SF Food Inspection Use Case

By: Aman Singh

Introduction

- Food inspection involves not only sampling and testing of end products but also assessing food centers to ensure compliance with food safety management systems. This minimizes the occurrence of public health food safety problems. Food inspection dates back to ancient times as part of the history of public health. The Food and Drug Administration (FDA) publishes the Food Code that sets guidelines and procedures to assist in food control jurisdictions. The Food Code provides a scientifically and legally backed basis for regulating the retail and food service industries.

Objectives

- The food inspectors can then have a “Critical first” inspection approach where the places that have been predicted to have critical violations are inspected first.
- Some of the factors that tend to predict critical violation include previous critical violations, high temperatures, nearby sanitation complains, nearby burglaries etc
- This report would be beneficial to public health specialists and every stakeholder working to alleviate public health concerns through preventive measures.
- It is not to introduce food inspection since these professionals are already carrying out food inspections in the relevant jurisdictions but to make the process more efficient.

Data Description

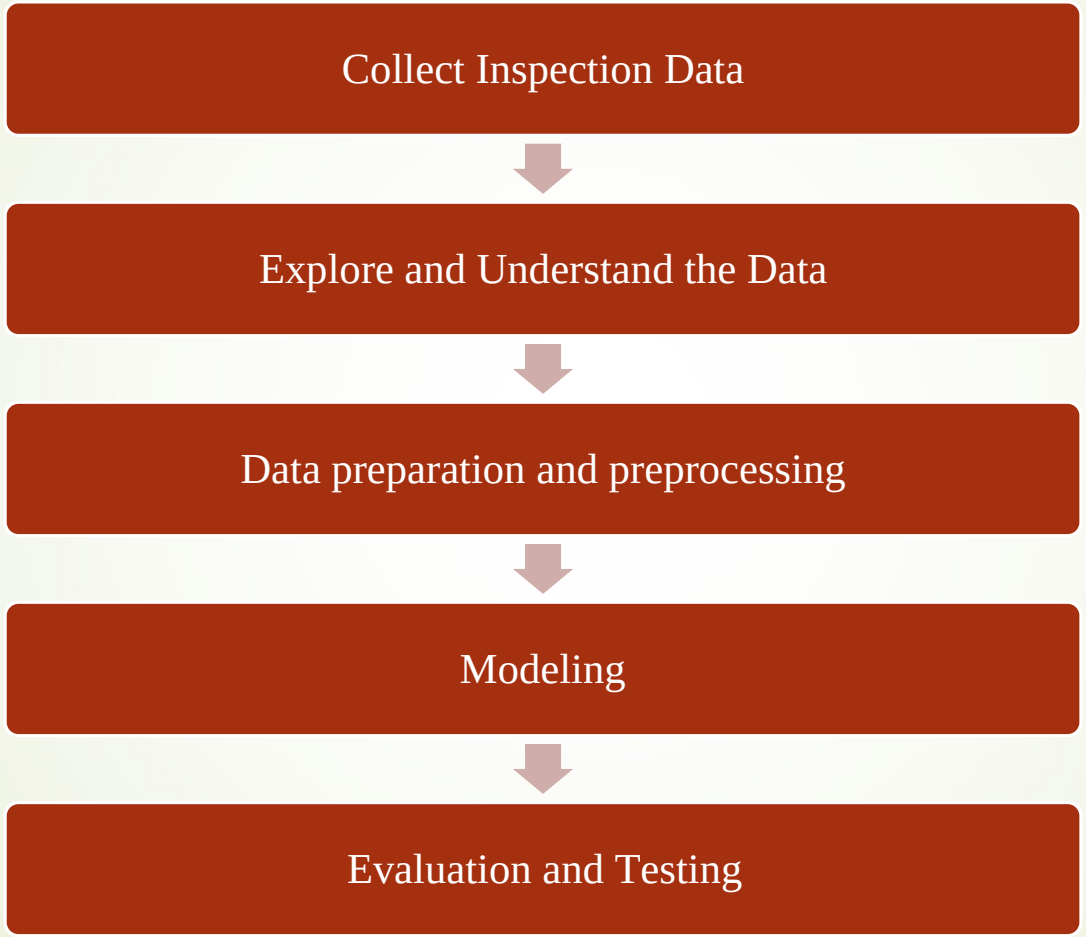
- After conducting an inspection of the facility, the Health Inspector calculates a score based on the violations observed. Violations can fall into:
- **High risk category**: records specific violations that directly relate to the transmission of food borne illnesses, the adulteration of food products and the contamination of food-contact surfaces.
- **Moderate risk category:** records specific violations that are of a moderate risk to the public health and safety.
- **Low risk category:** records violations that are low risk or have no immediate risk to the public health and safety.

Data Description

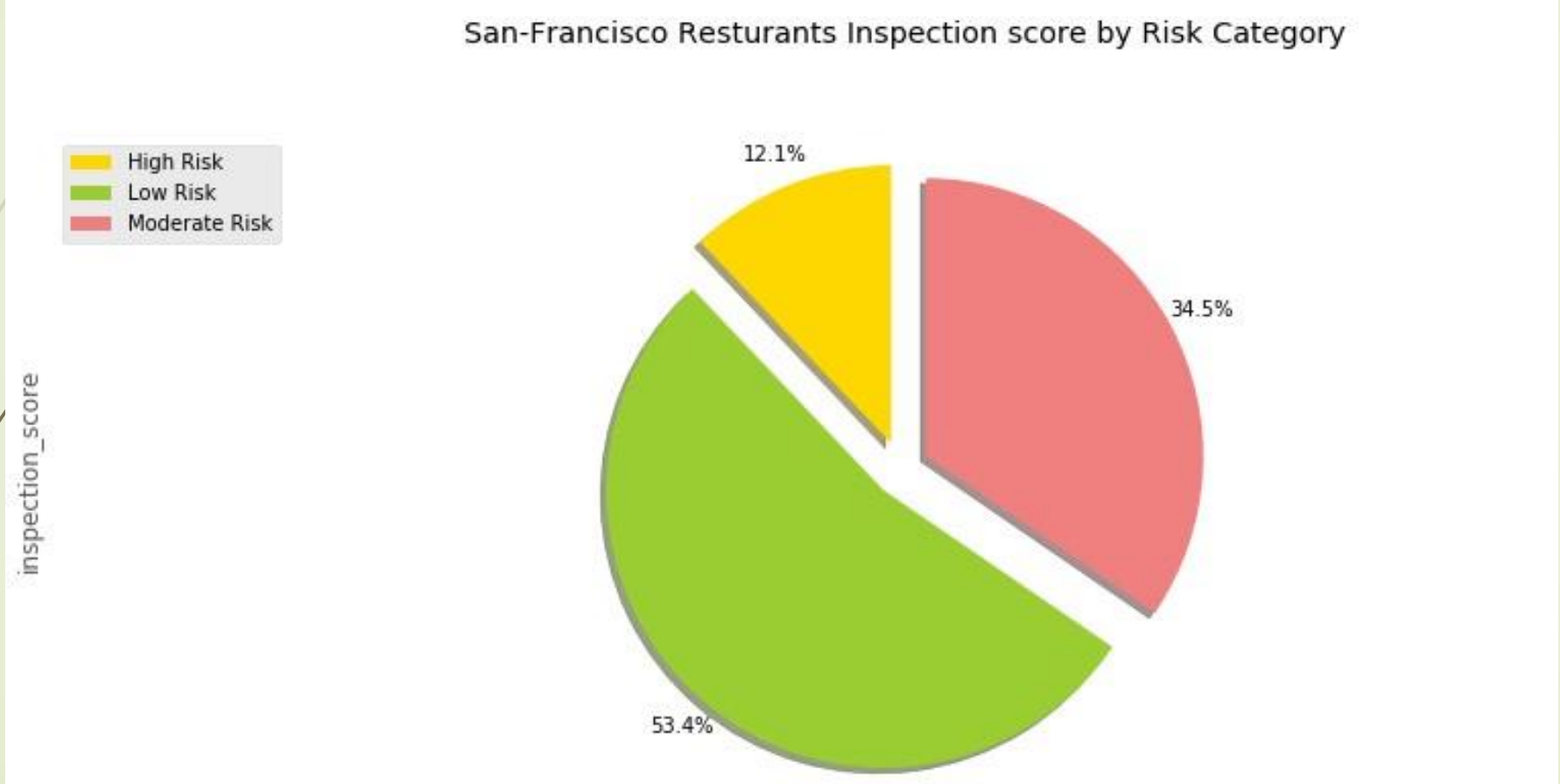
#	Feature Name	Description

1	business_id	Unique number used for identification of the business
2	business_name	Business Name
3	business_address	The address of the business
4	business_city	The City (here all records have the same city San-Francisco)
5	business_state	The state (here all records have the same state CA)
6	business_postal_code	Zip/postal code of the business
7	business_latitude	The latitude value of the business location
8	business_longitude	The longitude value of the business location
9	business_location	A tuple of the latitude and the longitude values
10	business_phone_no	Business phone number
11	inspection_id	Unique number that identifying the inspection case
12	inspection_date	The date of the inspection process
13	inspection_score	A score out of 100 that the business got after the inspection
14	inspection_type	Routine-Unscheduled, complaint, New ownership, new construction or Non-inspection site visit. In our dataset this feature has only one value “Routine-Unscheduled”
15	violation_id	Identification of violation
16	violation_description	Short description of the violation if any
17	risk_category	Classification of the business category, Low, Moderate or High Risk

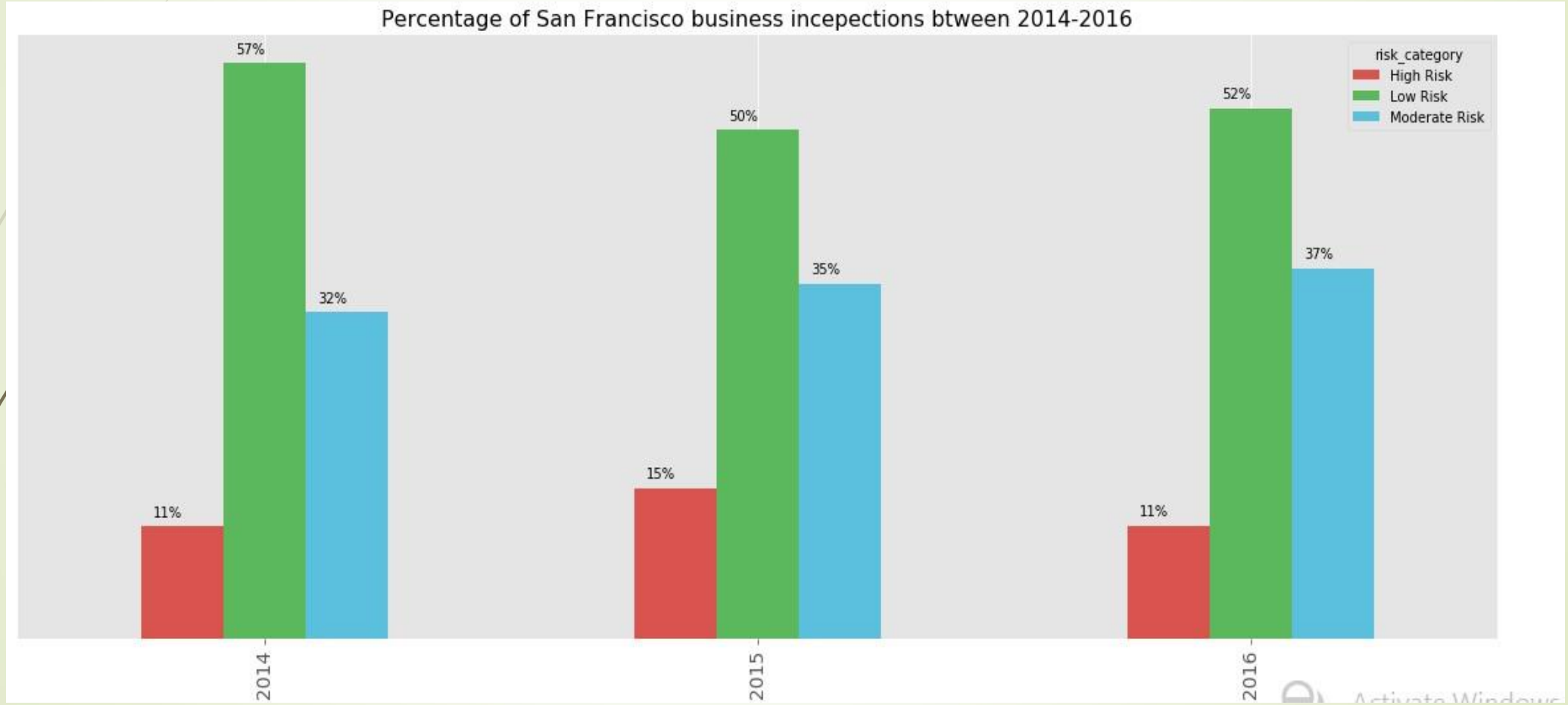
Methodology



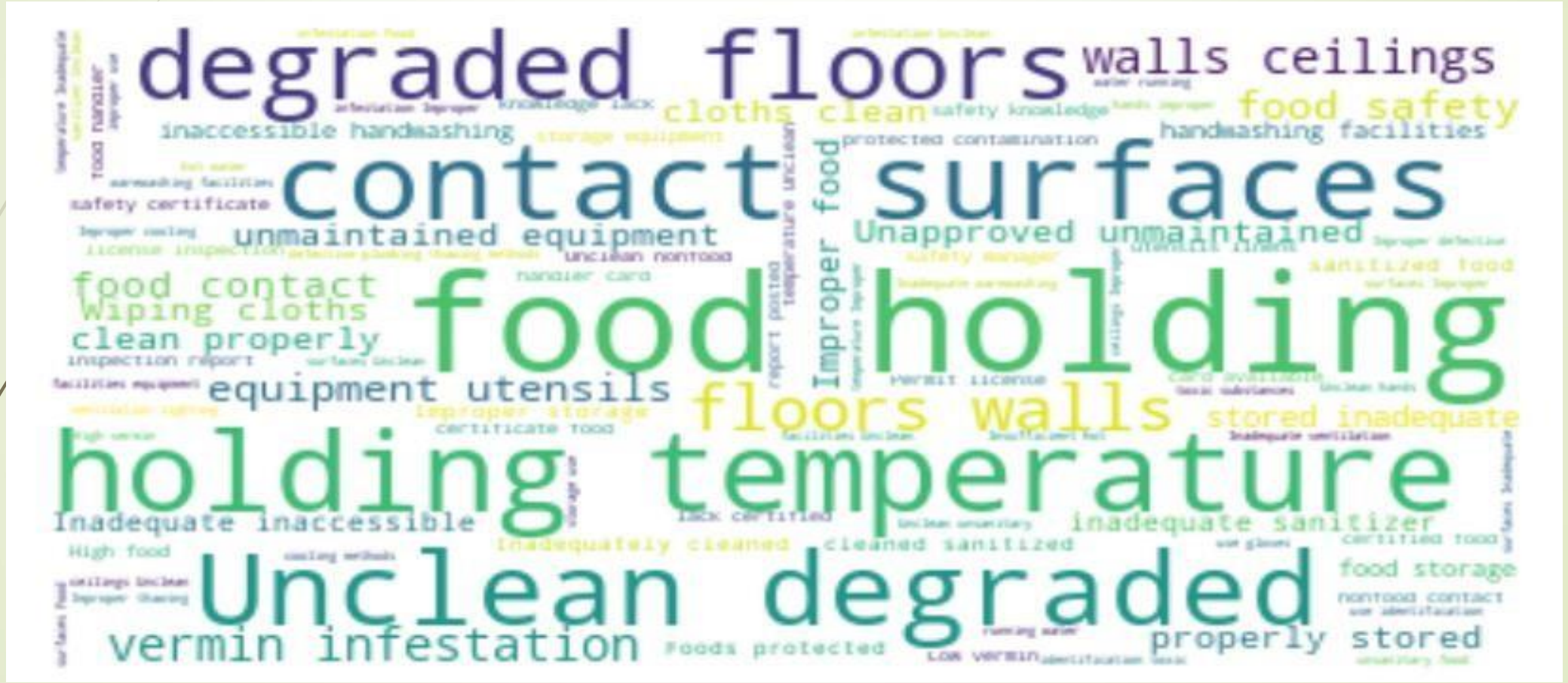
Results: **Inspections In general**



Results: **Inspections by year**

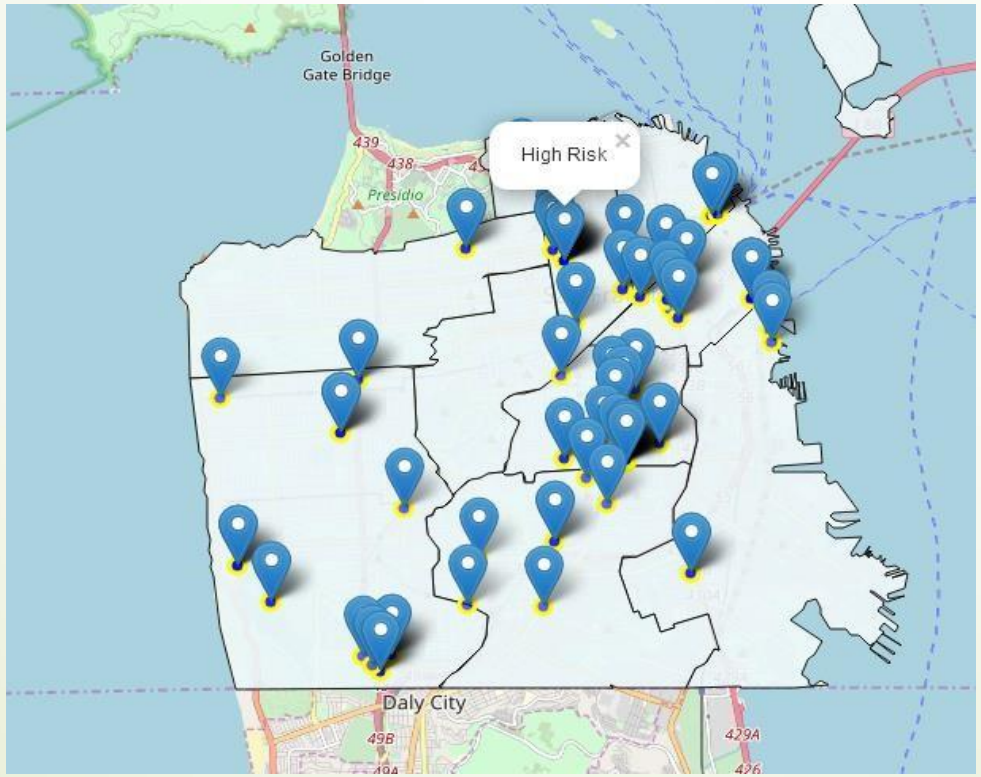


Violation description cloud words

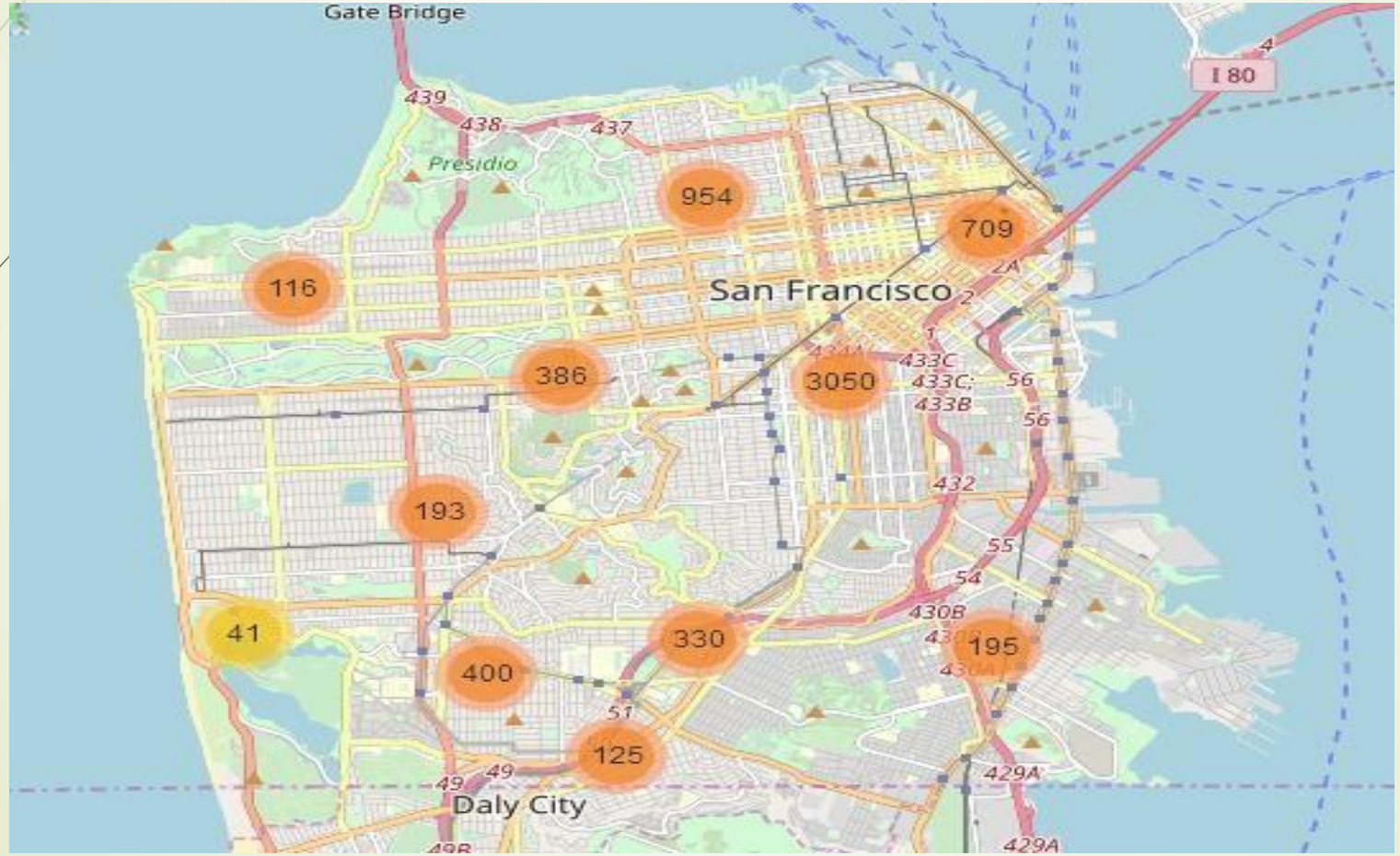


visualize the locations of the inspections

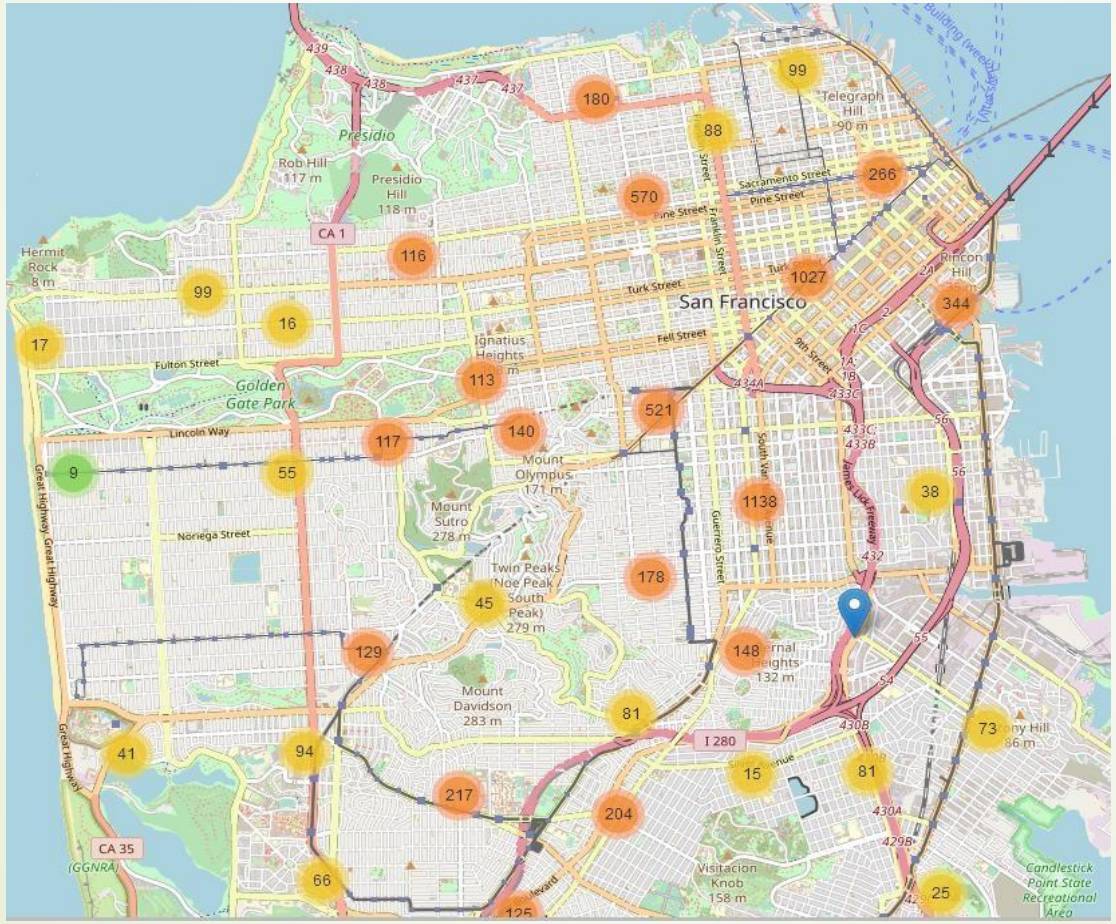
- **<u>In order to reduce computational cost, let's just work with the first 100 inspections in this dataset</u>**.



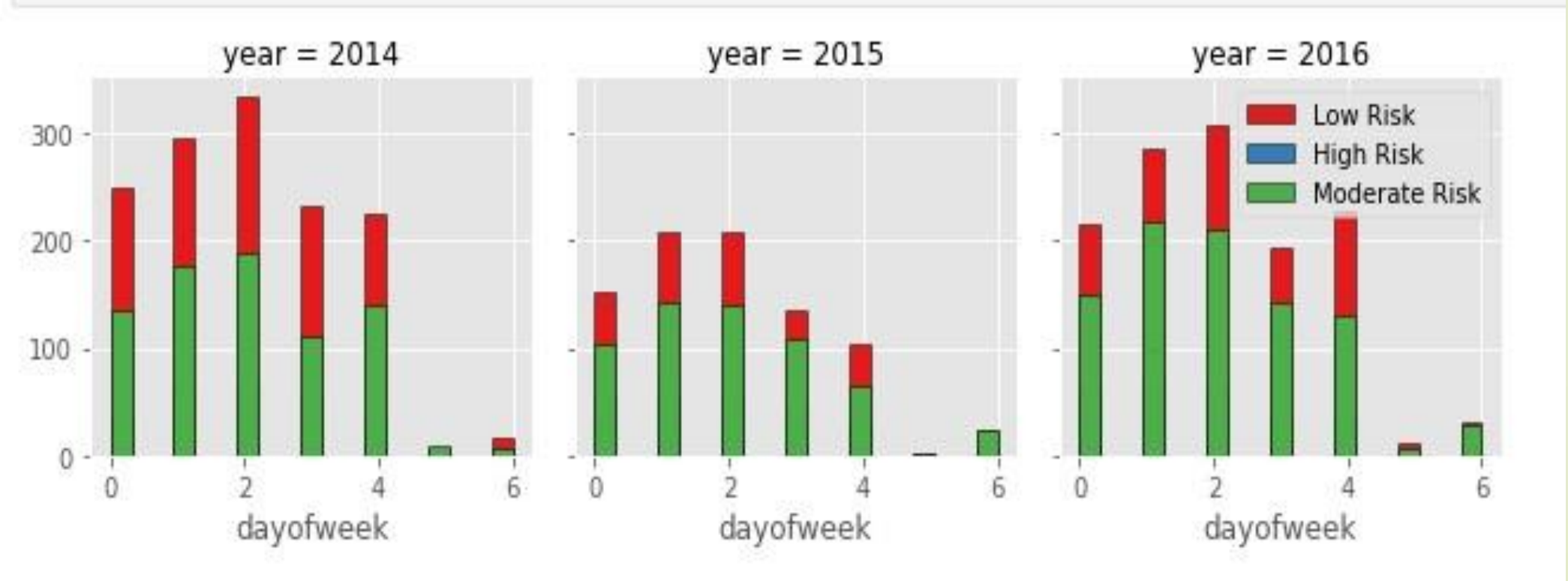
A clean and categorized copy of the map of San Francisco



A clean and categorized copy of the map of San Francisco (more details)



Inspection activities days of the week



Machine learning Algorithm results

	kNN	LR
Train set Accuracy	0.6332499518953242	0.5358860881277661
Test set Accuracy	0.5215384615384615	0.5246153846153846

F1 Accuracy	0.47777033142713926	0.36103702553753003

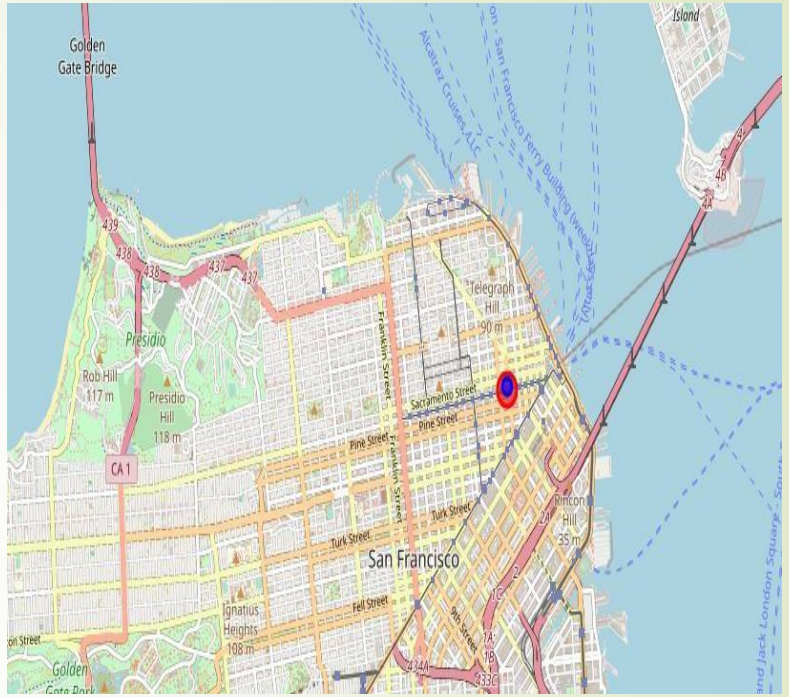
using Foursquare to analyze the neighborhood of the inspected businesses

- **In order to reduce computational cost, let's just work with the first 100 inspections in this dataset**

We choose "OMNI S.F. Hotel" it is Low Risk business inspection and it got 96 score

"OMNI S.F. Hotel" Venues data and location

```
[164]: {'meta': {'code': 200, 'requestId': '5be6fc239fb6b71ed1998912'},
        'response': {'venues': [{'id': '4a5ae9a1f964a520e0ba1fe3',
           'name': 'Omni San Francisco Hotel',
           'location': {'address': '500 California St',
            'crossStreet': 'at Montgomery St',
            'lat': 37.793119745957455,
            'lng': -122.4031025916338,
            'labeledLatLngs': [{'label': 'display',
              'lat': 37.793119745957455,
              'lng': -122.4031025916338}],
            'distance': 25,
            'postalCode': '94104',
            'cc': 'US',
            'city': 'San Francisco',
            'state': 'CA',
            'country': 'United States',
            'formattedAddress': ['500 California St (at Montgomery St)',
             'San Francisco, CA 94104',
             'United States']},
           'categories': [{'id': '4bf58dd8d48988d1fa931735',
             'name': 'Hotel',
             'pluralName': 'Hotels',
             'shortName': 'Hotel',
             'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/travel/hotel_',
              'suffix': '.png'},
             'primary': True}],
           'venuePage': {'id': '497545373'},
           'referralId': 'v-1541864483',
           'hasPerk': False}]}}
```



Foursquare API gives us we are able to only show one tip

- Example of one tip

```
[72]: {'meta': {'code': 200, 'requestId': '5be5bae71ed21905134bb134'},
       'response': {'tips': {'count': 46,
         'items': [{'id': '563f8028cd10133b6c53ec8c',
           'createdAt': 1447002152,
           'text': "Great hotel and they offer some exceptional last minute deals, so if you
      ery comfortable rooms",
           'type': 'user',
           'canonicalUrl': 'https://foursquare.com/item/563f8028cd10133b6c53ec8c',
           'photo': {'id': '563f8028cd10133b6c53ec8b',
            'createdAt': 1447002152,
            'source': {'name': 'Swarm for iOS', 'url': 'https://www.swarmapp.com'},
            'prefix': 'https://fastly.4sqi.net/img/general/',
            'suffix': '/2955249_JI2_JdY1_0Cernm9kZGJWofKLBnebpcTxe3d4E0g8rY.jpg',
            'width': 1440,
            'height': 1920,
            'visibility': 'public'},
           'photourl': 'https://fastly.4sqi.net/img/general/original/2955249_JI2_JdY1_0Cernm
           'lang': 'en',
           'likes': {'count': 0, 'groups': []},
           'logView': True,
           'agreeCount': 1,
           'disagreeCount': 0,
           'todo': {'count': 0},
           'user': {'id': '2955249',
            'firstName': 'Nadia',
            'lastName': 'IssaBella',
            'gender': 'female',
            'photo': {'prefix': 'https://fastly.4sqi.net/img/user/',
             'suffix': '/2955249-LEHE5CUVGIFEF4GG.jpg'}},
           'authorInteractionType': 'liked'}]}}}
```

Foursquare User

- We also examine the user who made that tip and we found that:
- She is female
- Here first name is Nadia and Last Name: IssaBella, Home City: Vaughan, Canada
- Nadia is very active in Foursquare as we can see she has 598 tips. Let us explore them.

Conclusion

- To promote health, stakeholders in the healthcare industry need to continuously innovate to make the process more efficient. In food inspection, technology can be used to predict a likely critical violation through the use of data analytics instead of inspecting every joint blindly

given the lack of enough manpower for this. The data used to predict critical violation include weather, crime and inspection data. Afterward, places data e.g. Foursquare is used to locate the food establishment for physical inspection