

Best Location to open an Italian Restaurant in San Francisco

1. Introduction

1.1 Background

We have decided to open a restaurant in San Francisco California, with an original menu and great service. The ideal place because there are a lot of owners that want to operate a first class, brand new, state of the art restaurant in San Francisco, one of our targets is to create a beautiful ambiance that fits the tenor of a great chef in a tony neighborhood that speaks to the San Francisco palate.

We have focus in two future problems for the succeed of the restaurant, the first is the cost opening price, so we need to locate it in a strategic location, and second for security, we are going to search for the neighborhoods and streets with the lower crime rate, without sacrificing possible good areas.

Opening up a high-end eatery in the City of San Francisco is no easy feat and not for the faint of heart. In this day and economic atmosphere, it is extremely difficult to put together restaurant deals that make sense financially. It has driven people out of the business and left some of us veterans working harder than ever before to improve food, service and overall value to keep up with the burgeoning competition.

1.2 Problem

In this project we are going to focus on find an optimal location for the restaurant, our principal problem is that there are a lot of restaurants in San Francisco, so find a location that are not already crowded with restaurant is one of our problems.

The second is to find an area with no Italian restaurants in vicinity, we will prefer locations near to the city center, in particular to Union Square that is a common place to visit in San Francisco.

Also we need to pay attention to the crime zone rate. We don't want to locate it in a dense crime area.

1.3 Interest

The people expected to invest to build out the restaurant are private investors and stakeholders that are expecting a good return on their investment.

2. Data acquisition and cleaning

2.1 Data sources

Based on the definition of our problem the factors what will influence our decision are the number of existing restaurants in the neighborhood, any type of restaurant, the distance to an Italian restaurant if any, the distance to our point of start Union Square, and the zone crime rate.

For centers of candidate areas, the date will be obtained using Google Maps reverse geocoding.

For the number of restaurants, their type and location in every neighborhood will be obtained using Foursquare API.

For the dataset of the crime rate in San Francisco we are going to use the provided in the seventh course Data Visualization with Python.

2.2 Data cleaning

Our start point was Union Square so use reverse geocoding to get the latitude and longitude of Union Square in San Francisco. 4km around this, we created a grid of cells covering our interest area. Those grid areas candidates were placed equally spaced around the center, to do this we needed to create our grid of locations in Cartesian 2D coordinate system, after that project those coordinates back to latitude and longitude degrees to be shown in Folium Map.

Then we made a request using Reverse Geocoding to get the address of the Union Square, and then made a request to obtain locations addresses.

After use FourSquare API to get info on restaurants in each neighborhood.

There were values that we did not need, so I decided to get the relevant information to our project.

Once we get the info, find the id for a specific category was a task to find the restaurants type, category, address, how near ere form our central point and look for the venues. Then we went over our neighborhood locations and get nearby restaurants locations, and separate this data for venue id, venue name, venue categories, venue latitude and longitude, if that restaurant was an Italian.

2.3 Feature selection

We find around one thousand only a 7.99 percent were an Italian restaurant, we get the whole restaurant info, also just the Italian restaurants, and the restaurant locations.

This way we conclude the gathering phase.

3. Methodology

Our methodology was, first collect the required data as location and type and category of every restaurant within 4km ratio from Union Square.

The second analysis in our analysis was the calculation and exploration of restaurant density across different areas in San Francisco, we made use of heatmaps to identify a few promising areas close the center with low number of restaurant in general.

The Heatmap was very usefully because with it we can get a very visual idea about the zones.

In third and final step we focused on most promising areas within those created clusters, finally we will take into consideration location with no more than two restaurants in radius of 250 meters, and locations without Italian restaurants in radius of 400 meters. And we will use k-means clustering of those zones to identify general zones which should be a starting point for final street level exploration.

4. Exploratory Data Analysis

Once we did a grid of candidates areas, we visualize them.

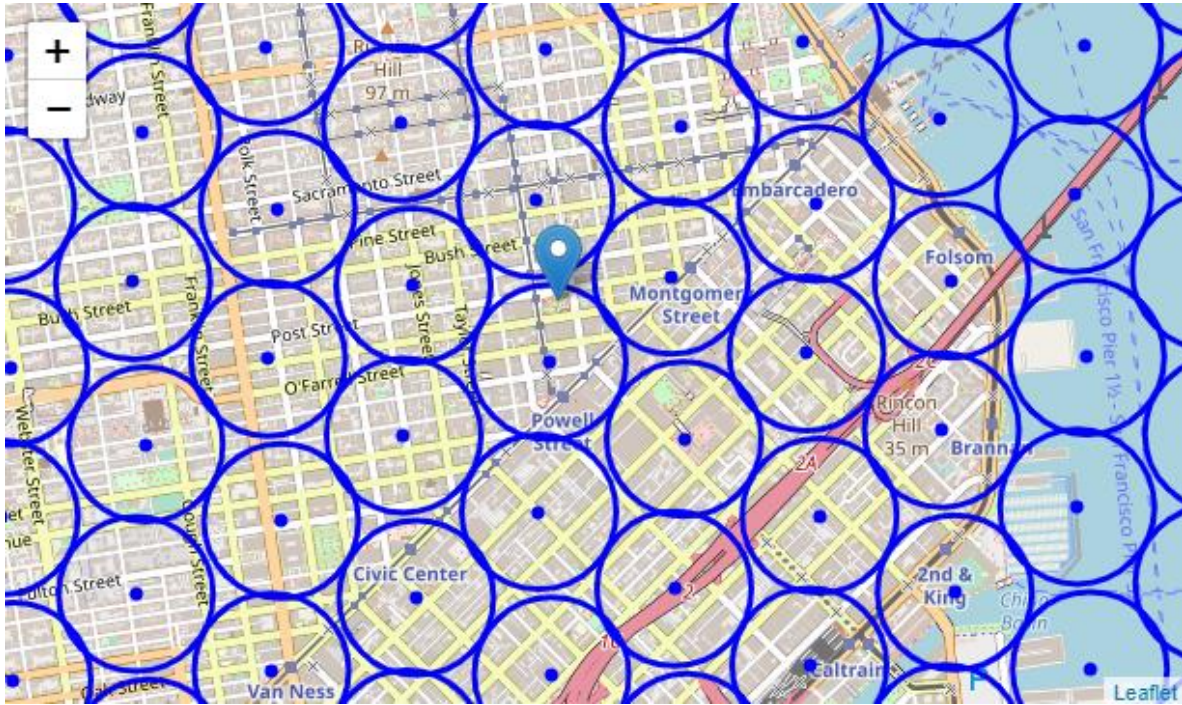


Figure 1. Candidate areas location

We used Foursquare and once we get all the restaurants in the area within a few kilometers from Union Square, we visualize them, the blue points are restaurants of every type, and red points are specifically Italian restaurants.

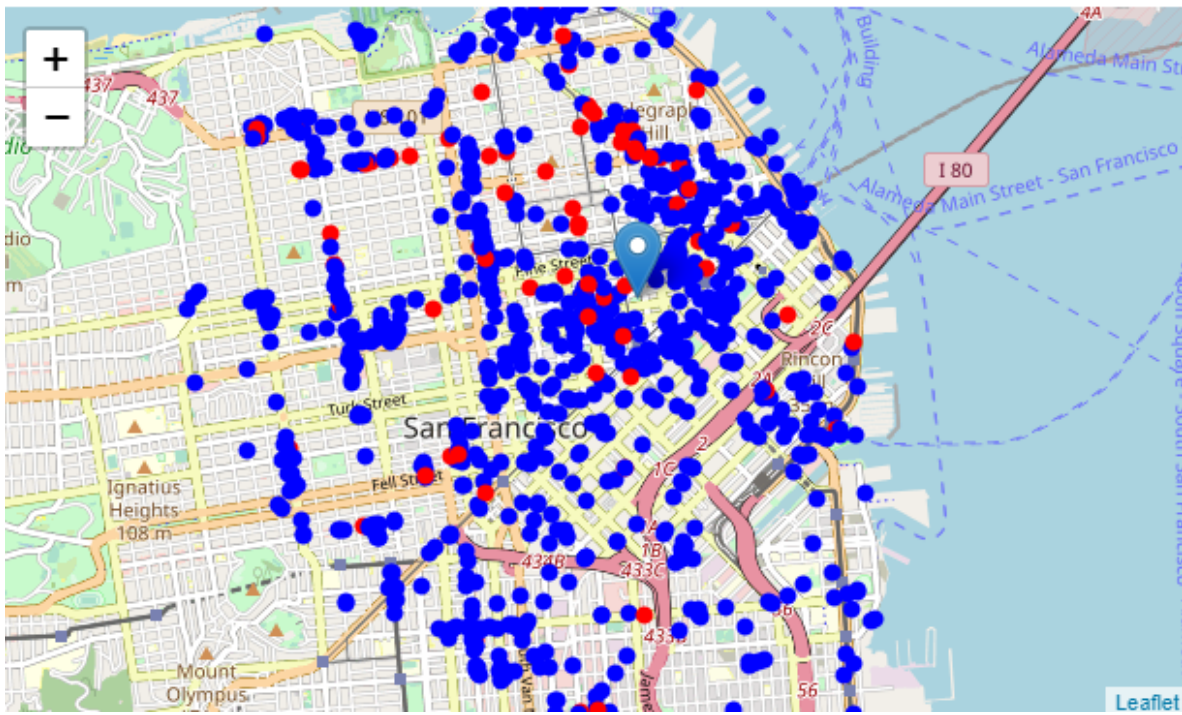


Figure 2. Restaurants around city, Italian and no Italian.

The average number of restaurant in every area with a radius of 300m is approximately 7.64, and the average distance to close Italian restaurant from each area center was of 700m, that is a little closer so we needed to filter our area carefully.

To get visual information to know denser areas, we use Heatmaps

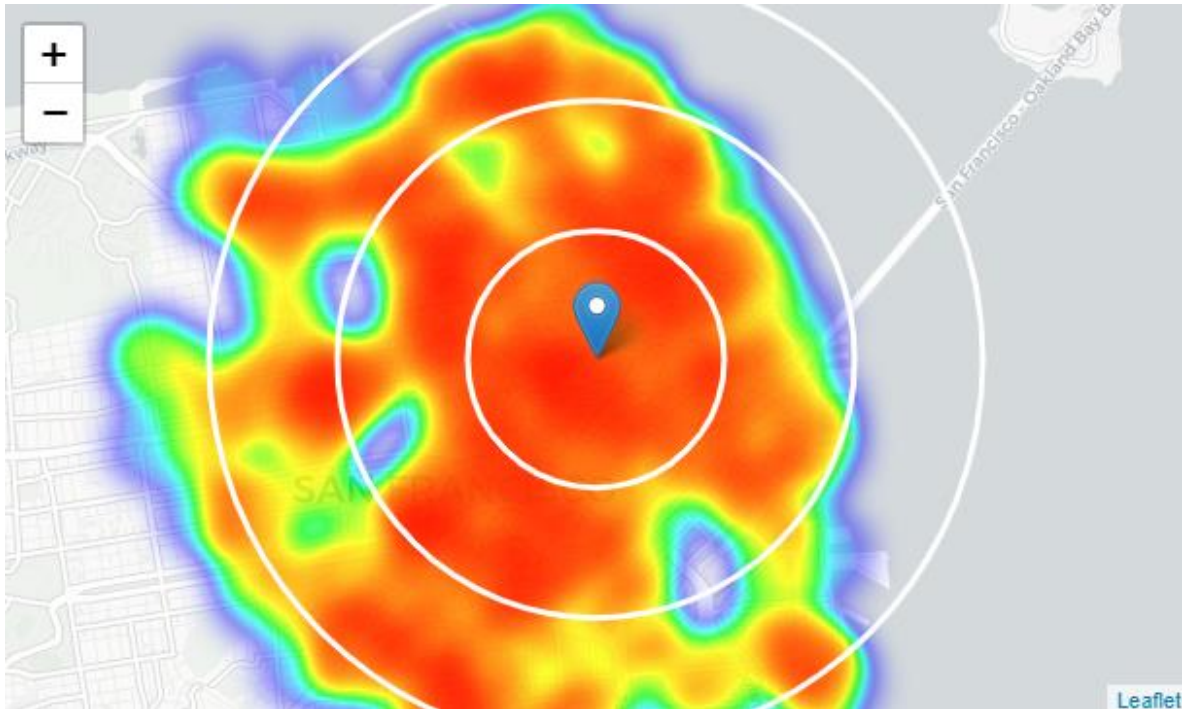


Figure 3. Heat map of all restaurant density

This tell us that we have a few pockets of low restaurant density, one of our possible candidates are located at the south, and west. So for get a better information we make a heat map of only Italian restaurants.



Figure 4. Italian restaurants heat map.

This Italian restaurant represents a subset of 8% of all restaurants in San Francisco, based on this we focus on south, since there are not a lot of Italian restaurants.

At San Francisco south, we find three candidates boroughs between them South Beach, West Soma and Hayes Valley.

South Beach is a neighborhood in San Francisco, California with a population of 5,041. South Beach is in San Francisco County and is one of the best places to live in California. Living in South Beach offers residents a dense urban feel and most residents rent their homes. In South Beach there are a lot of bars, restaurants, coffee shops, and parks. Many young professionals live in South Beach and residents tend to be liberal. The public schools in South Beach are highly rated.

Soma was once a light industrial and warehouse district. Today, it's one of the coolest neighborhoods in San Francisco. SOMA is packed with happening bars, boutique shopping, alternative culture, museums, art galleries, and tasty restaurants. It also has great parks and is conveniently located next to lots of public transportation.

Hayes Valley's transformation started back in 2003, when a committed group of citizens succeeded in taking down the Central Freeway that had divided their neighborhood. The efforts paid off massively: Since then, the area's seen major

investment from local brands and restaurants. The mix of creative boutiques and shops makes for a very stroll-able shopping neighborhood, not to mention what's become one of the best food scenes in the city.

Analysing deeper this information and more, we decide that one potential place could be South Beach, because it has a lot of hotels, which bring the most tourists.

So we focus on South Beach, since this is an area with a great quantity of tourists.

Every blue circle is a location possible for opening the new Italian restaurant.

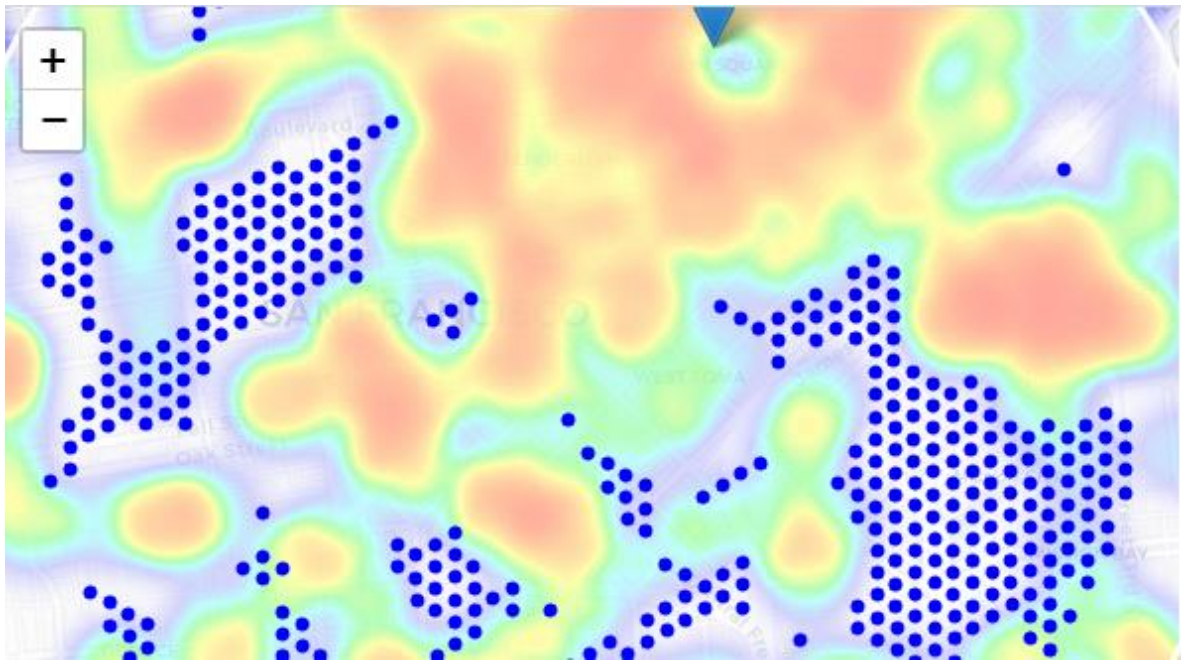


Figure 5. Possible locations for opening the new restaurant.

Filtering this and just plotting the candidate's areas and applying K-Means Clustering. We find that the center of the clusters are our new start point for the search of the perfect location.



Figure 6. K-means clustering centers.

5. Results

Finally, we use reverse geocode to those candidate area centers to get the addresses to be presented to the stakeholders.

But before to present that, we discard the areas with high crime rate.

The next figure represents the heat map of the high crime rate areas in San Francisco.

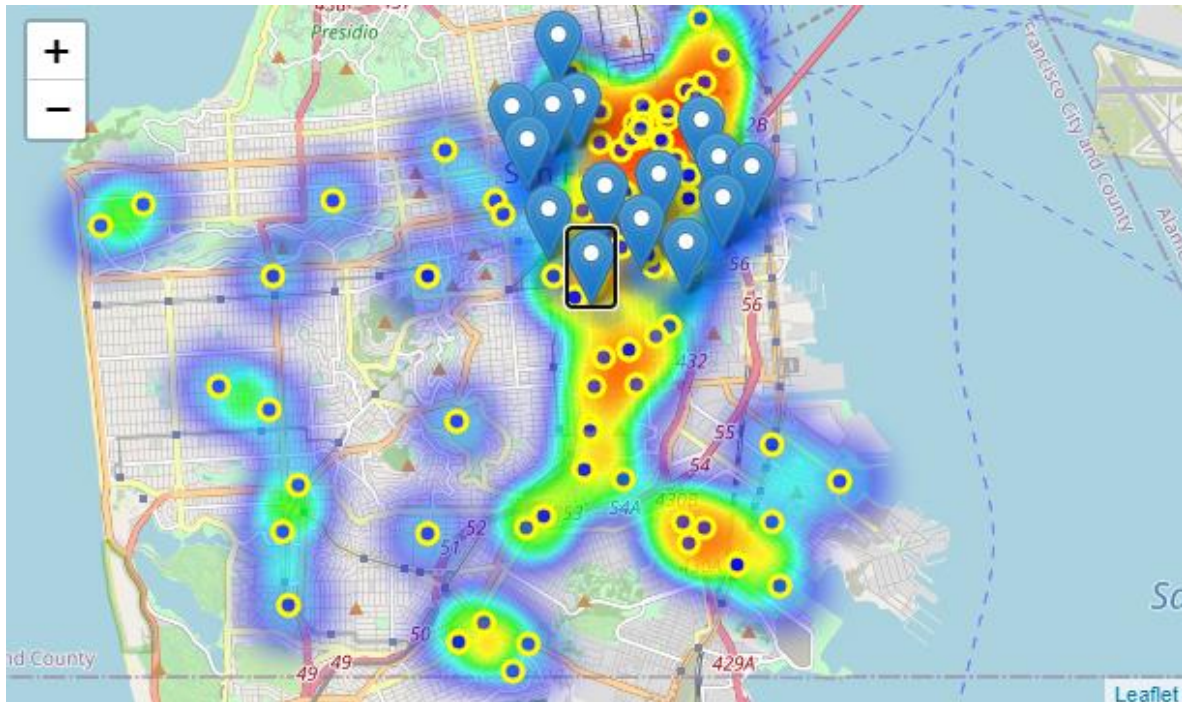


Figure 7. Dense areas of high most crime rate.

So our final results are the next.

```
[ '1675 Eddy St, San Francisco, CA 94115',
  '1200 7th St, San Francisco, CA 94107',
  '610 Long Bridge Street, San Francisco, CA 94158',
  '1141 Turk St, San Francisco, CA 94115',
  '1104 Bryant St, San Francisco, CA 94103',
  '2049 Sacramento St, San Francisco, CA 94109',
  '2145 18th St, San Francisco, CA 94107',
  '435 Townsend St, San Francisco, CA 94107',
  '295 Church St, San Francisco, CA 94114',
  '951 Eddy St, San Francisco, CA 94109',
  '951 Eddy St, San Francisco, CA 94109']
```

6. Discuss (observations, recommendations based on the results)

The previous analysis we did, show that although there is a great number of restaurant in San Francisco, there are pockets of low restaurant density at the south less than 3 km, which is in our focus area that was of 4km around. Highest concentration of restaurant was detected center, north, north-east from Union Square, that is why we focus to south and south-west areas, most of them corresponding to South Beach borough, some other possible candidates were located at south-east, but South Beach is more popular and attracts more

tourists.

After directing out attention to this narrow area of interest, we first created a dense grid of location candidates spaced 100m apart, those locations were then filters so that those with more than two restaurants in radius of 250m and those with an Italian restaurant closer than 400 were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of center of those zones were also generated using reverse geocoding to be used as starting point for more detailed local analysis.

Additional, we made a heat map of crime rate zones in San Francisco, 4 of the candidates for the restaurant were close to the concurrent centers of crime, so we decided to eliminate that possible location candidates.

Resulting in 11 zones containing largest number of potential new restaurants location based on the number, distance and crime rate to existing venues, distance to restaurants in general and Italian restaurant particularly. This does not imply that those zones are actually optimal locations for a new restaurant.

Purpose of this analysis was to provide information on areas close to San Francisco but not crowded with existing restaurants nearby. It is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reason which would make them unsuitable for a new restaurant regardless of lack of competition in the area.

The recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

7. Conclusion

The purpose of this project was to find the best location to open a new Italian Restaurant in San Francisco in order to aid stakeholders in narrowing down the search for optimal location. This was made by calculating density distribution from Foursquare data we have identified general boroughs that justify further analysis.

(South Beach), then generate a collection of locations which satisfy some basic requirements regarding existing nearby restaurants.

Clustering this location in order to create major zones of interest, and addresses of those zone centers were created to be used as starting point.

The final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and location in every one zone we recommended, taking in consideration additional factors like attractive of each location, could by the proximity to park, level of noise, proximity to major roads, availability, prices, social and economic behaviors of every neighborhood.

References

<https://insidescoopsf.sfgate.com/blog/2011/01/21/the-cost-of-opening-a-restaurant/>