Capstone Project - The Battle of the Neighbourhoods-Report

Applied Data Science Capstone by IBM/Coursera

Table of contents

Introduction: Business Problem

An investor is looking to open a new **Indian Style** restaurant in **San Francisco**, but he is not sure about the best location for his new venue. So, you got a call from him asking your input to help him choose the location. San Francisco is a very busy city, best known for tourist attractions and business innovation. Strolling around the city blocks, it is pretty easy to notice that the city already has a lot of restaurants in town. How should we proceed and decide the location?

In The Art of War, Sun Tzu has said: "Know the enemy and know yourself; in a hundred battles you will never be in peril." Following this line of thinking, the basic strategy here is then to know what are the most critical factors that contribute to the restaurant's profitability. According to a report https://www.foodnewsfeed.com/fsr/vendor-bylines/8-factors-choosing-new-restaurant-location by Tom Larkin published in the FSR magazine, these components stand out as the most important ones: visibility, parking, space size, crime rates, surrounding businesses and competitor analysis, accessibility, affordability, and safety. Using public datasets, we could actually address some of these considerations pretty straightforwardly.

Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighbourhood (any type of restaurant)
- number of and distance to Indian Style restaurants in the neighbourhood, if any
- distance of neighbourhood from city center

We decided to use regularly spaced grid of locations, cantered around city center, to define our neighbourhoods.

Following data sources will be needed to extract/generate the required information:

- Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
- number of restaurants and their type and location in every neighbourhood will be obtained using **Foursquare API**
- coordinates of San Francisco center will be obtained using **Google Maps API geocoding**

The city of San Francisco has maintained a large data repository hosted on the website DataSF (https://datasf.org/). From there, we can access the city's crime rate, housing price, and public parking information. Sounds good! So, let's proceed and see what insight can be extracted from the data. We will be working with two datasets:

Police Department Incident Reports: 2018 to Present

https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783

and

Assessor_Historical_Secured_Property_Tax_Rolls

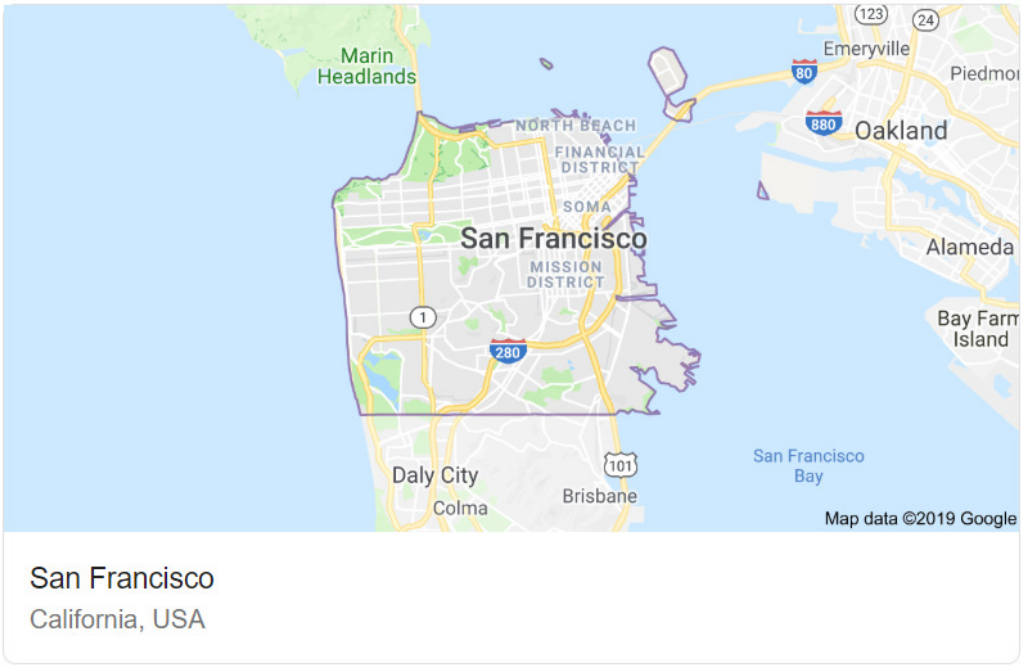
https://data.sfgov.org/Housing-and-Buildings/Assessor-Historical-Secured-Property-Tax-Rolls/wv5m-vpq2

The **Neighbourhood details of San Francisco** is available in this website

https://data.sfgov.org/Geographic-Locations-and-Boundaries/SF-Find-Neighborhoods/pty2-tcw4

Since these data files are in huge size, we cannot refer them as URL. We need to download them to our local system and should refer the file location in our program. Some of the files are in SHAPE files format and some other files are in CSV format. We are using both types of files in our project.

The Map of San Francisco:



Methodology

8 Factors for Choosing a New Restaurant Location

THESE ELEMENTS ARE CRITICAL FOR A NEW RESTAURANT'S SUCCESS.

Whether you're opening your first full-service restaurant, your second, or your 50th, it's important to understand what to look out for when choosing a new restaurant location. For seasoned restaurateurs, you may have a successful location where you are but how much of that success is inadvertently down to accidental—or purposeful—restaurant location choice? The answer may be it has everything to do with it. Here are tips on choosing just the right spot:

Visibility

You may have found the cutest, quaintest location for a restaurant in a historic downtown somewhere, but if you're tucked in a side street with little foot traffic, unless you have amazing marketing tactics, you may have the most Instagram- and Pinterest-worthy space with zero customers.
When choosing a space, you want somewhere that's visible. To determine visibility, look at foot and car traffic patterns. If people are walking by, who's walking by? Is it nannies with strollers? If your restaurant is not kid friendly, they aren't going to go inside. Are doctors and nurses passing by from a nearby hospital to grab a quick bite on their lunch breaks? What about local hipsters who want a trendy long lunch? Determine if nearby pedestrians are your target customers. If not, the space probably won't work for you.

Check different dayparts, too. Do these diners want a full-service restaurant for breakfast or lunch, or will you adapt for those times, offering faster service or a sandwich counter, or do you just want to open for fine dining in the evening? Will that daypart bring enough traffic?
For car traffic, can people pass your restaurant, and by the time they've seen it, make a last-minute choice to pull in and check it out? So many people drive by new places and think, oh, I must try that new restaurant, but customers need to be able to go there before you go out of business. How many restaurants have you been meaning to try only to see a "for rent" sign in the window six months later?

Parking

As much as no parking, parallel parking, and offsite parking may work for the street cafés of Europe and the country pubs in the UK, it doesn't work for American restaurants, with the exception of big cities. Whether you're opening fine dining or a popular chain, make sure the land you rent or buy is big enough to accommodate parking for your hungry clientele.

Space Size

The space you rent or buy needs to fit your size requirements. For every restaurant seat, you need about five square feet of kitchen space, so if you have eighty seats in your restaurant, your kitchen needs to have 400 square feet of space minimum. Larger spaces reduce workplace accidents by minimizing slips, trips, and falls, the most common workplace accident according to OSHA. You'll want to consider how big of a space you need for your restaurant requirements.

Crime Rates

Crime rates are unglamorous considerations, but if you place your restaurant in a crime-laden area, are your target customers going to visit? High crime rates can make potential customers uncomfortable, and if they feel they'll be mugged walking to their cars, it will only drive away business, no matter how legendary your coq au vin.

Surrounding Businesses and Competitor Analysis

You'll want to do your research surrounding businesses. Are they doing well? Is the area affluent? Is there enough room for your business? Also, you'll want to know what types of restaurants do

well in the area; however, you don't want to open a pizzeria if there are four in the area. Areas can only support so many of the same type of restaurant. What will distinguish any new restaurant is excellent service and consistently wonderful food.

You'll also want to know what the building was used for before you inhabit it. If the building wasn't a restaurant, will the conversion costs affect your bottom line? If it was a restaurant, was it there for twenty years with great success and the owners wanted simply to retire, or did it change hands five times? Find out why past restaurants were successful or not and if location played a role in these factors. Find out how location works for surrounding businesses, and will you be catering to the right crowd? For example, if there is a nightclub nearby, that will work if you're catering to the after-hours crowd when they've concluded a night of drinking.

Accessibility

Some restaurants find success in just-off-the-highway locations, or located near exits for those interstate travellers who need a bite to eat but don't want the usual fast-food restaurant. You'd have to keep peak times in mind for these kinds of locations, as well as customer demographic. Depending on your exact exit location, you may be limiting other types of traffic to your restaurant.

Affordability

Cost is always a bottom-line consideration for any business. If the rent or purchase of the space is more than you'll bring in each month in profits, that location is not feasible at that time. However, if you know that you'll generate business from that location, then you might consider it, but you'll need to be able to afford the upfront costs before you turn profits. Although some risks do pay off, you don't want to be at the point where you're struggling to cover basic costs. A killer location won't make up for driving your restaurant out of business.

Safety

Once you've landed your perfect space, keep in mind that OSHA reports that slips, trips, and falls are the most common workplace accident, accounting for 15 percent of all accidental deaths, and they are more prevalent in the restaurant industry. It's important to install slip-resistant flooring, have slip-resistant mats in the kitchen spaces, and provide your workers with highly-rated safe, slip-resistant footwear to help prevent those preventable slips and trips. Slips and trips could cost your company more in fines than renting a space, so make sure you consider all angles of safety when choosing the best space. If your space has its own equipment, make sure that all fryers are safe, ice machines aren't leaking, and there aren't any trip hazards in the kitchen.

When choosing the perfect location for your new place, make sure you're making the decision with your eyes wide open, informed with the knowledge of how location can affect your costs and profits.

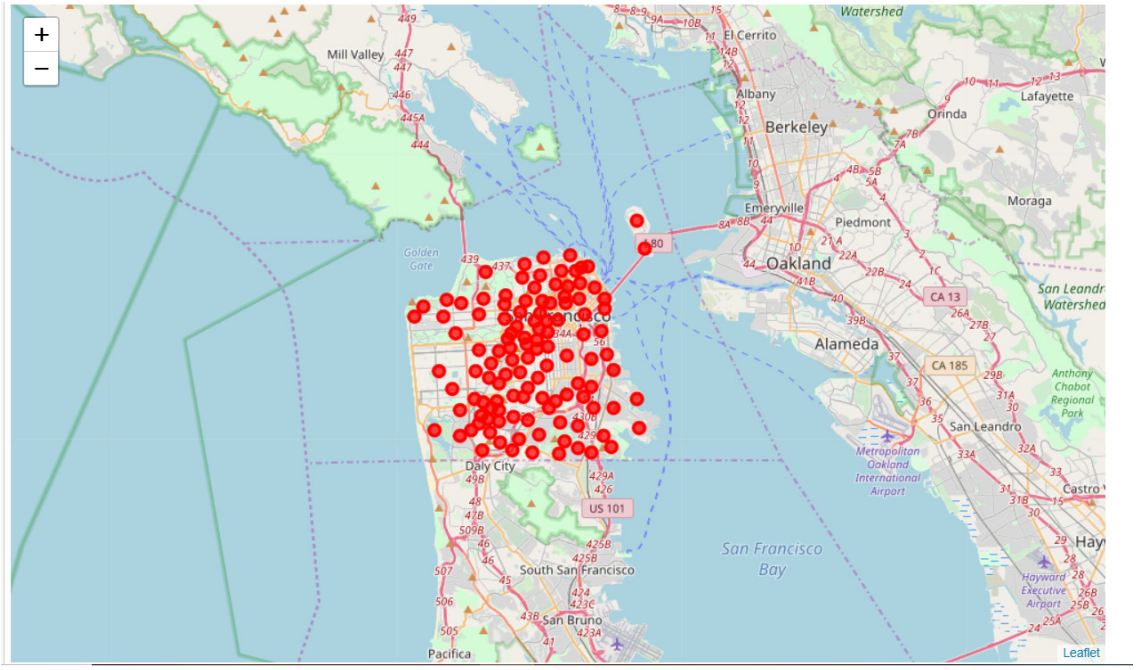
Analysis

In this Project we are analysing the following data to derive a conclusion
1) Crime data of San Francisco
2) Average House Pricing in the neighbourhoods of San Francisco
3) Location and Venue Analysis using Foursquare API
3) Competition from Opponent using K means Clustering algorithm
4) Parking Data available

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Seacliff	37.786521	-122.488899	China Beach	37.788090	-122.491186	Beach
1	Seacliff	37.786521	-122.488899	Seacliff	37.788259	-122.486401	Neighborhood
2	Seacliff	37.786521	-122.488899	Rochambeau Playground	37.784892	-122.484309	Playground
3	Seacliff	37.786521	-122.488899	Tal-Y-Tara Tea & Polo Shoppe	37.783710	-122.486750	Tea Room
4	Seacliff	37.786521	-122.488899	El Camino del Mar Trail	37.786568	-122.494399	Trail

Here we have used a pipeline to channel the json file from search to Pandas DataFrame. Note the last column is 'Venue Category.' The API requires GPS coordinates of the neighborhoods. The GeoDataFrame nbrhoods only defines polygons, but we could deploy the representative_point() function from GeoPandas to extract a location that is guaranteed inside each neighborhood. This is also how the location of popups is generated. The coordinates are recorded in the columns 'nbrhood Latitude' and 'nbrhoods Longitude' in the above dataframe.

Results and Discussion



Apparently, a lot of the neighborhoods are in the red cluster. When we actually look at the red cluster, it becomes clear that the most common venues in the neighborhoods are restaurants: bar, cafe/coffee shop, Chinese/Japanese/Korean/Italian/American restaurants. So yes, the clustered results are consistent with the impression: San Francisco indeed has lots of restaurants already! So where should we put our new restaurant? Well, by checking out the crime and housing price maps, it appears that the **North Beach** neighborhood might be a good candidate. This area is close to the city's water front attractions, so we expect the region to have a lot of foot and car traffic, i.e. good visibility. And the maps also indicate that **North Beach** has a relatively low crime rate in 2018 and a reasonable housing cost. By calling Foursquare API again and narrow our search to the food sector, we could obtain a detailed map of restaurants in North Beach as already discussed in the Project. Most of the venues are located in the blocks along the Columbus Avenue. There are also a few sitting along the Broadway.

Conclusion

So, this is like the first-order solution to the question “Where should we open a new restaurant in San Francisco?” Using public datasets, we are able to, at least partially, address a few factors we have mentioned at the beginning: crime rate and housing cost. We also carry out a very simple competitor analysis based on the distribution of restaurants in the chosen neighborhood. There certainly is lot of room for improvement. For example, we have yet to answer the question about parking space. DataSF has public parking datasets, which could help us understand the landscape of the parking space distribution. We may also use ParkWhiz (https://www.parkwhiz.com/) to help locate public as well as private parking spaces. The competitor analysis can also be refined by segregating the restaurants in North Beach into different categories. This information is extremely useful because we certainly don’t want too many competitors in the same sector. While the approach discussed here is primitive, it nevertheless showcases the usefulness of data analysis!

Data Analysis used for the Project:

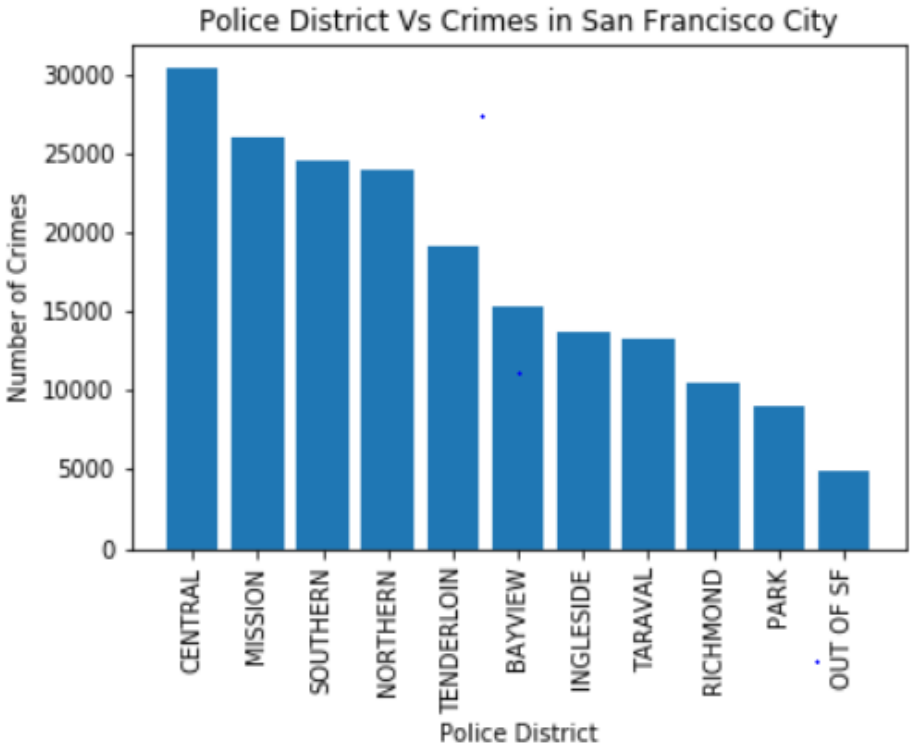


Figure 1 Police District Vs Number of Crimes

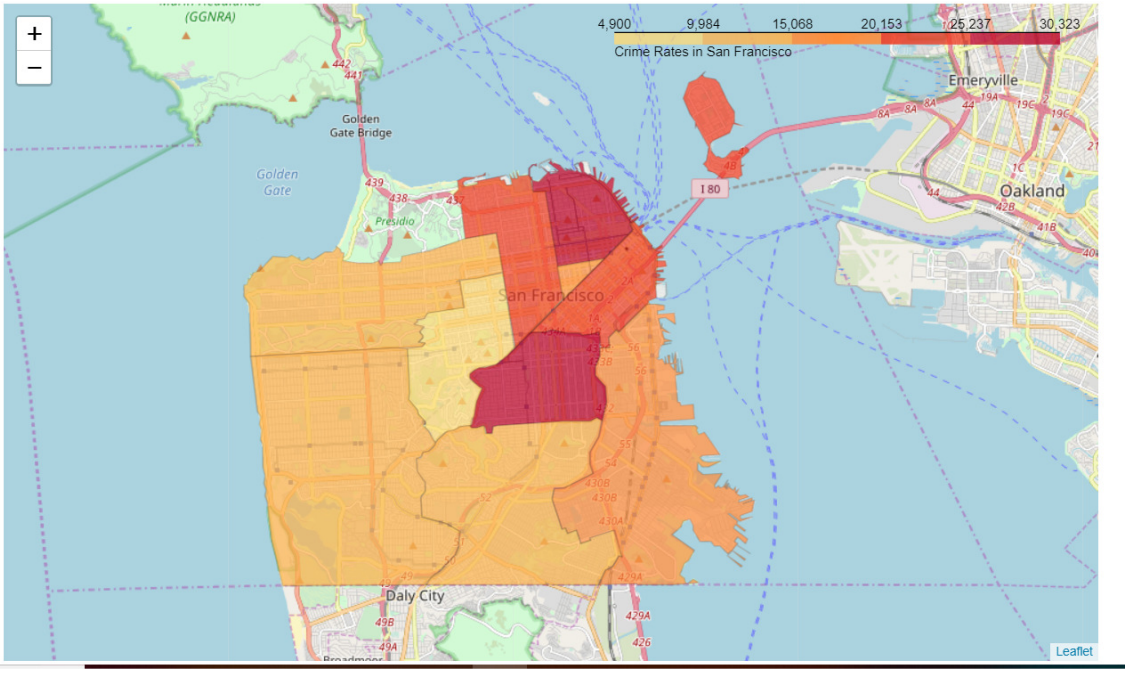


Figure 2 Pictorial Representation of Crime Rates in San Francisco

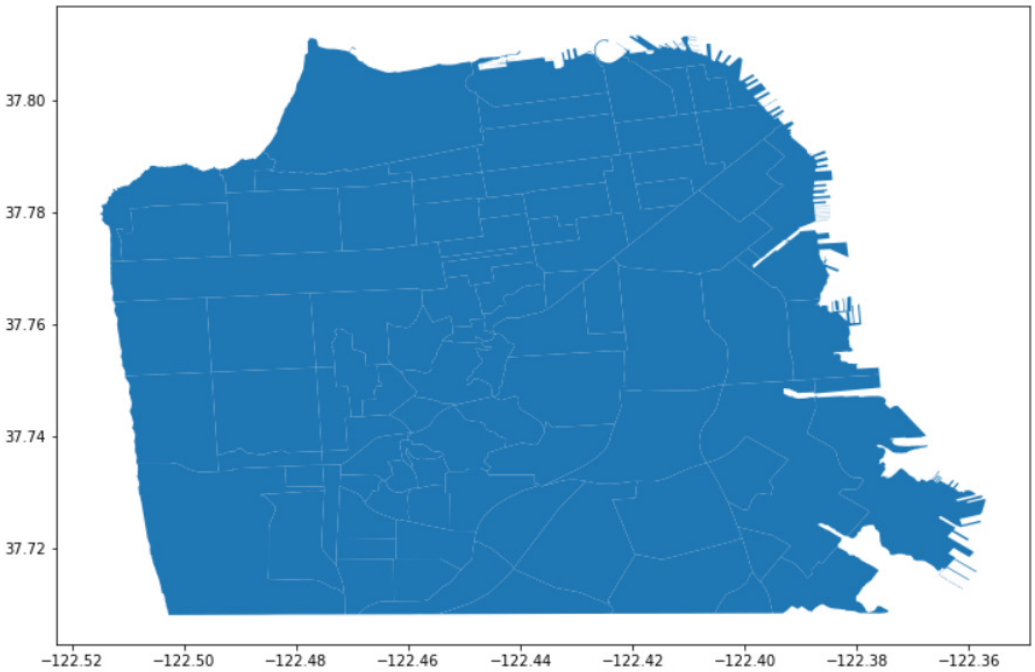


Figure 3 Neighbourhoods in San Francisco

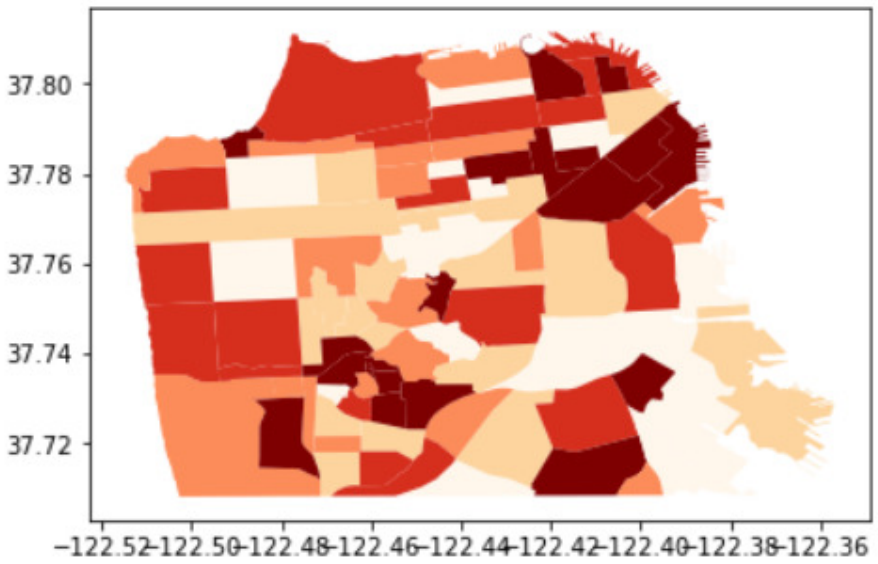


Figure 4 Crime Counts in Neighbourhoods in San Francisco

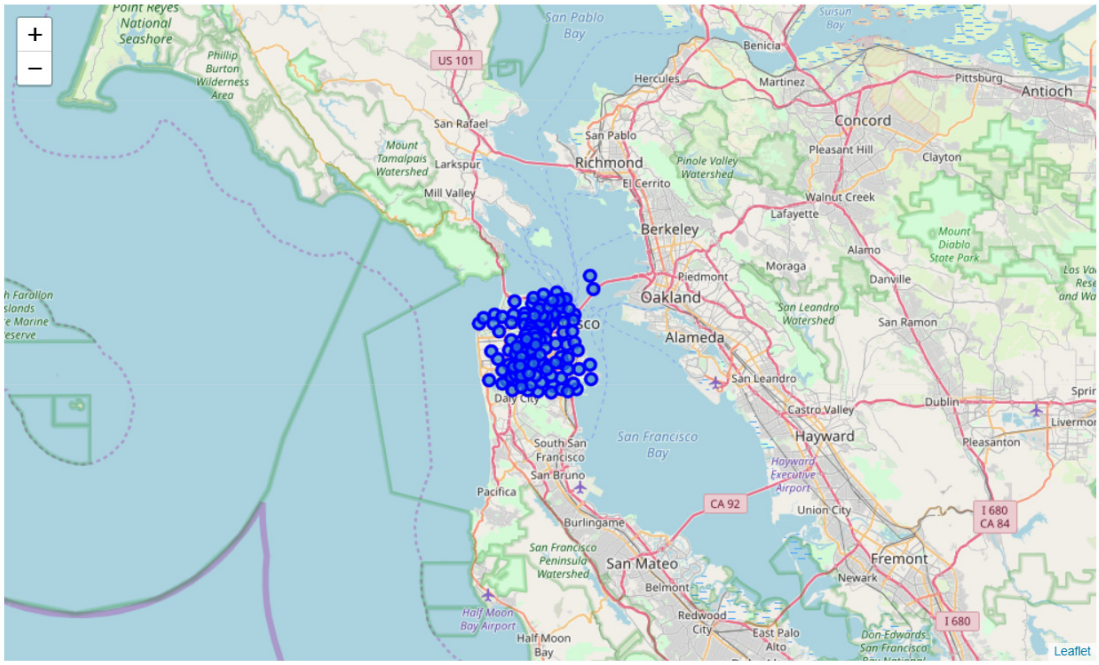


Figure 5 Neighbourhoods in San Francisco

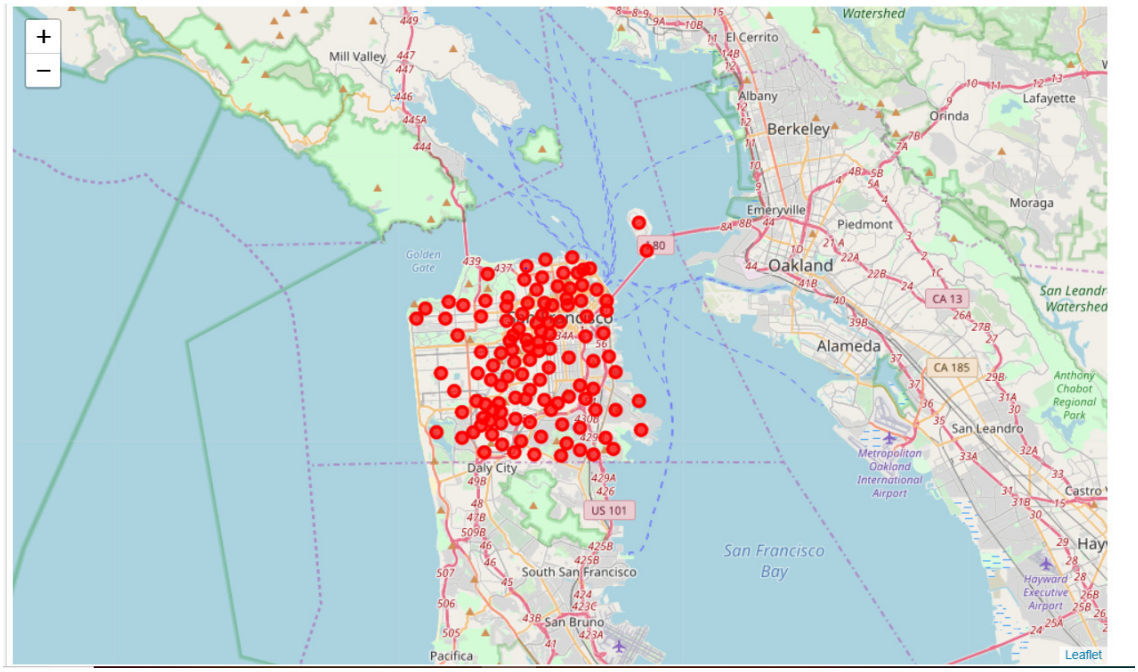


Figure 6 Analysis of Existing Competition in Neighbourhoods in San Francisco