# Understanding Factors Influencing Willingness to Ridesharing Using Big Trip Data and Interpretable Machine Learning

Ziqi Li[*1], Tianyang Xu[†2]

[1]School of Geographical and Earth Sciences, University of Glasgow
[2]Department for Geography and GIS, University of Illinois

January 12, 2022

**Summary**

Ridesharing, compared to traditional solo ride-hailing, can reduce traffic congestion, cut per-passenger carbon emissions, reduce parking infrastructure, and provide a more cost-effective way to travel. Despite these benefits, ridesharing only occupies a small percentage of the total ride-hailing trips. This study provides a reproducible and replicable framework that integrates big trip data, machine learning models, and explainable artificial intelligence (XAI) to better understand the factors that influence people's decisions to take or not to take a shared ride.

**KEYWORDS:** ridesharing, machine learning, GeoAI, XAI, ride-hailing

## 1. Introduction

Transportation network companies (TNCs), such as Uber and Lyft, provide ride-hailing services that have been a common mode of transportation in cities. According to a recent Juniper Research report published in December 2021, consumer spending on ride-hailing will approach US $937 billion by 2026, which is 50 times the total annual revenue of Transport for London, New York City's MTA, and Beijing Metro in 2021. While there is a large market for ride-hailing and it provides a convenient way to get around, it is also reported to have negative effects on cities. Research found that ride-hailing competes with urban public transport, increases vehicle miles travelled, intensifies pollution and traffic congestion (Erhardt et al., 2019). However, one type of service nested within ride-hailing that has been overlooked is the car-pooling style ridesharing such as Lyft Line and Uber Pool. Ridesharing matches multiple users in the same vehicle, with drivers picking up and dropping off passengers along the way. It is reported that when compared to solo ride-hailing, shared rides can reduce traffic congestion, cut per-passenger carbon emissions, reduce parking infrastructure, and provide a more cost-effective way to travel (Shaheen and Cohen, 2019). Despite these advantages, ridesharing is only available in a few cities, and it accounts for a small percentage of total ride-hailing trips (15–25% in cities such as Hangzhou, Chengdu, Toronto, and Chicago). There is substantial room to further adopt ride-sharing services to alleviate environmental and transportation issues. Understanding why people choose to or not to take a shared ride is essential to potentially promoting ridesharing in current and new cities.

Existing studies that discuss willingness to take shared rides have several drawbacks: 1) trip data is aggregated to a specific spatial and temporal resolution, resulting in a loss of detail in each individual trip record (e.g. Dean and Kockelman, 2021); 2) modelling approaches are based on linear assumptions through the use of linear models (e.g. Park et., 2018); 3) studies using machine learning models do not provide sufficient explanations (e.g. Hou et al., 2020). Consequently, the aim of this study is to address the above issues and to demonstrate a reproducible and replicable framework that integrates big trip data, machine learning models, and explainable artificial intelligence (XAI) to better understand the

[*] ziqi.li@glasgow.ac.uk
[†] tx14@illinois.edu

factors that influence people's decisions to take a shared ride. Here, we use the city of Chicago as an example, and the workflow can be replicated to other cities of interest where data are available.

## 2. Data and Model

The City of Chicago publishes Uber, Lyft, and Via ride-hailing travel records from November 2018 to the most recent. We used data from the entire year of 2019, excluding the time period affected by the COVID-19 service halt. Each trip record is timestamped and geocoded with the census tracts of drop-off and pick-up. Each trip has a binary variable that indicates whether the trip is requested as a shared trip or not, which will be used as our label in the model. Because of the matching availability, not all requested trips are shared. People's willingness to share a trip may be influenced by socioeconomics and the built environment of the census tract where the rider was picked up and dropped off; local weather at the time of request; and other trip attributes such as trip time, trip fare, trip distance, and pick-up and drop-off coordinates. Features are obtained from the US American Community Survey, the US Environmental Protection Agency's Smart Location Database, and NOAA Local Climatology. Because shared trips account for just 25% of total trips, the ride-share data was balanced using an under-sampling method. The final dataset, which has over 10 million records, was divided into 80/20 segments for training and testing, respectively. The Extreme Gradient Boosting (XGBoost) model was utilized, and its hyper-parameters were tuned using the hyper-opt optimization library in a 5-fold cross-validation. Then we used a local interpretable machine learning method called SHAP to explain and attribute the predictions (Lundberg and Lee, 2017) to each feature. The calculated SHAP value measures the probability of each trip being classified as a shared trip, which is a measure of willingness to share with respect to each feature. The general workflow is depicted in Figure 1.

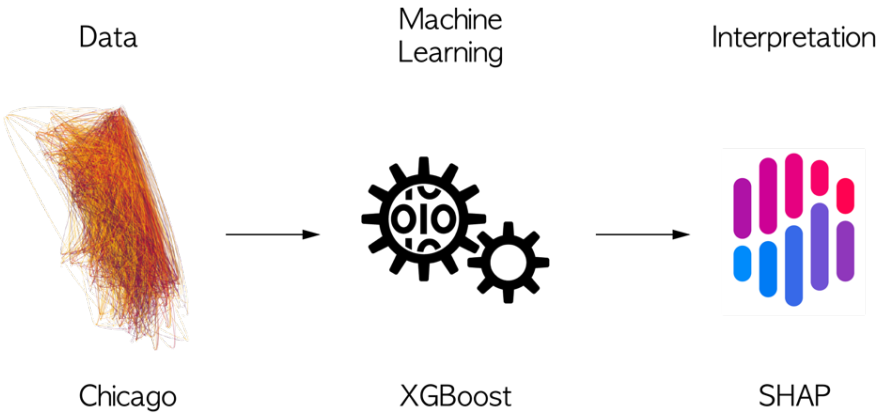


**Figure 1** A framework that integrates big trip data, machine learning model and explainable artificial intelligence.

## 3. Results

The overall model accuracy is 79%, with solo journeys classified at 86% accuracy and shared rides classified at 72% accuracy. Figure 2 plots some selected features against their calculated SHAP values, which shows the marginal relationship between the feature and the prediction. Points that beyond 0.5 indicates that the user is more likely to choose to share the ride. In Figure 2a, we can see that the trip fare has a non-linear relationship with the willingness to share a trip, that higher probability to share a ride is associated with low fare. Figure 2b shows that starting time, such as the hour, has an effect on willingness to share, with the highest probability related to 8 a.m. and 17 p.m., implying that users prefer to take shared rides for morning and afternoon commutes. Figure 2c shows that users heading to

the southwest-northeast are more likely to share. In terms of socioeconomics, the percentage of non-white population in the pick-up census tract positively correlates with willingness to share, whereas median household income negatively correlates. This suggests that users from non-white and low-income neighborhoods (<$75,000) are more likely to request shared rides. As shown in Figure 3f, Education also has a significant effect on willingness to share, and it appears as a non-linear relationship that people without degrees are more willing to take shared rides, but when education level increases further, people seem to have no preference.

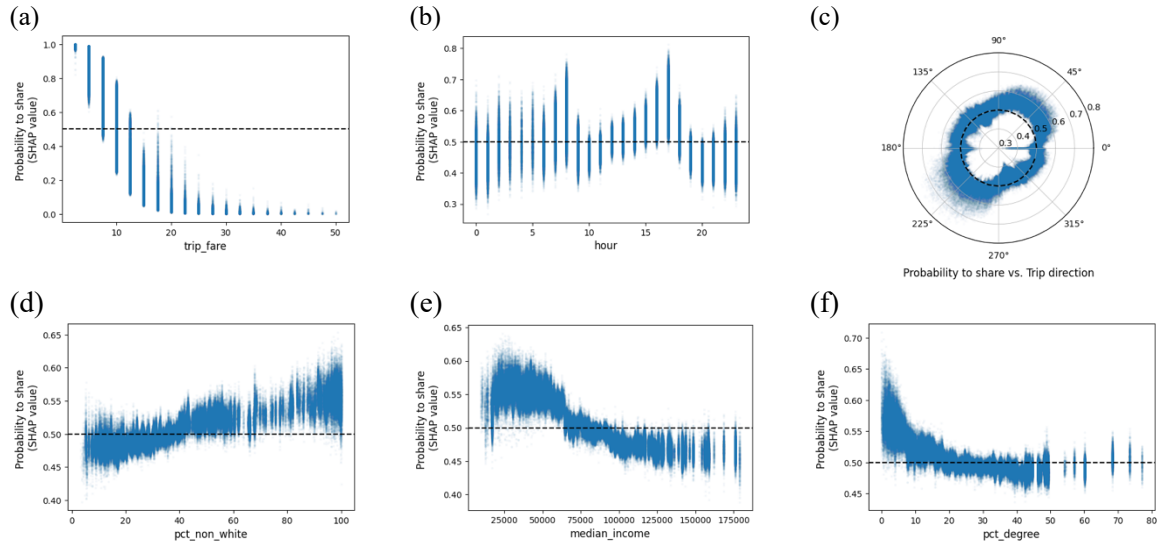


**Figure 2** Partial Dependence Plot of (a) trip fare, (b) starting hour, (c) trip direction, (d) percentage of non-White population in the pick-up census tract, (e) median household income in the pick-up census tract, and (f) percentage of people with a bachelor's degree in the pick-up census tract

For the influence of location in people's decision whether to request a share ride, we follow the approach introduced in Li (2022) to map the SHAP values of both longitude and latitude for pick-up and drop-off locations respectively. The resulting map (Figure 3) shows that the location effects in influencing people's willingness to share, after taking account of other factors. It appears that users who request trips from or to the west and northwest of Chicago are more likely to share.

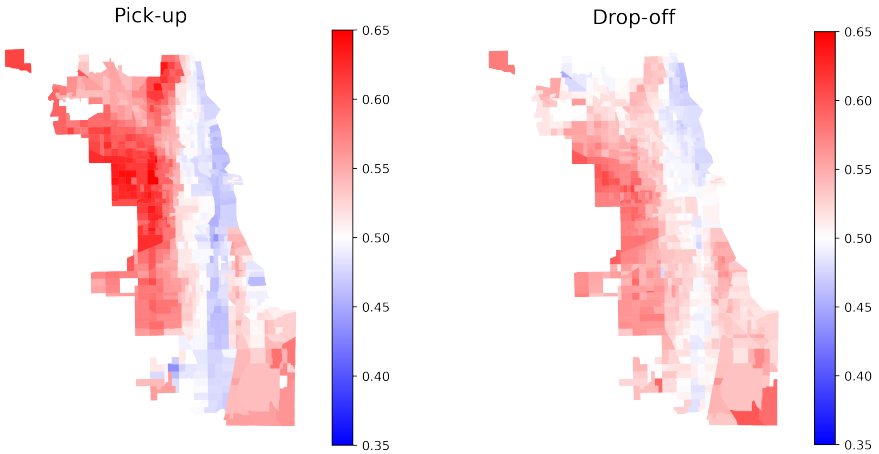


**Figure 3** The probability of people's willingness to share according to the pick-up (a) and drop-off (b) census tracts

## 4. Conclusion

This work presents a machine learning model based on more than 10 million trip records in the city of Chicago to understand users' willingness to share when requesting ride-hailing services. Local explainable AI method, SHAP, was used to interpret the ML model to identify key factors that influencing people's choice. Specifically, the cost of the trip remains the dominant incentive for people to share a ride with others, and it shows a negative non-linear relationship with the probability to share. Trip attributes such as distance, direction as well as temporal aspects are also identified to play a role in people's decision. There is a higher probability to share ride during commuting hours. For socioeconomics, users who requested trips from neighbourhoods with high percentage of non-white, low median household income, and low percentage of degrees are more likely to share the ride. Furthermore, there is also a geographical disparity that trips from or to the west of Chicago are more likely to be requested as a shared trip. This work helps to understand why people choose a shared ride over a solo ride; however, how shared rides may compete over public transportation or other shared mobility remains a further interesting topic of research. Also, future work will add point of interest (POI) data to increase the model predictability.

## References


Dean, M. D., & Kockelman, K. M. (2021). Spatial variation in shared ride-hail trip demand and factors contributing to sharing: Lessons from Chicago. *Journal of Transport Geography*, *91*, 102944.

Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion?. *Science advances*, *5*(5), eaau2670.

Hou, Y., Garikapati, V., Weigl, D., Henao, A., Moniot, M., & Sperling, J. (2020). Factors influencing willingness to pool in ride-hailing trips. *Transportation Research Record*, *2674*(5), 419-429.

Li, Z. (2022). An investigation of using SHAP to extract spatial effects from machine learning models.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768-4777).

Park, Y., Chen, N., & Akar, G. (2018). Who is interested in carpooling and why: the importance of individual characteristics, role preferences and carpool markets. *Transportation research record*, *2672*(8), 708-718.

Shaheen, S., & Cohen, A. (2019). Shared ride services in North America: definitions, impacts, and the future of pooling. *Transport reviews*, *39*(4), 427-442.


## Biographies

Ziqi Li is a Lecturer in GIScience at the University of Glasgow. His research interests broadly include spatial analysis and modelling, spatial statistical learning, interpretable machine learning, and their applications in multidisciplinary fields.

Tianyang XU is a graduate student in GIScience at the University of Illinois Urbana-Champaign. His research interests include spatial analysis and modeling, machine learning, and real-world applications of GIScience.