Generation of Tabular and Sequential Attributes in Synthetic Population with Enhanced Diversity and Feasibility

Farbod Abbasi[1], Zachary Patterson[2], and Bilal Farooq[3]

[1]PhD student in Information and Systems Engineering at Concordia University, Canada
[2]Professor at the Concordia Institute for Information Systems Engineering, Concordia University, Canada
[3]Associate Professor, Laboratory of Innovations in Transportation,Toronto Metropolitan University, Canada

Short summary

Synthetic population generation is essential for activity-based modeling, where individual level data are needed to simulate travel behavior. However, such data are often limited due to high costs and privacy concerns. A realistic synthetic population should capture both tabular and sequential attributes while preserving diversity (the ability to represent a wide range of plausible attribute combinations) and feasibility (the consistency of generated attributes with logically valid combinations). Despite recent advances in generative modeling, most existing approaches focus primarily on tabular attribute synthesis, and do not model sequential attributes. Moreover, diversity and feasibility of generated attribute combinations are often not explicitly addressed within the modeling framework. In this paper, we propose a unified generative framework for synthetic population generation that integrates tabular and sequential attributes. In the first stage, a Generative Adversarial Network (GAN) is used to learn the distribution of tabular attributes, augmented with an Inverse Gradient Penalty (IGP) regularization term to improve diversity while maintaining feasible attribute combinations. Since tabular attributes such as socio-demographic characteristics influence mobility patterns, the second stage uses a Transformer-based model to generates sequential attributes conditioned on the synthesized tabular attributes. The proposed framework is evaluated using distributional similarity, diversity, and feasibility metrics for tabular attributes, and unigram, bigram, and trigram pattern comparisons for sequential features. Results show that the regularized GAN significantly improves diversity coverage while maintaining high feasibility, and the Transformer model successfully reproduces dominant behavioral patterns observed in the real data.
Keywords: Population synthesis, Tabular attributes, Sequential attributes, Regularization term, Transformer, Generative adversarial networks

1 Introduction

Population synthesis is an essential component of activity-based modeling (Farooq et al., 2013). It generates disaggregate population that serve as inputs to activity-based models. Access to detailed mobility data is limited due to high costs and privacy concerns (Arkangil et al., 2022). As a result, reliable synthetic populations are needed to support travel demand modeling and policy analysis. A reliable synthetic population must capture tabular attributes and sequential behavioral features while preserving diversity and behavioral feasibility (Badu-Marfo et al., 2022; Kim & Bansal, 2023). Diversity refers to the ability to generate a wide range of plausible attribute combinations, while feasibility refers to the generation of valid and realistic samples that are consistent with the underlying population

Most existing synthetic population studies focus on tabular attribute synthesis. While these approaches aim to reproduce socio-demographic distributions, behavioral sequences are often excluded from the generation process. A limited number of studies have attempted to generate sequential attributes (Agrawal et al., 2025; Arkangil et al., 2022; Badu-Marfo et al., 2022); however, Transformer based architectures have not been applied in this context. Furthermore, to the best of our knowledge, no existing framework addresses sequence generation while also considering improvements in diversity and feasibility within the overall synthetic population.

The key challenge is to generate sequential behavioral features that are consistent with tabular attributes, while preserving overall diversity and feasibility in the synthetic population. Modeling these components together within a single framework is difficult, as tabular and sequential data have different structures and dependencies. In this study, we propose a two stage generative framework for synthetic population modeling. First, a generative adversarial network (GAN) (Goodfellow et al., 2014) is used to synthesize tabular attributes while enhancing diversity and feasibility. Second, a Transformer based model (Vaswani et al., 2017) generates sequential behavioral features (Mollajafari et al., 2025) conditioned on the tabular attributes.
The main contributions of this study are as follows:

- We introduce a unified synthetic population framework that integrates tabular and sequential attribute generation.
- We develop a Transformer-based model for conditional sequence attribute generation in the context of synthetic population modeling.
- We incorporate mechanisms to improve diversity and feasibility within the tabular generation process.

2 Literature Review

This section reviews existing research on synthetic population generation in transportation modeling. We first discuss traditional and deep learning approaches for tabular attribute synthesis, then examine recent efforts to incorporate sequential behavioral features. Finally, we review strategies for improving diversity and feasibility and identify remaining research gaps.

Tabular Synthetic Population Generation

Early synthetic population generation methods in transportation research primarily relied on statistical and probabilistic frameworks to reproduce socio-demographic distributions. Classical approaches such as iterative proportional fitting (IPF) were widely adopted to match aggregate marginal constraints (Zhu & Ferreira Jr, 2014). To overcome limitations of purely fitting techniques, simulation based and probabilistic models were introduced, including Markov Chain Monte Carlo (MCMC) methods (Saadi et al., 2016; Farooq et al., 2013) and Bayesian network formulations (Sun & Erath, 2015; Rahman & Fatmi, 2023).
Recent advances in deep generative modeling have increasingly influenced synthetic population research. Variational Autoencoders (VAE) (Kingma et al., 2013), and GAN (Goodfellow et al., 2014) have been applied to synthetic data generation to better capture complex joint distributions and address limitations of traditional fitting based approaches (Borysov et al., 2019; Garrido et al., 2020). More recently, copula-based generative frameworks have been proposed to better model dependency structures in tabular data (Jutras-Dubé et al., 2024). However, these deep learning approaches mainly focus on tabular attribute synthesis, and sequential features are generally not modeled. Diversity and feasibility are often discussed conceptually, but they are not directly addressed within the modeling framework.

Sequence Modeling in Synthetic Population Research

A limited number of recent studies have incorporated sequential feature generation within synthetic population frameworks using deep learning models. Arkangil et al. (2022) proposed a hybrid framework combining a Conditional GAN (CTGAN) for tabular attribute generation with a recurrent neural network (RNN) model for trip sequence synthesis. In their approach, tabular and sequential components are trained separately and later merged through an assignment procedure. Similarly, Badu-Marfo et al. (2022) introduced a Composite Travel GAN architecture that integrates GAN tabular generation with sequential trip modeling inspired by SeqGAN. In addition, recent work has explored deep generative approaches for trajectory and mobility sequence modeling using neural architectures tailored for spatio-temporal data (Agrawal et al., 2025).
Based on the existing literature, two main research gaps can be identified. First, although recent studies have introduced deep learning approaches for sequence generation within synthetic population frameworks, Transformer-based architectures have not been explored for conditional behavioral sequence modeling. Second, Kim & Bansal (2023); Abbasi et al. (2026) are the only study that explicitly addressed diversity and feasibility through model regularization, but their

framework does not include sequence generation. In contrast, studies that generate sequential features do not address diversity and feasibility. Thus, these two aspects have not been considered within a unified modeling framework.

3 Methodology

This study uses the 2018 Montreal Origin–Destination (OD) survey as a representative sample of the regional population. The dataset contains detailed travel information for 162,588 individuals, covering approximately 4% of the total population. For this study, we consider seven categorical attributes (35 classes in total) and three sequential attributes.

Problem Formulation

Let each individual in the population be represented by a pair (X, S), where $X \in \mathbb{R}^d$ denotes a vector of tabular attributes and $S = (s_1, s_2, \ldots, s_T)$ denotes a sequence of length T. The objective of synthetic population generation is to learn the joint distribution $P(X, S)$ from observed data and generate new synthetic samples $(\tilde{X}, \tilde{S})$ that preserve statistical realism, diversity, and behavioral feasibility.
Directly modeling the joint distribution of tabular and sequential attributes is challenging due to their different structures. To address this, we decompose the joint distribution as:

$$P(X, S) = P(X)P(S \mid X). \tag{1}$$

This formulation reflects the assumption that individual attributes influence behavioral patterns. Accordingly, the proposed framework models $P(X)$ using a generative adversarial network (GAN) and models $P(S \mid X)$ using a Transformer-based conditional sequence model.

Overall Framework

The proposed framework consists of two sequential stages. In the first stage, a GAN is trained to learn the distribution of tabular attributes. In the second stage, a Transformer model is trained to generate sequences attributes conditioned on tabular attributes.
Synthetic individuals are generated in two steps: (1) sampling synthetic tabular attributes from the trained generator, and (2) generating sequences attributes conditioned on those attributes.

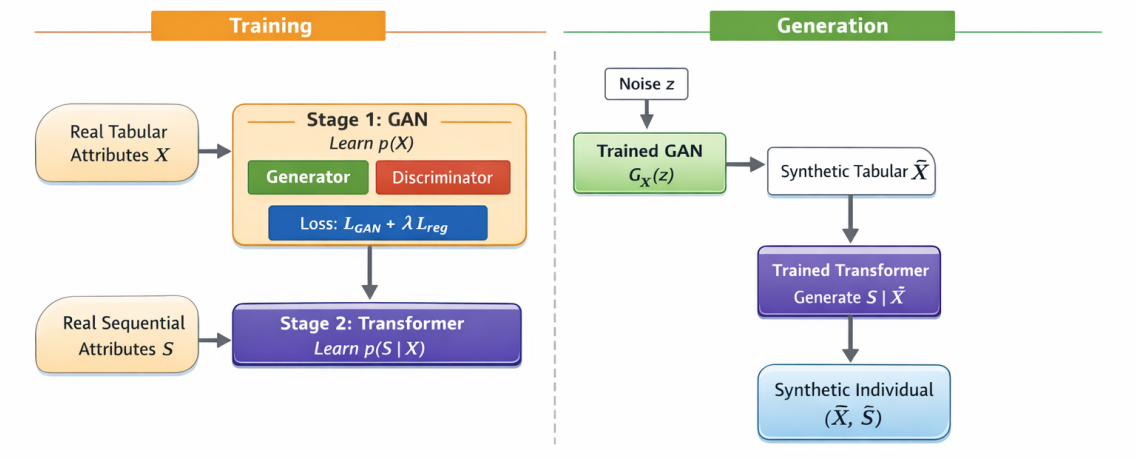


Figure 1: Two stage framework for synthetic population generation

Tabular Attribute Generation

To model the marginal distribution $P(X)$, we employ a GAN consisting of a generator G and a discriminator D. The generator maps a latent vector $z \sim \mathcal{N}(0, I)$ to synthetic tabular attributes:

$$\tilde{X} = G(z). \tag{2}$$

The discriminator attempts to distinguish between real samples X and generated samples $\tilde{X}$. The adversarial objective is defined as:

$$\min_G \max_D \ \mathbb{E}_{X\sim P_{data}}[\log D(X)] + \mathbb{E}_{z\sim P_z}[\log(1 - D(G(z)))]. \tag{3}$$

In synthetic population modeling, diversity refers to the generator's ability to represent heterogeneous socio-demographic attributes without collapsing to dominant modes, while feasibility refers to the production of realistic and structurally possible attribute combinations. To ensure that generated tabular attributes are both diverse and feasible, we incorporate a regularization mechanism directly into the generator objective.

To address these challenges, we use an Inverse Gradient Penalty (IGP) term (Li et al., 2023). The goal of the IGP loss is to encourage the generator to be sensitive to variations in the latent space. In an efficient generative model, small perturbations in the latent vector should produce meaningful changes in the generated attributes. If the generator ignores variations in the latent space, the generated samples become less diverse.

For two randomly sampled latent vectors z_1 and z_2, the IGP term measures the ratio between the distance in the output space and the corresponding distance in the latent space:

$$\mathcal{L}_{\text{IGP}} = -\mathbb{E}_{z_1,\, z_2}\left[\min\left(\frac{\|G(z_1) - G(z_2)\|}{\|z_1 - z_2\|},\ \tau\right)\right], \tag{4}$$

where τ is a predefined threshold that limits the maximum contribution of the ratio.

This formulation encourages the generator to produce more diverse outputs. The IGP term mitigates mode collapse and enhances diversity, by preventing the generator from mapping multiple latent vectors to similar outputs. At the same time, since the generator is trained adversarially on real data and the regularization term includes a threshold, the generated samples remain feasible. The final generator objective combines the adversarial loss with the IGP regularization term, where λ_{IGP} controls the strength of the regularization.

$$\mathcal{L}_G = \mathcal{L}_{GAN} + \lambda_{\text{IGP}}\mathcal{L}_{\text{IGP}}, \tag{5}$$

Conditional Sequence Generation and Transformer Architecture

In the second stage, we generate sequential attributes conditioned on tabular attributes. Each individual is associated with three aligned categorical sequences representing departure time, trip purpose, and travel mode.

We adopt a Transformer-based architecture to model the conditional distribution of behavioral sequences given tabular attributes. Unlike recurrent neural networks, Transformers rely on self-attention mechanisms that capture long range dependencies without sequential recurrence. This property is particularly advantageous for trip chain modeling, where decisions at a given trip index may depend on earlier trips.

In our implementation, tabular attributes are first embedded into a continuous representation and incorporated into the Transformer as a global conditioning context. The three behavioral components, departure time, trip purpose, and travel mode, are represented as token sequences. At each trip index t, the model jointly considers the previously generated components of all three aligned sequences.

Let s_t^{time} denote the departure time category at trip index t, s_t^{purpose} denote the trip purpose at trip index t, and s_t^{mode} denote the main travel mode at trip index t. The conditional joint distribution of the aligned sequences is modeled autoregressively as:

$$P(S^{\text{time}}, S^{\text{purpose}}, S^{\text{mode}} \mid X) = \prod_{t=1}^{T} P\big(s_t^{\text{time}}, s_t^{\text{purpose}}, s_t^{\text{mode}} \mid s_{<t}^{\text{time}}, s_{<t}^{\text{purpose}}, s_{<t}^{\text{mode}}, X\big)\,. \tag{6}$$

At each trip index t, the Transformer produces a shared hidden representation h_t. From this shared representation, three separate linear projection layers generate probability distributions over departure time, purpose, and mode, respectively:

$$P\big(s_t^{\text{time}} \mid s_{<t}^{\text{time}}, s_{<t}^{\text{purpose}}, s_{<t}^{\text{mode}}, X\big) = \text{Softmax}(W_{\text{time}} h_t), \tag{7}$$

$$P\big(s_t^{\text{purpose}} \mid s_{<t}^{\text{time}}, s_{<t}^{\text{purpose}}, s_{<t}^{\text{mode}}, X\big) = \text{Softmax}(W_{\text{purpose}} h_t), \tag{8}$$

$$P\big(s_t^{\text{mode}} \mid s_{<t}^{\text{time}}, s_{<t}^{\text{purpose}}, s_{<t}^{\text{mode}}, X\big) = \text{Softmax}(W_{\text{mode}} h_t), \tag{9}$$

where W_{time}, W_{purpose}, and W_{mode} are head-specific projection matrices corresponding to their respective vocabularies.
Let $\mathcal{K} = \{\text{time}, \text{purpose}, \text{mode}\}$. The negative log-likelihood training objective can then be written compactly as:

$$\mathcal{L}_{seq} = -\sum_{t=1}^{T} \sum_{k \in \mathcal{K}} \log P\big(s_t^k \mid s_{<t}^{\text{time}}, s_{<t}^{\text{purpose}}, s_{<t}^{\text{mode}}, X\big)\,. \tag{10}$$

This formulation explicitly reflects that, at each trip index, the model predicts three aligned behavioral attributes rather than a single token. The shared hidden representation h_t enables the model to capture interdependencies among departure time, purpose, and mode decisions, while the separate output heads allow each behavioral dimension to maintain its own categorical structure.

4 Results and Discussion

This section evaluates the realism of the generated population in two dimensions. First, we assess tabular attributes by comparing marginal and joint distributions, with attention to diversity and feasibility. Second, we evaluate sequential behavior by comparing unigram, bigram, and trigram patterns to ensure accurate reproduction of both frequencies and transition structures.

Tabular Feature Evaluation

To evaluate how generated tabular attributes match the real data, we examine marginal and joint distributions. Figure 2 compares the marginal probability distributions of several feature groups between the real dataset and the generated dataset. The synthetic data is produced using a 5% training sample from the OD dataset and achieves an SRMSE score of 0.028. As shown in the figure, the marginal distributions of the synthetic data closely follow those of the real data across all feature groups, indicating that the model successfully reproduces the overall population proportions. However, marginal and joint similarity does not ensure preserved diversity or feasibility. Therefore, we also evaluate the diversity and feasibility of the generated tabular attributes.

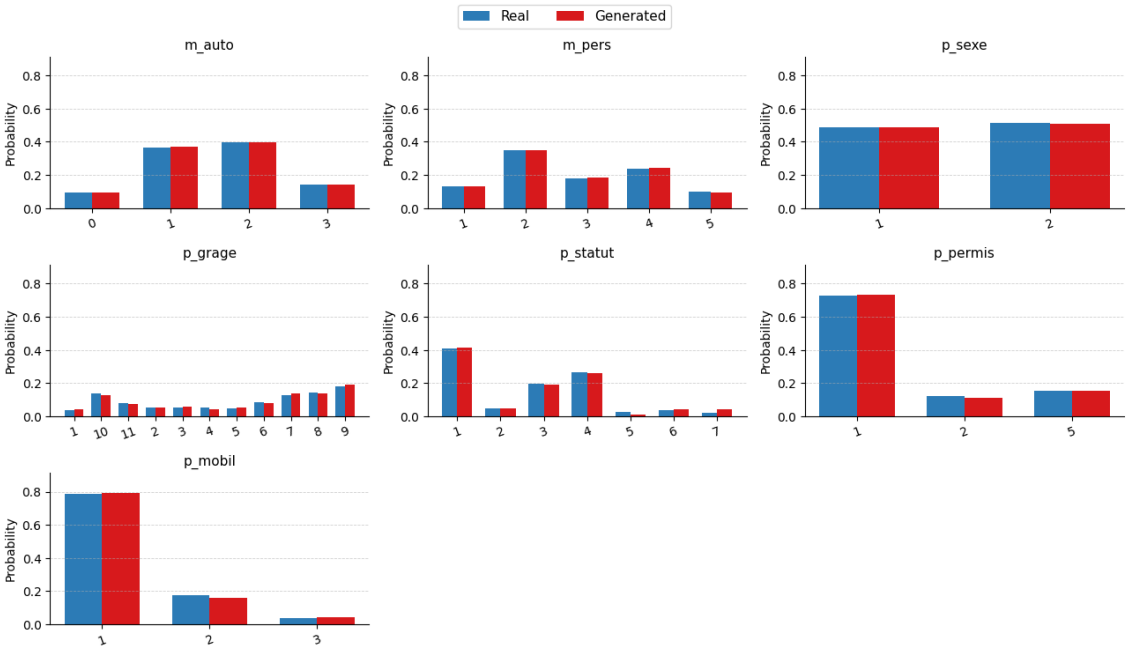


Figure 2: Marginal distribution comparison between real and generated data.

To assess these aspects, we compare the diversity and feasibility of synthetic populations generated with and without the proposed regularization term. This comparison helps determine whether the

regularization improves the realism of the generated tabular attributes by encouraging a broader range of valid combinations while maintaining feasibility.
As shown in Table 1, both models achieve low SRMSE values for the marginal, bivariate, and trivariate distributions, indicating that the statistical structure of the tabular attributes is well preserved. Based on these metrics, the vanilla GAN performs well.
However, matching distributions does not fully reflect the quality of the generated population. Two additional aspects are important: diversity and feasibility. Feasibility is high for both models, meaning that most samples representing plausible combinations and unrealistic combinations are rarely produced.
The main difference between the two models appears in terms of diversity. The vanilla GAN covers only 77.36% of the possible attribute combinations, suggesting that the generator tends to focus on a limited set of common patterns. In contrast, the proposed IGP GAN increases diversity to 95.41%, improving coverage by about 18 percentage. This result shows that the regularization term helps the model generate a wider range of valid attribute combinations while maintaining high feasibility and accurate distributional statistics.

Table 1: Evaluation of tabular feature generation.

Model	**Marg**	**Bivar**	**Trivar**	**Diversity**	**Feasibility**
Vanilla GAN	0.035	0.154	0.352	77.36%	97.74%
IGP GAN	0.028	0.102	0.286	95.41%	97.82%

Sequence Feature Evaluation

To evaluate the realism of sequential attributes, we compare structural patterns between real and synthetic sequences using unigram, bigram, and trigram statistics. Because the number of possible patterns can be large, we focus on the most representative ones by selecting the most frequent sequences from the real data until their cumulative frequency reaches a predefined threshold (e.g., 80% or 90%). These patterns represent the dominant sequential structures in the real dataset, and the same procedure is then applied to the generated data.
Using these sets of dominant patterns, we compute Precision and Recall to measure how well the generator reproduces the structure of the real sequences. Recall measures how many of the important patterns from the real data are recovered in the generated sequences, while Precision evaluates how many of the dominant patterns generated by the model also appear in the real dataset. Let R denote the set of dominant patterns extracted from the real data and G the corresponding set obtained from the generated data. Precision and Recall are defined as

$$\text{Recall} = \frac{|R \cap G|}{|R|} \tag{11}$$

$$\text{Precision} = \frac{|R \cap G|}{|G|} \tag{12}$$

Table 2 presents the evaluation of sequential features by comparing unigram, bigram, and trigram patterns between the real and generated datasets. The analysis is performed at two coverage levels (80% and 90%), focusing on the most representative sequence patterns. Precision and Recall measure how well the generator reproduces dominant patterns observed in the real data.
For all three sequence attributes, the unigram results show very high agreement between real and generated data. Both precision and recall values are close to or equal to one in most cases, indicating that the overall distribution of individuals is accurately reproduced. Moreover, the Top-k sets shows that the generated data closely matches the dominant categories observed in the real data, missing at most one category in the worst case. This means that the generator successfully captures the overall frequency of departure time, mode choice, and trip purpose.
At the bigram and trigram level, which capture transitions between consecutive categories, the results reveal an interesting behavior in terms of diversity and realism. For departure time, the generated dataset contains 45 unique bigram patterns compared to 28 in the real data. A similar trend appears at the trigram level, where the generated data includes 111 unique patterns while the real dataset contains 79. This indicates that the model increases diversity by generating additional combinations of departure time transitions. Importantly, this increase in diversity does not reduce

Table 2: Sequence pattern evaluation using Precision and Recall at 80% and 90% coverage.

				80% Coverage					90% Coverage				
Feature	**Level**	**Unique $_r$**	**Unique $_g$**	**TopK$_r$**	**TopK$_g$**	**Int**	**Precision**	**Recall**	**TopK$_r$**	**TopK$_g$**	**Int**	**Precision**	**Recall**
Departure Time	unigram	8	8	4	4	3	0.750	0.750	6	5	5	1.000	0.833
	bigram	28	45	9	8	7	0.875	0.778	11	11	10	0.909	0.909
	trigram	79	111	17	17	14	0.824	0.824	23	22	20	0.909	0.870
Mode Choice	unigram	18	18	4	4	4	1.000	1.000	6	6	6	1.000	1.000
	bigram	238	212	6	4	4	1.000	0.667	17	11	11	1.000	0.647
	trigram	875	670	15	9	9	1.000	0.600	46	29	28	0.966	0.609
Trip Purpose	unigram	14	14	5	5	5	1.000	1.000	8	7	7	1.000	0.875
	bigram	163	158	15	10	10	1.000	0.667	30	21	19	0.905	0.633
	trigram	1322	1002	115	101	76	0.752	0.661	226	212	166	0.783	0.735

realism. Precision remains high, reaching 87.5% and 90.9% at the bigram level and 82.4% and 90.9% at the trigram level for the 80% and 90% coverage thresholds.

For mode choice and trip purpose, the behavior is slightly different. In these cases, the generated data does not substantially increase the number of unique patterns, but the model still captures the dominant transition structures. In particular, it reproduces about 90% of the important transition patterns for mode choice and about 96% for trip purpose. One possible explanation is the larger number of categories for these attributes. Departure time has only 8 categories, while mode choice and trip purpose have 18 and 14, respectively. As the number of categories increases, the number of possible transitions grows rapidly, making it more difficult for the generator to explore and reproduce a wider range of combinations. As a result, increasing diversity becomes more challenging for attributes with many categories.

Overall, the model performs well at the bigram and trigram levels across all sequence features. Although the number of possible combinations increases substantially for mode choice and trip purpose, making the learning task more complex, the model maintains high precision, reaching one in many cases. This indicates that the generated patterns closely match those observed in the real data. Recall values also remain above 60%, showing that the model recovers a large portion of the important patterns even in more complex settings.

5 Conclusions

This paper introduced a unidief framework for synthetic population generation that models tabular and sequential attributes. The proposed approach combines a GAN for tabular attribute with a Transformer model for sequential attributes. In addition, a regularization mechanism was incorporated into the GAN to improve diversity while preserving feasibility in the generated tabular attributes.

The results show that the proposed framework successfully reproduces both the tabular attribute distributions and the behavioral patterns observed in real data. For tabular attributes, the model achieves strong marginal and joint distribution alignment while significantly improving diversity compared to a standard GAN without reducing feasibility. For sequential attributes, the unigram, bigram, and trigram analyses indicate that the model accurately captures dominant sequence patterns and transition structures. Even in complex settings with many possible combinations, the generated sequences remain highly consistent with the real data.

For future work, several extensions can be explored. First, instead of using a two-stage framework, future research could investigate architectures that generate tabular and sequential attributes jointly within a single model. Second, additional methods could be developed to further improve diversity in sequential features, particularly for attributes with many categories and possible transitions. Such improvements could further enhance the realism and applicability of synthetic populations for transportation planning and activity-based modeling.

Acknowledgements

This study is funded by the Canada First Research Excellence Fund under the Bridging Divides program.

References

Abbasi, F., Patterson, Z., & Farooq, B. (2026). Enhancing diversity and feasibility: Joint population synthesis from multi-source data using generative models. *arXiv preprint arXiv:2602.15270*.

Agrawal, P., Jayaraman, V., Oon, J., Ling, F., & Ramli, M. A. B. (2025). Generation of mobility patterns for private vehicles using multi-headed sequence generative adversarial networks. *Transportation Research Procedia*, *82*, 908–922.

Arkangil, E., Yildirimoglu, M., Kim, J., & Prato, C. (2022). A deep learning framework to generate realistic population and mobility data. *arXiv preprint arXiv:2211.07369*.

Badu-Marfo, G., Farooq, B., & Patterson, Z. (2022). Composite travel generative adversarial networks for tabular and sequential population synthesis. *IEEE Transactions on Intelligent Transportation Systems*, *23*(10), 17976–17985.

Borysov, S. S., Rich, J., & Pereira, F. C. (2019). How to generate micro-agents? a deep generative modeling approach to population synthesis. *Transportation Research Part C: Emerging Technologies*, *106*, 73–97.

Farooq, B., Bierlaire, M., Hurtubia, R., & Flötteröd, G. (2013). Simulation based population synthesis. *Transportation Research Part B: Methodological*, *58*, 243–263.

Garrido, S., Borysov, S. S., Pereira, F. C., & Rich, J. (2020). Prediction of rare feature combinations in population synthesis: Application of deep generative modelling. *Transportation Research Part C: Emerging Technologies*, *120*, 102787.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Jutras-Dubé, P., Al-Khasawneh, M. B., Yang, Z., Bas, J., Bastin, F., & Cirillo, C. (2024). Copula-based transferable models for synthetic population generation. *Transportation Research Part C: Emerging Technologies*, *169*, 104830.

Kim, E.-J., & Bansal, P. (2023). A deep generative model for feasible and diverse population synthesis. *Transportation Research Part C: Emerging Technologies*, *148*, 104053.

Kingma, D. P., Welling, M., et al. (2013). *Auto-encoding variational bayes.* Banff, Canada.

Li, Z., Usman, M., Tao, R., Xia, P., Wang, C., Chen, H., & Li, B. (2023). A systematic survey of regularization and normalization in gans. *ACM Computing Surveys*, *55*(11), 1–37.

Mollajafari, M., Patterson, Z., & Farooq, B. (2025). Lstm vs. distilbert: A comparative study for 2025 shl recognition challenge. In *Companion of the 2025 acm international joint conference on pervasive and ubiquitous computing* (pp. 1005–1011).

Rahman, M. N., & Fatmi, M. R. (2023). Population synthesis accommodating heterogeneity: A bayesian network and generalized raking technique. *Transportation research record*, *2677*(6), 41–57.

Saadi, I., Mustafa, A., Teller, J., Farooq, B., & Cools, M. (2016). Hidden markov model-based population synthesis. *Transportation Research Part B: Methodological*, *90*, 1–21.

Sun, L., & Erath, A. (2015). A bayesian network approach for population synthesis. *Transportation Research Part C: Emerging Technologies*, *61*, 49–62.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Zhu, Y., & Ferreira Jr, J. (2014). Synthetic population generation at disaggregated spatial scales for land use and transportation microsimulation. *Transportation Research Record*, *2429*(1), 168–177.