

---

# Impact of COVID-19 to NYC analysis based on OEM Notification Exploration

Group 10. Garlic

Yuexi Wu

Xuebin Zhang

Jiahui Wang

May 10, 2020

<https://github.com/Yuexi-Wu/NYC-Covid-Analysis>

---

# CONTENTS

<b>1. Abstract</b>	<b>3</b>
<b>2. Introduction</b>	<b>3</b>
<b>3. Data cleaning and Data summary</b>	<b>3</b>
3.1 Data Cleaning and Data Integration	3
3.1.1 Dataset Overview	3
3.1.2 Data Quality Discussion	4
3.1.3 Data integration	6
3.2 Data Summary	7
3.2.1 Notification each type since March	7
3.2.2 COVID-19 confirmed cases since March	7
<b>4. Methodology and Findings</b>	<b>9</b>
4.1 Fluctuation for each type of notifications during epidemic	9
4.2 Comparisons between 2019's first quarter and 2020's first quarter	12
4.3 Notifications related to COVID-19 exploration	15
4.3.1 Type distribution	15
4.3.2 Quantitative relationship with confirmed cases	15
4.3.3 Other Observations	16
<b>5. Conclusion</b>	<b>18</b>

---

# 1. Abstract

The purpose of this project is to find out whether the COVID-19 epidemic had brought any impact on the number of emergencies in different classifications in NYC and from the quantity fluctuation we could find out in which way the epidemic affects people life most in the city.

Dataset we are using is Emergency Notifications data posted by OEM. We selected data from year 2019 to 2020, check out whether there's fluctuations during the virus breakout.

## 2. Introduction

### Background

At the end of 2019, a novel coronavirus outbreak in Wuhan, a city in the Hubei Province of China, caused a cluster of pneumonia cases and rapidly spread throughout the world. The World Health Organization named the disease COVID-19, which stands for coronavirus disease 2019. As of 8 May 2020, over 3,870,237 cases have been identified globally in 177 countries with a total of over 270,198 fatalities.

The COVID-19 pandemic has a significant influence on people's life not only at an individual level but also a macro level. Data has captured many effects by comparing and analyzing the indicators before and during the pandemic. As NYU students, people's situation and daily life in NYC captured our attention. Also, New York as the most severe state is worthy of a further analysis.

### Research Question

This project aims to investigate the impact of COVID-19 on OEM Emergency Notifications. Analyzing the statistics of the overall notification and some specific emergencies associated with COVID-19 to reveal how life of NYC people changed during this epidemic.

## 3. Data cleaning and Data summary

In the data cleaning process, we used Socrata Open Data API and Pandas to access our dataset and got rid of blank or NaN data and data beyond the scope of our research. It will be explained specifically in the following.

### 3.1 Data Cleaning and Data Integration

#### 3.1.1 Dataset Overview

There are two major dataset we decide to employed. Although both of these datasets updating in realtime, we adopt data until 26/04/2020 for convenience.

1. **OEM Emergency Notifications**( emc) is a dataset includes messages sent with information about emergency events and important City services offered by OEM (<https://data.cityofnewyork.us/Public-Safety/OEM-Emergency-Notifications/8vv7-7wx3>)

Column name	Description	Type
<b>Record_ID</b>	record_id	object
<b>Date and Time</b>	Date and time that notification was sent	object
<b>NotificationType</b>	Notification type	object
<b>Notification Title</b>	Notification title	object, blank 1728
<b>Email Body</b>	text of notification	object, null 523

Totally 16872 entries

table 1: OEM Emergency Notifications dataset Info

2. **NYC-COVID** is a dataset includes daily counts of new confirmed cases, hospitalizations, and deaths. <https://github.com/nychealth/coronavirus-data/blob/master/case-hosp-death.csv>

Column name	Description	Type
<b>DATE_OF_INSERT</b>	Case fount by the date	date
<b>NEW_COVID_CASE_COUNT</b>	Cases are by date of diagnosis	int
<b>HOSPITALIZED_CASE_COUNT</b>	Hospitalizations are by date of admission	int
<b>DEATH_COUNT</b>	Deaths are by date of death	int

Totally 54 entries without null values

table 2: NYC-COVID dataset Info

### 3.1.2 Data Quality Discussion

#### 1. Data Cleaning

During exploring the two above datasets, we found some problems such as missing values (e.g. NULL, N/A, Blank, UNSPECIFIED etc) and duplicate entires and outlier that should be dropped.

**OEM** has 523 of them have a null value in the email\_body column. [blank] occurs 1728 times in the Notification title.

---

Check date range and find min value is 1900-01-01, which seems like a outlier.  
NYC-COVID has no missing values or outlier.

## 2. Data Labeling

Since we do not need the whole raw data, it should be remarked and renamed.

For instance, 'Date and Time' is renamed as 'Date' and was kept only the date part because this project do not plan to use time.

**The result re-arrange for OEM is below:**

New Name	Type	Explanation
Date	datetime	kept only the date part because this project do not plan to use time.
Notification Type	category	one type has occurred at least 16 times
Count	int	count each notification type occurred times by date

table 3: OEM dataset info after adjustment

'Record ID' and 'Notification Title' are dropped because no significant effect.

Check duplicated entries and keep the first one. And did not find duplicated entry.

Check date range and find min value is 1900-01-01, which seems like a wrong value.

Count the number of occurrences of different notification types. Found there has no wrong value, and the least type has occurred 16 times.

Count the number of occurrences of different notification title. [blank] occurs 1728 times. Convert [blank] to np.NaN

Count the number of occurrences of different notification types. Found there has no wrong value, and the least type has occurred 16 times.

Count the number of occurrences of different notification title. [blank] occurs 1728 times. Convert [blank] to np.NaN

**The result re-arrange for NYC-COVID is below:**

New Name	Type	Explanation
Date	datetime	reformat the date the same as OEM.date
Newcase	int	count each new COVID cases by date

table 4: NYC-COVID dataset info after adjustment

---

Drop HOSPITALIZED\_CASE\_COUNT and DEATH\_COUNT

Rename DATE\_OF\_INSERT as Date and rename NEW\_COVID\_CASE\_COUNT as Newcase.

Check the range of date, min is 2020-03-03, max is 2020-04-25.

### 3.1.3 Data integration

To do integration, we select the entries after 2020-03-02 and join these two datasets together. Part of the joined table are shown below

---

	Date	Newcase	notificationtype	count
0	2020-03-03	2	Aerial (Fly-Over)	1
1	2020-03-03	2	Mass Transit Disruption	1
2	2020-03-03	2	Mass Transit Restoration	1
3	2020-03-03	2	Road Closure	5
4	2020-03-03	2	Weather	1
5	2020-03-04	5	Environmental	1
6	2020-03-04	5	Mass Transit Disruption	4
7	2020-03-04	5	Mass Transit Restoration	3
8	2020-03-04	5	Missing Adult	1
9	2020-03-04	5	Public Health	1
10	2020-03-04	5	Road Closure	3
11	2020-03-04	5	Utility	1
12	2020-03-04	5	Weather	2
13	2020-03-05	3	Public Health	1
14	2020-03-05	3	Road Closure	6
15	2020-03-06	7	Mass Transit Disruption	2
16	2020-03-06	7	Public Awareness	2
17	2020-03-06	7	Road Closure	2
18	2020-03-06	7	Utility	1
19	2020-03-06	7	Weather	2
20	2020-03-07	7	Aerial (Fly-Over)	2
--	----	--	- . . . .	-

## 3.2 Data Summary

### 3.2.1 Notification each type since March

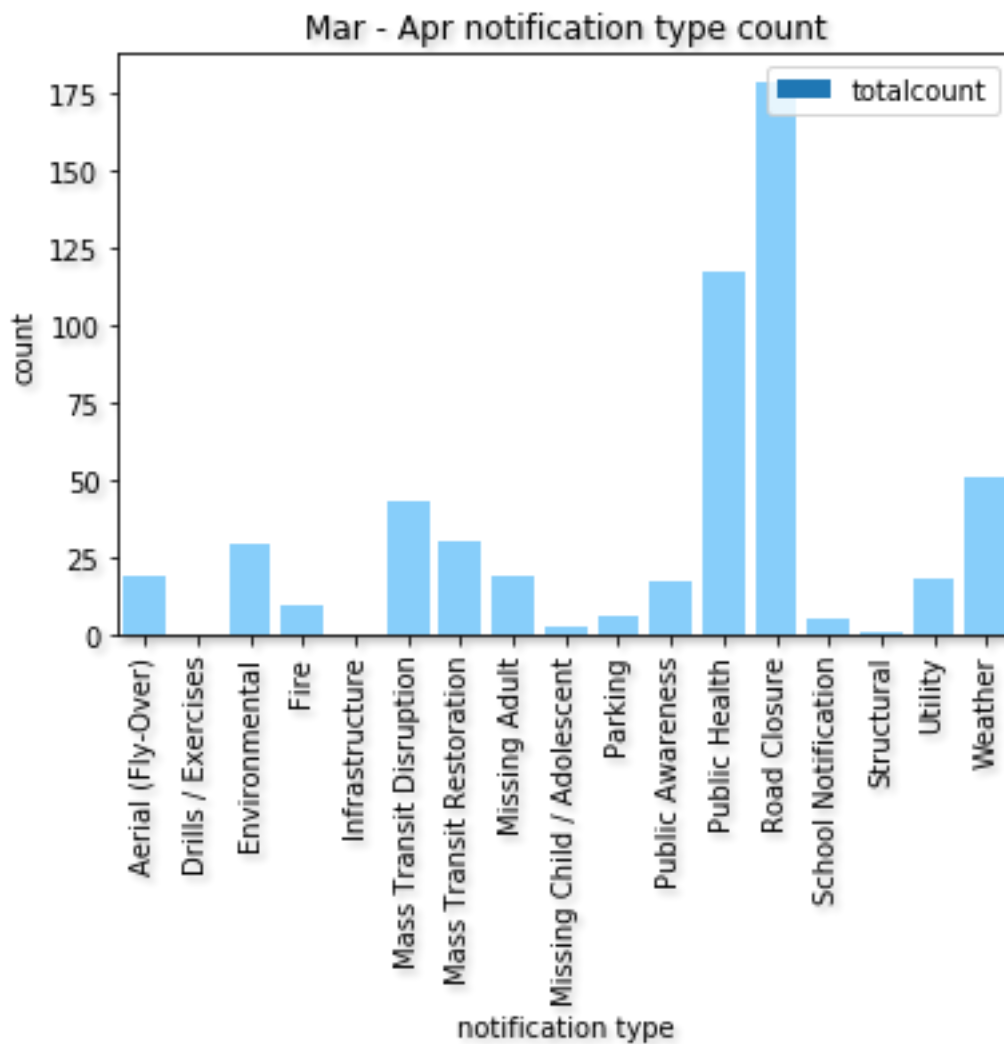


chart 1: Notification each type since March

From the figure above, we can see since the virus broke out in NYC, **Road Closure** type still takes the first place, which is 179 and **Public Health** is following closely with the number of 117. Types like **Drills / Exercises** and **Infrastructure** have zero notifications. **Missing Child / Adolescent** and **Parking** also take a small number which is 2 and 6.

### 3.2.2 COVID-19 confirmed cases since March

Following is the line chart shows the trend of confirmed Covid-19 case in NYC

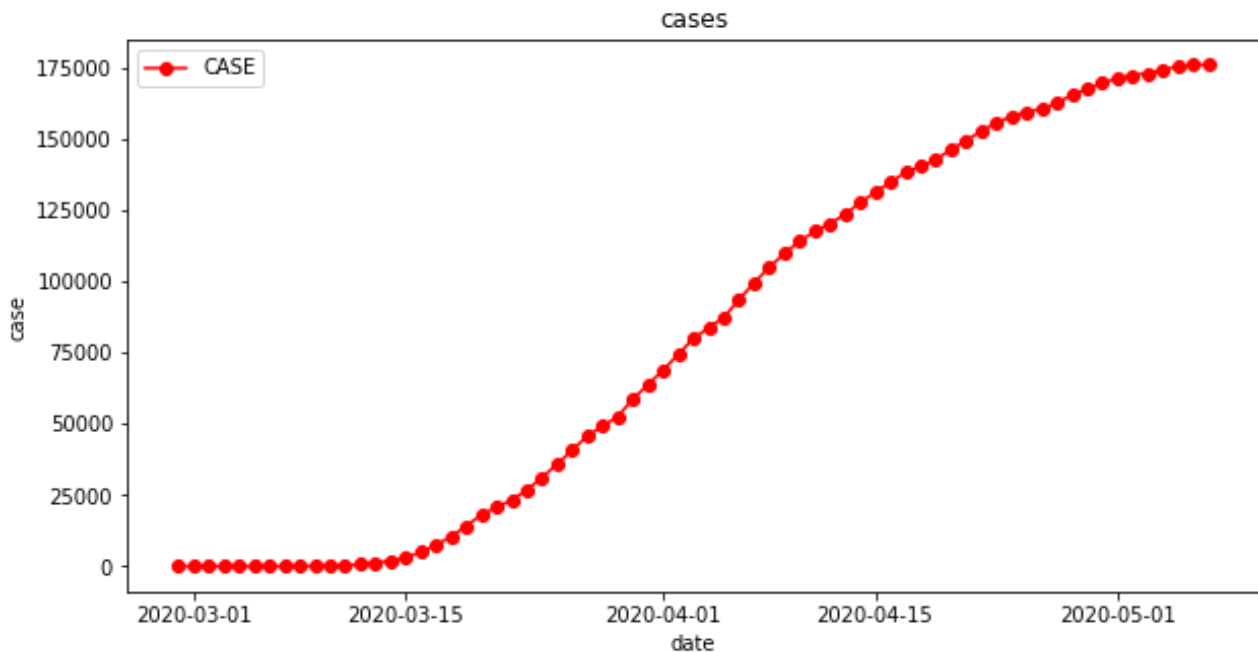


chart 2: Trend of Confirmed COVID-19 in NYC since March

Chart 2 above directly presents an exponentially growth trend of COVID-19 confirmed cases in New York City during March and April. It gradually flattens when step into May.

Also, when looked into the new confirmed case each week as shown in below (chart 3), there is a rapid increasing poly line since mid March and then drop down after mid April. Thus, it can be conclude that the outbreak of COVID-19 in NYC is during March and April, which laid a time foundation for our subsequent research.

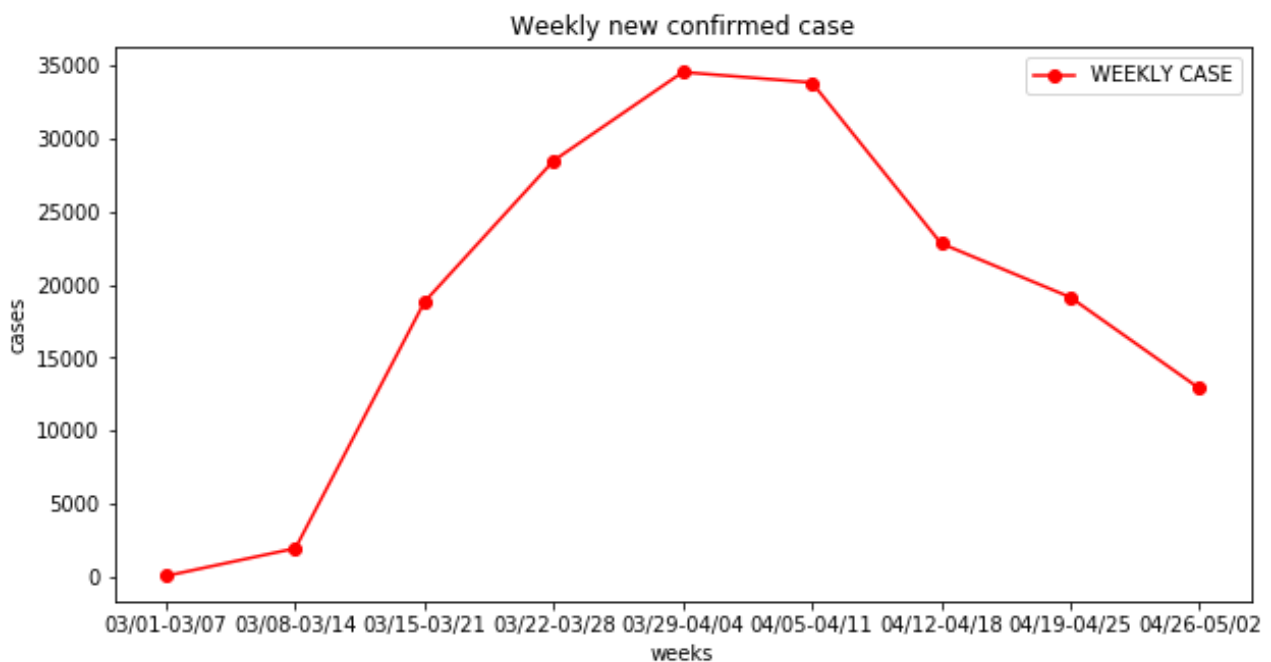


chart 3: Weekly new confirmed case since March



## 4. Methodology and Findings

### 4.1 Fluctuation for each type of notifications during epidemic

#### 4.1.1 How COVID-19 affected the number of different types of notification monthly

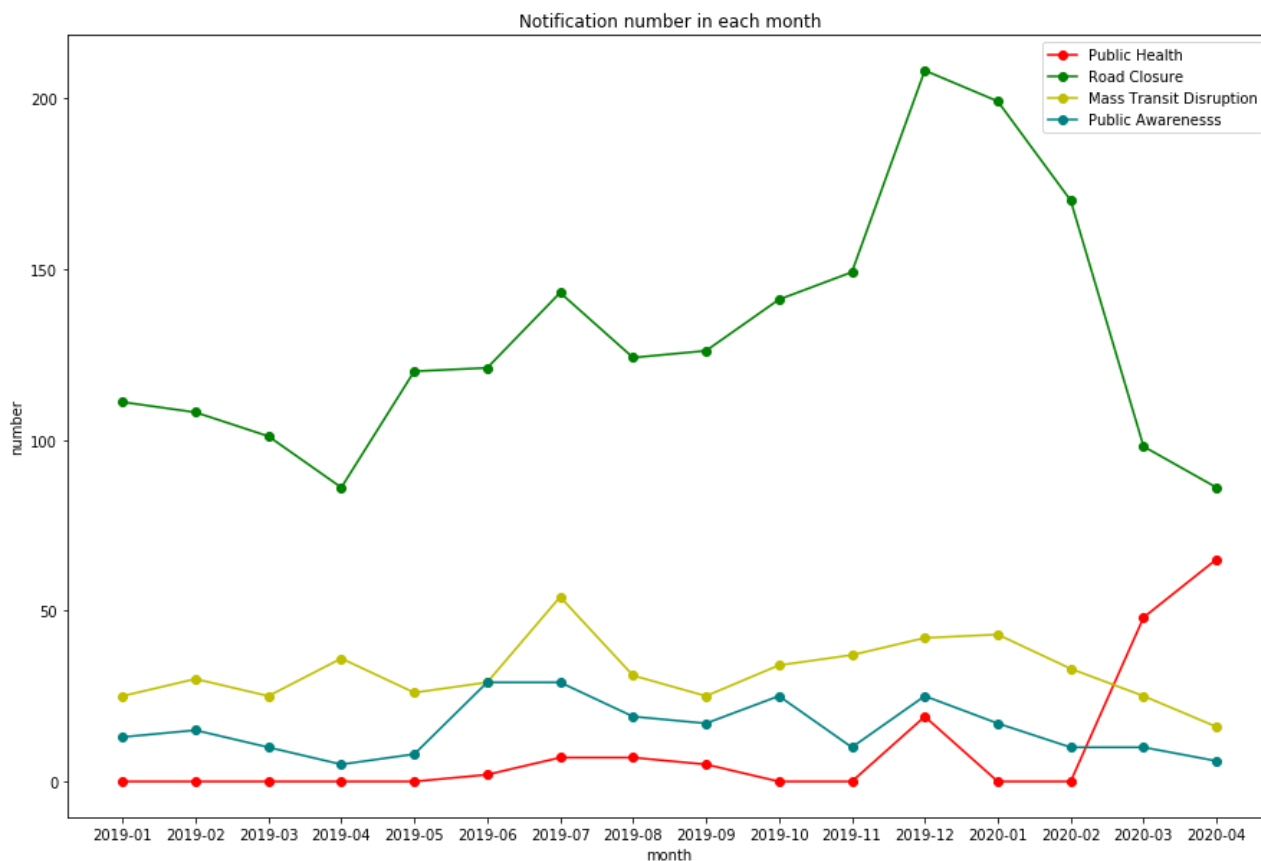


chart 4: Notification number in each month

This line chart shows the numbers of different types of notification from January 2019 to April 2020.

Chart 4 shows that the number of *Road Closure* has decreased dramatically since January 2020. Moreover, *Public Health* began to rise significantly in February 2020, and far more than the same period in 2019 because of COVID-19. 95.45% of public health notification titles have 'COVID-19'. Also, the number of *Mass Transit Disruption* and *Public Awareness* starts to decrease since January 2020 and December 2019, and are less than the same period last year.

#### 4.1.2 How COVID-19 affected the number of different types of notification monthly

Aggregating different types of notification is adopted because there are few emergency notifications per day and more than 60 days need to be analyzed.

First, from the chart 1 in Data summary, it is pronounced that **Road Closure** is always accounted for the largest proportion in all of the notification types before March 2020. However, as the outbreak of COVID-19 shown in chart 3 - a surge in the number of newly confirmed cases from 3/9/2020 to 3/22/2020, **Public Health** surpass **Road Closure** to become the largest proportion from 3/22/2010 to 3/28/2020 shown in the following chart 5. Continually, in the next seven days, Public Health and Road Closure are almost equal, which are far more than other types.

Second, as chart 3 shows a downtrend of new confirmed cases from 4/6/2020, the proportion of **Public Health** shown in chart 5 starts to decline at the same time. In conclusion, **Public Health** correlated to new confirm COVID-19 cases positively.

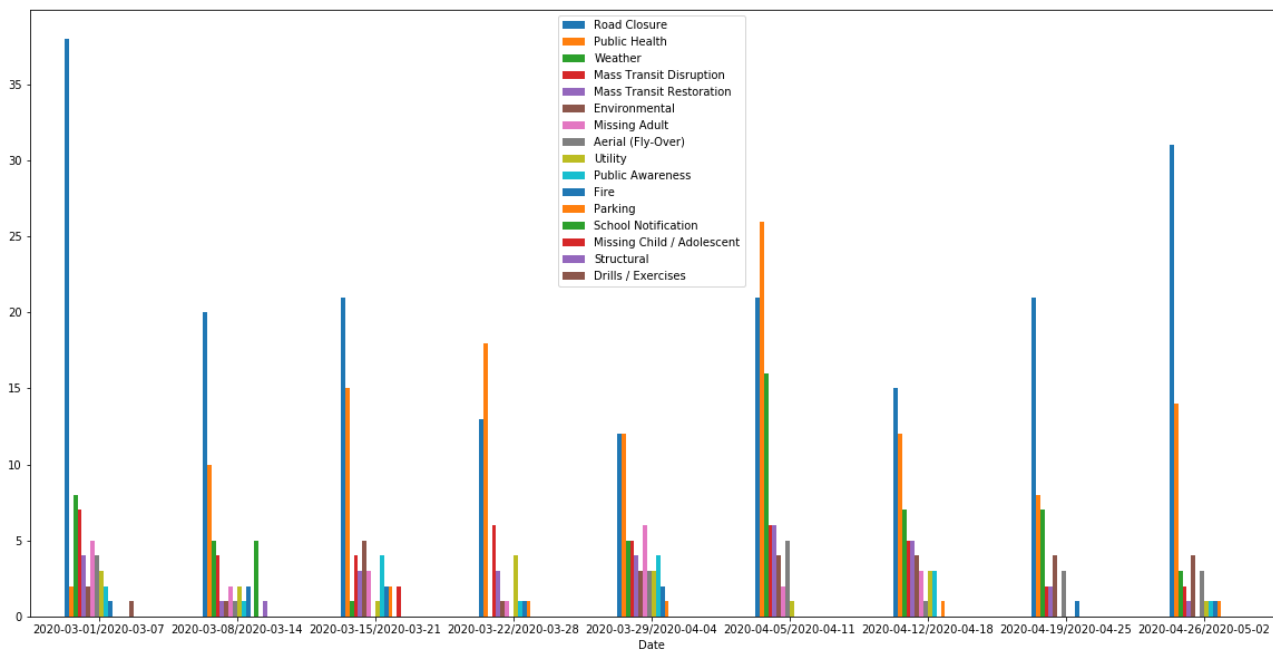


chart 5: Weekly notification type quantity comparison from 2020-03-01 to 2020-05-02

#### Challenge here:

In order to compare the trend of different types of emergency notification every seven days, at first we drew a histogram chart of every seven days separately. However,

---

we found it not distinct for comparison, besides the x-axis coordinates are always in a different order in those charts, which made the comparison even harder.

Because the number of emergency notifications are different every 7 days, so we can not put them in one data frame and use pandas to draw the histogram chart. So we need to get the number of different types of emergency notification in the data frame in advance and store them in a new data frame and use the start date and end date of the seven days as the index. And draw a chart of the frequency of each notification type every 7 days.

## 4.2 Comparisons between 2019's first quarter and 2020's first quarter

### Proportion of each notification type in the first quarter, 2019

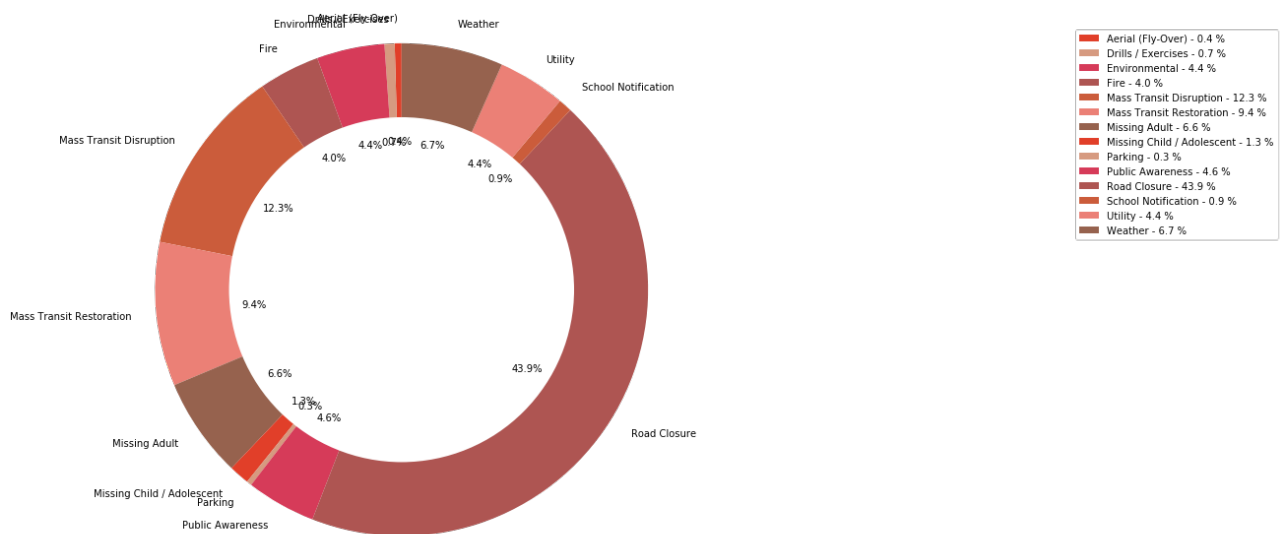


chart 6: Proportion of each notification type in the first quarter, 2019

The figure above shows how much percentage each type takes in all the notifications in 2019.01- 2019.04, we can see that **Road Closure** takes the most, which is 43.9%, **Mass Transit Disruption** takes the second place, which is 12.3%, **Mass Transit Restoration** takes the third, **Missing Adult** takes 6.6% and **Public Health** is 0 which is not even on the graph.

## Proportion of each notification type in the first quarter, 2020

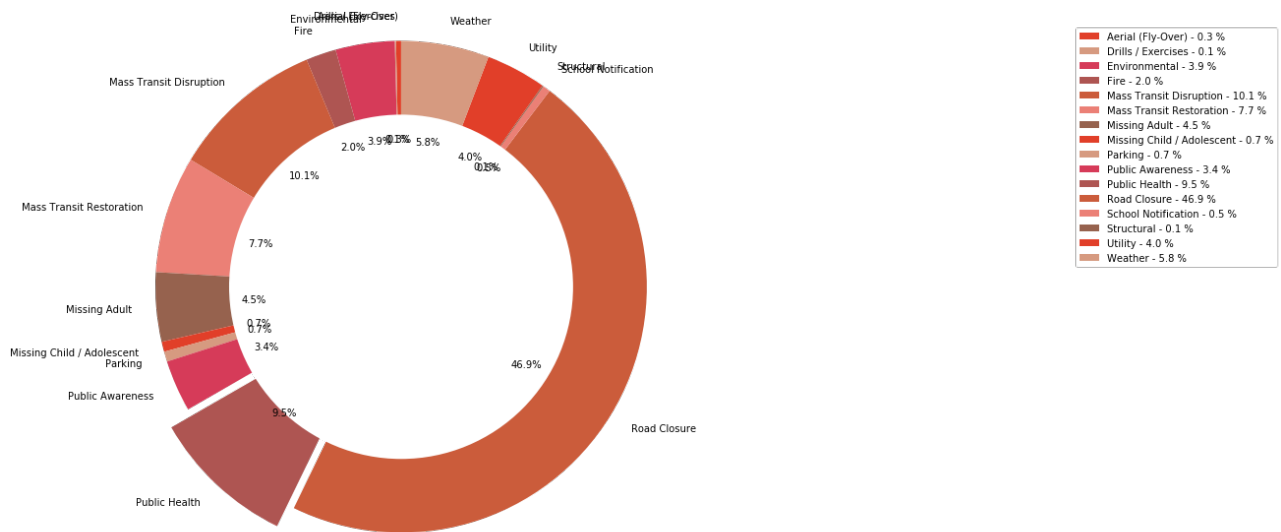


chart 7: Proportion of each notification type in the first quarter, 2020

And here is the type proportion in 2020.02-2020.04, we could tell that **Road Closure** still takes the largest proportion and **Mass Transit Disruption** and **Mass Transit Restoration** still follows. However, **Public Health** comes forth and takes the third place, taking 9.5%, which also fits the conclusion in chart above. **Missing Adult** drops to 4.5%.

Thus we can derive a table showing how the number of each type of notification changes from 2019's first quarter to 2020's first quarter.

	notificationtype	total_count_20	total_count_19	difference
1	Drills / Exercises	1	6.0	-5.0
2	Environmental	47	40.0	7.0
3	Fire	24	36.0	-12.0
5	Mass Transit Disruption	122	111.0	11.0
6	Mass Transit Restoration	93	85.0	8.0
7	Missing Adult	55	59.0	-4.0
8	Missing Child / Adolescent	8	12.0	-4.0
9	Parking	8	3.0	5.0
11	Public Health	115	0.0	115.0
12	Road Closure	567	395.0	172.0
13	School Notification	6	8.0	-2.0

14	Structural	1	0.0	1.0
15	Utility	48	40.0	8.0
16	Weather	70	60.0	10.0

table 5: Notification comparison between 2019's first quarter and 2020's first quarter

And we can put the number of notifications in 2019 and in 2020 in the same graph so that we can see a more intuitive comparison:

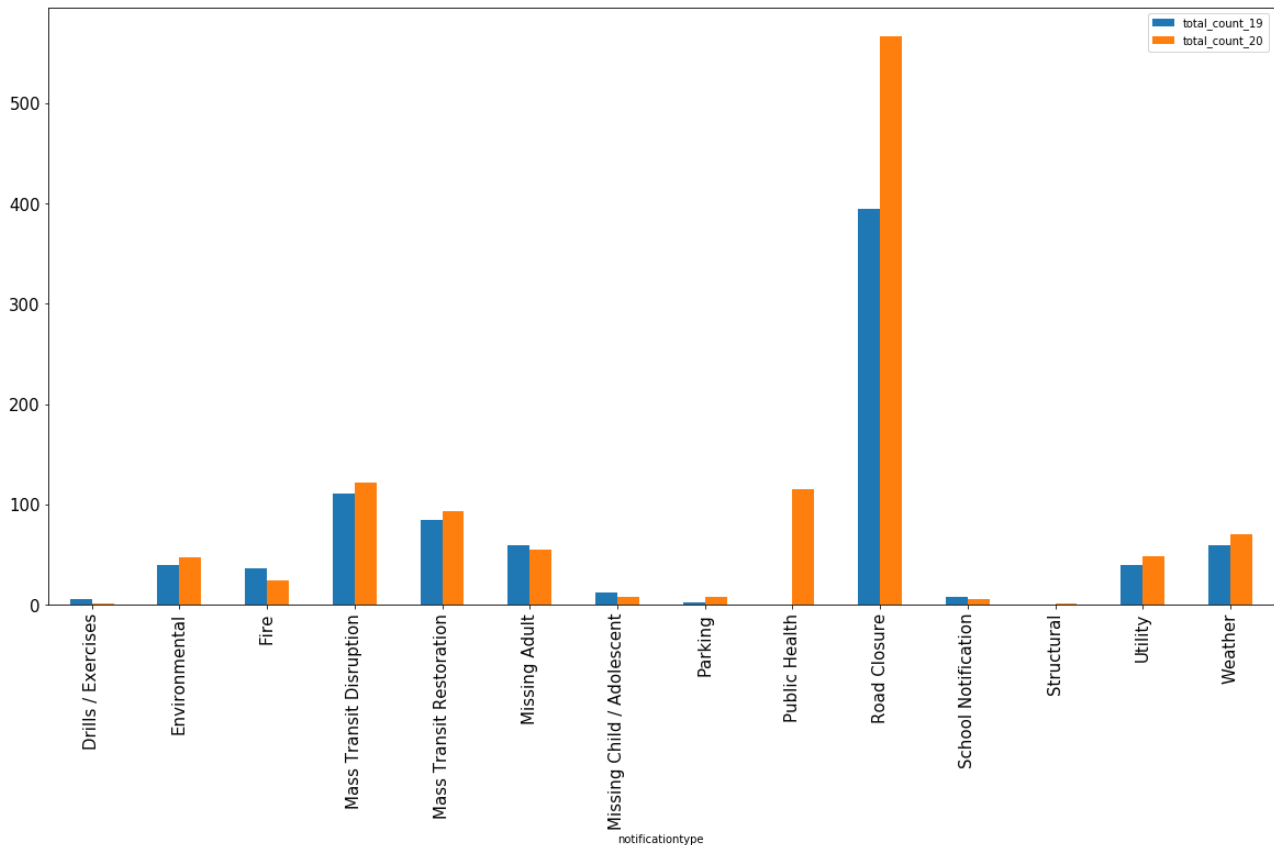


chart 8: Notification quantity comparison between 2019's first quarter and 2020's first quarter

From the graph we can see notifications in types like *Environmental*, *Mass Transit Disruption*, *Mass Transit Restoration*, *Parking*, *Public Health*, *Road Closure*, *Utility*, *Weather* and *Structure* increased. Especially, *Public Health* and *Road Closure* rise dramatically. Meanwhile, *Drills/Exercise*, *Fire*, *Missing Adult*, *Missing Child / Adolescent* and *School* Notifications drops.

---

## 4.3 Notifications related to COVID-19 exploration

### 4.3.1 Type distribution

Since the data summary that draws a basic overview of the COVID-19 and the OEM notification, it can be concluded that “Public Health”, “Road Closure” and “Mass Transit Disruption” had a significant change. Therefore, we decided to further explore notification contents

To look for notifications related to COVID-19, we searched keywords like ‘COVID19’, ‘COVID-19’, ‘Coronavirus’ or ‘COVID’ in notification title and email body.

Surprisingly, it appears that not all COVID related notifications are in Public Health type, showing the types and the distribution.

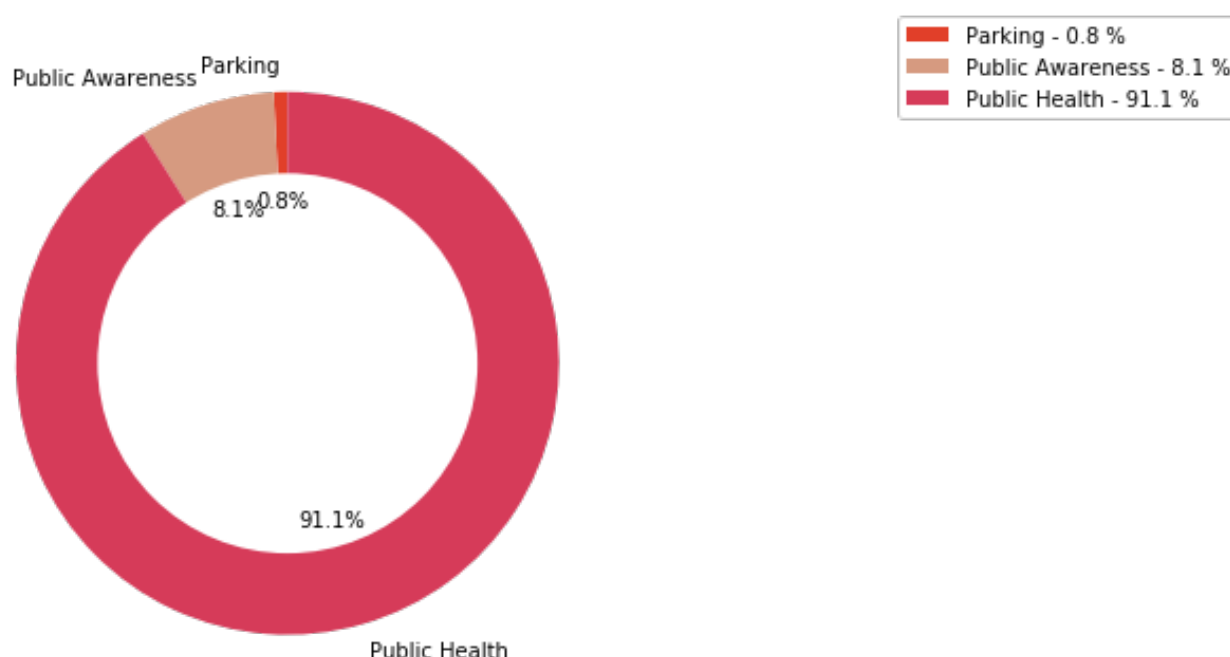


chart 9: COVID-19 keywords distribution in notification type

There are three types of notifications had ever mentioned COVID-19 and 91.1% of them are in *Public Health* type, 8.1% of them are *Public Awareness*, and only 0.8% are *Parking* type.

### 4.3.2 Quantitative relationship with confirmed cases

Whether there is any connection between confirmed cases and notifications that related to COVID-19 keywords?

We organized data by week since 2020-02-28 (the day NYC reported the first case) and visualized them respectively in bar charts.

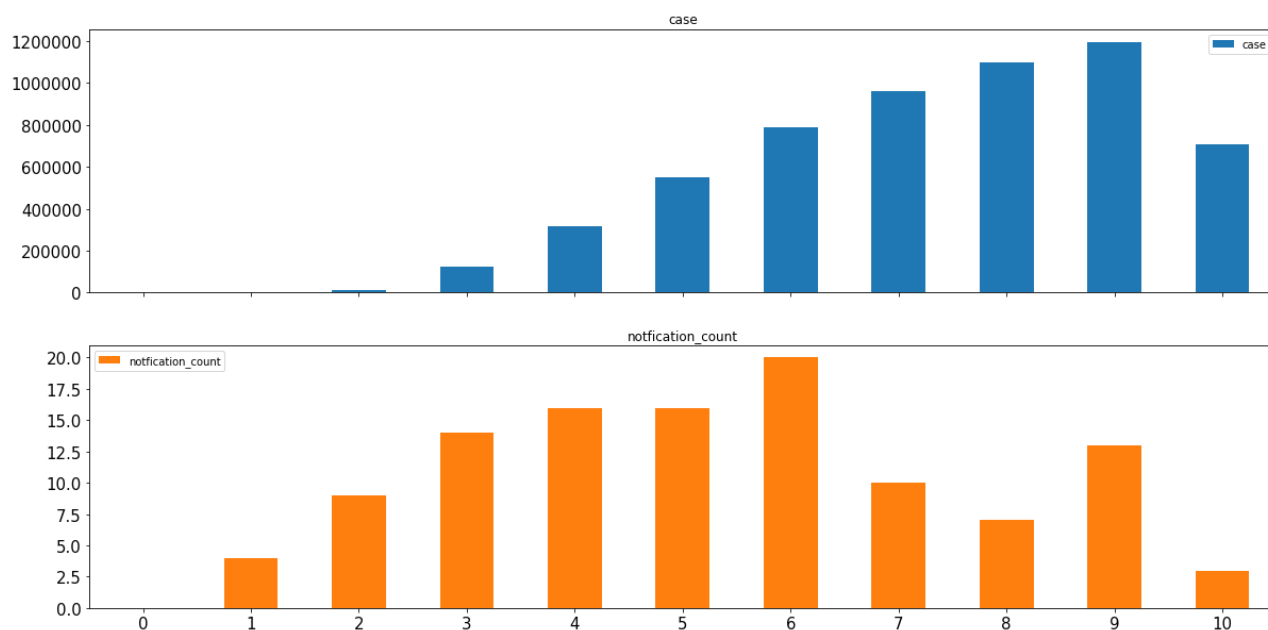


chart 10: Quantitative relationship between confirmed cases and associated notifications

In The first several weeks, with the number of case increasing exponentially, the number of COVID-notifications also increases, however after the sixth week, the notification quantity drops slightly.

### 4.3.3 Other Observations

**What is the proportion of COVID related notifications in all Public Health notifications since the break out?**

From the dataset, we can derive there are 117 notifications in type *“Public Health”* since the epidemic and thereinto there are 112 of them mentioned COVID related keywords, thus the proportion is 95.73%.

**The mayor of New Jersey Michael Melham believes he had COVID-19 back in November last year and it noticed that there are public health notifications in 2019.12, so are those notifications about coughing or fever?**

No. We filtered our data and found out that all *Public Health* notifications in December last year were about a sewage incident in Queens, nothing is about flu or cough.



---

## Are there any COVID-related notifications about daily life or welfare?

We search for keywords like 'buy', 'shop', 'grocery', 'food', 'meal', 'free', 'delivery' and 'price' in our COVID data, and found 29 notifications contains at least one or more keywords above, which takes 26% proportion.

## Are there any COVID-related notifications about facial covering or social distancing?

There are 13 notifications about telling people to cover their faces and 5 about must keep social distance.

Besides those, there are 4 notifications about violence during epidemic and 20 about helping the community like donating or volunteering.

## Does the COVID-19 bring any impact on the traffic according to the notifications?

We can see from notifications that almost all the road closure can be ascribed to vehicle collisions, thus we search for "collisions" in the email body, we found the number of collision in the first four months in 2019 and 2020 are as below:

```
monthly(2019)
```

```
Jan:53  
Feb:64  
Mar:50  
Apr:50
```

```
monthly(2020)
```

```
Jan:119  
Feb:112  
Mar:61  
Apr:47
```

Even though the number of collisions in March and April, 2020 are still more than in 2019, but we can see that it drops almost 1/2 of Jan and Feb, 2020. Thus we can conclude that due to COVID-19, the number of vehicle collisions decreased a lot, maybe because of people started to stay at home instead of driving around a lot.

---

## 5. Conclusion

**After the analysis of our data, we found following facts about emergency notifications in 2020 NYC:**

Road Closure takes the largest proportion of all notifications and Public Health comes next. Drills / Exercises and Infrastructure takes the least, there were even no emergencies in these types.

**Some interesting findings from comparison among 2019 and 2020:**

After the outbreak of coronavirus, we found the number of Public Health emergencies increased significantly, which had never published in 2019's first quarter, now it takes the third place in proportion of all notifications.

Besides, in March, the number of school notifications also increased due to the suspended class policy, and the number of Mass Transit Disruption drops may also due to much fewer people taking public transportation due to the epidemic.

By exploring notifications related to COVID-19, we can also find that big portion of Public Health notifications are about the coronavirus and some of them are about daily life and some of them are asking for helping the community.

From the road closure notifications, we can see that the number of vehicle collisions in March and April, 2020 had dropped about a half compared to January and February this year. Thus we can conclude people started to stay at home instead of driving around.