IE6600_Sec05_Group_16_Hackathon

Akilesh Vishnu Mohan Raj & Ravi Patel

2022-10-16

Contents :

1 Problem Statement

2 About the Dataset

3 Output

```
    3.1 Total Number of Restaurants in San Francisco Analysis
    3.2 Risk Categories Analysis
    3.3 Geographical  Mapping
```

4 Conclusion

5 References

1 : Problem Statement

The San Francisco Department of Public Health (SFDPH) is composed of various subdivisions that work together to serve the city of San Francisco. SFDPH primary aim is to be protecting, prioritizing, and promoting the health of all the citizens in San Francisco.

A report and scoring system for inspections have been created by the Health Department. The Health Inspector assigns a score to the facility based on the violations discovered during the inspection. Examples of violations include group with a high risk: records specific violations related to the contamination of surfaces where food encounters it, the adulteration of food products, and the spread of food-borne illnesses. records specific violations that pose a moderat e threat to the health and safety of the general public. low risk category: records violations that pose no immediate danger to the public's health or safety or have a low risk. This dataset contains the score card that the inspector will issue, which is kept at the food establishment and is accessible to the general public.

We find this information of great value for making us understanding and underlying issues and factors behind the inspection and their risk categories the SFDPH gets. It Is known that the department of public health have a lot of inspections regarding the complaints, change in ownership and much more in restaurants that is why understanding how the inspections are making the restaurant in future, are they having more inspections in the future also.

Restaurant Scores - LIVES Standard The Health Department has developed an inspection report and scoring system. 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 tothe 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.The score card that will be issued by the inspector is maintained at the food establishment and is available to the public in this dataset.

San Francisco's LIVES restaurant inspection data leverages the LIVES Flattened Schema (https://goo.gl/c3nNvr (https://goo.gl/c3nNvr)) which is based on LIVES version 2.0, cited on Yelp's website (http://www.yelp.com/healthscores (http://www.yelp.com/healthscores)).

---

2 : About the Dataset

Restaurant Scores dataset contains all the restaurants, their name, their location, if they had inspections, their risk categories after inspection and much more. The information here is collected by the San Francisco Department of Public Health.

Some of the variables in the dataset are as follows:

• Business name, id, and location
• Inspection date, id, and type • Violation id, and description • Risk category…etc

Source: https://data.sfgov.org/Health-and-Social-Services/Restaurant-Scores-LIVES-Standard/pyih-(https://data.sfgov.org/Health-and-Social-Services/Restaurant-Scores-LIVES-Standard/pyih-) qa8i/data(Found via SFO Open Data)

---

3 : Output

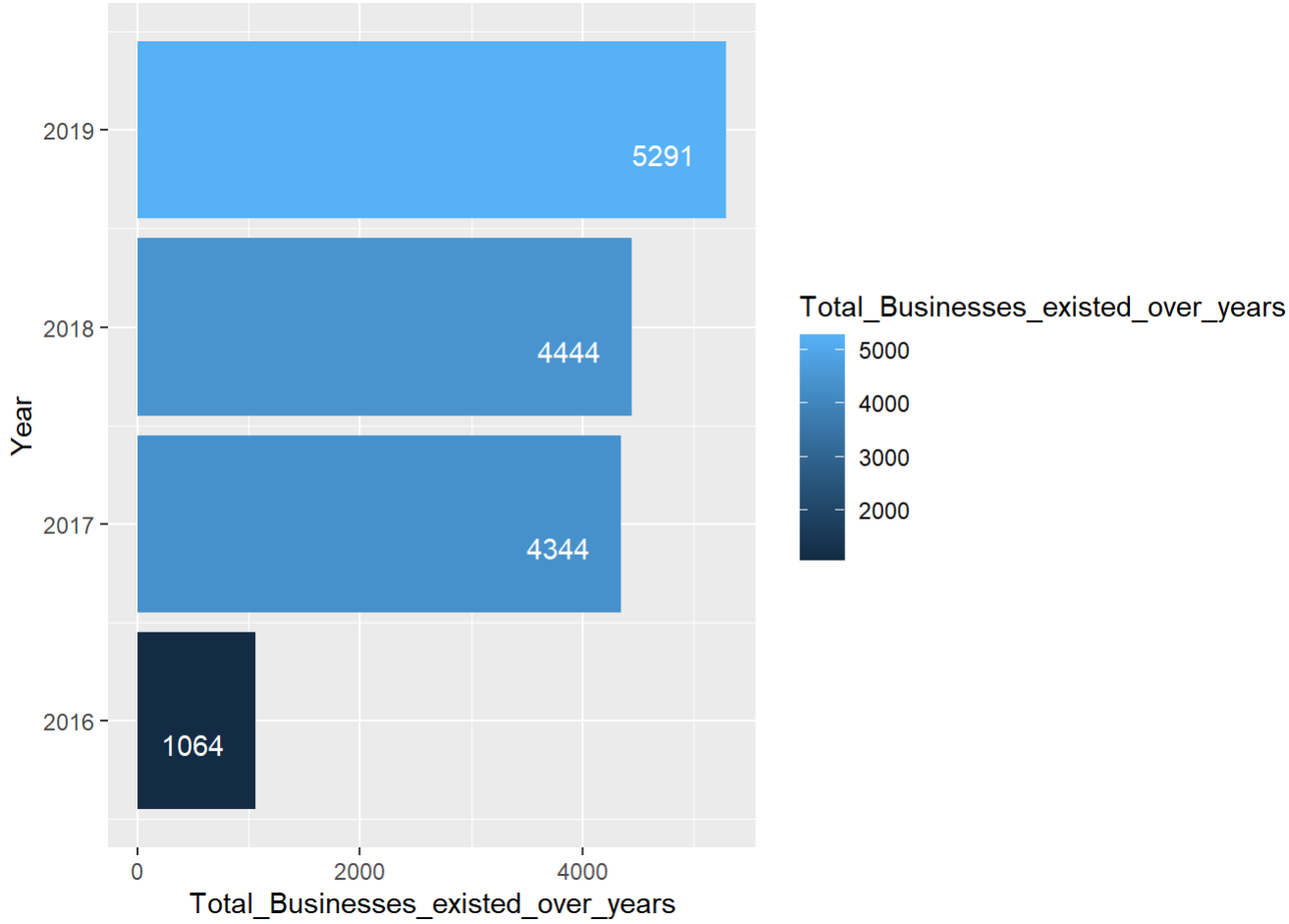
Section 3.1 Total Number of Restaurents inSan Francisco:-

Analysing the DataSet with the Total Number of Restaurents in San Francisco and the number of Violations over the Years

Plot 1

Firstly Let us Observe the total number of business that existed in the State of California, more in accordance with the city San Francisco as available in the San Francisco Department of Public Health (SFDPH) is composed of various subdivisions that work together to serve the city of San Francisco.

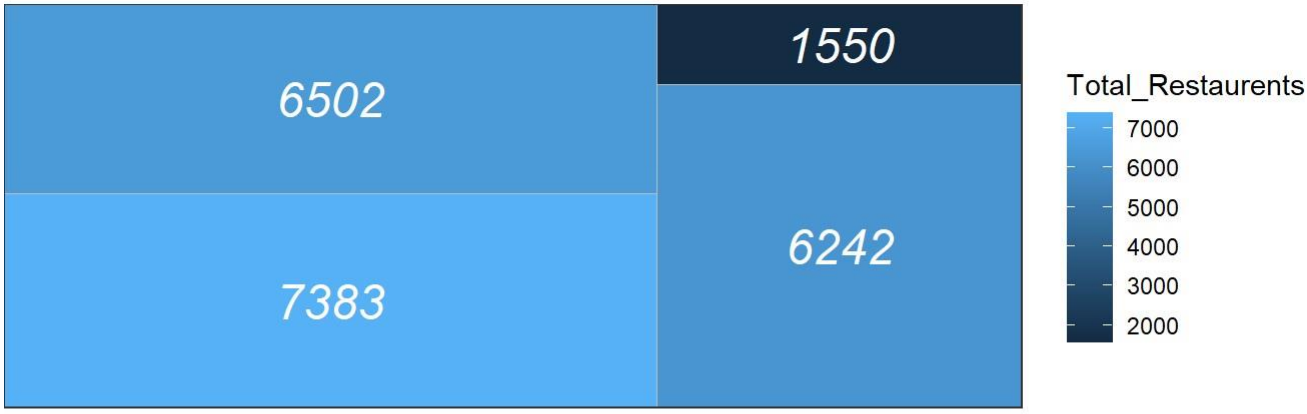
Thus the below Bar plot says the total number of Restaurants that is present in the city of San Francisco during the years 2016, 2017, 2018, 2019. Now this is a very important data point to analyse as only after getting this data we will be able to do further analysis on the no of restaurants that have violated the standards and the risk categories that it has been awarded for the respective years.



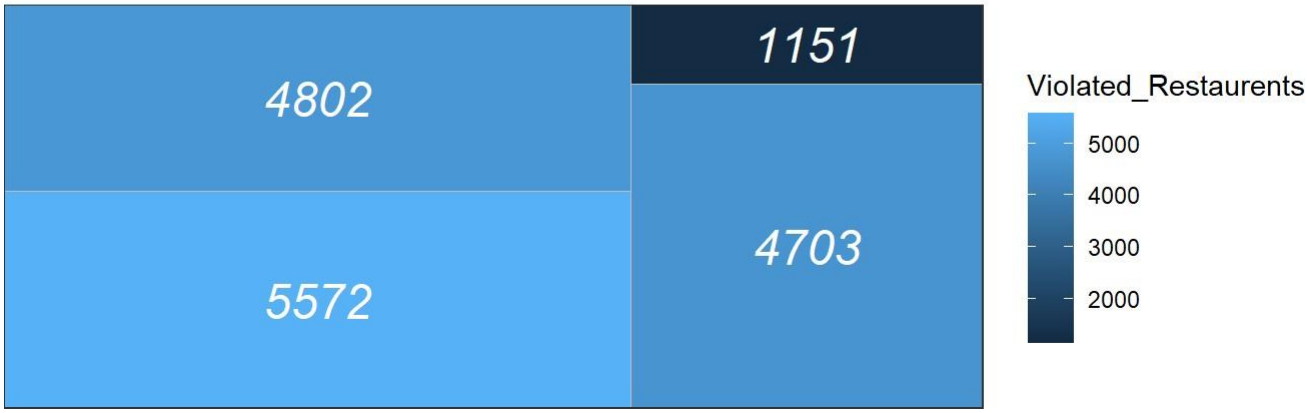
Plot_2 :

Now it becomes very crucial to understand and analyse the ratio of total restaurants to the no of restaurants that have violated the standards. So the below Tree map (1) On the top shows the total number of restaurants across the years (2016-2019) and the total violations categorized from (2016-2019)

Total Restaurants in SF

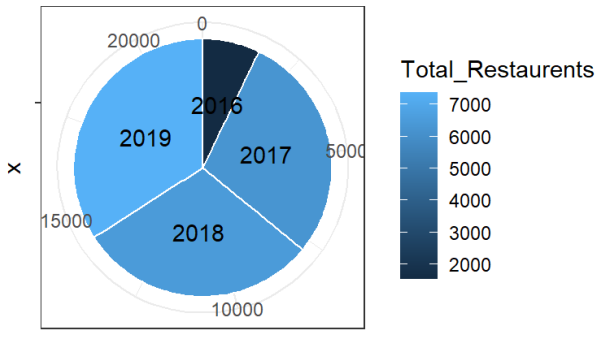


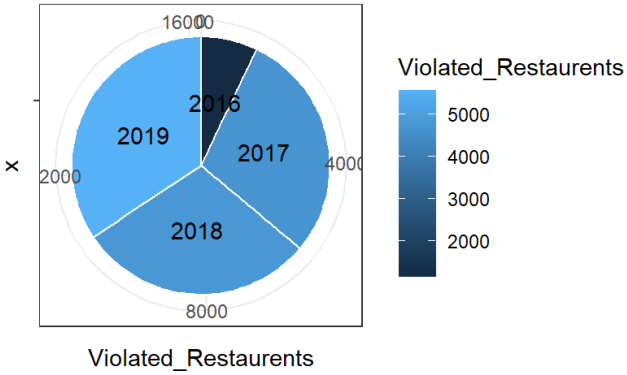
Restaurents that have Violation codes in SF



Plot 3

The below Pie chart gived more context to the one observed earlier through the Tree map by giving the Total Restaurants in San-Francisco Vs Total Violated Restaurants in San Francisco from the years (2016-2019).





Plot 4

Interesting Analogy by Taking the percentage of the total Restaurants vs Violated restaurants to find some trend and we can figure out that every successive year it is getting low and the next year it is raising and we can say that the restaurants becomes reluctant and do not maintain proper measures once after the first year and after further scrutinize the next year they improve in quality

```
## # A tibble: 4 × 4
## # Groups:   Year [4]
##    Year  Total_Restaurents  Violated_Restaurents  Percentage_Violated
##   <dbl>              <int>                 <int>                <dbl>
## 1  2016               1550                  1151                 74.3
## 2  2017               6242                  4703                 75.3
## 3  2018               6502                  4802                 73.9
## 4  2019               7383                  5572                 75.5
```

Some important Questions that have arised during the analysis and the answers for the are:

A. Have the Restaurant Count increased over the Years?

ans) Yes, The restaurants count have drastically inclined over the years.

B. Have the inspections reduced due to the new restaurants coming in?

ans) No, from the analysis it is clear that despite the count of new restaurants the inspections has stayed thesame and the Government is doing well to keep up with the quality standards.

Section 3.2:- Risk Categories Analysis :

Deep Analysis on the risk Categories of violations over years

This section is a very crucial part of the analytics as it is the deciding factor if the restaurant is a good restaurant or a bad restaurant . This analysis is done based on a lot of factors like below.

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.

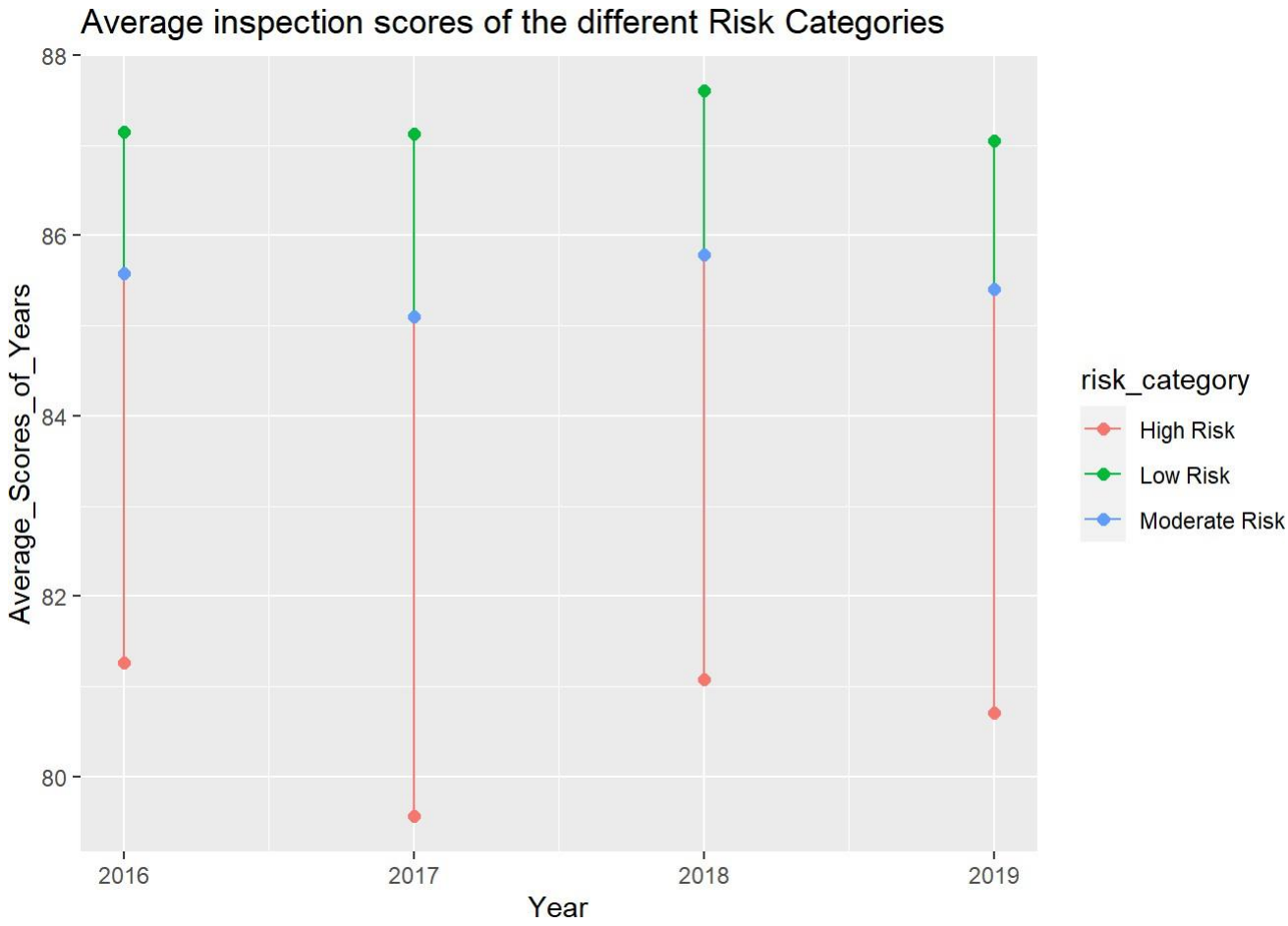
Low risk category: records violations that are low risk or have no immediate risk to the public health and safety.

Plot 5

The below Graph shows us the Average Violation scores of the various risk categories over the years (2016-2019).

Average Score is the score that lands with a geom_point.

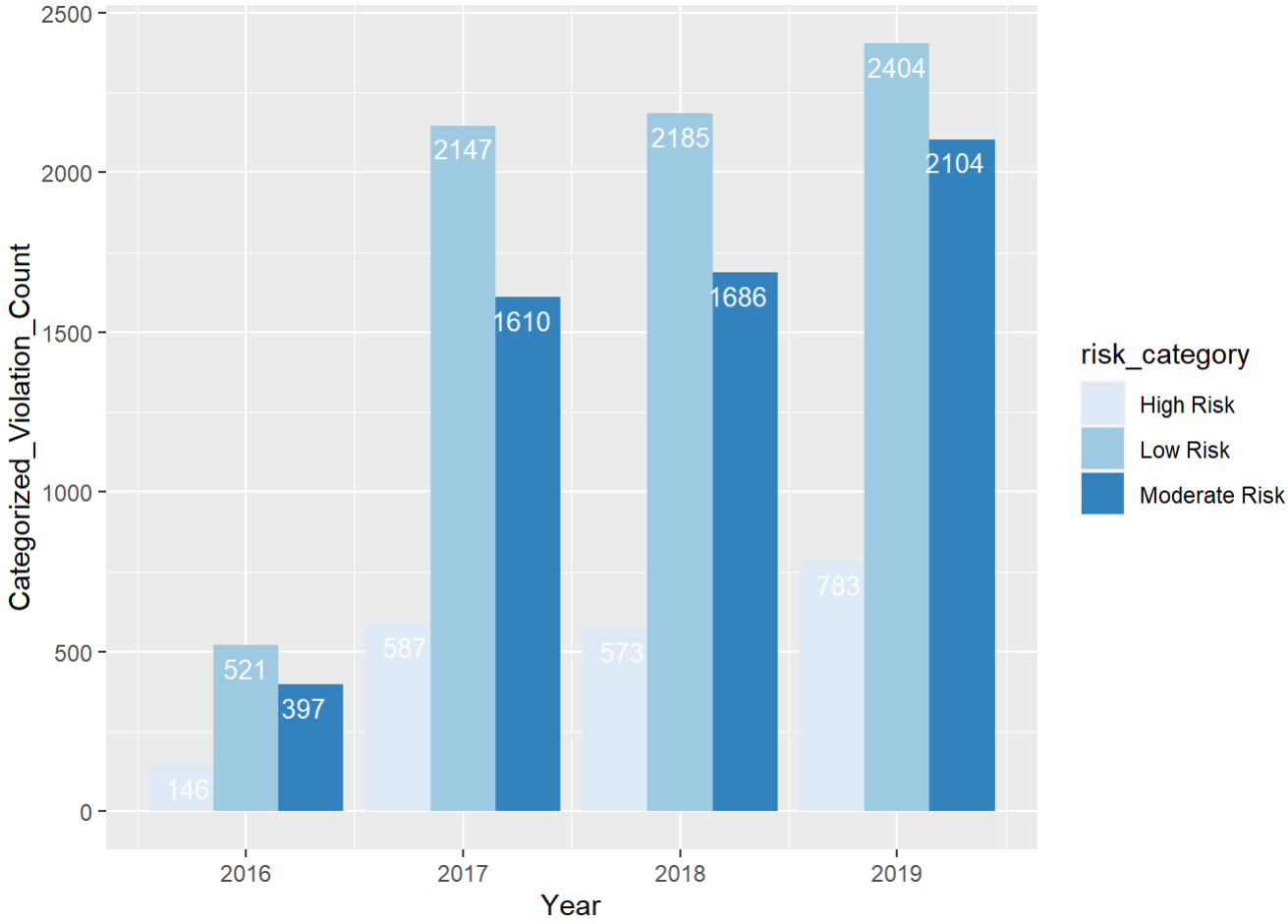
The Risk categories is categorized on the legend.



Plot 6

The Below Plot helps us understand how many violations have been categorized for the restaurents in the city of San Francisco based on the Risk Category Vs Years.

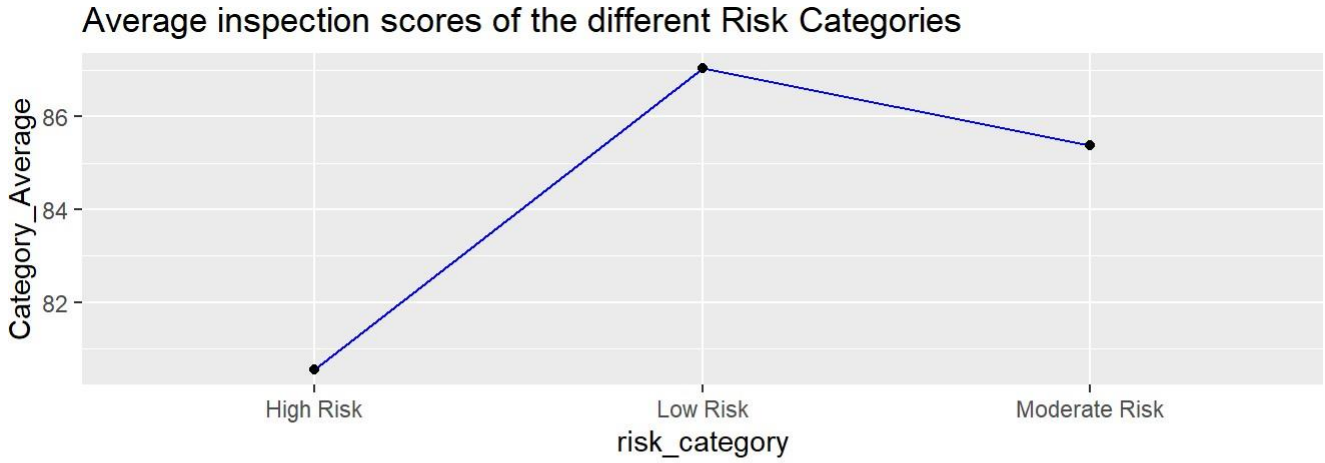
So based on the understanding on the below Groups Bar plot we can interpret how many High risk to Low risk restaurants were classified and which year the performance was bad and do deep analysis on that particular year.



Plot 7

Average Categorical Inspection Scores over the years which is the expected metrix for quantifying the volume of violation over the various Risk Categories.

We have acheived in computing this with a line chart and an Area Plot.



Some important Questions that have arised during the analysis and the answers for the are:

A. Have The High Risk Categories increased/Decreased over years?

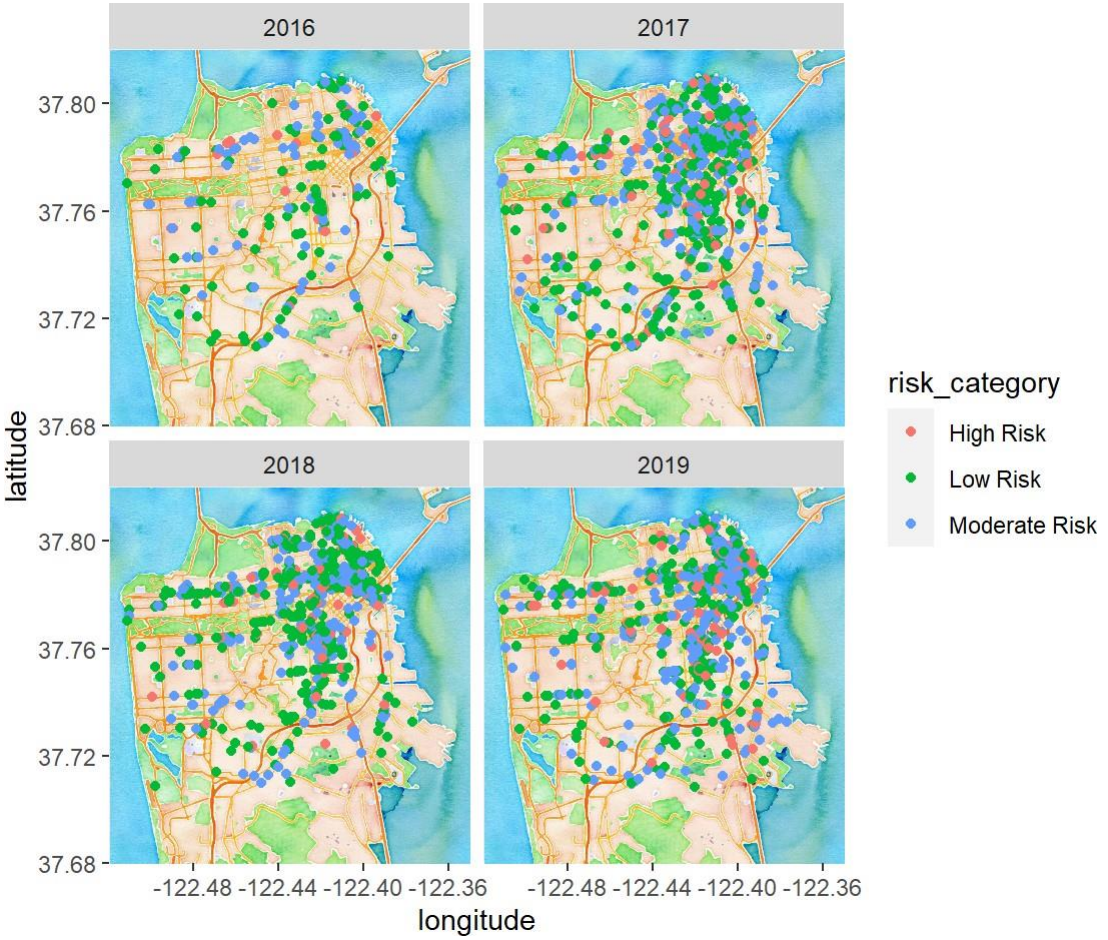
ans) There is no Drastic change in the High risk category in a progressive mannar. But one interesting observation is that every Successive year there is a Dip and Increase in the next successive year. From this we can Infer that the Restaurants are doing good once inspection is done but the next year they are missing some quality standards which is resulting in the Hike in High risk ratio.

Section 3.3: Geographical Mapping of the Violations wrt Risk_Category

Plot 8:

All Years Risk wise Map:

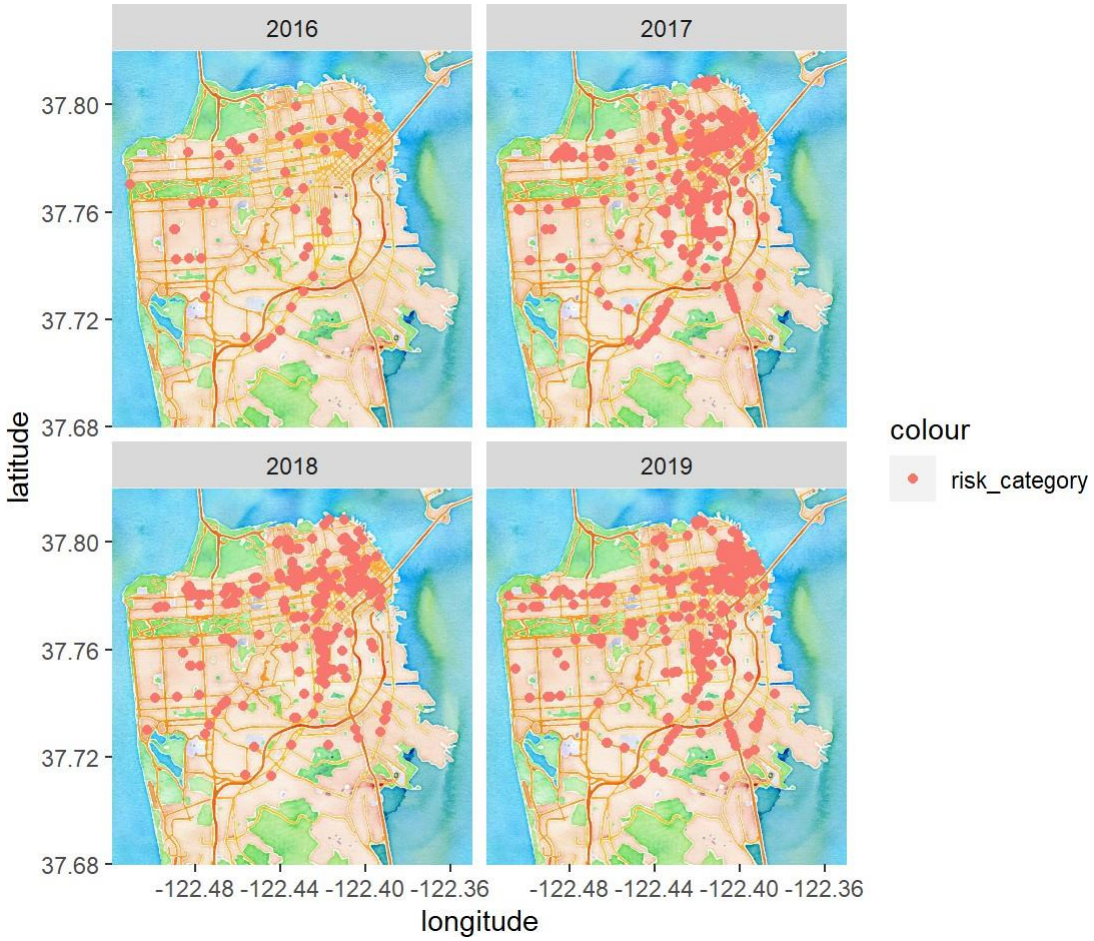
San Francisco Geographical map of total number of Violations during the years of (2016-2019). This involves all the risk Categories from low risk, moderate risk and High risk hilighted as scatters in the Map with the colours green , blue and red as hilighted in the Legend. These colours represent the severity of the violation.



Plot 9

All Years only HIGH Risk wise Map:

The below Geographical map of San Francisco gives our more clear emphasis on the High risk categories of restaurants over years (2016-2019). Now this is very important for the analysis as we have a perfect idea on the location of the High risk restaurents.



Some important Questions that have arised during the analysis and the answers for the are:

A. Are the Restaurants equally spread across San Francisco:

ans) No, The restaurants count is more on Towards the Downtown City areas and keeps continouesly increasing in those areas.

Conclusion

Over this research, we were able to learn a lot of things how the Department of Public Health in San Francisco works and how are they able to inspect the restaurants,reasons for inspection, if there are risky restaurants then they provide violation description and categorized them in risk categories.We firmly believe that the insights and the visualization provided by data analysis can help the restaurants and customers about the quality of a restaurant and give improvement measures and thus improve quality of Food,Service and customer Satisfaction

The first thing that we discovered from the data was that major of the restaurant had a lot of inspections may be in different branches and each year the number of business keep on increasing on exponential scale and a lot of restaurant had opened up new branches in different locations.

Section - 3.1:

The First plot tells us that the total number of business in San Francisco keeps on increasing every year.The data is shown from 2016-2019 years.This was the basis for us to further dive into data and analysis it thoroughly.

The Second plot and Third plot tells us the comparison between the total number of restaurants and the number of restaurants that have violated using pictorial representation in the form of tree map and pie chart over the years.

Section - 3.2:

The Fourth plot tells us the comparison between the percentage of restaurants and the violated restaurants. One can see here that if the restaurant has violated in the current year then it will improve its service quality in the coming year

The Fifth plot tells us the average inspection score of the restaurant in each year the data has been recorded.

The Sixth plot tells us the categorized violation count for each risk category in each year.

The Seventh plot tells us the difference between the average inspection scores for different risk categories and the stacked are graph tell us the comparison between different risk categories and the inspection score

Section - 3.3:

The Eight plot tell us the total number of violations in each year with each risk categories that is scattered across the geographical representation of San Francisco. Also we have two Geographical Graphs one that sums all the available restaurants across all risk categories and the next gives more context on the restaurants with High Risk.

The Department of Public Health of San Francisco were trying their best to inspect the restaurant that were violating guidelines and taking strict actions as much as possible. We also learned that the department was listening and coorporating with the citizins if they had issues regarding the restaurants. We can Infer that every Successive year there is a Dip and Increase in the next successive year. From this we can Infer that the Restaurants are doing good once inspection is done but the next year they are missing some quality standards which is resulting in the Hike in High risk ratio.

References

DataSet: https://data.sfgov.org/Health-and-Social-Services/Restaurant-Scores-LIVES-Standard/pyih-(https://data.sfgov.org/Health-and-Social-Services/Restaurant-Scores-LIVES-Standard/pyih-) qa8i/data(Found via SFO Open Data)