chinese Restaurant	Latin American Restaurant	Bank	Sandwich Place	Grocery Store	Pizza Place	Vietnames Restaura
5	Financial District/South Beach	Coffee Shop	Café	Food Truck	Juice Bar	Art Gallery	Salad Place	Bar	Italian Restaurant	Sandwich Place	Gy

Figure 4. The neighborhoods grouped as the 1st cluster containing SoHo along with the top 10 venues in each area

We are going to assume that by only looking at neighborhoods sharing the the 1st most common venue with SoHo, we can both minimize the long list of neighborhoods that make up the first cluster from the K-Means algorithm while also carefully working to find the most similar neighborhood to SoHo.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
19	Russian Hill	Italian Restaurant	Park	Coffee Shop	Playground	Pizza Place	Café	Dive Bar	Sushi Restaurant	Chocolate Shop	Rock Club
22	Nob Hill	Italian Restaurant	Hotel	Coffee Shop	Café	Bar	Wine Bar	French Restaurant	Yoga Studio	American Restaurant	Hotel Bar
41	SoHo	Italian Restaurant	Hotel	Boutique	Coffee Shop	Clothing Store	Mediterranean Restaurant	Women's Store	Men's Store	Sushi Restaurant	Dessert Shop

Figure 5. All neighborhoods that share the same 1st most common venue with New York City's SoHo

As shown above in *Figure 5*, if we take a look at all neighborhoods in San Francisco that share the 1st most common venue with New York City's SoHo, the list dwindles down to Russian Hill and Nob Hill.

The first observation to note is the fact that both Nob Hill and SoHo have "Hotel" as the 2nd most common venue in their respective areas. Further comparing the three neighborhoods, we can see that SoHo has "Coffee Shop" as it's 4th most common venue, which also resides in both Russian Hill and Nob Hill, both as the 3rd most common venue.

Last, SoHo and Nob Hill share common ground on European cuisine, with Mediterranean restaurants residing as the 6th most common venue for SoHo and French restaurant ranking as the 7th most common venue for Nob Hill. SoHo and Russian Hill also share an affiliation for Sushi restaurants, where the category is ranked number eight for Russian Hill and number nine for SoHo.

All that being taken into account, it appears Nob Hill shares the most common characteristics with New York City's SoHo neighborhood. This will be the top suggestion we provide Jimmy De Luca as the place he should consider moving to. However, when calculating the neighborhoods where Jimmy should open up his Italian restaurant in the next step, we will also consider Russian Hill as a viable neighborhood to move to so Jimmy is presented with a variety of options.

Let's now find the top five closest neighborhoods to both Nob Hill and Russian Hill as possible contenders for locations where Jimmy could open up his Italian restaurant by using a distance formula and leveraging the latitude and longitude coordinates. The results for these two tables are shown in *Figure 6* and *Figure 7*, respectively.

	Neighborhood	Latitude	Longitude	Dist. to Nob Hill
0	Russian Hill	37.800073	-122.417094	44.954766
1	Chinatown	37.794301	-122.406376	56.684316
2	Tenderloin	37.784249	-122.413993	57.989573
3	North Beach	37.801175	-122.409002	64.130234
4	Japantown	37.785579	-122.429809	104.582943

Figure 6. The top five closest neighborhoods to Nob Hill, the 1st choice recommended for Jimmy to move to

	Neighborhood	Latitude	Longitude	Dist. to Russian Hill
0	Nob Hill	37.793262	-122.415249	44.954766
1	North Beach	37.801175	-122.409002	51.783916
2	Chinatown	37.794301	-122.406376	77.294286
3	Tenderloin	37.784249	-122.413993	102.745466
4	Marina	37.799793	-122.435205	114.846588

Figure 7. The top five closest neighborhoods to Russian Hill, the 2nd choice recommended for Jimmy to move to

Now that these two tables have been derived, we can answer the following questions by leveraging the Foursquare API:

1. How many restaurants are within each of these areas?
2. How many restaurants are in fact Italian restaurants?

3. Which neighborhood has the highest number of restaurants and the highest ratio of Italian restaurants to total number of restaurants?

Figure 8 and *Figure 9* shows the results after answering the above questions and attaching the respective columns to the previously-created dataframes.

	Neighborhood	Latitude	Longitude	Dist. to Nob Hill	Total No. of Restaurants	No. of Italian Restaurants	Italian Restaurant Frequency
0	Russian Hill	37.800073	-122.417094	44.954766	7.0	0.0	0.000000
1	Chinatown	37.794301	-122.406376	56.684316	49.0	1.0	0.020408
2	Tenderloin	37.784249	-122.413993	57.989573	38.0	0.0	0.000000
3	North Beach	37.801175	-122.409002	64.130234	46.0	11.0	0.239130
4	Japantown	37.785579	-122.429809	104.582943	48.0	0.0	0.000000

Figure 8. Nob Hill's closest neighborhoods and their respective restaurant densities

	Neighborhood	Latitude	Longitude	Dist. to Russian Hill	Total No. of Restaurants	No. of Italian Restaurants	Italian Restaurant Frequency
0	Nob Hill	37.793262	-122.415249	44.954766	10.0	1.0	0.100000
1	North Beach	37.801175	-122.409002	51.783916	46.0	11.0	0.239130
2	Chinatown	37.794301	-122.406376	77.294286	49.0	1.0	0.020408
3	Tenderloin	37.784249	-122.413993	102.745466	38.0	0.0	0.000000
4	Marina	37.799793	-122.435205	114.846588	45.0	2.0	0.044444

Figure 9. Russian Hill's closest neighborhoods and their respective restaurant densities

4. Results and Discussion

Our analysis shows that for both Russian Hill and Nob Hill, there is a clear outlier that fits the question we sought to answer: Which neighborhood has the highest number of restaurants and the highest ratio of Italian restaurants to total number of restaurants?

Before looking at the results, let's summarize exactly how we were able to get to this point in our analysis. We first started out by determining which neighborhoods in San Francisco share the most similarity to New York City's SoHo neighborhood, where Italian chef Jimmy De Luca currently lives. The Foursquare API was then used to determine the top venues within each neighborhood and the K-Means clustering algorithm was then applied to group these neighborhoods by venue similarity. Once accomplished, it was determined that both Nob Hill and Russian Hill are two great contenders for neighborhoods in San Francisco Jimmy should highly consider moving to.

The next task at hand was to help Jimmy decide where he should open his next Italian restaurant. Three criteria guided us on finding the most optimal location:

1. Closeness to his new San Francisco neighborhood (either Nob Hill or Russian Hill)
2. Neighborhoods with a high concentration of restaurants (indicating a demand for food exists)
3. Neighborhoods with a high concentration of Italian restaurants (indicating a market exists for Italian food)

To address the first criterion, a distance formula was used (leveraging the latitude and longitude coordinates) to arrive at a dataframe with the top five closest neighborhoods to both Nob Hill and Russian Hill. Addressing the second and third criteria, the Foursquare API helped determine the total number of restaurants that exist within each area (using a radius of 175 meters) as well as how many Italian restaurants encompass this total. An Italian Restaurant Frequency calculated column was also included, allowing us to quickly determine which area has the highest concentration of Italian restaurants.

Turning to the results, it is clear that North Beach is the best choice in both the Nob Hill table and the Russian Hill table. Let's look at the numbers to understand why. First observing the Nob Hill table, we can see in the Total No. of Restaurants column there is a lack of restaurants in Russian Hill, the closest neighborhood to Nob Hill. The remaining four neighborhoods all have a relatively high number of restaurants, but after observing both the No. of Italian Restaurants column and the Italian Restaurant Frequency column we can see that North Beach clearly has both the highest number of Italian restaurants in its area (11 to be exact) as well as the highest concentration (standing at nearly 25%).

Similar results can be found if we now turn our attention to the Russian Hill table. Again, looking first at the Total No. of Restaurants column, Nob Hill clearly dwarfs the other four neighborhoods in total restaurant count and so can be quickly eliminated from the list. Now turning to the No. of Italian Restaurants and Italian Restaurant Frequency column, it is clear that North Beach once again is the obvious winner among the five contenders due to the fact that, as observed in the Nob Hill table, it has the highest number and greatest density of Italian restaurants.

5. Conclusion

The purpose of this project was to find an area in San Francisco, CA for Chef Jimmy De Luca to move to that has similar characteristics to his current home: the beautiful Soho neighborhood in New York City. This was accomplished with the help of the Foursquare API, used to gather the ten most popular spots around each San Francisco neighborhood and using the K-Means algorithm to cluster the neighborhoods based on venue similarity. This led us to find that Nob Hill and Russian Hill share the most common characteristics to New York City's SoHo neighborhood. These are the two neighborhoods we suggest Jimmy should consider moving to in order to feel more at home.

After we found two contenders for Jimmy's next home, we used a distance formula to calculate the closest neighborhoods to him where he could possibly open up his next hit Italian restaurant. To determine which of these areas might provide Jimmy's restaurant with the highest chance of success, we leveraged the Foursquare API to look for the nearby neighborhoods that have a high number of restaurants and a high number of Italian food restaurants, which should be a strong indication that food (specifically, Italian food) is in demand in the area. This led us to find that North Beach fits the description for both the aforementioned criteria. North Beach has a relatively high number of total restaurants compared to other contenders (standing at 46 restaurants within a 175 meter radius from the epicenter) and the highest concentration of Italian restaurants making up these venues (rounding to roughly 25% of total restaurants).

We believe this data-driven approach should provide Jimmy De Luca sufficient reasoning as to why he should consider Nob Hill or Russian Hill as his next place to live and why North Beach could likely grant him success if he opens up his next Italian restaurant there. Many other factors (including crime rate and educational opportunities for his kids) were not in scope for this analysis and should be considered when Jimmy finalizes his plans to move across the country.