# Editor’s comments:

When submitting your revision, we need you to address these additional requirements.

## Please ensure that your manuscript meets PLOS ONE's style requirements, including those for file naming. The PLOS ONE style templates can be found at

[https://journals.plos.org/plosone/s/file?id=wjVg/PLOSOne\_formatting\_sample\_main\_body.pdf](https://urldefense.com/v3/__https:/journals.plos.org/plosone/s/file?id=wjVg*PLOSOne_formatting_sample_main_body.pdf__;Lw!!KGKeukY!kucfxaRme5LxitUh5oRFWfbWbKThYIMPJSo5NhND4z1Uk5sRRVTtOjrKWSj7QRjyxMk$) and

[https://journals.plos.org/plosone/s/file?id=ba62/PLOSOne\_formatting\_sample\_title\_authors\_affiliations.pdf](https://urldefense.com/v3/__https:/journals.plos.org/plosone/s/file?id=ba62*PLOSOne_formatting_sample_title_authors_affiliations.pdf__;Lw!!KGKeukY!kucfxaRme5LxitUh5oRFWfbWbKThYIMPJSo5NhND4z1Uk5sRRVTtOjrKWSj7y-wtUI0$)

## In ethics statement in the manuscript and in the online submission form, please confirm that only publicly available data have been used in your study, and that no users' personal information have been accessed or collected.

Yes, all the data used by this paper are publicly accessible. We upload

## We note that you have stated that you will provide repository information for your data at acceptance. Should your manuscript be accepted for publication, we will hold it until you provide the relevant accession numbers or DOIs necessary to access your data. If you wish to make changes to your Data Availability statement, please describe these changes in your cover letter and we will update your Data Availability statement to reflect the information you provide.

## Thank you for stating the following in the Financial Disclosure section:

"The author(s) received no specific funding for this work."

We note that one or more of the authors are employed by a commercial company: "Transit App Inc.,"

a) Please provide an amended Funding Statement declaring this commercial affiliation, as well as a statement regarding the Role of Funders in your study. If the funding organization did not play a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript and only provided financial support in the form of authors' salaries and/or research materials, please review your statements relating to the author contributions, and ensure you have specifically and accurately indicated the role(s) that these authors had in your study. You can update author roles in the Author Contributions section of the online submission form.

Please also include the following statement within your amended Funding Statement.

“The funder provided support in the form of salaries for authors [insert relevant initials], but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the ‘author contributions’ section.”

If your commercial affiliation did play a role in your study, please state and explain this role within your updated Funding Statement.

b)  Please also provide an updated Competing Interests Statement declaring this commercial affiliation along with any other relevant declarations relating to employment, consultancy, patents, products in development, or marketed products, etc.

Within your Competing Interests Statement, please confirm that this commercial affiliation does not alter your adherence to all PLOS ONE policies on sharing data and materials by including the following statement: "This does not alter our adherence to  PLOS ONE policies on sharing data and materials.” (as detailed online in our guide for authors [http://journals.plos.org/plosone/s/competing-interests](about:blank)) . If this adherence statement is not accurate and  there are restrictions on sharing of data and/or materials, please state these. Please note that we cannot proceed with consideration of your article until this information has been declared.

Please include both an updated Funding Statement and Competing Interests Statement in your cover letter. We will change the online submission form on your behalf.

Please know it is PLOS ONE policy for corresponding authors to declare, on behalf of all authors, all potential competing interests for the purposes of transparency. PLOS defines a competing interest as anything that interferes with, or could reasonably be perceived as interfering with, the full and objective presentation, peer review, editorial decision-making, or publication of research or non-research articles submitted to one of the journals. Competing interests can be financial or non-financial, professional, or personal. Competing interests can arise in relationship to an organization or another person. Please follow this link to our website for more details on competing interests: [http://journals.plos.org/plosone/s/competing-interests](about:blank)

## Please amend either the title on the online submission form (via Edit Submission) or the title in the manuscript so that they are identical.

## Please remove your figures from within your manuscript file, leaving only the individual TIFF/EPS image files, uploaded separately.  These will be automatically included in the reviewers’ PDF.

## Please ensure that you refer to Figure 4 in your text as, if accepted, production will need this reference to link the reader to the figure

## We note you have included a table to which you do not refer in the text of your manuscript. Please ensure that you refer to Table 1 in your text; if accepted, production will need this reference to link the reader to the Table.

**Comments to the Author**  
  
1. Is the manuscript technically sound, and do the data support the conclusions?  
  
The manuscript must describe a technically sound piece of scientific research with data that supports the conclusions. Experiments must have been conducted rigorously, with appropriate controls, replication, and sample sizes. The conclusions must be drawn appropriately based on the data presented.

Reviewer #1: Yes

Reviewer #2: Yes

2. Has the statistical analysis been performed appropriately and rigorously?

Reviewer #1: No

Reviewer #2: Yes

3. Have the authors made all data underlying the findings in their manuscript fully available?  
  
The [PLOS Data policy](https://urldefense.com/v3/__http:/www.plosone.org/static/policies.action*sharing__;Iw!!KGKeukY!kucfxaRme5LxitUh5oRFWfbWbKThYIMPJSo5NhND4z1Uk5sRRVTtOjrKWSj7fQGGbiY$) requires authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception (please refer to the Data Availability Statement in the manuscript PDF file). The data should be provided as part of the manuscript or its supporting information, or deposited to a public repository. For example, in addition to summary statistics, the data points behind means, medians and variance measures should be available. If there are restrictions on publicly sharing data—e.g. participant privacy or use of data from a third party—those must be specified.

Reviewer #1: No

Reviewer #2: Yes

4. Is the manuscript presented in an intelligible fashion and written in standard English?  
  
PLOS ONE does not copyedit accepted manuscripts, so the language in submitted articles must be clear, correct, and unambiguous. Any typographical or grammatical errors should be corrected at revision, so please note any specific errors here.

Reviewer #1: Yes

Reviewer #2: No

5. Review Comments to the Author  
  
Please use the space provided to explain your answers to the questions above. You may also include additional comments for the author, including concerns about dual publication, research ethics, or publication ethics. (Please upload your review as an attachment if it exceeds 20,000 characters)

Reviewer #1: This paper studied an interesting and timely research question regarding the transit demand change during the COVID-19 pandemic. The authors employed the data from Transit App to capture transit demand and derived various indexes to describe the change patterns. Overall, this study offers timely data analytics to monitor transit demand during COVID-19. However, there are still several notable concerns with this paper. Detailed comments follow:  
  
The methodological contribution of this paper is limited. Most analyses conducted in this study are descriptive, and the whole paper lacks convincing and strict model build and description:

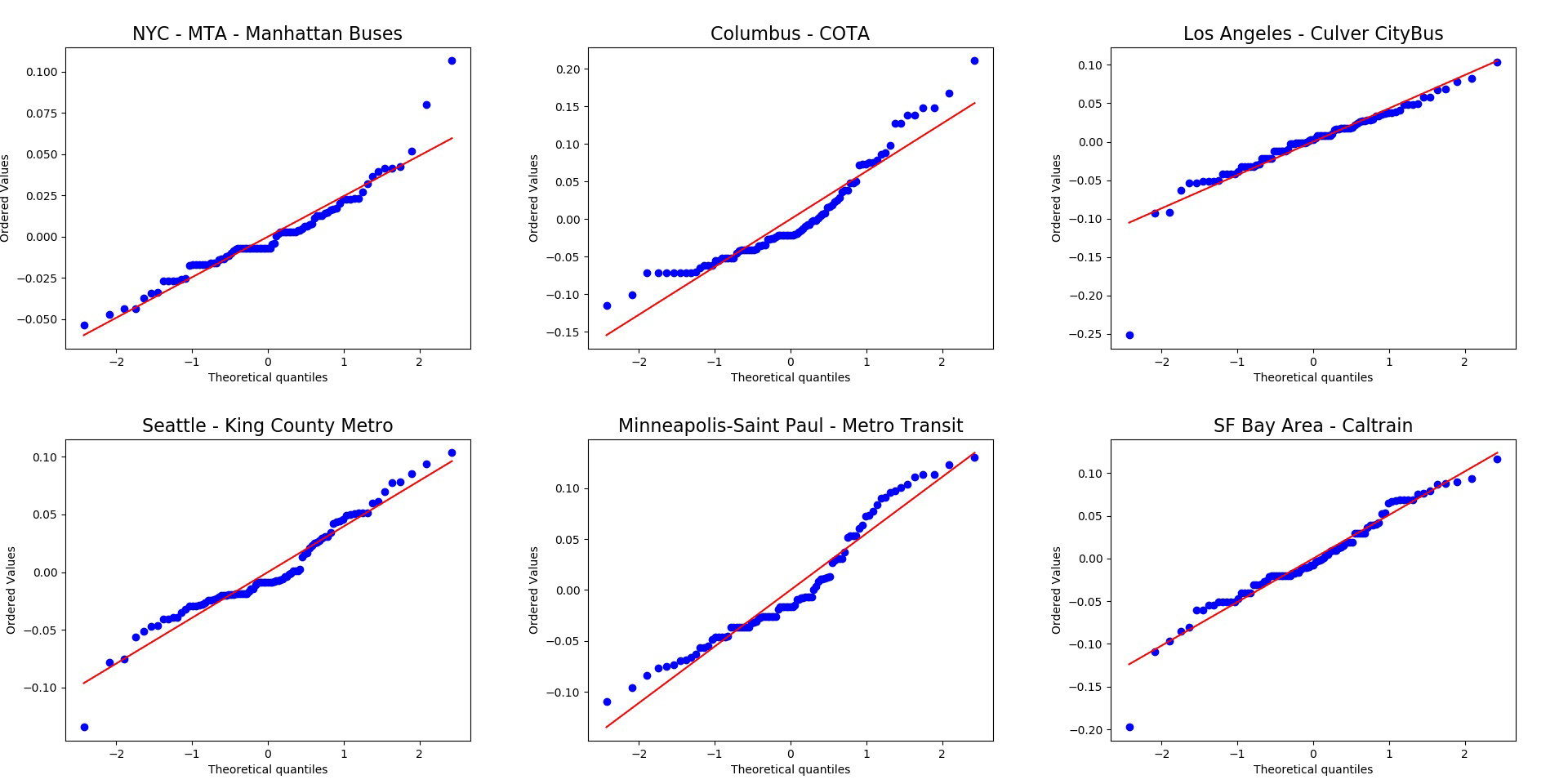
**Response**: we appreciate the comment and we admit that the methodological contribution of this paper is indeed limited. The techniques used by the paper are not new.

1) The authors employed a logistic function to fit the transit demand data for each transit system. First, the logistic function is quite different from the logistic model, the authors should be careful when describing their methods. Second, I failed to found any results of the logistic functions. The authors should at least give a summary of the fitting accuracy and statistical significance of the logistic functions for different transit systems.

Response: This is a reasonable suggestion and we made corresponding changes and clarifications correspondingly:

First, the logistic function is indeed different from the logistic model. We changed all “logistic model” to “logistic function” in section 2.2 to avoid confusions.

Second, we agree with the comment and we added three test to show logistic function’s goodness of fitting: R-squared, Shapiro-Wilk test p-value, and QQ plots. We added corresponding method explanations in section 2.2; we also added a new section to interpret the logistic function fitting results in section 3.1. The median of all model’s R-squared is 0.969 and 5% percentile is 0.92, which shows a very high fitting accuracy. Shapiro-Wilk test shows that 30 of 119 systems’ residual cannot reject the normality assumption. However, considering the sensitivity of Shapiro-Wilk test for large sample size (>50), we moreover used Q-Q plots to test the normality of the residuals. The Q-Q plots show that the results show that each system’s actual quantiles are very close to the theoretical normal distribution quantiles (we show some typical plots in Picture 1). Although many systems’ residuals do not pass the strict normality test due to outliners and test’s high sensitivity, their Q-Q plot still indicates their normality as shown in Picture 1. We can conclude that logistic function can properly fit the transit demand data with very high fitting accuracy.



Picture 1: QQ plots of some transit systems.

2) When modeling the factors related to floor value, some essential variables are missing. For example, the population density, the job density, and the factors related to transit accessibility (for example, the number of transit stations in each city. The data can be derived from OSM POI). The authors should do more literature review regarding the built environment and public transit to understand which covariates are essential.

We appreciate the comment and added these proposed factors to the model. We used population density and employment density (employed civilian population 16 years old and over) of the county-equivalent. It turns out the population density and job density are highly correlated with the ratio of working from home, therefore we did not add the two variables to the final model duo to multicollinearity. It is also very intuitive: the industries that can work from home are naturally rooted in metropolitans, high-tech centers, and university cities, which generally have higher population and employment density.

Pop density and job density are classic measure for transit determinants, however, In terms of the decline of transit demand, Work for home is a direct measure; the other two are indirect. Better fit.

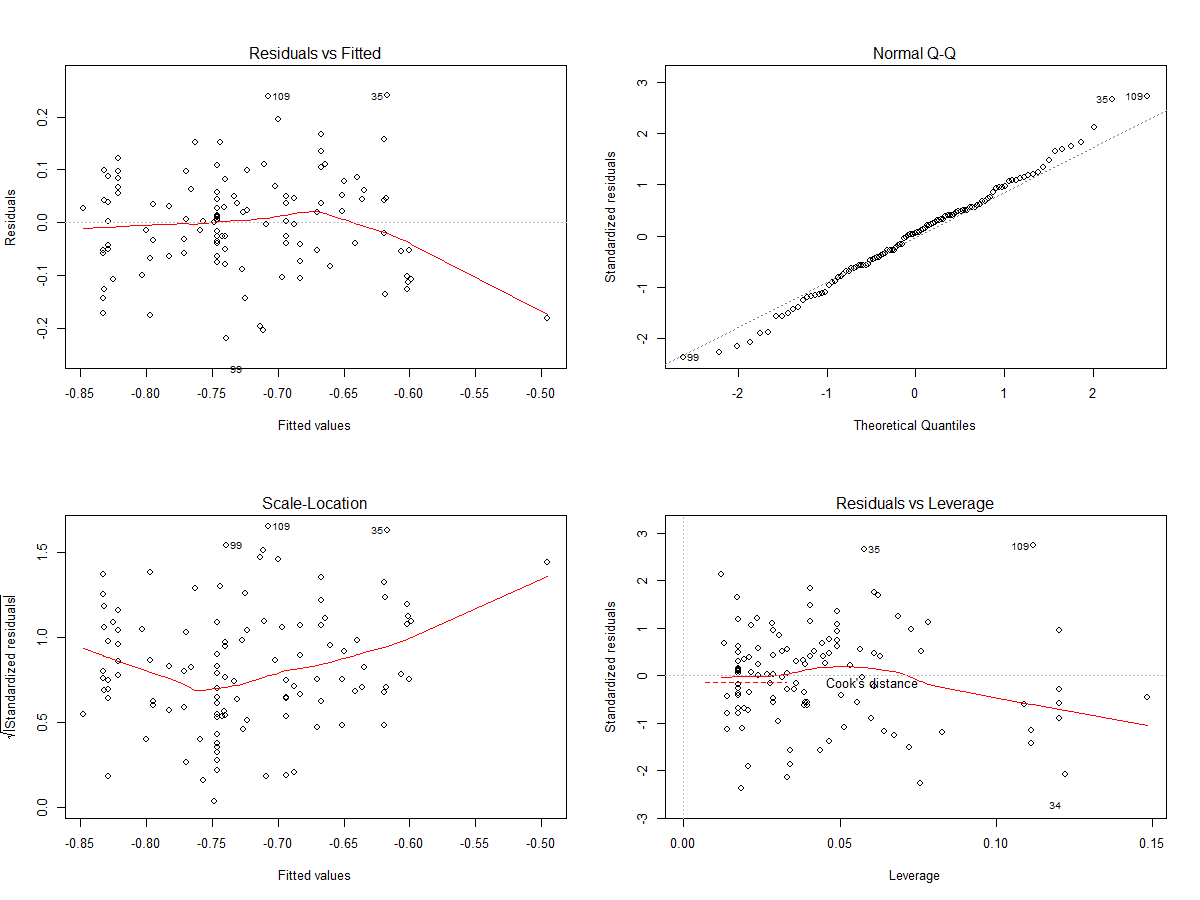
We also added the Transit Score into the model. Transit Score uses an algorithm to summarize transit accessibility and the relative usefulness of transit routes [1]. The model shows that Transit Score is not significant with the other three factors (p-value = 0.1426).

We added more references to the section 2.2.1 to moreover justify the section of independent variables. Similar to what we argued in the section 3.2.4, the results show that the usage rate during the pandemic is more relevant to economic and demographic factors such as race, job, and age composition than other built environment factors. On the other hand, the status of city facilities and amenities would not change within a short period, therefore it will not be a major factor affecting people’s decision. For example, when a work was deciding whether to work from home or stop using transit to the workplace in March, her/his company’s work-from-home decision (employment) and whether she/he has a car (vehicle ownership) are the top two factors to be considered. The environment and city facilities could hardly impact the decision; we could also see the regression results support this conclusion.

This conclusion also reveals the reason why COVID19 is different from traditional modal shift and transit usage studies: the propulsion of the shift is completely different. Traditional shift is generally long-term and largely affected by city facilities, such as building area, commercial establishment, service facility, attraction, accessibility, and road density [2,3]. However, COVID19 is a short-term disruption that is driven by a public health event. A sudden disruption’s impact is hardly discussed by former literatures and this paper could be one of the first attempts to solve this issue. We also added it to the potential future direction.

3) Is the simple linear model appropriate to fit the floor value? Do the data meet the normality assumption? How to handle spatial auto-correlations? The authors should address these issues before using an OLS model.

**Normality**: The Shapiro-Wilk test shows that we cannot reject the normality assumption for the dependent variable (W = 0.99015, p-value = 0.5917). We also mentioned that the residuals of the model are normally distributed in section 3.2: Picture 2 shows the four indicators of regression model assessment. The Q-Q plot (top right) shows that the residual generally follows the normality assumption, which is also proven by the model’s F-test p-value. From these results, we can conclude that the model meets the normality assumption.



Picture 2: four indicators of the model

**Autocorrelation**: we calculated the univariate Moran’s I with a Queen spatial weight. The Moran’s I is 0.24. Based on this fact, we conducted the spatial lag model regression with the same weight. Tab 1 shows the results: the spatial weight item is not significant. This could be because the other independent variables also have spatial autocorrelation effect, therefore the effect is offset for the residual. Therefore, the results do not support using the spatial autoregressive model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std.Error | z-value | Probability |
| Spatial weight | 0.24499 | 0.133449 | 1.83584 | 0.06638 |
| Intercept | -0.38183 | 0.186735 | -2.04479 | 0.04088 |
| Ratio of working from home | -0.40732 | 0.123835 | -3.28919 | 0.001 |
| Ratio of population over 45 years old | 0.80443 | 0.257634 | 3.12237 | 0.00179 |
| Ratio of African American | 0.359243 | 0.08014 | 4.48267 | 0.00001 |
| Coronavirus Google trend | -0.00437 | 0.001919 | -2.27779 | 0.02274 |

Tab 1: spatial lag model output summary.

**Other assumptions**: Picture 2 (top left) shows the model generally holds linearity and homoscedasticity assumption (the results could be better without an outlier). Moreover, Table 1 (in the paper) shows that each independent factor’s variance inflation factor (VIF) is very small, which means there is little lingering multicollinearity. Picture 2 (bottom right) also shows that there is no leverage points in the model and data.

In conclusion, we can conclude that OLS model is appropriate to model the base value (floor value in the last draft) with statistically significant results.

4) Why the authors only build a model for floor value, while ignoring the other indexes like cliff and floor points, response intervals, the decay rate?

**Response**: We did these analyses.

There are several reasons:

1. The length of paper is a major factor when we were deciding whether to put contents in the paper. The last draft is ~8300 words; we also need to add the proposed contents by the referees. Therefore, we decided not to put these analyses in the paper to keep a manageable paper length.
2. The regression results of the four dependent variables are less interesting and informative, therefore we intended to be more selective. This could be because: the three time measures (cliff, floor points, and response intervals) mostly depends on the temporal development of the pandemic, including the testing process and media. These factors could be highly random and non-linear, therefore making them less relevant to the local demography and built environment. In contrast, base/floor value is a much more important and robust measure than the other four, therefore we chose to focus on the most informative measure.
3. We did not ignore other indexes. Instead of using a regression model, we chose to focus on their own spatiotemporal patterns and some more interesting correlation results for the four measures. For example, in section 3.3 to section 3.4 we presented the hyperbolic relationship between decay rates and cliff points because of its good accuracy and theoretical support.

The visualization part is insufficient also. At least two figures are important but missing. First, a figure of the transit demand varying patterns across the study period. Second, a figure visualizing the observed data versus the fitted data using the logistic function. The indexes like floor value, cliff and floor points, response intervals, the decay rate, can also be annotated in the figures.

**Response**: we thank the reviewer for the good suggestion. We added the two visualizations in

Some other minor comments:  
1) The authors should involve a proofreader to improve writing. Many words are unprofessional and hard to understand. For example, the floor value mostly means the closest integer less than or equal to a given number, rather than the lowest plateau value the authors want to express.

**Response**: We also changed the name “floor value” and “floor point” to “base value” and “base point”. We also did a proofread

2) The holidays should be excluded from the study periods due to the unusual human mobility patterns.

**Response**: this is a good question. However, the research time span is from February 15th to May; the only national holidays is the President’s day (Feb 17th). It is not a major holiday that could vastly impact human mobility like Christmas and New Year’s Day. Moreover, the transit demand data are normalized and adjusted by historical data; therefore, even if the mobility is impacted, the impact is normalized therefore comparable to other normal days.

3) The authors should also report the variables with insignificant P-values in Table 1.

**Response**: There is

4) In Line 353, why does the ratio of female have high multi-collinearity with the ratio of African Americans?

**Response**: this may be because of African Americans’ demographic structure in the studied cities, which is dominantly large cities. Picture 3 shows the scatter point plot between the two factors. The

Picture 3: the scatter point plot between ratio of African American and ratio of female.

Reviewer #2: This interesting paper investigates the impacts of the COVID-19 pandemic of public transit ridership across major systems in the US. The main data supporting this research are provided by the Transit app. It is a very timely effort, focusing on an important topic with a strong tie to the society. Methods are adequate. But I have several concerns for the authors to address.  
  
1. Introduction section is not very motivating. For example, why is it important to study the changes in public transit ridership, along with some metrics like the floor value and so on? What knowledge do we gain from this? How can this knowledge be beneficial to the society? The authors could’ve done a better job discussing these points.

2. My major concern is that there is no lit review in this paper. Without discussing previous studies of relevant scopes, how can we know the research gap and the contributions of this work? It is important to add such a section to back up your ideas.

**Response**: this is a good question and we also share the concern too. The reasons why we did not include a dedicated literature review section are:

1. PLOS ONE’s requirement. PLOS ONE requires authors to organize the papers as: Introduction + Materials and Methods + Results + Discussion and Conclusions. Please refer to the official guidelines: <https://journals.plos.org/plosone/s/submission-guidelines>.
2. The topic is unprecedented. Before the occurrence of COVID-19, there were very few studies that investigated the impact of a pandemic on the public transit system. This could be because of two reasons: 1) the lack of accessible empirical transit demand/ridership data; 2) there were very few widespread pandemic such as COVID-19 in the North America.

Therefore, instead of dedicating a whole section for the literature review, we made references in the introduction part as background and in each corresponding section. We introduced two papers based on Taiwan’s SARS and South Korea’s MERS pandemic and their experience about the pandemic impact on the transit systems. have yet to find any references about the impact on

3. Variables. The authors should justify why some variables are selected. I am concerned about a few varaibles. One such variable, for example, is the occupation type factor. As described in lines 168-169, “Information, Financial activities, and professional and business service” were selected and adopted in the model. The assumption, as detailed in lines 164-165 and line 169, is that these types of workers are more likely to work from home during this pandemic and thus areas with more of these workers are more likely to experience a greater hit in ridership. This assumption/assertion is somehow problematic. I think these subgroups are less likely to use public transit but instead rely more on private vehicles before this pandemic. That said, they may not be an important component to the typical ridership. Therefore, looking at communities with higher percentage of these workers for examining sudden ridership change is less convincing.  
In addition to the variables already included in the model, I think the number of homeless people should be considered. Homeless people are more likely to take/occupy public transit, especially in large cities like NYC. As this particular subgroup of population reportedly has higher infection risk, the related transit systems may be affected more severely. This can also be related to the awareness factor discussed in the paper.

Response: this is a good question. We added some clarification in section …

**Occupation type**. This is a very reasonable comment.

1. Many past survey and research results concluded that the privileged population (such as high-income population and the four mentioned industries employees) is an important component of the transit ridership.
2. Empirical results support the assumption. If the privileged “choice riders” are indeed very few, the regression analysis should not see a significant regression between the floor/base value and the ratio of “choice rider”.

**Homelessness**. This is a good question. We first collected the homeless population data from National Alliance to End Homelessness [4] and conducted correlation and regression analysis between the floor/base value and the factor “Homeless per 10000 people in the general population”. Picture 4 is the scatter point between the floor/base value and the number of homeless people per 10000 people. Although it is statistically significant, the coefficient is less than 0, which means the more homeless people, the less people will continue to use transit during the pandemic. This is contradicting to the assumption that homeless people is a significant factor for the transit decline.

Meanwhile, several statistics also show that homeless people are not a major component of transit ridership. For example, in the New York City, one of the cities with highest homeless ratio, numerous local news and city data reported that there are 2000 – 3000 homeless people relied on subway system daily. NYC subway’s average workday ridership in 2018 is 4602905; the average daily ridership from March to May 2020 is 1094822.

Picture 4: scatter point plot between floor/base value and homeless per 10000 people.

4. Provide more details. Throughout the paper, the authors claimed that the Transit app is a widely used app. The only statements related to this is in lines 99-101—“the app covers over 200 cities aournd the world with … download on…” This is insufficient to back up the point that it is a widely used app, and thus leading me to question the representativeness of the data. As the study area is the US, so the authors should provide more details about the user coverage and usage stats (ideally some comparisons with other competitors for showing its market share) to define how “widely” it is being used in the US.  
More details about methods/analyses. Section 2 describes the analyses/methods, but I find it a bit loosely connected. More details should be provided to better connect these steps and help readers get the full picture.  
I think it would be great if every city in the maps is labeled.  
  
5. Figure 1. Why COVID curve (orange) is more prominent than the typical curve (blue)?

6. There are many typos and formatting issues in the paper, making it difficult to read. The language should be improved.