VISITING RESTAURANTS IN SAN FRANCISCO DURING THE COVID 19 PANDEMIC

Matias Garib

July 27th, 2020

INTRODUCTION

Background

- Covid 19 has had tremendous impact on our way of life and hygiene has become a fundamental aspect everywhere.
- Many day-to-day activities have been affected by the virus and going out to dinner is for many the most sorely missed thing to do.
- People will have to choose wisely where to go in order to avoid getting infected.

Problem

People want to start going out and visiting restaurants, but they want to visit places with the best hygiene practices. The questions we want to answer, for the city of San Francisco, is: **Which are the safest neighborhoods to go out to eat?**

DATA

Data Sources

Three main data sets from two sources:

- Foursquare
- Hygiene Inspections of SF restaurants (SF Government)
- Realtor Data of SF neighborhoods (SF Government)

DATA

Feature Selection

Foursquare

- Venue (name)
- Venue_category
- Venue_latitude
- Venue_longitude

SF Hygiene Inspections

- Business_name
- Business_Address
- Inspection_Date
- Inspection_Type
- Violation_Description
- Risk_Category
- Inspection_Score

DATA

Cleaning

From the 53.000 inspections data, the cleanup consisted of:

- Approximately 12.000 rows with inspection_score = 'NaN' values were dropped.
- All the data with inspection_score = 100, perfect score, has risk_category value equal to 'Nan'. So, a new category was included and named 'No Risk'.
- The data was ordered by Inspection date and only the most recent inspection was considered for restaurants that were inspected more than once Approximately 34.000 rows where dropped.

Merging

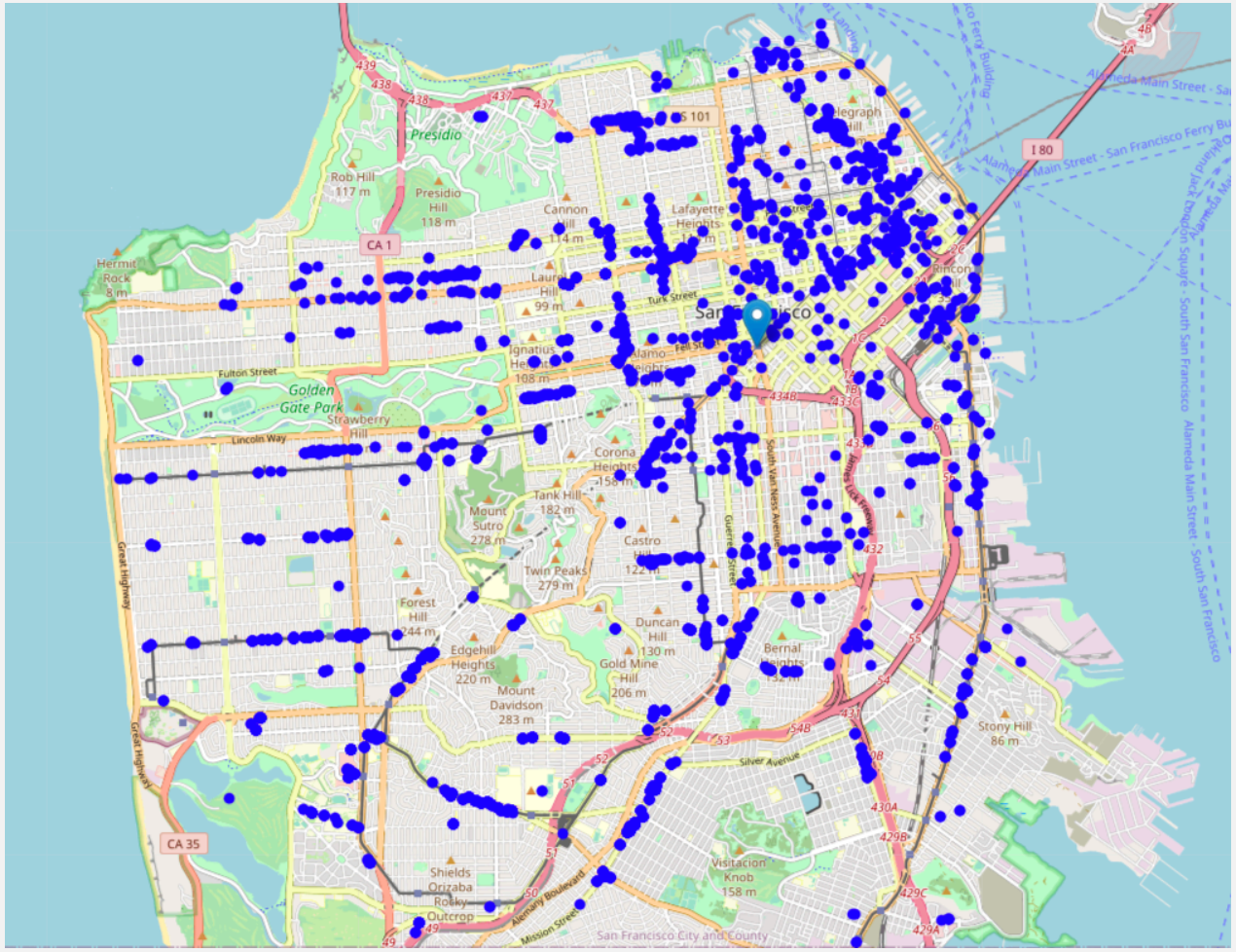
- Merge using Fuzzy Pandas

DATA

	Neighbourhood	Venue	Venue_Latitude	Venue_Longitude	Venue_Category	business_name	business_address	inspection_date	inspection_type	violation_description	risk_category	inspection_score
0	Alamo Square	Little Star Pizza	37.777489	-122.438281	Pizza Place	Little Star Pizza	846 Divisadero	2018-04-24T00:00:00.000	Routine - Unscheduled	No Violation	No risk	100
1	Alamo Square	Brenda's Meat & Three	37.778265	-122.438584	Southern / Soul Food Restaurant	Brendas Meat & Three	919 DIVISADERO ST	2019-03-13T00:00:00.000	Routine - Unscheduled	Unapproved or unmaintained equipment or utensils	Low Risk	92
2	Alamo Square	The Mill	37.776425	-122.437970	Bakery	The Mill	736 DIVISADERO St	2019-04-11T00:00:00.000	Routine - Unscheduled	Unapproved or unmaintained equipment or utensils	Low Risk	88
3	Alamo Square	Jane the Bakery	37.783797	-122.434283	Bakery	Jane the Bakery	1875 Geary Blvd	2019-07-03T00:00:00.000	Routine - Unscheduled	Unclean or unsanitary food contact surfaces	High Risk	87
4	Alamo Square	The Progress	37.783745	-122.432972	American Restaurant	The Progress	1525 Fillmore St	2019-02-14T00:00:00.000	Routine - Unscheduled	Moderate risk food holding temperature	Moderate Risk	90
...	...	...	...	...	...	...	...	...	...	...	...	...
5240	Nob Hill	Osso Steakhouse	37.791447	-122.413530	Steakhouse	Osso Steakhouse	1177 California St	2019-06-03T00:00:00.000	Routine - Unscheduled	Inadequately cleaned or sanitized food contact...	Moderate Risk	96
5241	Nob Hill	Batter Bakery	37.789551	-122.420776	Bakery	Batter Bakery	1517 Pine St	2018-08-21T00:00:00.000	Routine - Unscheduled	Wiping cloths not clean or properly stored or ...	Low Risk	98
5242	Nob Hill	Nobhill Pizza & Shawerma	37.790767	-122.419747	Pizza Place	Nobhill Pizza & Shawerma	1534 California St	2019-09-23T00:00:00.000	Routine - Unscheduled	High risk food holding temperature	High Risk	93
5243	Nob Hill	Kasa Indian Eatery	37.789655	-122.420449	Indian Restaurant	Kasa Indian Eatery	4001 18th St	2019-09-23T00:00:00.000	Routine - Unscheduled	Insufficient hot water or running water	Moderate Risk	86
5244	Nob Hill	Golden Horse Restaurant	37.790860	-122.417340	Chinese Restaurant	Golden Horse Restaurant	1060 Hyde St	2018-01-23T00:00:00.000	Routine - Unscheduled	Foods not protected from contamination	Moderate Risk	79

Master Data Frame (5216,12)

DATA



Data points mapped out

METHODOLOGY

1. *Geographic Visualizations*:

We first want to understand the big picture and the best way to do this is without any numbers. The output from this first part will aim to locate where the best/worst restaurant zones are using both heat a choropleth maps. If there are clearly defined zones of better performance, this analysis should give us those hints.

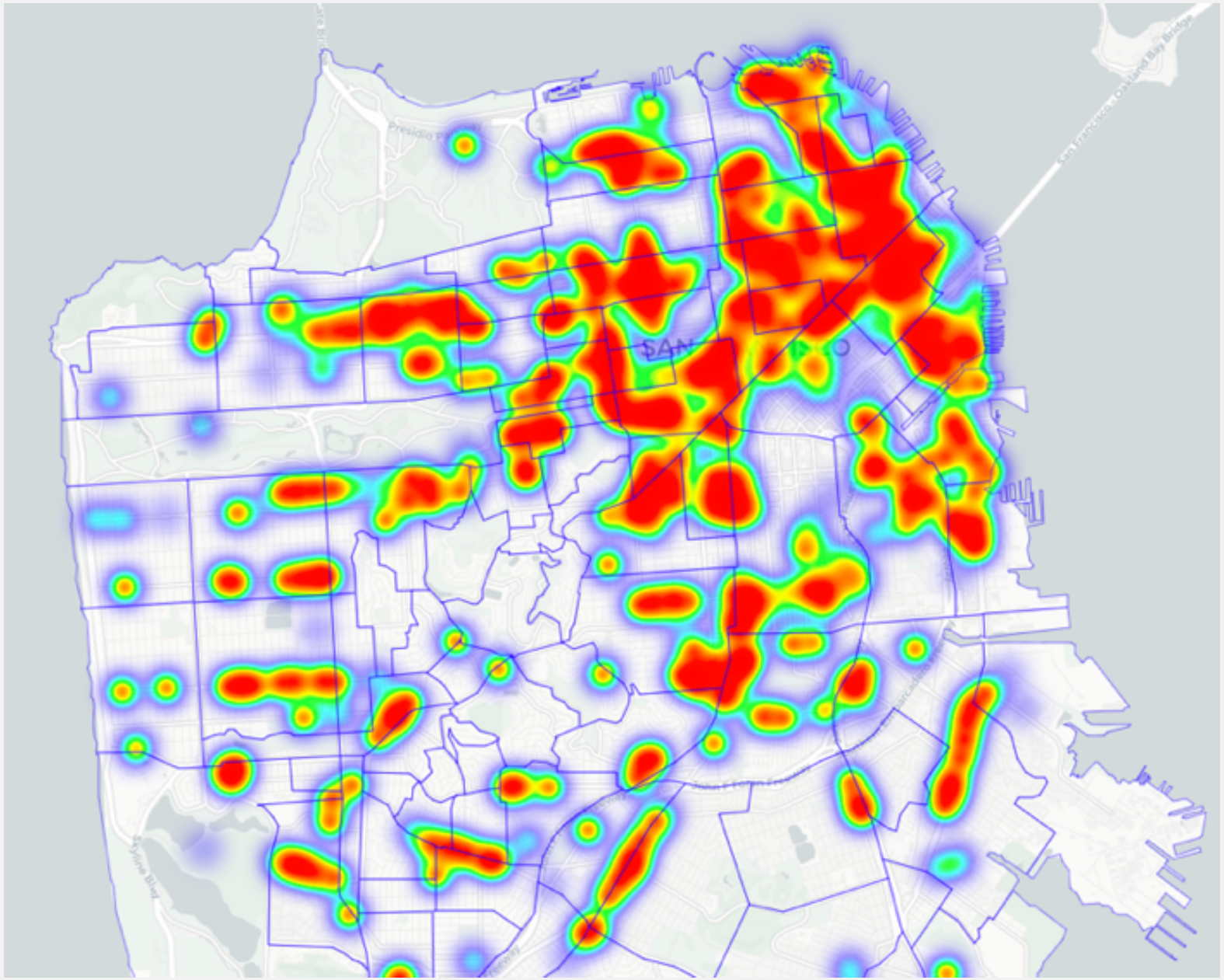
2. Clustering and Analysis:

This second part will mainly focus on separating the good/bad neighborhoods and understanding how, numerically, they differ. Two clusters will be used as we expect to get "GO" and "DON'T GO" groups of neighborhoods. The analysis will then focus on the following criteria to choose which cluster is best:

RESULTS AND DISCUSSION

All Restaurants Heat Map:

- Restaurant dense areas are focused in the NE part of SF
- Restaurant density is focused around financial district
- Park areas present the least restaurant dense areas

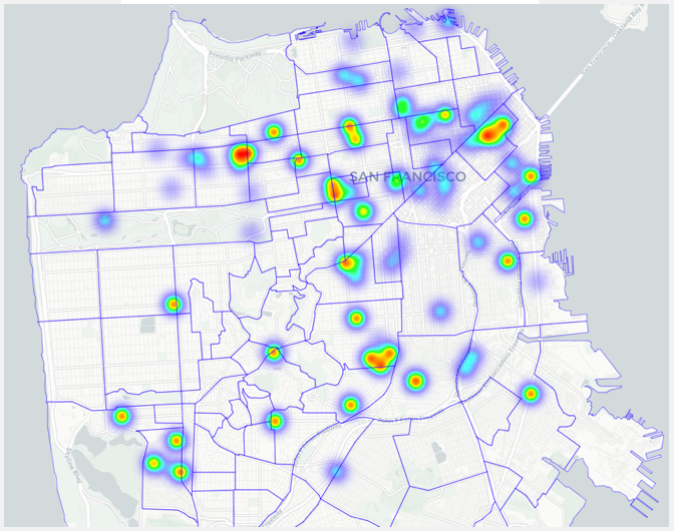


RESULTS AND DISCUSSION

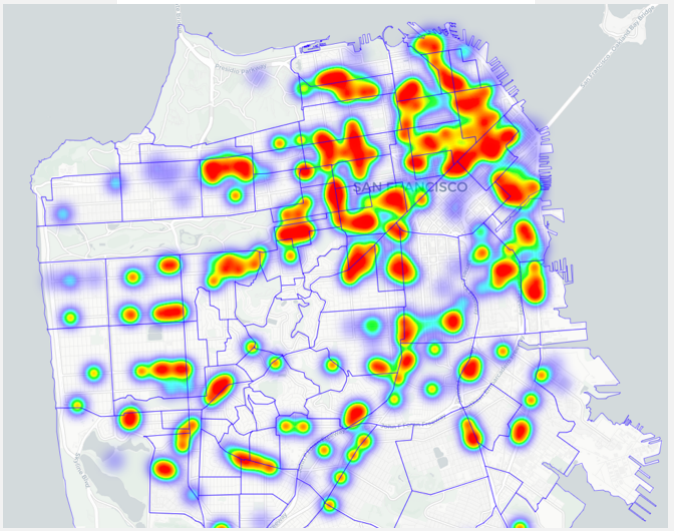
Restaurants Heat Map for each Risk Category:

- Most restaurants are categorized either as Low or Moderate risk restaurants
- No Risk category has less coverage compared to the others

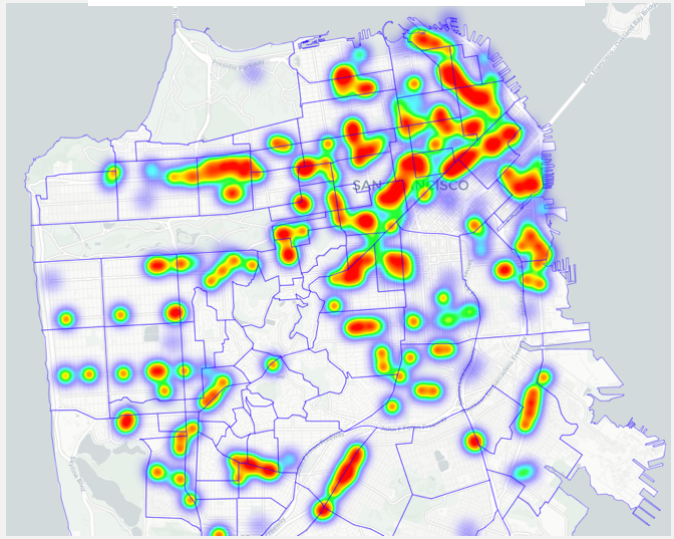
No Risk restaurants



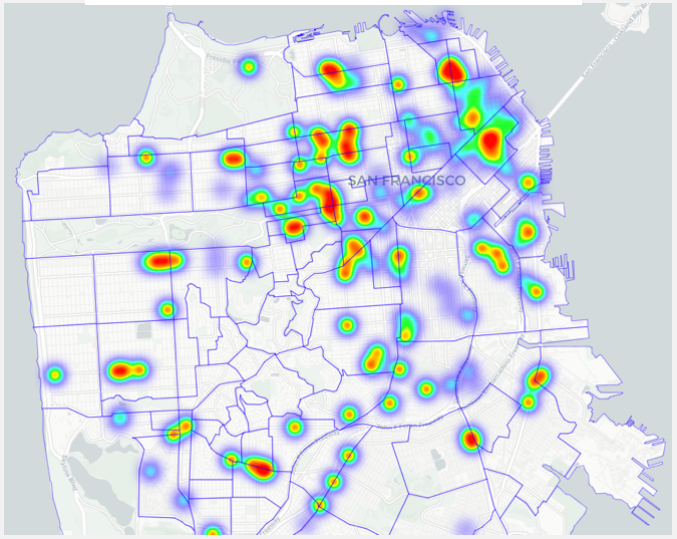
Low Risk restaurants



Moderate Risk restaurants



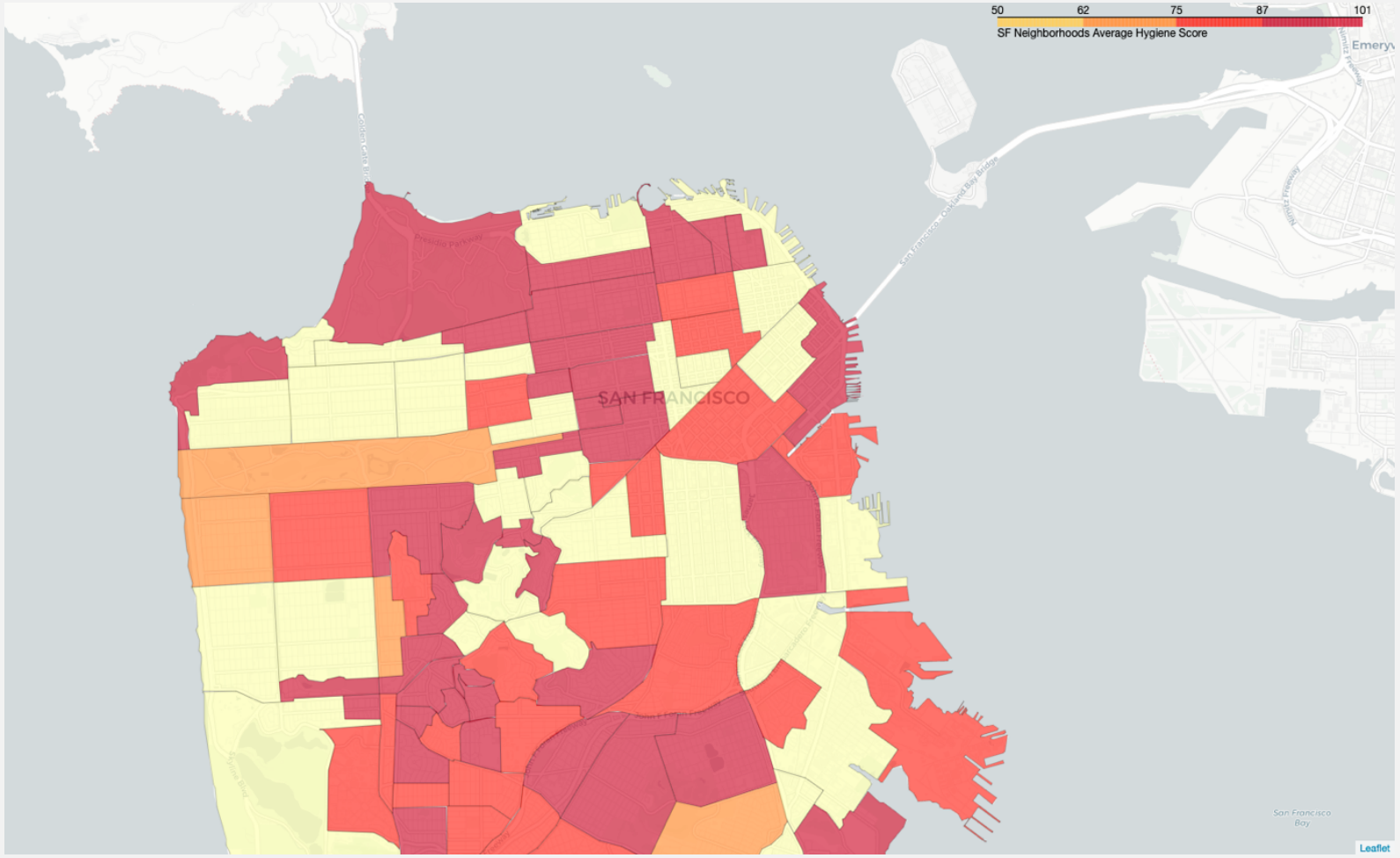
High Risk restaurants



RESULTS AND DISCUSSION

Average Inspection Score per Neighborhood:

- Darker red neighborhoods (higher scores) tend to be located NE.
- As one moves S-SW, lighter reds and orange start to appear, suggesting that areas with less restaurants also have worst average inspection scores

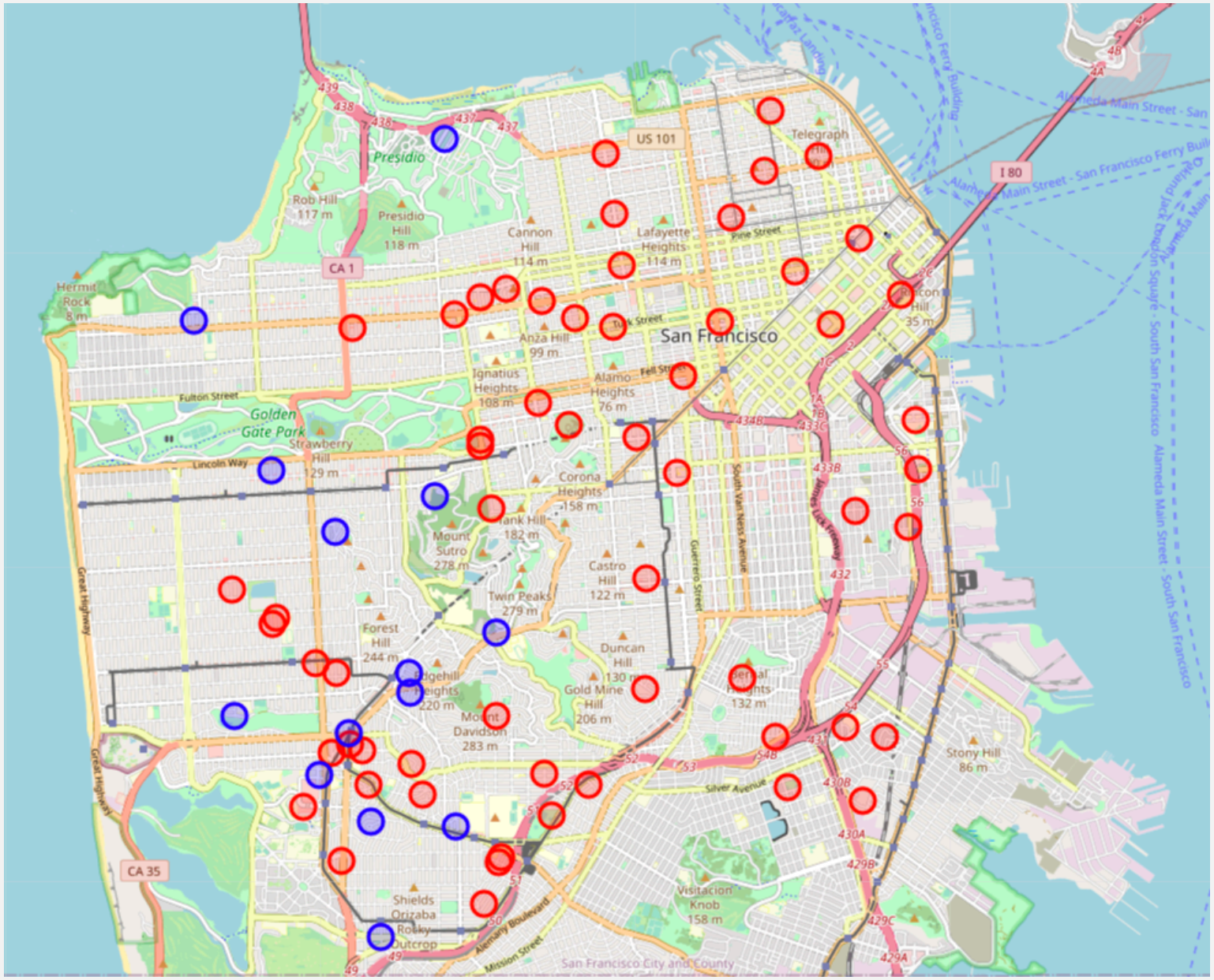


Note: Yellow neighborhoods correspond to those which the Foursquare API couldn't get any data

RESULTS AND DISCUSSION

Clustering

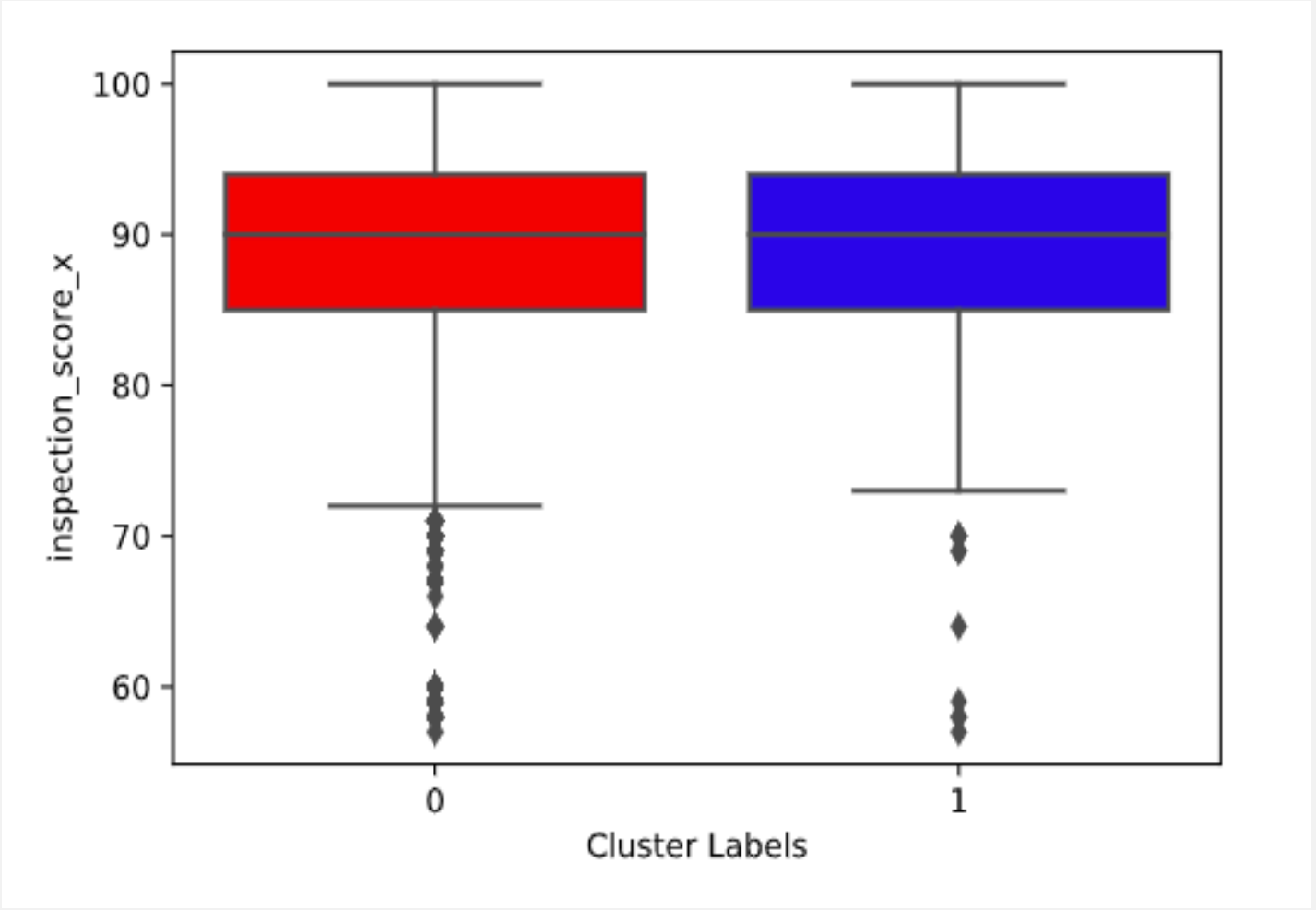
- Two Clusters: high restaurant density areas and lower restaurant density areas.



RESULTS AND DISCUSSION

Cluster's inspection scores comparison

- Almost equal distribution of inspection scores
- Inspection scores not the best indicator to differentiate both clusters



RESULTS AND DISCUSSION

Cluster's other metrics compared:

- Similar means, maximum and minimum values.
- Average number of restaurants per neighborhood differ
- Red cluster has higher percentages of No Risk and Low Risk restaurants and lower percentages of Moderate Risk and High-Risk restaurants

Metric	Red Cluster (0)	Blue Cluster (1)
# of Neighborhoods	59	14
Average # Restaurants per Neighborhood	82	28
Average Inspection Scores	89	89
Max Inspection Score	100	100
Min Inspection Score	57	57
No Risk restaurants (%)	6%	4%
Low Risk restaurants (%)	43%	41%
Moderate Risk restaurants (%)	37%	45%
High Risk restaurants (%)	14%	10%

CONCLUSION

- **Most restaurants in SF are located towards the NE part of the city.**

The financial district and its surroundings have the highest restaurants concentrations. This concentration is reduced as one move S-SW to residential areas, usually near parks.

- **Most restaurants are considered to have either Low or Moderate Risk of infection.**

When going out to eat in SF you should expect mostly Low to Moderate risks of catching COVID19. It's also concerning that there are more High-risk restaurants than No Risk restaurants.

- **When looking at the city as a whole, there is no clear distinction between good and bad areas, you can find everything everywhere**

Probably the most relevant conclusion. In which neighborhood a restaurant is located is not a good indicator of how clean it is.

CONCLUSION

- **A better variable to determine Low or No risk of contamination is how restaurant-dense the neighborhood is.**

You can expect high restaurant density neighborhood to be less risky than lower restaurant density neighborhoods.

- **When choosing where to go out to dinner, choose the neighborhoods in the red cluster.**

Following the last conclusion, neighborhoods with more restaurants around will probably have better hygiene standards.

FUTURE CONSIDERATIONS

- Would be great to have all neighborhoods covered in the Foursquare API
- Comparing restaurant categories (i.e Chinese, American, fast food, etc.), their inspection scores and risk categories could be very insightful
- Using the Foursquare premium API, reviews could be scrapped to find any more suggestions of high/low hygiene standards