The effect of COVID-19 on the NYC taxi demand

Introduction

COVID-19, as an epidemic with a wide impact around the world, is closely related to our life with the recent. To keep up with the current events, I would like to explore the effect of it on the For-Hire industry, which to be specific, people's demand for taxi in New York City. This report will examine the effect of COVID-19 by comparing the trip volumes, trip trend, taxi demand, trip distribution, drivers earnings and industry coverage before and during the epidemic. Also, I will explore the different between several of the most common used vehicle types, which are medallion (yellow) taxis, street hail livery (green) taxis and for-hire vehicles (FHVs) (TLC Trip Record Data).

Data Wrangling

According to the purpose of this report, I initially choose data from January-June, 2019 and January-June, 2020 and from datasets for Yellow Taxi, Green Taxi and FHV (It should be noted that HFHV datasets are not used as all attributed we need is in FHV datasets). Generally speaking, the chosen of months and years is to ensure that the experimental data could be least influenced by any lurking factors. As COVID-19 approximately began at 2020, we can roughly regard Year 2020 as the 'COVID-19 Year'. Moreover, 2019, as the closest year of 2020, can avoid the gap caused by different years as much as possible. Similarly, the same months are chosen to avoid the differences by seasons or off-season travel. Therefore, I initially present the line chart to check the veracity of the chosen data (Fig 1). Average daily trips for each months are chosen. It justifies the through of taxi's demand shows sign at February and touches bottom at April, then rising modestly from April to June. Therefore, data after February is worth paying more attention to.

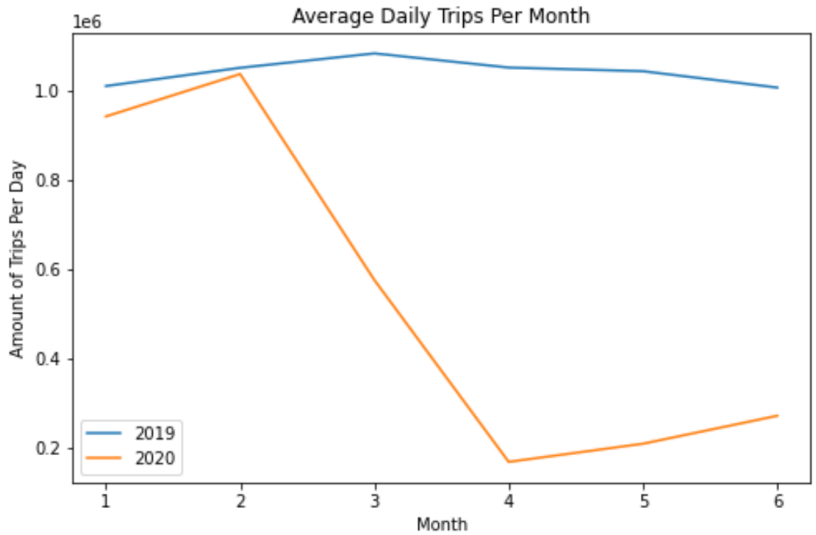

Fig. 1

As what I mentioned before, in order to analyze the changes the taxi recommend, the following attributes are used.

[1]. *PULocationID & DOLocationID*

These attributes are IDs for pickup location and drop off location corresponding to the "taxi zone" file. In this file, we can find 'ID'(as index), 'Borough', 'Zone' and 'Sevice_zone' in order to locate the taxi pickup or drop off location with the zone on the map. These attributes are assumed to use for the trip volumes comparison between zones. Also, same zones will be compared through time (2019-2020) to verify the effect of COVID-19. The attributes are common for all sectors thus we can also compare between different vehicles. It is worth mentioning that they are the only attributes we use for FHV thus for other attributes, we can use only the summary data from "monthly report" or the data from Yellow taxi and Green taxi.

[2]. *total_amount*

This attribute records the total amount of all tolls passenger paid in trip. It will be used to measure the driver earnings and how much drivers loss in the epidemic. It is essentially stressed that the attribute could only be used to measure the driver earningss but not the net income of drivers. We are not considered the cost like car maintenance, license application expense, fuel expense and etc.. What we mainly consider is how much less drivers earn because of COVID-19.

[3]. *Trips Per Day*

This attribute records the summary data of the trips per day by each month. It is supposed to provide information about the comparison between sectors and comparison between month to month for each sector.

[4]. *lpep_pickup_datetime, tpep_pickup_datetime & pick_up_time*

These attributes provides daily time information in the form of 'YY/MM/DD Hour:Minute:Second AM/PM'. They are for Green Taxi, Yellow Taxi and FHV correspondingly. They are used to group data in any time span needed.

Besides feature selection, other data pre-processing steps are also applied. For attribute '*Trips Per Day*' and '*Vehicles Per Day*' (in Monthly Data Report), whose datatype is *string*, they are transferred into *int* datatype for further plotting. Missing data in the initial dataset is dropped as losing information in any area makes the data useless, while for datasets after preprocessing, there are still some data processing steps specifically for certain preprocessed datasets and plots, which will be introduced later in the visualization part when the datasets are used.

There is no what is called 'outlier'. Any data items which are regarded as uncommon provides variety to the data. All data is checked to be in the datatype corresponding to the datatype it ought to be. For instance, for attribute *PULocationID,* the datatype is *int*, thus any data in this area are checked to be integer datatype. Moreover, data is checked to be in the region it ought to be. For datetime attributes (lpep_pick_up_datetime, tpep_pick_up_datetime and pick_up_time), data is remained only if it conforms to the year specified by the dataset, i.e. if the dataset is for year 2019, data in it can only from 2019/01/01 12:00:00AM to 2019/12/31 11:59:59PM; for ID attributes (PULocationIID and DOLocationID), data need to be not negative; for quantity attributes (trip_distance and total_amount), data is not negative. Inconsistent data are deleted as they are not able to provide valid information.

Visualization

Average Daily Trips Per Month by Sector

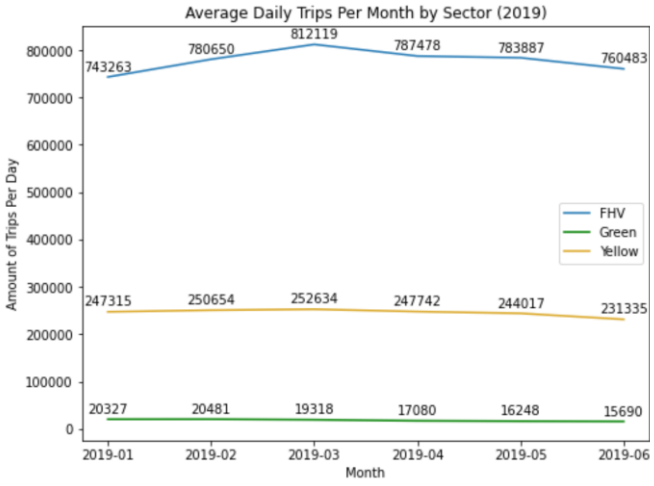

Fig. 2(1)

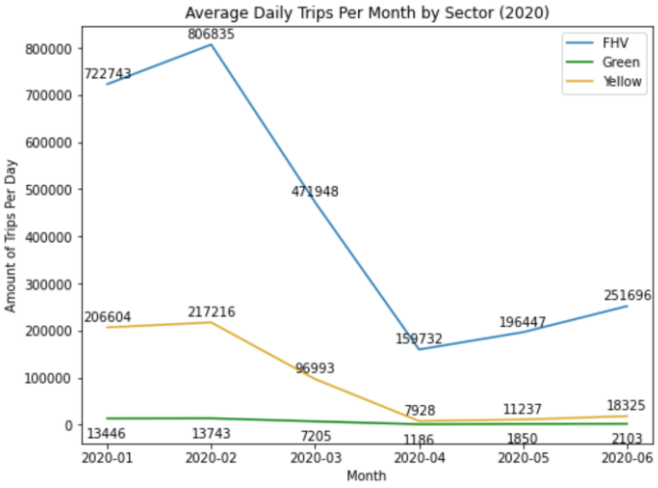

Fig. 2(2)

Fig. 2(1) and Fig. 2(2) displays the trip declines. The epidemic has much effect on the demands for Yellow Taxis and High-Volume For-Hire Vehicles (FHVs). In January and February, the amount of trips for FHVs are above 700,000, and the amount of trips for Yellow Taxis are about 200,000, which are quite similar to the situation in 2019. From Fig. 2, it may look like no impact on Green Taxis, but it is not the fact.

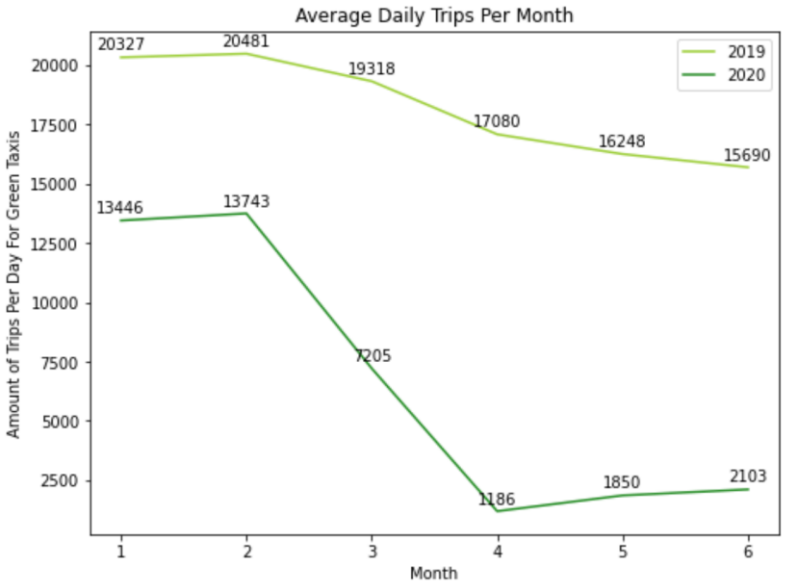

Fig. 3

Fig. 3 presents it has already had a gap in January and February. It may not the effect of the epidemic as a decreasing trend occurs in 2019. Nonetheless, a sharp decline occurs from March. Therefore, Fig.2 and Fig.3 displays the decline in all three sectors beginning at February and reached the bottom in April (160,000, 1,100, 8,000 respectively). From April, there is a slightly increase.

Trip trends

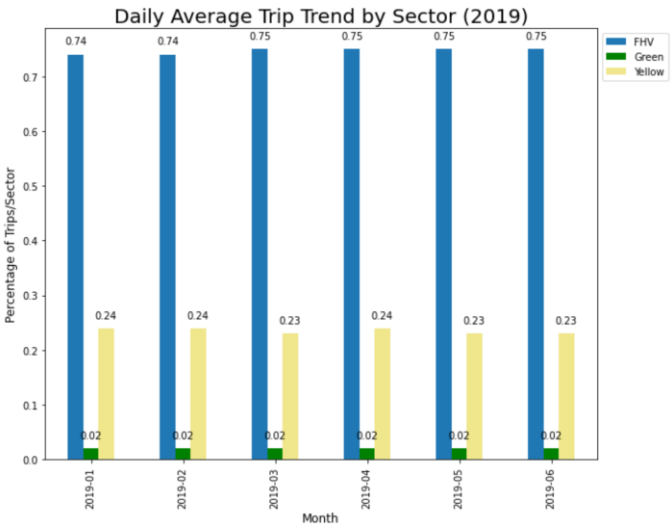

Fig. 4(1)

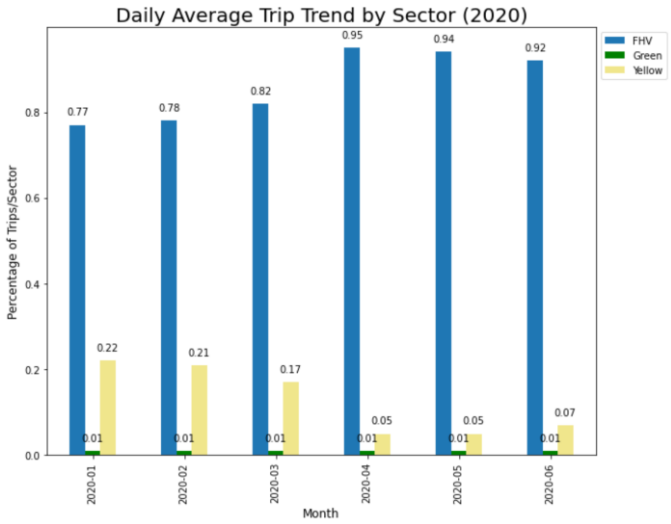

Fig. 4(2)

From Fig. 4(1) and Fig.4(2), the change of taxi demand structure is exposed. In 2019, the demand structure is at a stable state, with FHV accounting for about 75%, Yellow Taxi for 23% and Green Taxi for 2%. Whereas in 2020, from March, an increasing number of people prefer FHVs rather than the others. From February to March, FHV takes 4% of the demand of Yellow Taxi (78%, 1%, 21% to 82%, 1%, 17% respectively). From March to April, the demand of FHV rises again. It accounts for 95% of the market, increasing 13% from last month and keeps stable at about 94%. At the same time, the demand of Yellow Taxi decreased from 17% to 5% and Green Taxi remains at 1%.

Demand of taxis in each zone

The demand of taxis in New York City are measured separately for both pickup locations and dropoff locations. In this part, Green Taxi, Yellow Taxi and FHV are combined, the overall quantity of pickups or dropoffs for each zone are measured according to its location ID. Unlike the processing of initial data, in this part we combine the map file (file in .shp format) with the data about the amount of pickups/dropoffs. Missing data are not handled as they indicate zones uncovered. On the map, uncovered zones are surrounded by light blue edge and filled with grey color with slash label on it. Fig. 5(1) is an example for the uncovered zone symbol. Sequential colormap is used to represent the magnitude of pickups/dropoffs. It changes the lightness. With quantity increase, the color incrementally saturated. It should be noted that we use an opposite qualitative colormap for the ratio plots (Fig. 8). In these plot, we measure the decline of pickups/dropoffs in each region by ratio for Year 2020 compared with Year 2019. Therefore, larger magnitude represents for sharped decline in that zone. To be more specifically, more the measured value close to 0, lighter the filled color is. Also, we use quantile (<-90%, -90%-70%, -70%-50%, >-50%) to discrete quantities. Fig. 5(2) is an example pf the colormap.

Fig. 5(1) Symbols for uncovered data

Fig. 5(2) Colormap

In addition, only data for April, 2020 is used to represent situation in 2020 as April is the worst of epidemic brought to taxi industry, while the average of first six months of 2019 are used as the situation of 2019 to provide generalization of data.

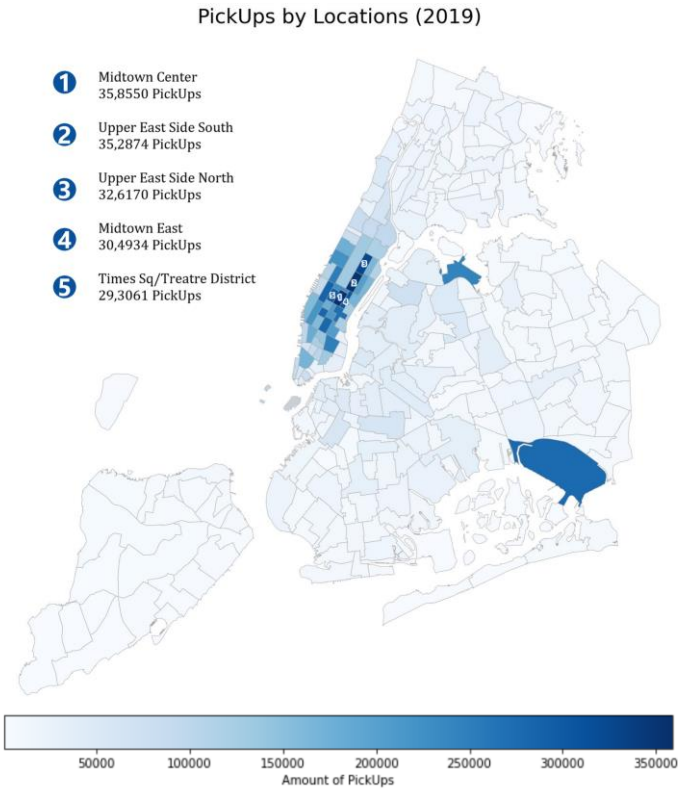

Fig. 6(1)

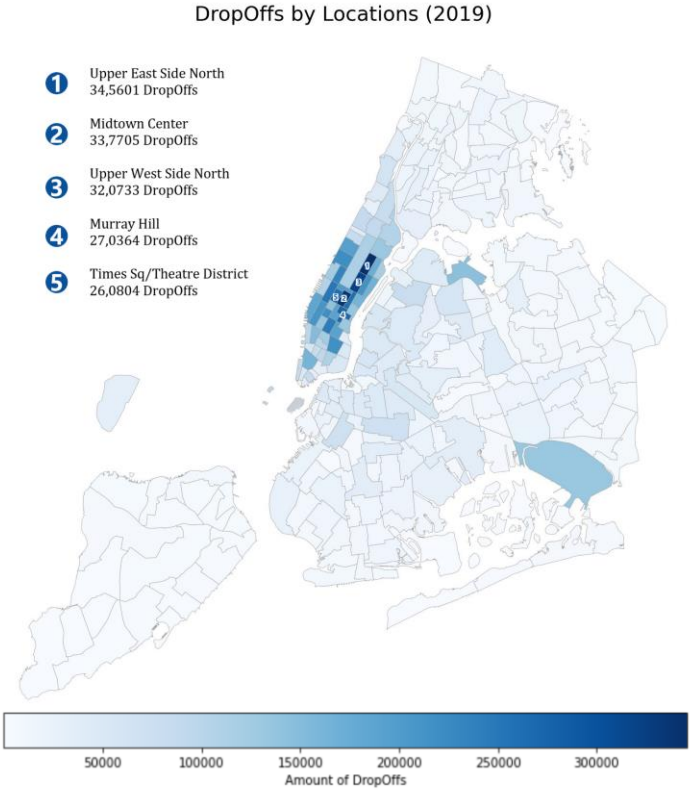

Fig. 6(2)

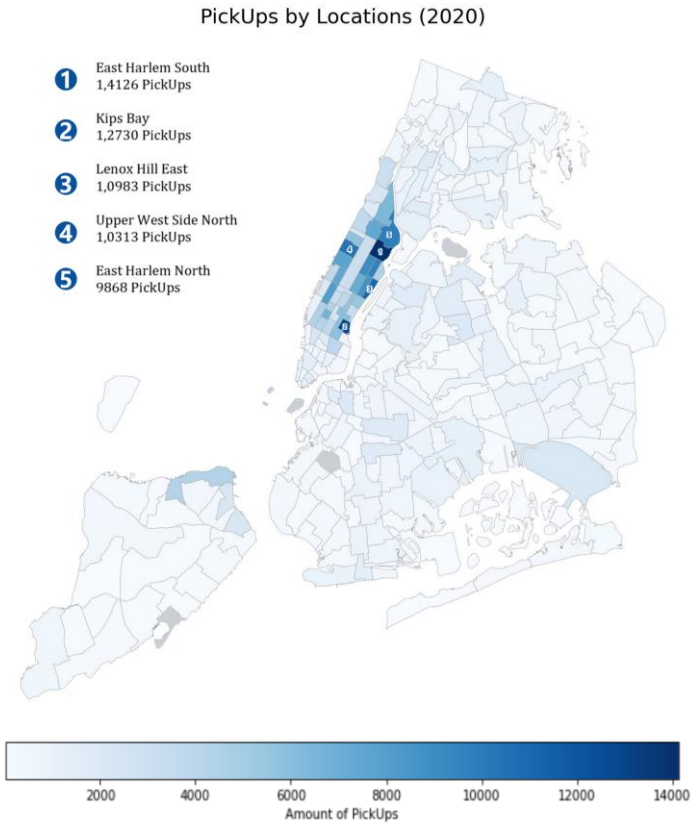

Fig. 7(1)

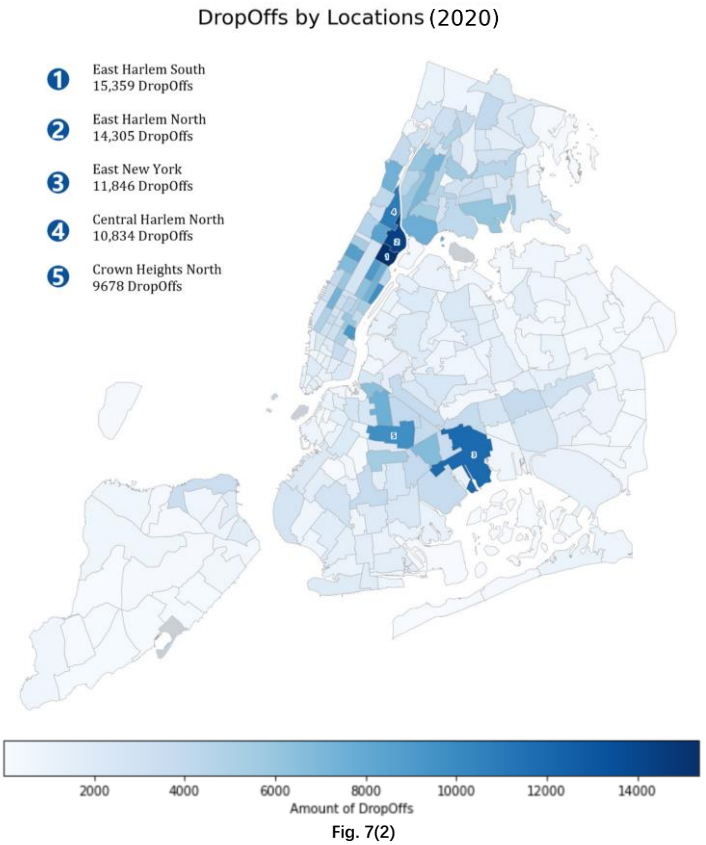

Fig. 7(2)

By each month, there are more than ten millions people in average to take a taxi in 2019. Both pickups and dropoffs are most common in Manhattan. Midtown Center, Upper East Side North and Time Sq/Theatre District are zones in top 5 of both, while Upper East Side South and Midtown East are in top 5 of pickups only and Upper West Side North and Murray Hill are in dropoffs only . Rikers Island, Great Kills Park, Jamaica Bay and Freshkills Park are the zones people least to go. Moreover, Governor's Island/Ellis Island/Liberty Island is a zone no one live and go. The common zones for pickups and dropoffs are quite similar, indicating the mobile population between zones are basically equal.

While in April of 2020, there are only about three hundred and fifty thousand trips, which are 30% of 2019. The most popular zones for pickups and dropoffs are quite different at this time. Only East Harlem South and East Harlem North are in top 5 zones for both. Pickups are often needed in Manhattan as 2019, while dropoffs tend to happen in both Manhattan and Brooklyn. Also, Bronx is another common place people tends to go. Besides Governor's Island/Ellis Island/Liberty Island, no one takes a trip from and to Great Kills Park and Rikers Island.

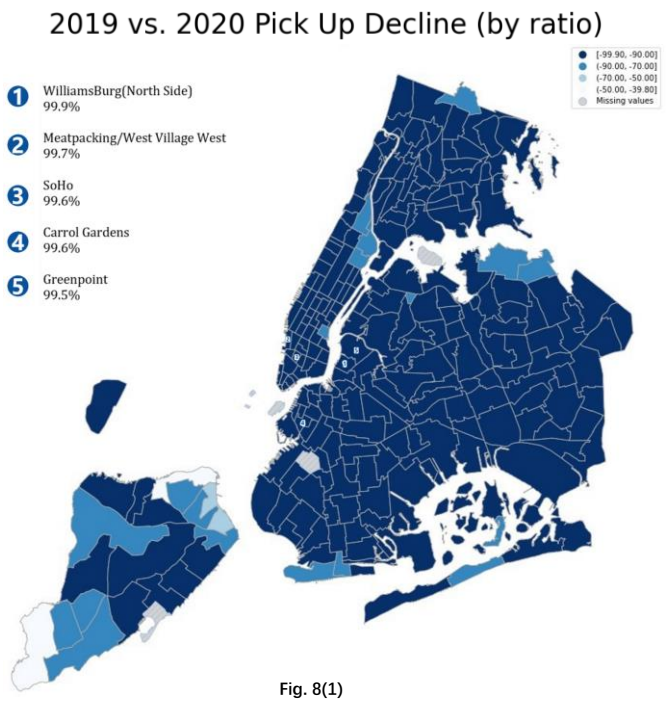

Fig. 8(1)

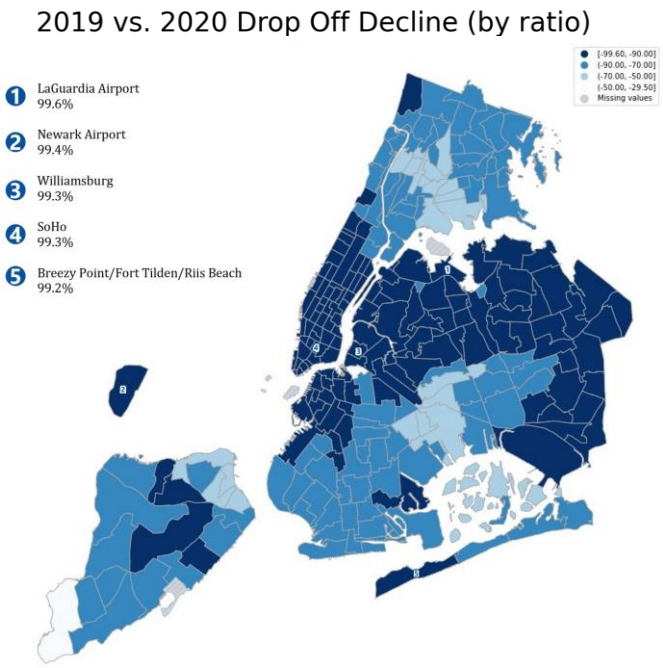

Fig. 8(2)

By the effect of COVID-19, a large amount of people prefer not to take a trip. 241 out of 263 zones reduce above 90% about the amount of pickups compared to data in 2019. Only two zones reduce less than 50% pickups. Most reduce for pickups are happened in Manhattan and Brooklyn. For dropoff locations, 143 zones reduce above 90% and 261 zones reduce more than 50%. Most reduce are happened in Manhattan. It is worth mentioning that the amount of people to go to the airports reduce severely. The reduce ratio for three airports are 99.6%, 99.4% and 99.2%, ranking at number 1, 2 and 6 in the list of dropoff decline measured by ratio.

Trip Distribution

The distribution of trips is measured by the amount of pickups within Boroughs.

Fig. 9(1)

Fig. 9(2)

FHVs, Green Taxis and Yellow Taxis serves New York City's borough differently. In 2019, Manhattan takes up the most distribution of trips compared with other boroughs for all sectors (44.6%, 34.4% and 91.8% respectively), while in 2020, the most distribution for FHV is Queens. Besides, FHV serves much more on Queens, Brooklyn and Staten Island. Meanwhile, Green Taxis distribute more on Manhattan than 2019 and Yellow Taxis distribute similar as 2019.

Driver Earning

In the above parts, we analyze the people's demand on taxis before and within the COVID-19. We find that COVID-19 seriously influence people's desire to travel thus influencing the need of taxis. In this part, we will analyze on the drivers earning and justify whether the epidemic impact on drivers earning and whether the demands of taxis related to drivers earning.

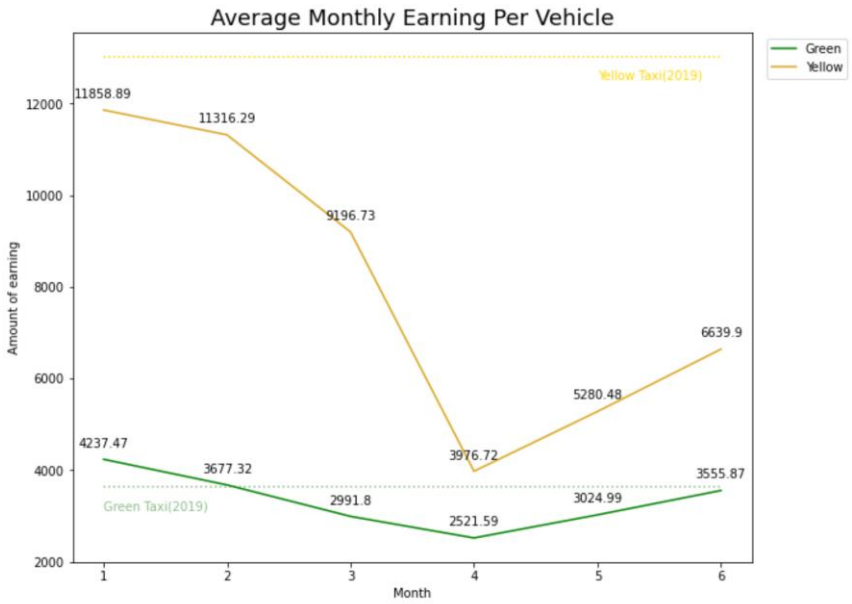

Fig. 10

As data about the amount of driver are not provided, we assume a vehicle represents for a driver. Unfortunately, data about amount of vehicle are not provided for FHV, only Yellow Taxi and Green Taxi datasets for 2020 are processed to compare with the average monthly earning per vehicle in 2019. Also, as I mentioned before, the measure of earning is from the total amount drivers receive of trips. It doesn't matter since only the changes of the earning are what we are looking for. Fig. 10 shows clearly the severe influence of epidemic on driver's income.

As shown in Fig.10, both sectors experiences a sharply decrease from February and a slightly rebound from April. The change pattern is quite similar to the trip trend. Therefore, drivers earning are highly correlated with the amount of trips. The reduction of economic activities directly effects drivers earning. The earning of 2019 are used as baseline. From March to June, the monthly earnings for Green Taxi driver of 2020 are respectively 17.5%, 30.5%. 16.6% and 1.9% lower than them in 2019. Green Taxi driver earnings seem like return to pre-pandemic level. However, things are different when comes to Yellow Taxi driver. Driver's earning are 29.4%, 69.5%, 59.5%, 49.0% corresponding to months from March to June. Although driver's earning has a slightly increase, there is still a considerable earning gap for Yellow Taxi drivers.

From Fig. 10, we can see the trend of drivers earning is similar to the trend of the amount of trips (Fig. 2 & Fig. 3), thus the drivers' earning is positively correlated with the amount of trips.

Summary

COVID-19 has inevitably impact on the taxi industry in Ney York City. This report is a summary of its impact on taxi industry by analyzing people's demands (trip volumes, trip trend, taxi demand and trip distribution) and drivers earnings based on Green Taxi, Yellow Taxi and FHV data from 2019 and 2020. The report is for government, for-hire industry and anyone who is interested on COVID-19.

Based on the report, we can say that COVID-19 has already influenced the industry in any aspect and will keep impact for some time. In such situation, we hope the government and stakeholders could provide some supportive policies like financial support, emergency loans and mental health support as well as providing new job opportunities like delivery business in order to relieve taxi drivers' financial and mental pressure.

Bibliography

TLC Trip Record Data. (n.d.). Retrieved September 05, 2020, from https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page