Covid_Food_Business_Team2

Question:

1.How many businesses have permanently closed during Covid-19? How many are open? How many are temporarily closed? (based on yelp and google places data)
2.How much activity is happening at different types of businesses by restaurant type (e.g., Italian vs. Thai vs. pizza, etc.), by neighborhood or zip code? (use safe graph data)? Do hours of operation seem to impact overall activity/ success?
3.What has been the relative impact on government policies (e.g., sidewalk dining) or assistance programs?

Dataset:

1.Google Places
2.City of Boston Certified Businesses
3.Food Establishment Inspections - Datasets - Analyze Boston
4.Active Food Establishment Licenses - Datasets - Analyze Boston
5.Monthly Places Patterns (aka "Patterns") Feb 2020 – Present(safegraph)

Data Collection

We have collected data from Google Places API to get the information about business status of restaurants. And we have also combined the data from google places API with Food Establishment Inspections datasets to check the inception status and violation status of restaurants in the Boston area. (The Active Food Establishment Licenses dataset and Food Establishment Inspections have overlaps, after analyzing we choose Inspections datasets) To analyze the impact of pandemic, we extract data from safegraph. Unfortunately, we are not able to get more details about restaurants' business like the sales or the interests, so we do the analysis mainly based on the popularity for different regions.

In addition, the google place API can only return 60 pieces of restaurant information. Therefore, we divide latitude and longitude into 10 different values to cover boston area, then reconstruct 10*10 points to request Google Place API. Finally, we got 1367 restaurant information after removing repeated restaurants.

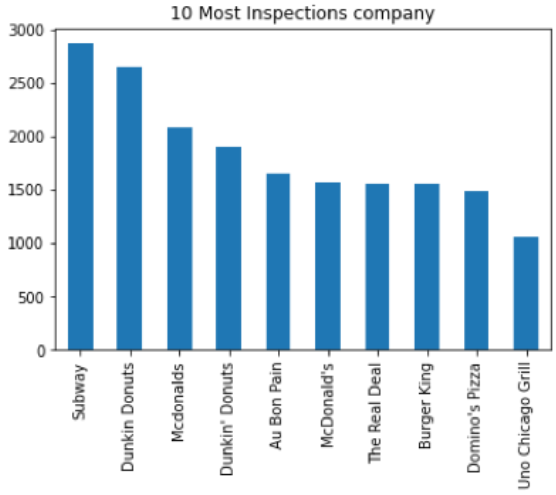
Code explanation

If you want to use Google Place API, you need to create a Google Cloud Plate account and then enable Google Place API then generate credentials and keys. Getting specific step by step instructions click here https://cloud.google.com/apis/docs/getting-started

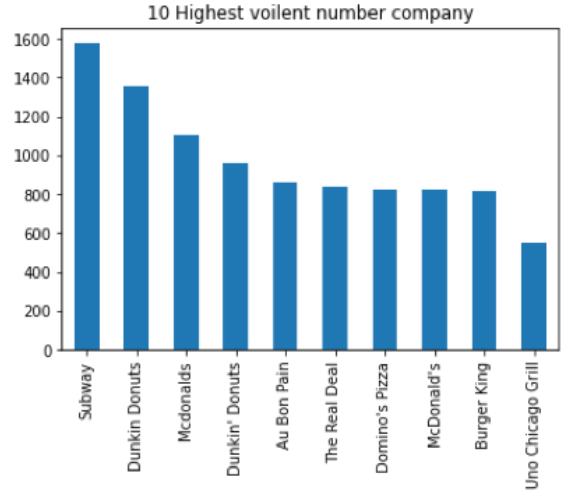
After credentials and keys are generated, requesting 6000 restaurant information then removing those restaurants with the same id. Finally, we got 1367 restaurant information. First we extract valid and useful data from safegraph for each month from February to October, and save it as json form which we show in the “processsed_data” folder. Later we read the json file combined with MA's geographic information producing the visitors flow CSV dataset. Then, we utilized the new Choropleth class producing the heatmap of October’s traffic for different zones in the Boston area based on different postal codes. Furthermore, we plot each month’s Visitors flow in the Boston Area with different postal regions. Moreover, we did different month traffic analysis from Monday to Sunday and different days (from Monday to Sunday) analysis from February to September. We will get a clear understanding when seeing the plot. At last, we plot the covid_19 cases trend for different months, correlated with the popularity of postal code. We can have a better understanding of the relation between covid_19 and restaurant activity.

Data Analysis

First, we extract the data from food_inspection and google place API, we find out the frequency of company’s inspection and violation, we plot out the top 10 of them. We can find out the companies like “subway”, “Dunkin Donuts' ', “McDonalds” all have high frequency being inspected and violated.

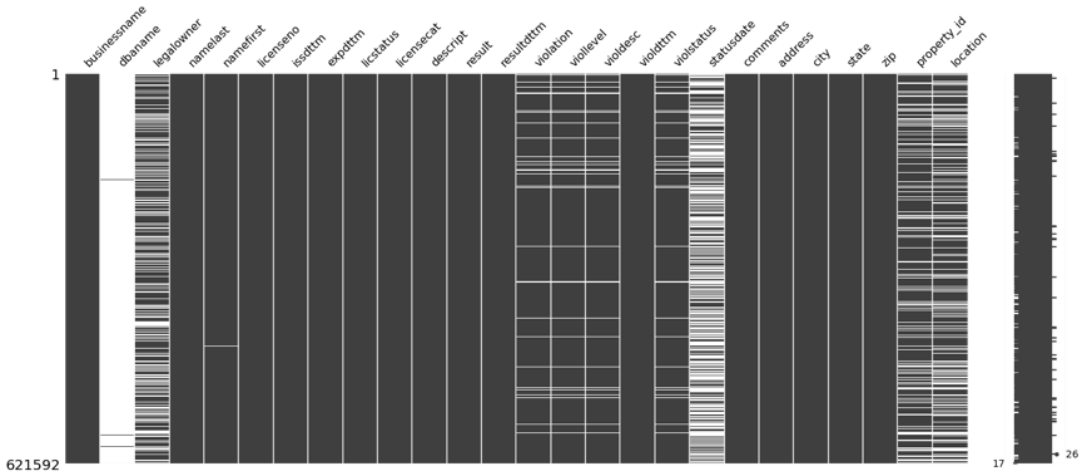


Picture1: inspection



Picture2:violation

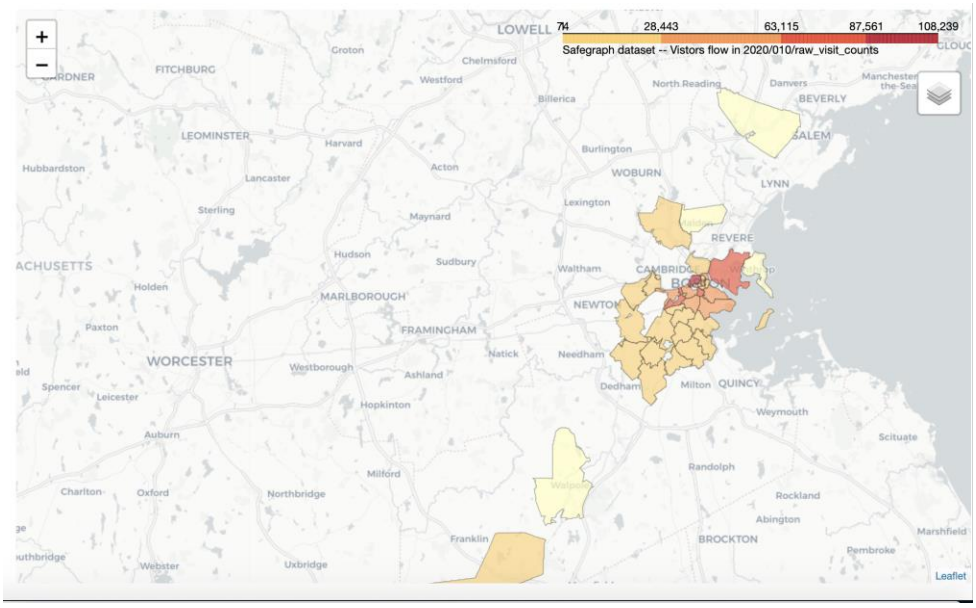
In order to help us do analysis for the combination of googleplace_Api data with the above dataset, we plot the data distribution out to clearly see which data has the most missing part and which are mostly valid. By doing this, we want to check if the business status has something to do with its violation, but after comparison, we can’t make a conclusion for this. In picture3, black means the data is valid, white one means the missing part.



Picture3: Dataset quality analysis

And based on the google place API, we can find out the answer to problem 1, we can see which companies are closed and which companies are open. We store the output file as csv and upload the GitHub.

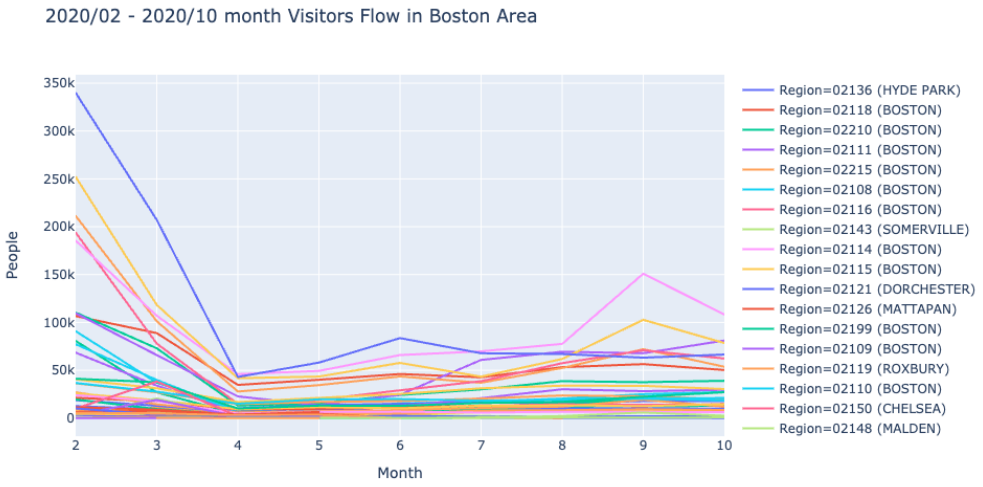
In order to explore different region's business activities and its correlation with covid_19, we mainly run the analysis based on safegraph. For the most recent data we can get, we plot the distribution heatmap for October in Boston. The deeper the color is, the more popularity this region has.



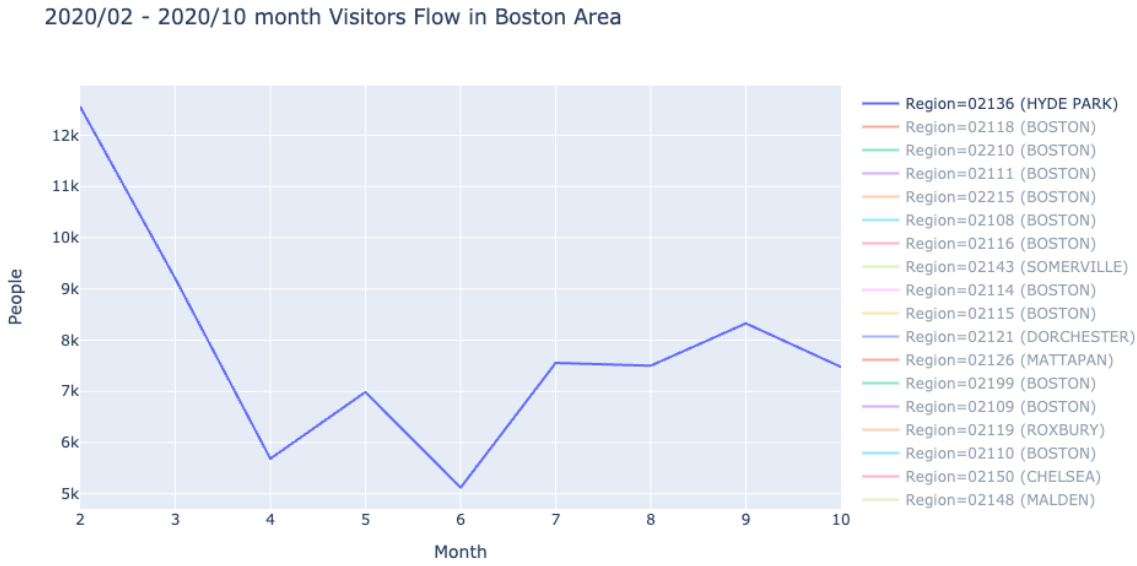
Picture4: Heatmap in Boston region

From picture4, we can obviously find out the regions besides downtown have the most popularity, and those suburbs have relatively lower popularity.

Then we plot the visitor flow for different regions in Boston from February to October, in google colab, we can select the specific area to see its trend clearly like picture6.

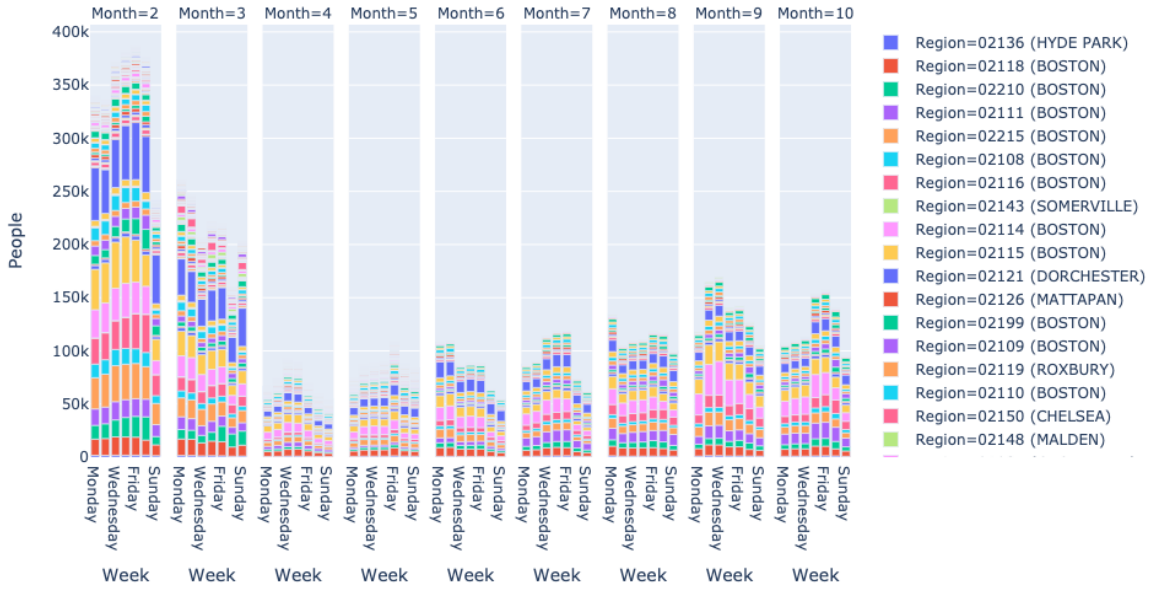


Picture5: monthly visitor for different region

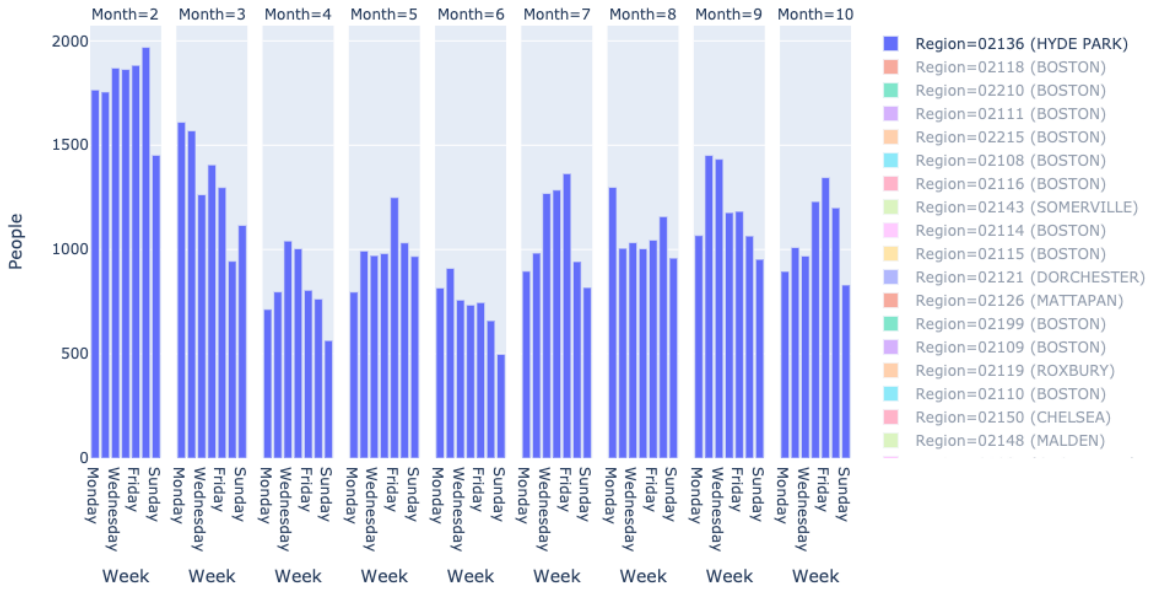


Picture6: HYDE PARK(02136) monthly visit

Furthermore, we did the traffic analysis and visualization from each month's Monday to Sunday. And we plot it in a bar and add them together to show the trend of popularity as an integrity. Same as above and below, we can select one specific region and have a clear observer for that region like Picture 8.



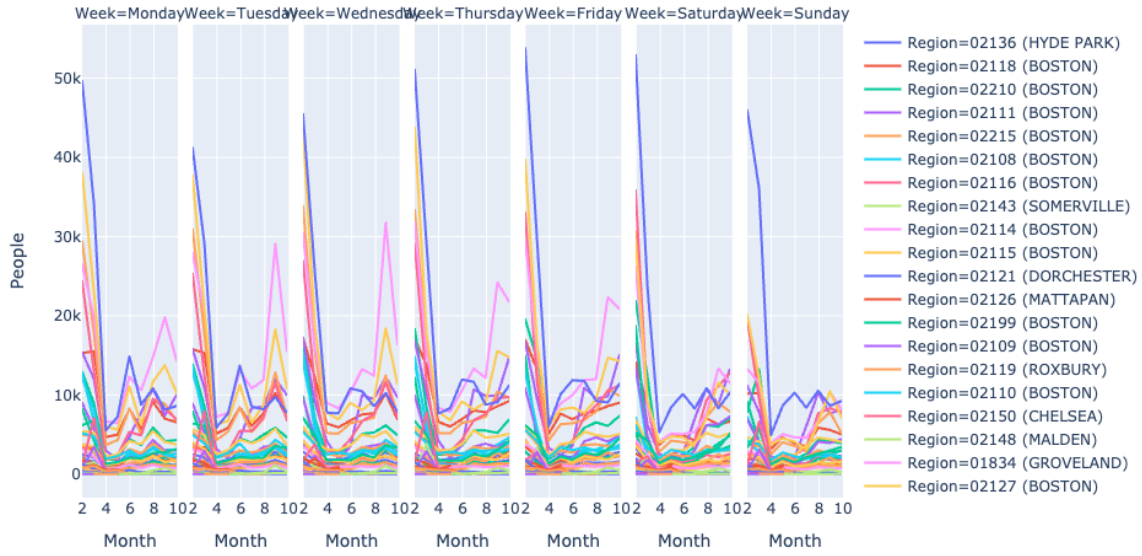
Picture7: Specific month's popularity for every day in the week



Picture8: HYDE PARK(02136) popularity

From picture7, we can find out every Friday's popularity for dominates, and every Sunday has the lowest popularity in the week.

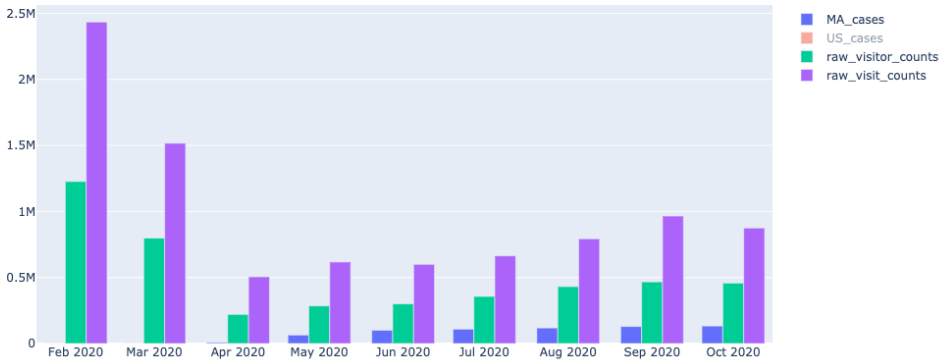
Similarly, we did the traffic analysis and visualization for each month differently from Monday to Sunday. This time, we plot it as a line chart to see the curve like picture9.



Picture9: Specific day's popularity analysis for every month

For picture4 to picture 6, we use "raw_visit" in safegraph as data to do the analysis, which shows the visit flow during a month's time range. For picture7 to picture 9, we use "popularity_by_day" as data for analysis, which shows a mapping of day of week to the number of visits on each day during the date range.

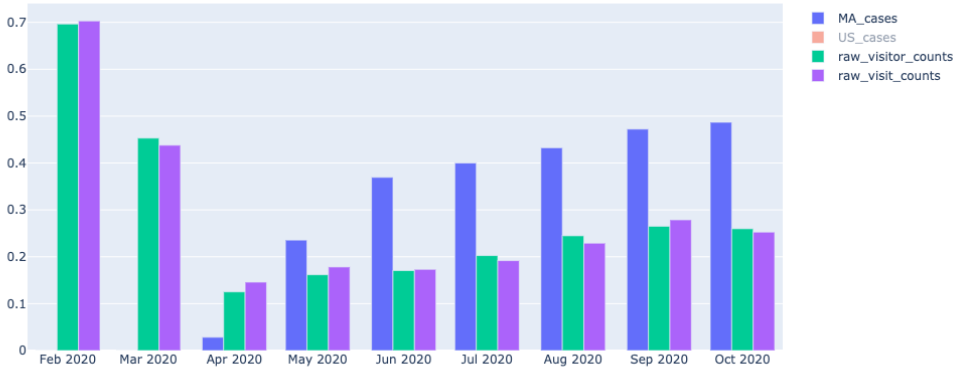
Because of the limitation of the dataset, we are not able to analyze the impact for different types of restaurant but we can see the impact in general. In the figure below we can see that the number of visitors has reduced a lot after the pandemic happened.



Picture10:before normalization

Picture 10 shows the covid_19 cases relationship with visit, before normalization, we can't see how covid_19 affected business activity. But we find something interesting which is that for each month, raw_visit is twice as much as raw_visitor, based on the definition, raw_visitor is

the number of unique visitors during the time range. So we can make an approximate conclusion that each visitor will take a visit twice a month on average.



Picture11: After normalization

Picture 11 shows the plot after normalization, at this time, we can clearly see the relation between MA_cases with raw_visit. We see covid_19 start exposing during April, that's the time the government banned all the activities and business, so we have the lowest visit at that time. But after May, the time the government reopened business and activities, visits began increasing and recovering, but more apparently, covid_19 cases increased incredibly faster. And it keeps this rising pace until now.

From February to April the number of visitors has reduced 80%. Meanwhile, according to the policy of the MA state government, the governor announced the dine in is not allowed since March 15th. The impact for restaurants is huge at this time. But we have noticed that the number of visitors has increased after May. By analyzing the relevant news, we found that the government has announced reopening of the MA on May 18th. But the restaurants are still not allowed to provide dine service in this phase. And on June 8th the MA entered the first stage of Phase 2 of its reopening plan. In this phase, the outdoor seating of restaurants is finally allowed.

At the same time, we can notice that the number of visitors are gradually recovered. And on June 22nd the indoor dining restarted in MA. There's an obvious increase in popularity, however, once we normalize the data we can find out that the confirmed cases have increased incredibly faster after reopening starts. And on November 2nd, the state government requires restaurants to stop providing table service after 9:30 pm. So, we think that the business of restaurants cannot go back to normal as long as the pandemic cannot stop.

And the reopening policy needs to be reconsidered. Because the reopening policy will cause the deterioration of the pandemic without doubt. Once the pandemic deteriorated the policy will try to stop reopening which is like an endless cycle. So, if the government can insist on stay-at-home advisory for longer time the situation now might be different.

Conclusion

In conclusion, historical restaurant data is not available online, thus can not answer how many restaurants permanently closed during Covid-19. However, we request 1636 restaurant information by Google API on 11/11/2020. At that time, there were 94 restaurants permanently closed. Close rate is 5.745%. If we can get the close rate before Covid-19 period, Then we can compare them to conclude the degree of impact on the restaurant by Covid-19. Among permanently closed restaurants, the rate of having a violent record to no violent record is about 22.34% which is very close to that of open restaurants (23.74%). Therefore, there is no evidence that violent records lead to those restaurants permanently closed. For the activity during Covid-19, the number of visitors dramatically decreased in the earlier Covid-19 outbreak. In Picture 5, it shows the number of visitors Categorized by zip code in detail. In addition, government policy (Stay-at-Home) limited people having activity outside especially in April, 2020, which can be reflected in the figure 9.