Reviewers' comments:

Reviewer 1: 1. The study is very limited in contribution

2. what is a validation step for the synthetically generated data?

3. How many dataset was synthetically generated? what is the ratio of original dataset to synthetically generated dataset.

Reviewer 2: This research proposes a framework for estimating a 'synthetic' population as a proxy for on-demand delivery demands. The authors primarily use three data sources: ACS-PUMS, PSRC, and NHTS, focusing mainly on demographic information. They employ Iterative Proportional Fitting (IPF) and machine learning regression methods, calculating what they term 'unique imputation factors' (essentially, calculating the mean). While the research topic is worthwhile, the article needs significant improvement in both methodology and readability. It also appears to be misaligned with the focus of this journal. I recommend rejection at this point, but I encourage the authors to either restructure the article or seek a different platform for their work. I hope my comments are helpful.

Major comments:

1. The definition of a synthetic population is unclear. Your literature review discusses how other studies have used synthetic populations but fails to explicitly define what you mean by the term.

2. Existing research has already attempted to measure synthetic populations using mobile phone or GPS signals, often referring to it as 'human dynamics' or 'floating population'. The studies have used such populations as proxies for transit demand. How does your work relate to, or differ from, the previous efforts?

3. You mention that NHTS and PSRC each have their own limitations, but you do not elaborate. Additionally, there is a discrepancy in the data years referenced (2017, 2019, and 2021) on page 11, even though the timeframe for the two datasets appears to be the same on page 5.

4. Is this method of calculating unique imputation factors widely accepted? Should the use of weights in the calculations be considered?

5. For the gradient boosting regression analysis, you conducted 5-fold validation. It's unclear whether the validation and test datasets are the same. Furthermore, how were these datasets partitioned? Shouldn't multiple sample sets be tested across more than one iteration?

6. The discussion section is inadequate. While you mention contributions in calculating uncertainties and conducting comparative studies, the logical links in the latter sections are lacking. What are the implications? Why do you think that's happening in your estimation?

Minor comments:

1. Include a map of the study area and compare the coverage of different data sources.

2. Avoid overusing the terms 'this' and 'these' as they can make your statements ambiguous. Please elaborate for clarity.

3. On page 5, where are the acronyms for the explanatory variables derived from? Are they from ACS-PUMS, PSRC, or NHTS?

4. Adding cross-tabulation could enhance the relevance of the information in Table 1.

5. Please clarify the data sources in Figure 1; which data comes from NHTS and which from PSRC?

6. Figures 3 and 4 lack direct explanations; please include these for clarity.

7. It would be beneficial to elaborate descriptively on how dummy variables were addressed in your analysis.

Reviewer 3: This paper highlights a pertinent concern: on-demand delivery services. It introduces a solution by suggesting demand prediction to enhance efficiency and provide data-driven insights for decision-makers. The study is focused on constructing a predictive framework for estimating delivery demand, which leverages statistical methods to generate synthetic populations (i.e., iterative proportional fitting (IPF)) and employs the Gradient Boosting Machine (GBM).

The primary issue with the proposed methodologies lies in their lack of comprehensive explanations. The text fails to elucidate why each methodology is considered appropriate, resulting in an insufficiently logical presentation of their suitability. Similarly, the explanation provided for the model selection is inadequate. If the authors intended to emphasize the model's high accuracy, it would have been advantageous to incorporate the author's viewpoint on the potential factors impacting this predictive precision. In the context of on-demand delivery prediction, the population synthesis constitutes the foundational basis of the entire process. Thus, it is crucial to address prominent issues like the zero-cell problem and the unbiased problem, which hinder the distortion of estimated joint distributions. Specifically, the credibility of the generated data should be a focal point in the population synthesis task. Given the availability of open-source tools like PopulationSim,

which offers validation tools for marginal distributions alongside diversity and feasibility through optimization techniques, the proposed methodology falls short in terms of convincing accuracy, even when considering spatial analysis aspects.

The exposition concerning the utilization of multiple datasets is challenging to comprehend. With four distinct data sources (i.e., ACS, PUMS, NHTS, PSRC) involved, the specific chosen variables of interest within the two selected datasets and the constitution of the original dataset remain ambiguous. Furthermore, the text does not provide an explanation for TABLE 1, rendering its contents unclarified. The rationale behind the selection and use of these attributes and datasets for on-demand delivery prediction lacks a clear connection. Particularly, the treatment of delivery data as a predictive variable lacks elucidation, contributing to an overall poor understanding of the methodology.

In the results section, the proposed bias square fails to exhibit a significant difference between the GBM and Random Forest models. This absence of a clear distinction raises questions about the appropriateness of selecting GBM for this task. The lack of explanation for the close proximity of values between the models further compounds this issue. The insufficiency in addressing the gap for selecting an appropriate tool is evident. The exploration of various population synthesis methods remains unaddressed, with focus solely on the IPF method. Additionally, the provided text inadequately explains the proposed methodology. For instance, in the population generation stage, the absence of an explanation for the algorithm 1 of the IPF method leaves readers without insight into the rationale for measuring error via the Euclidean distance between the original matrix M and the generated matrix N. Moreover, there exists uncertainty regarding how the subsequent steps of predicting

on-demand delivery are substantiated by the population synthesis. The lack of detailed justification for incorporating population synthesis into this task is conspicuous in both the literature review and methodology sections.

Certain content needs to be relocated to different sections. For example, the elucidation of the PSRC dataset within the Robustness and Validation section should be transferred to the Data Source section. Additionally, ensure that all Figures and Tables are appropriately referenced in the provided text. While offering estimations for on-demand delivery services presents a valuable avenue for data-driven analysis and prospective insights in city planning, it is advisable to undertake further refinement of this paper. Enhancements in writing style and content strength are essential to bolster the clarity and effectiveness of the presentation, ultimately contributing to its readiness for publication.

Reviewer 4: Please provide suggestions/comments to the author for improving the paper here.

The authors address the important topic of estimating on-demand residential deliveries. They correctly state the scarcity of data and the necessity to compensate for lack of observed data with estimations as well as the value of predictive models. Given the importance of population synthesis for the task, the literature review on the topic probably should be substantially broadened. Integration of disparate data sets such as a household survey from an MPO, PUMS and NHTS probably needs more explanations. The choice of the independent variables is not explained, and it is not clear how model specification were derived and how authors arrived at that set of explanatory variables. I would add more details specific to the research, results, implementation and previous relevant work.

A few specific comments:

Page 3 line 22

"from which the American Community Survey tables were generated."

Suggest rephrasing, the ACS PUMS files are "a set of records from individual people or housing units, with disclosure protection enabled".

Page 4 line 4

Population Synthesizers emerge as an inherent part of transportation activity-based models. There is a substantial body of literature reflecting this fact. It is not a must in the context of this paper, but it would have been beneficial to mention the research. See, for example, Vovsha, P. et al., New Features of Population Synthesis. TRB 94th Annual Meeting Compendium of Papers.

Years are missing in the first five references.

Page 11 line 29 - reference is made to section 5.2, but there is no section 5.2 in the paper. Was the text copied from somewhere else? Than the proper reference should be provided.