1. **Introduction & Business Problem :**

   As of 2020,City of San Francisco has the highest salaries, disposable income, and median home prices in the world at $1.7 million, as well as the highest median rents. San Francisco Bay Area in Northern California serves as a global center for high technology, innovation, venture capital, and social media. It is multicultural. It provides lot of business oppourtunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

   Business Problem

   The City of San Francisco is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history. Sushi restaurants have become so popular in the United States now it seems that there is one on every corner, not only in major cities but also in smaller cities. Starting a sushi restaurant can be a great business opportunity, but you need to distinguish yourself from others to enjoy long-term success.

   My client wants to open his business in San Francisco area, so I focus on that borough during my analysis. We define potential neighborhood based on the number of sushi bars which are operating right in each neighborhood.

2. **Data**

   Data 1: The number of restaurants in the neighbourhood and types of restaurant. Number of and distance to Italian restaurants in the neighbourhood. Distnace of neighbourhood from city center.

Data2: 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 neighborhood will be obtained using Foursquare API

coordinate of San Francisco center will be obtained using Google Maps API

3. Methodology

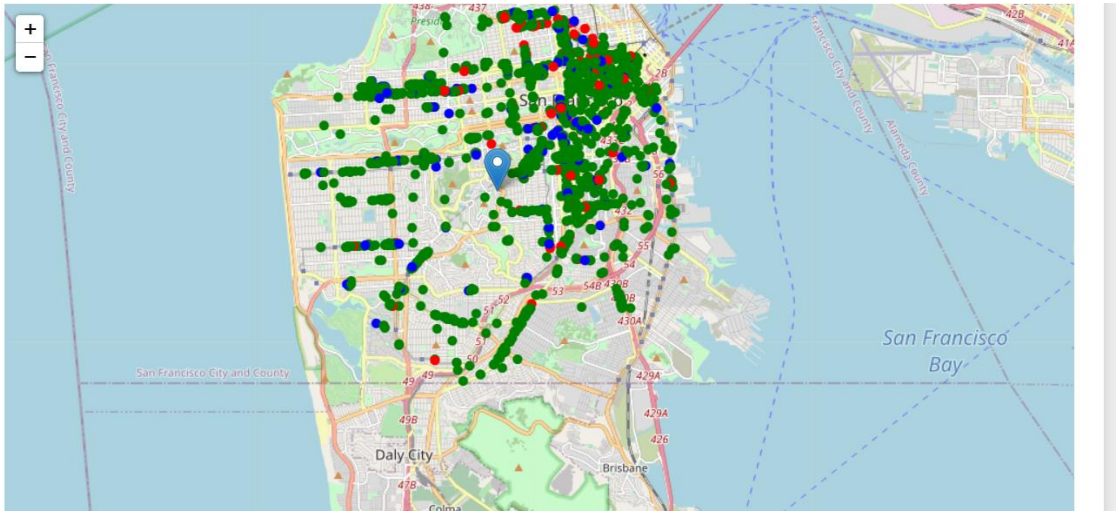
In this project we will direct our efforts on detecting areas of San Francisco that have low restaurant density, particularly those with low number of Italian/Sushi restaurants. We will limit our analysis to area ~6km around city center.

In first step we have collected the required data: location and type (category) of every restaurant within 6km from San Francisco center (Corbett Ave). We have also identified Italian/Sushi restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of San Francisco - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Italian/Sushi restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Italian/Sushi restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

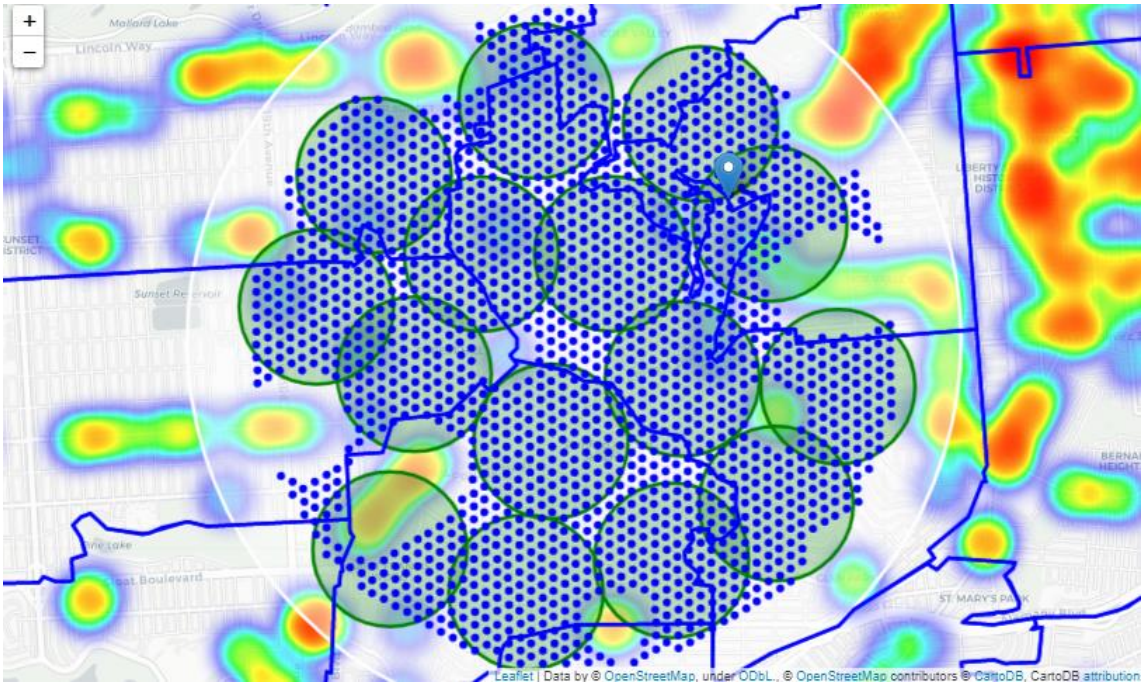
Italian/Sushi restaurants SFO



Then use this feature to group the neighborhoods into clusters K-means clustering algorithm will be used to complete this task. And also, the Folium library to visualize the neighborhoods in SFO and its emerging clusters.

4. Results

K-mean Cluster Using K-mean to clustering data area with less number of Italian/Sushi bars



5. Discussion

Results and Discussion

```
Addresses of centers of areas recommended for further analysis
==============================================================

9 Turquoise Way, San Francisco, CA 94131, USA      => 1.4km from Corbet
Ave
35 Medical Center Way, San Francisco, CA 94122, USA => 1.6km from Corbe
t Ave
255 San Marcos Ave, San Francisco, CA 94116, USA   => 2.7km from Corbet
Ave
70 Los Palmos Dr, San Francisco, CA 94127, USA     => 2.8km from Corbet
Ave
40 Worth St, San Francisco, CA 94114, USA          => 0.4km from Corbet
Ave
1631 12th Ave, San Francisco, CA 94122, USA        => 2.7km from Corbet
Ave
490 Colon Ave, San Francisco, CA 94127, USA        => 3.4km from Corbet
Ave
99 Uranus Terrace, San Francisco, CA 94114, USA    => 0.6km from Corbet
Ave
677 Duncan St, San Francisco, CA 94131, USA        => 1.7km from Corbet
Ave
18 Greenview Ct, San Francisco, CA 94131, USA      => 1.0km from Corbet
Ave
968 Portola Dr, San Francisco, CA 94127, USA       => 2.3km from Corbet
Ave
5705 Diamond Heights Blvd, San Francisco, CA 94131, USA => 2.3km from C
orbet Ave
1000 Clarendon Ave, San Francisco, CA 94131, USA   => 1.9km from Corbet
Ave
55 Santa Ana Ave, San Francisco, CA 94127, USA     => 3.7km from Corbet
Ave
2027 16th Ave, San Francisco, CA 94116, USA        => 3.2km from Corbet
Ave
```

Our analysis shows that although there is a great number of restaurants in San Francisco (~2000 in our initial area of interest which was 12x12km around Corbet Ave), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north and west from Corbet avenue.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Corbet Ave) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Italian restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Berlin center but not crowded with existing restaurants (particularly Italian) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. 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.

6. Conclusion

Purpose of this project was to identify San Farncisco areas close to center with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location

(proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.