

CS506 Final Project Report:

Boston Restaurants During Covid

Team 1

Ganghao Li
Zhou Shen
Zhenfei Yu
Kefan Zhang

1. Background & Targets

As the Covid-19 attacked the world in 2020, the restaurants in Boston were struggling to survive this difficult time. The Mayor's Office of Economic Development was looking for some conclusions about the restaurants during Covid, supported by solid data.

Main questions we are expected to answer are:

- How many businesses have permanently closed during Covid? How many are open? How many are temporarily closed?
- 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?? Do hours of operation seem to impact overall activity/ success?
- What has been the relative impact on government policies or assistance programs?

Mainly we are focusing on answering them with the data we fetched and tools we use on those data, but we may also analyze the data in many other different ways to come up with other useful conclusions.

2. Resources (APIs & Datasets)

- Google Maps Places API

Google Maps Places API provides static data at the time when we started collecting information. It provides answers to below questions: location of restaurants, current business status of restaurants, whether restaurants are permanently closed, number of total ratings and ratings from users. To access it, we need to get a Google Maps Places API key.

- **Safegraph**

Safegraph provides data of Place foot-traffic and demographic aggregations that answer: how often people visit, where they came from, where else they go, and more. Available for ~3.6MM POI in the USA. To access it, we need to set up an AWS client and get an access key.

- **Active Food Establishment license dataset**

The Health Division of the Department of Inspectional Services (ISD) creates and enforces food safety codes to protect public health. This dataset contains a list of restaurants that met the City's standards to become licensed food service establishments.

3. Code instruction

- **Data preparation and clearing**

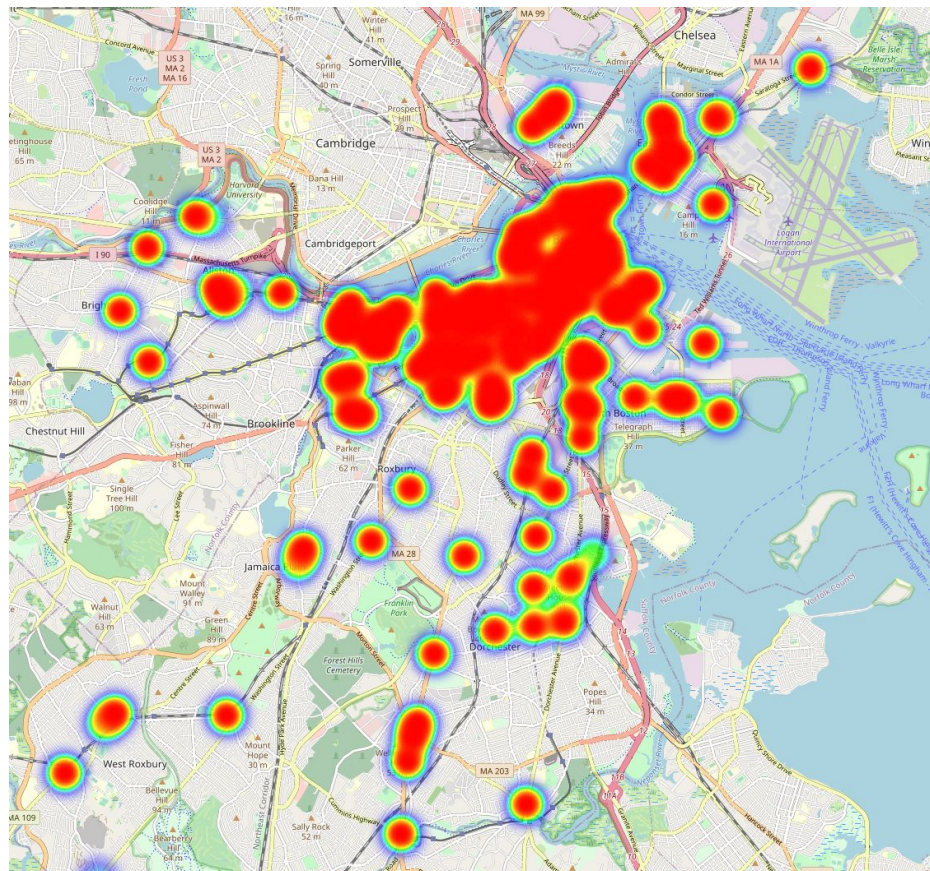
Safegraph provides us data of monthly visits to all of the visiting places in the United States, so we first extracted data about Boston from it based on the city name column. Then we join what we get from safegraph with the active food license table, which generates the visiting information to Boston restaurants from January 2020 to September 2020.

The nine datasets are the main sources that our observation and analysis is based on. Here is an example of January data, which includes columns like restaurant name, address, city, postal_code, raw_visit_counts and etc.

	name	street_address_x	city	region	postal_code	brands	date_range_start	date_range_end	raw_visit_counts	raw_visitor_counts	...	CITY	State	LICSTATUS
0	Ilona	783 Tremont St	Boston	MA	2118	NaN	2020-01-01T00:00:00-05:00	2020-02-01T00:00:00-05:00	506	282	...	Roxbury	MA	Active
1	Juice Press	225 Franklin St	Boston	MA	2110	Juice Press	2020-01-01T00:00:00-05:00	2020-02-01T00:00:00-05:00	249	93	...	Boston	MA	Active
2	Boloco	283 Longwood Ave	Boston	MA	2115	Boloco	2020-01-01T00:00:00-05:00	2020-02-01T00:00:00-05:00	499	254	...	BOSTON	MA	Active
3	Starbucks	360 Huntington Ave	Boston	MA	2115	Starbucks	2020-01-01T00:00:00-05:00	2020-02-01T00:00:00-05:00	110	54	...	BOSTON	MA	Active
4	Starbucks	300 Longwood Ave	Boston	MA	2115	Starbucks	2020-01-01T00:00:00-05:00	2020-02-01T00:00:00-05:00	3478	1559	...	BOSTON	MA	Active

● HeatMap

We also use the latitude, longitude and the raw visits from restaurants per month, in above filter safegraph DataFrame to illustrate this heatmap, which can reflect the popularity of the restaurants and variety of different regions.

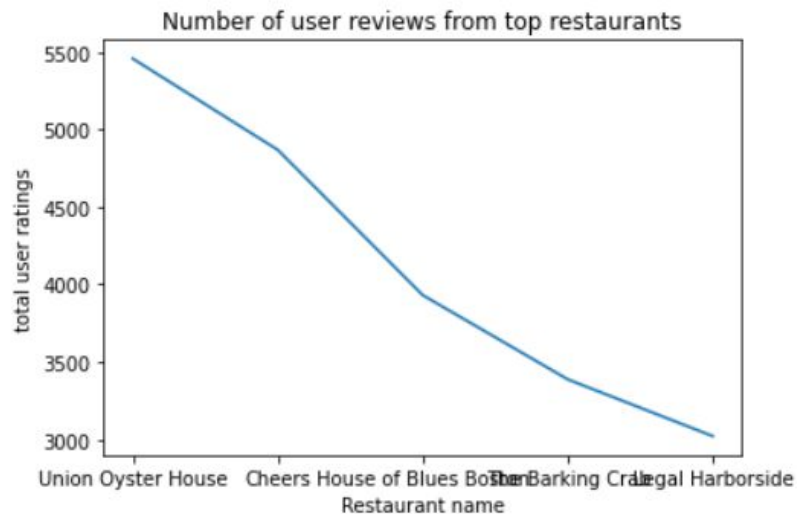


- Data analysis and visualization

We categorized data by different postcodes and different areas in Boston, and analyzed the relationship between monthly mean visits and different post codes, between monthly mean visits and time period.

4. Data visualization

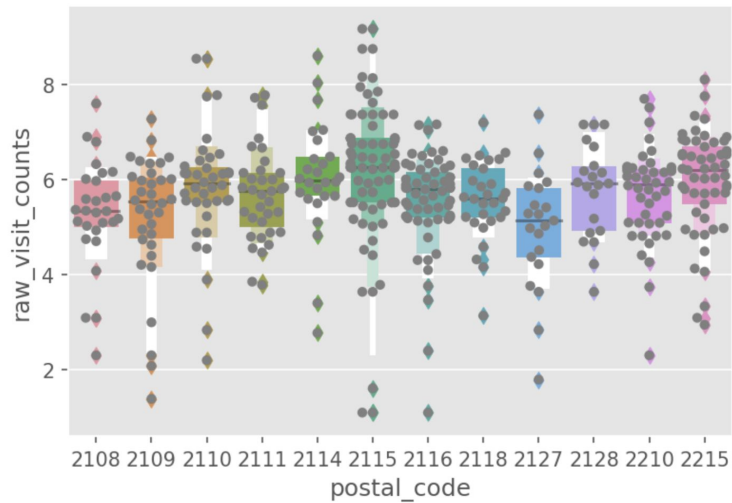
- Number of user reviews from top restaurants from Google Maps Places API



This Figure shows that the number of user ratings is different among the restaurants, and the span is relatively large, the highest and lowest has 2,000 ratings gaps.

- Relationship between visits and different areas (by postal code/CITY/brand of restaurant)

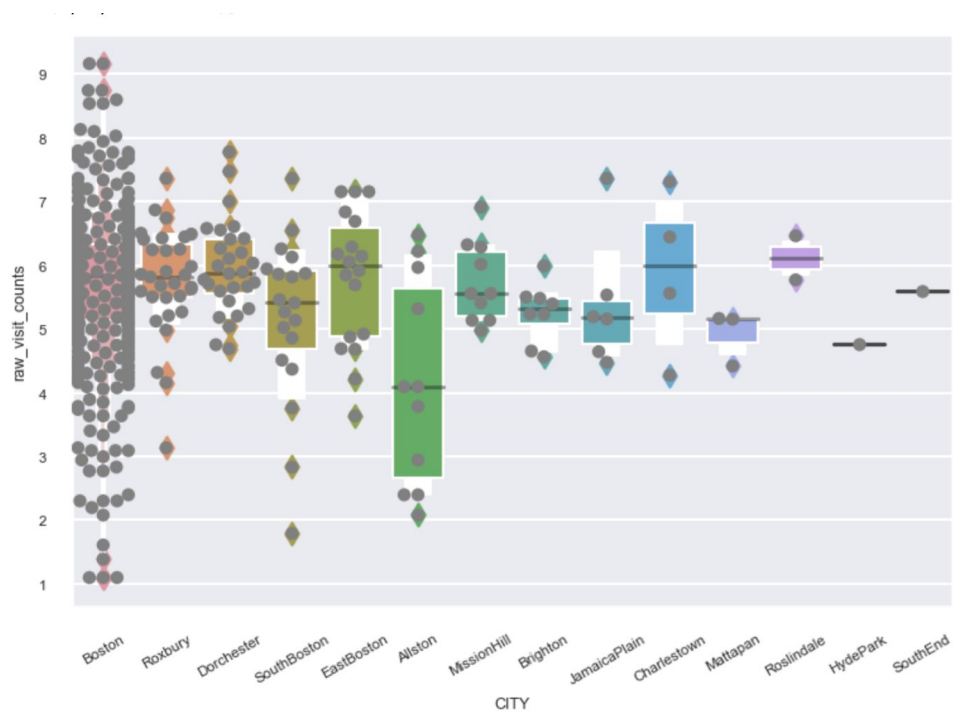
By postal code



The X axis is the postal code, and the y axis is the raw visits take Logarithm operation (to display the graph as the better visual format).

This proves that restaurants in different postal codes have the different distribution of raw visits. We can see that restaurants in 02115 have the highest and lowest visits, but most of them in this region are at relatively high visits.

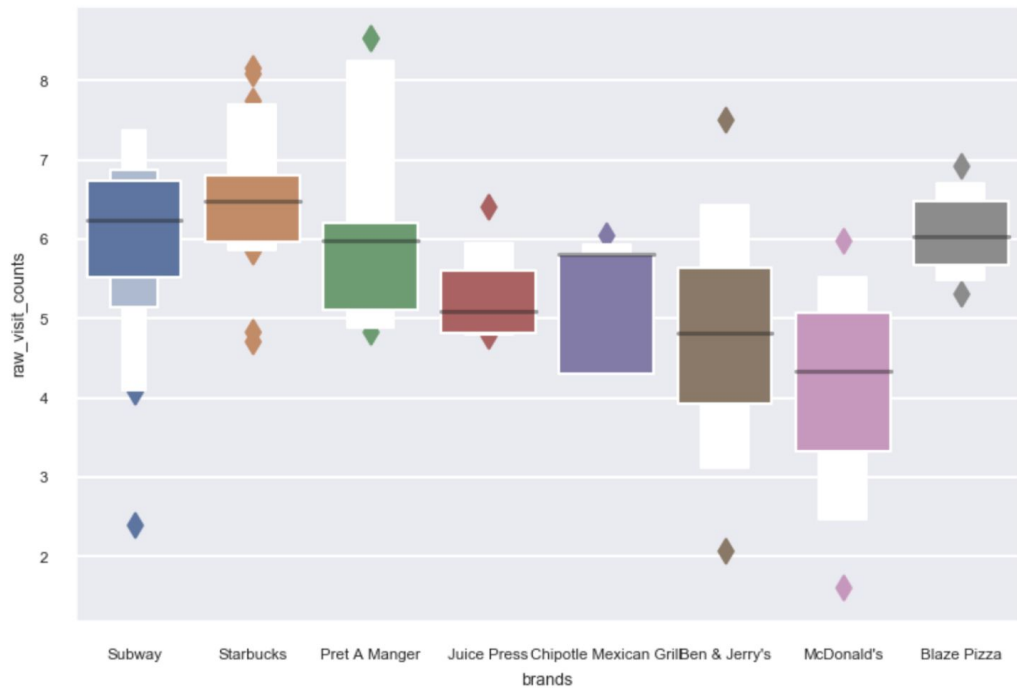
By city



The X axis means the name of the district and Y axis is the same as above.

This proves that most restaurants are centered in the Boston District, and most of the raw visits are from 16 ~ 256 per day. It strengthens the conclusion that different locations do impact the restaurant's visits.

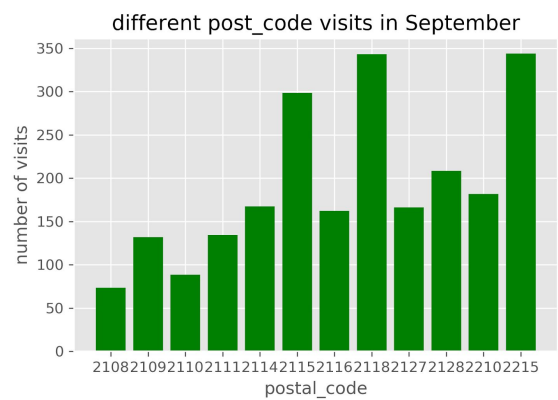
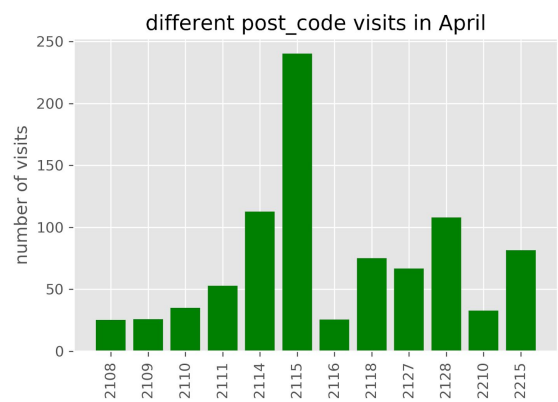
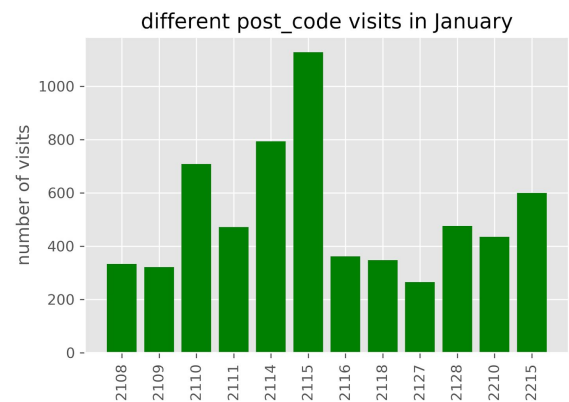
By brands of restaurants



The X axis represents the brand of restaurants like Subway, Starbucks, Five guys and so on. It shows that different brands have different distributions of customer visits. Besides, it also displays that even restaurants are the same brand, their visits still vary a lot, which we can conclude from Subway, Ben & Jerry's.

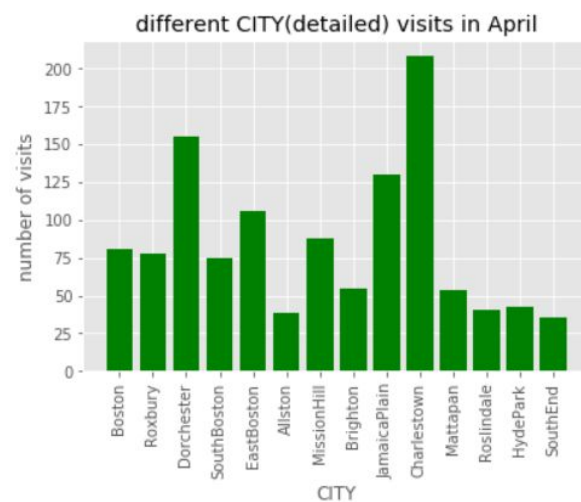
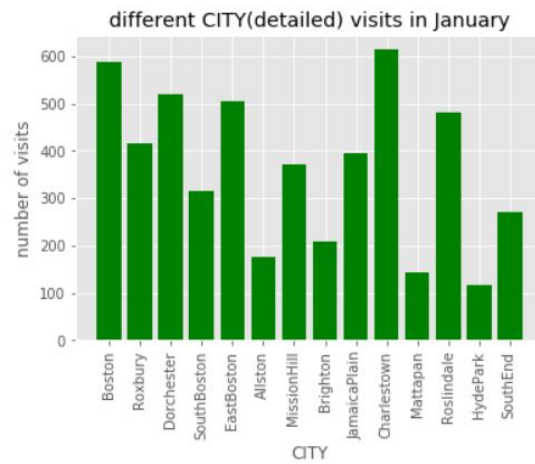
- Number of visits to different CITY/postal code in different months

By postal code



These 3 graphs shows each postal code groups' visits in January, April, and September.

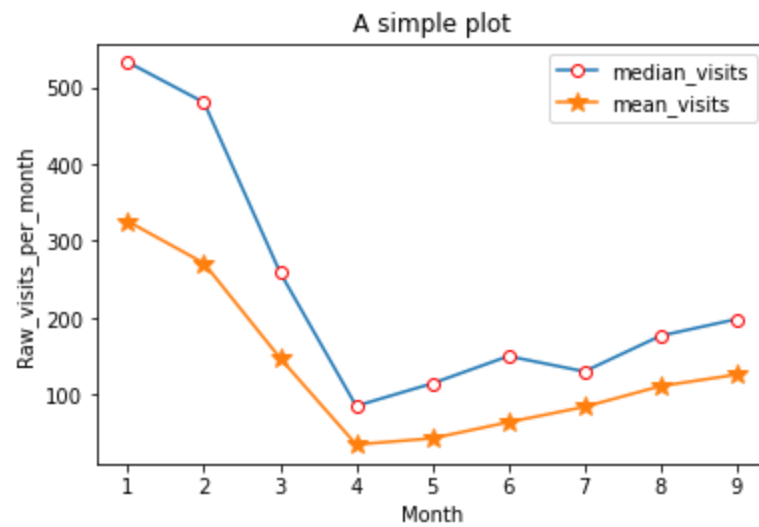
By CITY



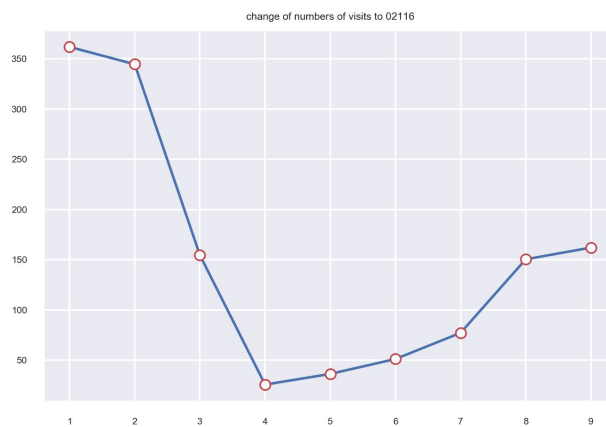
Above 2 graphs show each postal code group's visits in January and April..

- Change of numbers of visits to different postal code/CITY

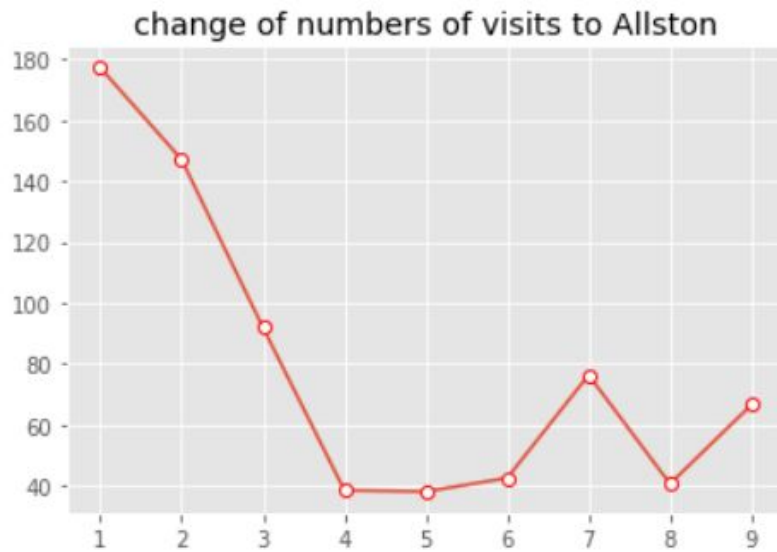
General Trend of all the restaurants



By different postal code

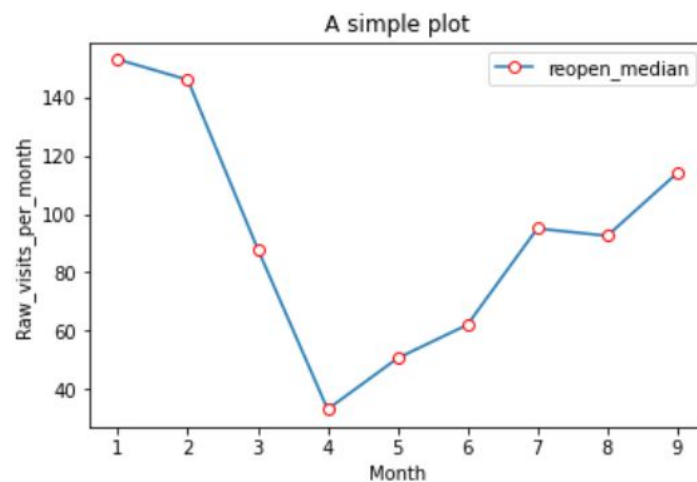


By CITY - Allston



It proves that although the general trend reduced sharply and then increased slowly, but actually different location, different brand, the visits trend is not always the same as the general trend. Some of them will increase relatively more from April to September, and some will decrease in August while others still keep rising.

- Restaurants' visits with Governments Support Policies



This image shows the restaurants which receive the government support policies. We can see that in April, the visits reduce to the lowest at less than 40 visits /

day. After these restaurants receive the support, their visits start to rise at stable speed.

5. Conclusion

From the above data visualization, some conclusions are as follows.

- (1) With the time changes, the total visits of all the restaurants reduce sharply from January to April, and then it increases slowly from April to September. That means firstly, each restaurant has suffered a lot from Covid-19, because people gradually noticed the harm of Covid -19 and the Government also Strongly suggested citizens not go outside if not necessary. But after April, people started to be familiar with covid and were not as terrified as before, so they began to eat in the public areas, but the total visits is still at relatively low number
- (2) Through the graph of relationship of visits and different areas (by postal code/CITY/brand of restaurant), we can see that the total visits amount and distribution is different according to above elements, which means the location, City Region, brand are also the key factors to the restaurants business.

6. Limitations

- (1) We could not obtain the feature which can presents each restaurants' operation hours, we found that feature of each restaurant is null
- (2) Most Apis is not open-source and free, so there are few resources available from the third party apis
- (3) Because the data format in each datasets is different, so after we merge the same item, there are only few elements left

7. Roles and responsibilities of the team

Ganghao Li:

Mainly work on obtaining data from safegraph, data cleaning, visualization, heatmap and analysis, writing reports.

Zhou Shen:

Mainly work on obtaining data from Google Maps Places API, visualization (relationship between visits and different components), analysis and completing reports.

Zhenfei Yu:

Mainly work on data cleaning and filtering from safegraph and license dataset, visualization of covid impact on restaurants categorized by postcode, and writing reports.

Kefan Zhang:

Mainly work on obtaining data using Yelp API, exploiting government policy datasets (for example Reopen Boston Fund and Small Business Relief Fund), integrating them with datasets fetched from SafeGraph and find useful common tuples, merging useful columns, removing duplicates and so on.