# Exploring Food Safety Violations in San Francisco Neighborhoods

## Lenin Mendoza

## May 4th, 2020

## 1. Introduction

### 1.1. Background

Public perceptions about food risks shape personal and household behaviors and create demand for or acceptance of governmenental actions related to food safety (See Box 2-1). Perceptions of foodborne hazards by the public often differ from those of the scientific community. For many years, risks posed by chemicals in food have concerned the consumer more than the expert, while the reverse is true for microbial hazards (Wolf, 1992). Those disparate views seem to arise from the values, needs, and priorities that the different sectors apply as they judge the acceptability of risks. It has been suggested that the criteria used by the public to judge the acceptability of risks include such risk attributes as familiarity, choice (whether the risk is voluntary or imposed), controllability, memorability, dread, immediacy, detectability, and equity of the distribution of risks and benefits. The levels of dread and the degree to which a risk is voluntary culminate in an "outrage" factor, which has been used to predict the acceptability of a risk *(Groth, 1991; NRC, 1989; Sandman, 1987; Scherer, 1991; Slovic, 1986, 1987; Slovic et al., 1979).*

Cultural factors also play a role in public responses to food safety issues. Public acceptance of a risk and demands for protection have been described as related to values and views associated with such things as freedom of choice, government regulation, understanding of technology, credibility of science, preference for homemade or natural foods, and attribution of risks to fate *(Dale and Wildavsky, 1991; Douglas, 1985; Douglas and Wildavsky, 1982; Fitchen, 1987; NRC, 1989).*

### 1.2. Problem

A group of investors is exploring to invest and start a Food Safety consulting firm in the city of San Francisco, CA. In order to make sure the right focus is applied on the launch, marketing and hiring for the firm I have been tasked with exploring data available that could be relevant to assist in this project.

### 1.3 Interest

Data analysis and exploration will be used to answer the following questions and assist on the business processes described in the problem section:

- What are the most common violations in this city?
- Can neighborhoods be clustered based on violation types to ensure effective deployment of marketing and specialists/packages?
- What are the most common food venue types in the neighborhoods?

**2. Data Sources and Cleaning**

**2.1. Data Sources**

**Neighborhood Names and Zipcodes:**

San Francisco names and Neighborhoods will be scraped from the Healthy San Francisco Website. Using the *Beautiful Soup* library in Python.

Source: http://www.healthysf.org/bdi/outcomes/zipmap.htm

**Latitude and Longitude for Neighborhoods:**

*Uszipcode* python library will be used to extract neighborhood locations using their zipcode search engine.

Source: https://pypi.org/project/uszipcode/

**Restaurant Violations:**

The Restaurant Scores dataset available at the Data SF website will be utilized for recorded violations. The dataset contains records from October 2016 to November 2019.

Source: https://data.sfgov.org/Health-and-Social-Services/Restaurant-Scores-LIVES-Standard/pyih-qa8i

**Foursquare API:**

The Foursquare API will be utilized to identify dominant restaurant types in the neighborhoods, helping us get a sense of the market we would be entering.

Source: https://foursquare.com/

**2.2. Data Cleaning**

Actions taken on each dataset:

**Neighborhood Names and Zipcodes:**

- None

**Latitude and Longitude for Neighborhoods:**

- The longitude for the Outer Richmond neighborhood returned by *uszipcode* is incorrect. I have resorted to Google Maps to find the correct one and updated the data frame accordingly.

**Restaurant Violations:**

- Dropped all rows older than January 1st 2019
- Removed columns not relevant to our analysis
- Kept only data for neighborhoods in San Francisco City

**Foursquare API:**

- Filtering was done at the API level, through the Food Category ID (*4d4b7105d754a06374d81259*)

### 3. Exploratory Data Analysis

Data exploration will be used to answer the different business problems. We will cover each of the different topics.

### 3.1. Violation types and frequency

The resulting dataset contains a total of 62 violation types and 11,270 occurrences.

Violations by risk level are:

| Risk | Count | % of Total |
|---|---|---|
| Low | 5,102 | 45.3% |
| Moderate | 4,470 | 39.7% |
| High | 1,698 | 15.0% |
| **Total** | **11,270** | |

Listed below are the Top 15 violations in the period of analysis along with their risk category. This top 15% covers ~72% of total infractions. Developing solutions for this common problems should help us capture the interest of a large portion of the market.

| Description | Risk | Total |
|---|---|---|
| Inadequately cleaned or sanitized food contact surfaces | Moderate Risk | 961 |
| Unapproved or unmaintained equipment or utensils | Low Risk | 896 |
| Unclean or degraded floors walls or ceilings | Low Risk | 847 |
| Moderate risk food holding temperature | Moderate Risk | 766 |
| Inadequate and inaccessible handwashing facilities | Moderate Risk | 760 |
| Foods not protected from contamination | Moderate Risk | 560 |
| Wiping cloths not clean or properly stored or inadequate sanitizer | Low Risk | 526 |
| High risk food holding temperature | High Risk | 447 |
| Moderate risk vermin infestation | Moderate Risk | 443 |
| Improper food storage | Low Risk | 408 |
| Unclean nonfood contact surfaces | Low Risk | 405 |
| Low risk vermin infestation | Low Risk | 359 |
| Unclean or unsanitary food contact surfaces | High Risk | 350 |
| Food safety certificate or food handler card not available | Low Risk | 348 |

### 3.2. Top food venue types by neighborhood

Below most common food venue types by neighborhoods. Varied cuisine is found in San Francisco, Cafés, Italian, Chinese and Mexican Restaurants stand our particularly and should be of interest for marketing purposes.

| Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue |
|---|---|---|---|---|
| Bayview-Hunters Point | Bakery | Food | Vietnamese Restaurant | Fried Chicken Joint |
| Castro/Noe Valley | Deli / Bodega | Mediterranean Restaurant | Thai Restaurant | Pizza Place |
| Chinatown | Café | Chinese Restaurant | Sushi Restaurant | Italian Restaurant |
| Haight-Ashbury | Bakery | Café | Indian Restaurant | Burrito Place |
| Hayes Valley/Tenderloin/North of Market | Sandwich Place | Vietnamese Restaurant | Thai Restaurant | Deli / Bodega |
| Ingelside-Excelsior/Crocker-Amazon | Pizza Place | Chinese Restaurant | Mexican Restaurant | Vietnamese Restaurant |
| Inner Mission/Bernal Heights | Mexican Restaurant | Deli / Bodega | Italian Restaurant | Café |
| Inner Richmond | Sushi Restaurant | Vietnamese Restaurant | Pizza Place | Japanese Restaurant |
| Lake Merced | Café | Food Truck | Sandwich Place | Pizza Place |
| Marina | Italian Restaurant | French Restaurant | American Restaurant | Thai Restaurant |
| North Beach/Chinatown | Italian Restaurant | French Restaurant | American Restaurant | Thai Restaurant |
| Outer Richmond | Café | Sushi Restaurant | Chinese Restaurant | Pizza Place |
| Parkside/Forest Hill | Chinese Restaurant | Café | Pizza Place | Sandwich Place |
| Polk/Russian Hill (Nob Hill) | Mexican Restaurant | Sushi Restaurant | Diner | Italian Restaurant |
| Potrero Hill | Food Truck | Café | Pizza Place | Sandwich Place |
| South of Market | Café | Sandwich Place | Food Truck | Restaurant |
| Sunset | Chinese Restaurant | Bakery | Vietnamese Restaurant | Szechuan Restaurant |
| Twin Peaks-Glen Park | Food | Mexican Restaurant | American Restaurant | Thai Restaurant |
| Visitacion Valley/Sunnydale | American Restaurant | Vietnamese Restaurant | Fried Chicken Joint | Ethiopian Restaurant |
| Western Addition/Japantown | Mexican Restaurant | Chinese Restaurant | Sushi Restaurant | Snack Place |

### 3.3. Neighborhood Clustering

Using the *k-means* clustering algorithm it was possible to cluster all SF neighborhoods based on violations. See defining characteristics by cluster below:

**Cluster 1:**

| | Neighborhood | 1st Most Common Violation | 2nd Most Common Violation | 3rd Most Common Violation | 4th Most Common Violation | 5th Most Common Violation | 6th Most Common Violation | 7th Most Common Violation | 8th Most Common Violation | 9th Most Common Violation | 10th Most Common Violation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | Castro/Noe Valley | Wiping cloths not clean or properly stored or ... | Unclean or degraded floors walls or ceilings | Inadequately cleaned or sanitized food contact... | Moderate risk food holding temperature | Inadequate and inaccessible handwashing facili... | Foods not protected from contamination | Unapproved or unmaintained equipment or utensils | Moderate risk vermin infestation | Low risk vermin infestation | High risk food holding temperature |
| 16 | St. Francis Wood/Miraloma/West Portal | Foods not protected from contamination | Unclean or degraded floors walls or ceilings | Moderate risk food holding temperature | Inadequately cleaned or sanitized food contact... | Unapproved or unmaintained equipment or utensils | Wiping cloths not clean or properly stored or ... | Improper thawing methods | Unclean nonfood contact surfaces | Inadequate and inaccessible handwashing facili... | Low risk vermin infestation |
| 17 | Twin Peaks-Glen Park | Foods not protected from contamination | Wiping cloths not clean or properly stored or ... | Unclean or degraded floors walls or ceilings | Moderate risk food holding temperature | Unapproved or unmaintained equipment or utensils | Food safety certificate or food handler card n... | Low risk vermin infestation | Unclean or unsanitary food contact surfaces | Inadequately cleaned or sanitized food contact... | Improper thawing methods |
| 18 | Lake Merced | Wiping cloths not clean or properly stored or ... | Inadequate and inaccessible handwashing facili... | Foods not protected from contamination | Unclean or degraded floors walls or ceilings | Permit license or inspection report not posted | Inadequately cleaned or sanitized food contact... | Moderate risk food holding temperature | Improper thawing methods | Improper or defective plumbing | Unclean or unsanitary food contact surfaces |

**Cluster 2:**

| | Neighborhood | 1st Most Common Violation | 2nd Most Common Violation | 3rd Most Common Violation | 4th Most Common Violation | 5th Most Common Violation | 6th Most Common Violation | 7th Most Common Violation | 8th Most Common Violation | 9th Most Common Violation | 10th Most Common Violation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | Castro/Noe Valley | Wiping cloths not clean or properly stored or ... | Unclean or degraded floors walls or ceilings | Inadequately cleaned or sanitized food contact... | Moderate risk food holding temperature | Inadequate and inaccessible handwashing facili... | Foods not protected from contamination | Unapproved or unmaintained equipment or utensils | Moderate risk vermin infestation | Low risk vermin infestation | High risk food holding temperature |
| 16 | St. Francis Wood/Miraloma/West Portal | Foods not protected from contamination | Unclean or degraded floors walls or ceilings | Moderate risk food holding temperature | Inadequately cleaned or sanitized food contact... | Unapproved or unmaintained equipment or utensils | Wiping cloths not clean or properly stored or ... | Improper thawing methods | Unclean nonfood contact surfaces | Inadequate and inaccessible handwashing facili... | Low risk vermin infestation |
| 17 | Twin Peaks-Glen Park | Foods not protected from contamination | Wiping cloths not clean or properly stored or ... | Unclean or degraded floors walls or ceilings | Moderate risk food holding temperature | Unapproved or unmaintained equipment or utensils | Food safety certificate or food handler card n... | Low risk vermin infestation | Unclean or unsanitary food contact surfaces | Inadequately cleaned or sanitized food contact... | Improper thawing methods |
| 18 | Lake Merced | Wiping cloths not clean or properly stored or ... | Inadequate and inaccessible handwashing facili... | Foods not protected from contamination | Unclean or degraded floors walls or ceilings | Permit license or inspection report not posted | Inadequately cleaned or sanitized food contact... | Moderate risk food holding temperature | Improper thawing methods | Improper or defective plumbing | Unclean or unsanitary food contact surfaces |

## Cluster 3:

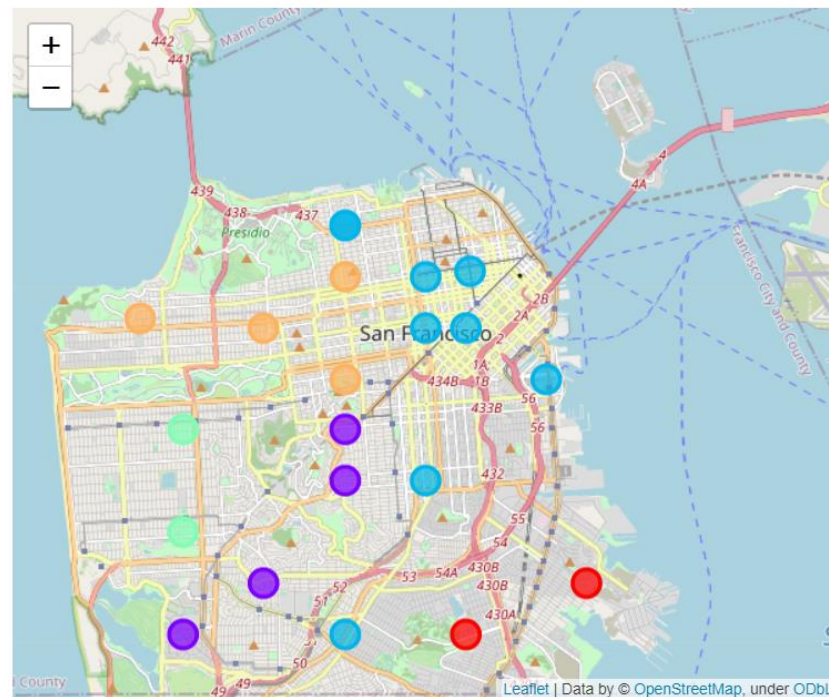| | Neighborhood | 1st Most Common Violation | 2nd Most Common Violation | 3rd Most Common Violation | 4th Most Common Violation | 5th Most Common Violation | 6th Most Common Violation | 7th Most Common Violation | 8th Most Common Violation | 9th Most Common Violation | 10th Most Common Violation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Hayes Valley/Tenderloin/North of Market | Inadequate and inaccessible handwashing facili... | Moderate risk food holding temperature | Inadequately cleaned or sanitized food contact... | Unapproved or unmaintained equipment or utensils | Unclean or degraded floors walls or ceilings | High risk food holding temperature | Wiping cloths not clean or properly stored or ... | Unclean nonfood contact surfaces | Improper food storage | Improper cooling methods |
| 1 | South of Market | Unclean or degraded floors walls or ceilings | Inadequately cleaned or sanitized food contact... | Moderate risk food holding temperature | Unapproved or unmaintained equipment or utensils | Inadequate and inaccessible handwashing facili... | High risk food holding temperature | Foods not protected from contamination | Moderate risk vermin infestation | Wiping cloths not clean or properly stored or ... | Food safety certificate or food handler card n... |
| 2 | Potrero Hill | Moderate risk food holding temperature | Inadequately cleaned or sanitized food contact... | Unapproved or unmaintained equipment or utensils | Inadequate and inaccessible handwashing facili... | Unclean or degraded floors walls or ceilings | Foods not protected from contamination | Food safety certificate or food handler card n... | Wiping cloths not clean or properly stored or ... | High risk food holding temperature | Inadequate food safety knowledge or lack of ce... |
| 3 | Chinatown | Inadequately cleaned or sanitized food contact... | Inadequate and inaccessible handwashing facili... | Unapproved or unmaintained equipment or utensils | Moderate risk food holding temperature | Improper food storage | Unclean or degraded floors walls or ceilings | High risk food holding temperature | Wiping cloths not clean or properly stored or ... | Foods not protected from contamination | Low risk vermin infestation |
| 4 | Polk/Russian Hill (Nob Hill) | Unapproved or unmaintained equipment or utensils | Inadequate and inaccessible handwashing facili... | Inadequately cleaned or sanitized food contact... | Moderate risk food holding temperature | Wiping cloths not clean or properly stored or ... | Unclean or degraded floors walls or ceilings | Foods not protected from contamination | Improper food storage | Unclean nonfood contact surfaces | High risk food holding temperature |
| 5 | Inner Mission/Bernal Heights | Inadequately cleaned or sanitized food contact... | Unclean or degraded floors walls or ceilings | Moderate risk food holding temperature | Unapproved or unmaintained equipment or utensils | Unclean nonfood contact surfaces | Moderate risk vermin infestation | Inadequate and inaccessible handwashing facili... | High risk food holding temperature | Foods not protected from contamination | Inadequate food safety knowledge or lack of ce... |
| 6 | Ingelside-Excelsior/Crocker-Amazon | Inadequately cleaned or sanitized food contact... | Unapproved or unmaintained equipment or utensils | Unclean or degraded floors walls or ceilings | Inadequate and inaccessible handwashing facili... | High risk food holding temperature | Moderate risk food holding temperature | Improper food storage | Foods not protected from contamination | Inadequate food safety knowledge or lack of ce... | Unclean hands or improper use of gloves |
| 14 | Marina | Food safety certificate or food handler card n... | Unapproved or unmaintained equipment or utensils | Unclean or degraded floors walls or ceilings | Inadequately cleaned or sanitized food contact... | Wiping cloths not clean or properly stored or ... | Inadequate and inaccessible handwashing facili... | High risk vermin infestation | Unclean nonfood contact surfaces | High risk food holding temperature | Foods not protected from contamination |
| 19 | North Beach/Chinatown | Inadequately cleaned or sanitized food contact... | Inadequate and inaccessible handwashing facili... | Unapproved or unmaintained equipment or utensils | Moderate risk food holding temperature | Unclean or degraded floors walls or ceilings | Improper storage of equipment utensils or linens | Foods not protected from contamination | High risk food holding temperature | Low risk vermin infestation | Improper food storage |

## Cluster 4:

| | Neighborhood | 1st Most Common Violation | 2nd Most Common Violation | 3rd Most Common Violation | 4th Most Common Violation | 5th Most Common Violation | 6th Most Common Violation | 7th Most Common Violation | 8th Most Common Violation | 9th Most Common Violation | 10th Most Common Violation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | Parkside/Forest Hill | Unclean or degraded floors walls or ceilings | Foods not protected from contamination | Inadequate and inaccessible handwashing facili... | Inadequately cleaned or sanitized food contact... | Improper food storage | Moderate risk vermin infestation | High risk vermin infestation | Unclean nonfood contact surfaces | Unapproved or unmaintained equipment or utensils | Moderate risk food holding temperature |
| 13 | Sunset | Unclean or degraded floors walls or ceilings | Foods not protected from contamination | Inadequate and inaccessible handwashing facili... | Moderate risk food holding temperature | Wiping cloths not clean or properly stored or ... | Inadequately cleaned or sanitized food contact... | Improper food storage | High risk food holding temperature | High risk vermin infestation | Moderate risk vermin infestation |

**Cluster 5:**

| | Neighborhood | 1st Most Common Violation | 2nd Most Common Violation | 3rd Most Common Violation | 4th Most Common Violation | 5th Most Common Violation | 6th Most Common Violation | 7th Most Common Violation | 8th Most Common Violation | 9th Most Common Violation | 10th Most Common Violation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | Western Addition/Japantown | Unapproved or unmaintained equipment or utensils | Inadequately cleaned or sanitized food contact... | Moderate risk vermin infestation | Unclean or degraded floors walls or ceilings | Inadequate and inaccessible handwashing facili... | Wiping cloths not clean or properly stored or ... | Moderate risk food holding temperature | Low risk vermin infestation | Foods not protected from contamination | Improper food storage |
| 10 | Haight-Ashbury | Unapproved or unmaintained equipment or utensils | Unclean or degraded floors walls or ceilings | Moderate risk food holding temperature | Moderate risk vermin infestation | Inadequately cleaned or sanitized food contact... | Inadequate and inaccessible handwashing facili... | Unclean or unsanitary food contact surfaces | Food safety certificate or food handler card n... | Foods not protected from contamination | Improper storage of equipment utensils or linens |
| 11 | Inner Richmond | Unapproved or unmaintained equipment or utensils | Inadequately cleaned or sanitized food contact... | Unclean or unsanitary food contact surfaces | Inadequate and inaccessible handwashing facili... | Moderate risk food holding temperature | Moderate risk vermin infestation | Unclean or degraded floors walls or ceilings | Improper food storage | Foods not protected from contamination | Improper thawing methods |
| 12 | Outer Richmond | Inadequately cleaned or sanitized food contact... | Unapproved or unmaintained equipment or utensils | Inadequate and inaccessible handwashing facili... | Moderate risk food holding temperature | Improper thawing methods | Wiping cloths not clean or properly stored or ... | Improper food storage | Unclean or unsanitary food contact surfaces | Moderate risk vermin infestation | Low risk vermin infestation |

After performing the algorithm we proceeded to plot the clusters in the map. Interestingly enough the clusters align geographically which should help us tailor marketin and pricing strategies in an effective manner:

4. **Conclusions**
   - Identifying most common violations in the city has been possible with the dataset available. Overall business strategy should be planned around the most common ocurrences. This data should be used to drive the marketing strategy along with areas of focus when hiring specialists to be deployed to the different neighborhoods
   - Neighborhoods have been clustered based on most common violations and there seems to be a geographical alignment between them (worth digging deeper into possible causes). In any case these geographical alignment can help us tailor the marketing strategy in these neighborhoods in an effective manner
   - Varied cuisine is found in San Francisco. Most common venues are Cafés, Italian, Chinese and Mexican Restaurants. Tailoring our offering to these common venues can help us increase market share

5. **Additional considerations**
   - COVID19 has impacted all industries and restaurants are not the exception, in fact, they have been particularly impacted. Special consideration to the new business landscape needs to be taken, understanding possible regulatory changes as well as financial struggles from restaurant owners which can decrease the appetite for specialized services
   - Despite the fact filtering has been made to select 'Food' places in the Foursquare API. This API is not necessarily complete as not all restaurants will be registered and additionally other places of interest (e.g. School Cafeterias) will not be returned. Therefore, further market research can be done around other venue types. Out of scope for this exercise
   **Additional considerations**

6. **Link to Code:**
   The code used to perform this analysis can be found at :
   https://www.kaggle.com/madoxxx/exploring-food-safety-violations-in-san-francisco