Efficiency and equality of the multimodal travel between public transit and bike-sharing accounting for multiscale

Abstract

As a supplement to the existing public transit system, bike-sharing is considered an effective solution to the "first mile" and "last mile" of travel. While many stakeholders believe that multimodal travel between public transit and bike-sharing can improve urban accessibility and sustainability, few studies have assessed the impact of bike-sharing on existing public transportation systems in terms of efficiency and equality. This research uses three months of mobile phone location data and about 140 million bike-sharing trips (origin-destination, OD) data from Shenzhen, China, to analyze first mile and last mile bike-sharing multimodal travel and measure the impact of bike-sharing on the existing public transportation system in terms of efficiency and equality at different scales. The research finds that bike-sharing is less effective in improving the operational efficiency of urban public transport and creates new inequalities at both global and local scales of the urban public transport system. Bike-sharing is only effective in tiny areas of the city and specific modes (subway-bike-sharing) and does not benefit groups with low socioeconomic levels and those living in edge areas of the city. Improving the equity and accessibility of public transportation is a key factor towards promoting sustainable urban development, and the analysis of this study on multimodal travel efficiency and inequality of bike-sharing can provide helpful insights for future sustainable urban planning.

Keywords: Multimodal travel, Bike-sharing, Efficiency and equality, Public transit system, Multiscale

1. Introduction

As an emerging mode of transportation, bike-sharing has garnered significant attention globally, being recognized as an eco-friendly and health-promoting solution to mitigate traffic congestion and diminish pollution emissions (Otero et al., 2018; Cerutti et al., 2019). Concurrently, bike-sharing serves as a complementary option to public transportation, offering a practical solution to the challenges of "first mile" and "last mile" connectivity (Shaheen and Chan, 2016).

While the popularity of bike-sharing has brought many social and environmental benefits, the unbalanced distribution and low utilization rate in bike-sharing have also caused new urban problems (Zhang et al., 2019). For example, shared bikes are in short supply at some bus stops, while many shared bikes occupy public space at some subway stations and obstruct the flow of pedestrian traffic sometimes leading to accidents (e.g., involving blind passengers). Improving the utilization rate of bike-sharing and its integration with the existing public transport system is an important and urgent issue.

Contemporary practices and studies indicate that the advantages of bike-sharing might be overstated (Hosford and Winters, 2018; Bauman et al., 2017; Hoffmann, 2016; de Chardon, 2019). Various studies have shown that bike-sharing users are predominantly male, affluent, healthier, younger, and well-educated, catering mainly to an already privileged demographic in urban centers (Pellicer-Chenoll et al., 2021; Duran et al., 2018; Bauman et al., 2017; Ricci, 2015) and primarily facilitate an already privileged population in increasingly exclusive urban cores (Hu et al., 2022; de Chardon, 2019). What is more, bike-sharing trips have mainly replaced walking, private cycling and public transport, but surveys of European and American cities have found that this substitution effect is likely to be negligible (Bauman et al., 2017; Ricci, 2015; de Chardon, 2019). Congestion caused by bike-sharing is also expected to reduce the efficiency of urban commuting (de Chardon et al., 2017; Castillo-Manzano et al., 2015). Thus, the question if bike-sharing is improving urban transport systems, especially public transport systems, in terms of equality and efficiency remains controversial.

China has become the world's largest bike-sharing market(Gu et al., 2019), driven by local governments and large amounts of capital, but it has also resulted in a massive waste of resources and a large amount of urban public space being taken up (Sun et al., 2023; Ma et al., 2018). However, most Chinese academics remain optimistic about bike-sharing and have focused their research on increasing bike-sharing usage and the factors that influence it (Gao et al., 2021; Li et al., 2020; Zhang et al., 2015). In recent years, many local governments, such as Shanghai, Guangzhou and Shenzhen, have introduced restrictive policies to control the unwarranted expansion of bicycle sharing and set penalties for operators (Hu and Creutzig, 2022). The bike-sharing market in China is becoming more orderly (Wang and Sun, 2022), and bike-sharing usage in China continues to increase following the lifting of the COVID-19 lockdowns (IDTP, 2020). Nevertheless, research has shown that the development of dockless bike-sharing projects is "mainly supply-driven by operators rather than by user demand or triggered by government policy" (Gu et al., 2019). Whether this private sector-driven travel mode is equitable and efficient for everyone remains a question.

With restrictive policies in place by local governments, what impact does bike-sharing in China have on the existing public transport system and will it make a difference? This is an important question to be answered. Firstly, the current mismatch between the supply and demand of shared bikes limits the adoption of bicycle sharing by users, reduces the connectivity of different public transportation modes, and reduces the potential for reducing urban emissions. Secondly, the benefits of bike-sharing are controversial as well because existing studies are based on spatial and temporal analysis at different scales (Li et al., 2020). Wang et al. (2022)have been implementing a data-driven method to show that bike-sharing increases public commuting efficiency in the city center, while Wu and Kim (2020) have used data from the United States, Canada and China concluding that bike-sharing in peripheral urban areas lacks connectivity to public transport networks, leading to low accessibility. Therefore, a multi-scale approach is needed to respond to the existing controversy. What is more, knowing what factors affect the spatial distribution imbalance of shared bicycles and the efficiency of multimodal travel can determine the placement and planning of bike-sharing, which will help build a sustainable transportation system and a livable city (Gallotti and Barthelemy, 2014).

Existing studies on bike-sharing-public transit are based on the premise that bike-sharing makes urban public transportation more equitable and efficient and that the benefits are homogeneous (Radzimski and Dzięcielski, 2021). Furthermore, these studies tend to focus on specific modes of multimodal travel, such as bike-sharing with subway feeders (Guo et al., 2021), and lack detailed distinctions between different modes of travel. More importantly, data limitations often make it difficult for researchers to obtain actual multimodal travel data, making it difficult to assess in detail the role bike-sharing plays in existing public transportation systems.

The primary objective of this paper is to analyze whether bike-sharing can enhance the efficiency and equality of urban public transportation. We focus on the "first mile" and "last mile" aspects of multimodal travel involving bike-sharing and public transit systems in Shenzhen, China. As conceptualized in Fig. 1, the "first mile" refers to the segments of a journey where a commuter uses a shared bike to reach a public transit station from their residence or workplace, while the "last mile" pertains to the journey from the transit station to their residence or workplace. To achieve this, we amalgamate two extensive datasets: a four-month record of 1.4 billion bike-sharing trips and a three-month aggregated dataset of 769,164 jobs-housing commuting ODs derived from mobile phone location data. Through strict and careful data processing, these two independent datasets are integrated to compute the probability of multimodal travel using shared bikes within grids. Concurrently, we employ community detection algorithms to assess the efficiency and equality of bike-sharing multimodal travel at both local and global scales. Lastly, we leverage interpretable machine learning techniques, specifically gradient boosted decision trees (GBDT), to delve deeper into the

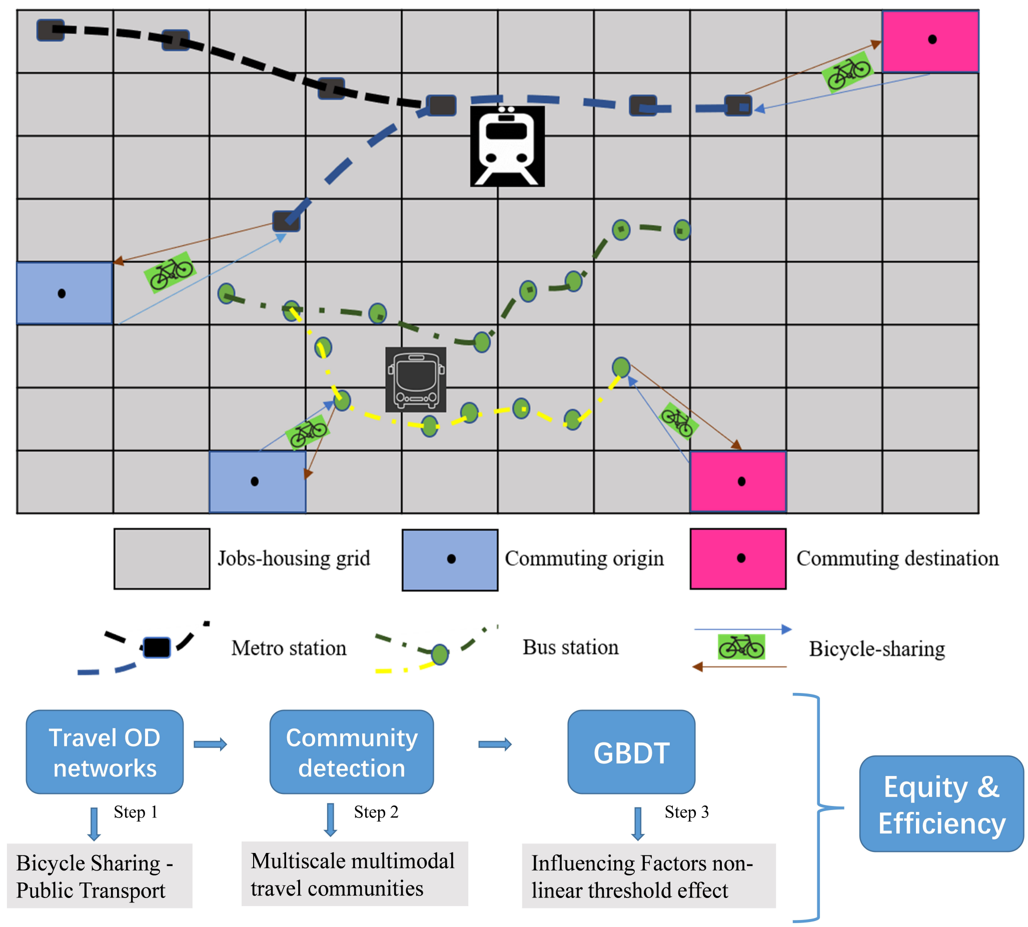

Figure 1: Research Framework. The above diagram conceptualizes bike-sharing-public transit multimode travel, and the following diagram is the analysis flow of this study.

nonlinear threshold effects of socio-economic factors influencing bike-sharing multimodal travel. These comprehensive approaches allow us to understand the multifaceted impact of bike-sharing on the existing public transportation system in terms of efficiency and equality.

The structure of this paper is outlined as follows: Section 2 provides a review of the relevant literature. In Section 3, we introduce the data mining methods and analytical techniques employed. Section 4 delves into the efficiency enhancement of bike-sharing multimodal travel and examines equality across various scales. Section 5 presents the conclusions and engages in a discussion on the findings, followed by suggestions for future research.

2. Literature review

This section will start with the multimodal travel of bike-sharing and public transportation, exploring their roles and impacts in cities, then delve into the relationship between job-housing commuting and spatial multiscale effects, and finally discuss the efficiency and equality issues of multimodal travel. Through this series of progressive reviews, we aim to provide readers with a comprehensive perspective to understand the role of bike-sharing in public transportation and the research gap, and finally propose this research approach.

2.1. Bike-sharing-public transit multimodal travel

Bike-sharing is seen as an emerging mode of transportation to solve the "first mile" and "last mile" travel problems of urban public transportation(Yu et al., 2021) and has become a key supplement to modern urban public transportation systems (Cheng et al., 2021). They provide passengers with a seamless connection, allowing them to easily travel from home/workplace to the nearest public transportation stations, or from the public transportation stations to their destination. This integration offers urban residents a more convenient and efficient multimodal travel option (Ricci, 2015; Fishman et al., 2014). Therefore, many studies have analyzed the factors influencing the use of shared bicycles and their connection with subway and bus systems (Yang et al., 2018; Caggiani et al., 2020; Li et al., 2020; Guo et al., 2021). These studies are generally based on the analysis of shared bicycle trajectory data from individual cities, such as Montreal (Faghih-Imani et al., 2014), Shanghai (Li et al., 2020), Seoul (Park and Sohn, 2017), Brasilia (Cerutti et al., 2019) and Barcelona (Faghih-Imani et al., 2017). Faghih-Imani et al. (2014, 2017) studied the effects of meteorological conditions, time of day, bicycle infrastructure, land use, and built environment on bicycle sharing use through mixed linear models. Most contemporary studies indicate that built environment and land use characteristics are crucial for bicycle use (Caulfield et al., 2017). For example, a high mix of land uses (Yang et al., 2021), the convenience of public transportation (Guo et al., 2021; Liu et al., 2022), and more supportive bicycle facilities (Lin et al., 2017) all promote the use of shared bicycles.

On this basis, by coupling bike-sharing data with public transportation stops, scholars further studied the impact of the built environment near public transportation stops (usually within 100 to 500 meters) on shared bicycle use and connection with public transportation (Ma et al., 2019; Guo and He, 2020; Guo et al., 2021; Liu et al., 2022), and proposed planning suggestions to increase this bike-sharing-public transit multimodal mobility (Caggiani et al., 2019; Saltykova et al., 2022). Studies have found that public transportation stops located in urban commercial centers with highly mixed land use often have more bike-sharing use (Fu et al., 2023). The above studies not only imply that we need to pay attention to the impact of the built environment and land use when analyzing shared bike-sharing-public transit multimodal travel, but also reflect the interactive relationship between the existing public transportation system and bike-sharing (Liu et al., 2022), and have sparked scholars' interest in multimodal travel research between bike-sharing and public transportation (Olafsson et al., 2016; Cheng et al., 2019; Guo et al., 2021).

Jäppinen et al. (2013) analyzed the impact of bike-sharing on the existing public transportation system through simulation and found that it can reduce public transportation travel time and improve public transportation accessibility. They proposed that bike-sharing should be regarded as part of urban public transportation and emphasized the need for integration with existing public transportation. Yang et al. (2018) found that shared bicycles can reduce passengers' average travel time, improve the efficiency of urban public transportation networks, and effectively alleviate the uneven spatial distribution of traffic flow in urban public transportation networks by constructing a multimodal travel network. In addition, some scholars have proposed route planning and station location models for bike-sharing to make this multimodal travel mode more efficient and equitable (Caggiani et al., 2020; Cheng et al., 2019).

In empirical terms, Wang et al. (2022) used mobile phone data from Beijing to model commuting modes and found that bike-sharing reduced commuting time, improved workplace accessibility, and significantly reduced horizontal and vertical inequalities in commuting time and workplace accessibility at both the individual and spatial levels. Kapuku et al. (2021) predicted the performance of multimodal travel with and without bike-sharing using machine learning, and by constructing a comparison of multimodal travel with and without bike-sharing, found that bike-sharing can effectively improve mobility. These studies are based on single data sources, simulating and analyzing the impact of bike-sharing on urban commuting efficiency and equality under ideal conditions, but these models often overlook complex urban commuting conditions, which can only be reflected by actual multimodal travel data. Therefore, to analyze the real impact of bike-sharing multimodal travel on public transportation commuting, data on jobs-housing commuting needs to be combined.

2.2. Jobs-housing commuting and multiscale effects

Jobs-housing commuting accounts for a large part of urban public transportation travel (Wu and Hong, 2017), and bike-sharing mainly meets users' commuting needs for the "first mile" and "last mile" (Ricci, 2015; Chen et al., 2022). A survey based in Shanghai, China found that after the emergence of bike-sharing, the proportion of cyclists commuting increased significantly from 21.9% to 30.9% (Jia and Fu, 2019), and the use of shared bicycles is also mainly concentrated during commuting hours (morning and evening peaks) (Li et al., 2020). Jobs-housing commuting reflects the stable daily travel pattern in cities (Hu and Wang, 2016), so when analyzing bike-sharing-public transit multimodal travel, special attention needs to be paid to the jobs-housing commuting mode. However, existing empirical studies on bike-sharing rarely consider the time and distance of jobs-housing commuting, and commuting time and distance are core factors affecting transportation mode choices (Redmond and Mokhtarian, 2001). Studies have found that the riding time (distance) range of bike-sharing users is concentrated at 2.5-10 minutes (500-2000 m) (Guo et al., 2021), but how the overall commuting time and distance of multimodal travel affect bike-sharing use remains to be explored. And commuting time and distance involve the core feature of travel, which is the spatial scale issue.

Although existing literature focuses on the impact of the built environment and mixed land use on bike-sharing use, it overlooks the spatial scale, a factor that may have a significant impact on multimodal travel. Firstly, the uneven spatial distribution of urban public transportation systems will affect the adoption of multimodal travel (Yu et al., 2021), so the efficiency and fairness of bike-sharing multimodal travel need to be analyzed from a spatial perspective. Secondly, spatial non-stationarity means that multiscale shared bicycles will affect public transportation at different scales (Yao and Kim, 2022); perhaps it only improves the efficiency of public transportation on a smaller spatial scale, such as in urban commercial centers with high bike-sharing deployment density, but on a larger scale and for groups living in non-core urban areas, it creates new social injustices (de Chardon, 2019). Therefore, we need to evaluate the efficiency and equality of bike-sharing multimodal travel from a multiscale spatial perspective.

2.3. Efficiency and equality of bike-sharing multimodal Travel

Most of the existing research on bike-sharing multimodal travel is based on the premise that bike-sharing enhances the efficiency of the existing public transportation system, and this gain is homogeneous in space (Wang et al., 2020; Guo and He, 2020; Yu et al., 2021). Only a few studies have evaluated the efficiency and fairness of multimodal travel (Jäppinen et al., 2013; Lu et al., 2018; Yang et al., 2018; Eren and Uz, 2020). Studies have found that the emergence of bike-sharing can reduce the time of public transportation commuting (Jäppinen et al., 2013) and can effectively enhance the accessibility of public transportation (Lu et al., 2018; Chen et al., 2020), benefiting more residents and improving the equality and sustainability of urban public transportation (Ricci, 2015).

However, some scholars believe that the oversupply of bike-sharing results in a lot of resource waste and greenhouse gas emissions (Wang and Sun, 2022), and the congestion of public spaces caused by it also reduces the efficiency of urban traffic operations (De Chardon et al., 2016). More importantly, some studies show that the improvement of urban public transportation efficiency by bike-sharing may have been exaggerated (Koglin and Mukhtar-Landgren, 2021; de Chardon, 2019; Castillo-Manzano et al., 2015; Audikana et al., 2017), and bike-sharing largely benefit the affluent elite and central urban areas, creating new social inequality for vulnerable groups and urban fringe areas (Ricci, 2015; Chen et al., 2020; Eren and Uz, 2020).

This paper believes that the main reason for these conflicting views is that existing research focuses on short-distance multmodal travel (Yang et al., 2018) and multi-modal travel near subway stations (Chen et al., 2020; Guo and He, 2020), without truly constructing a bike-sharing-public transit integrated travel chain, and lacks multiscale comparative analysis. Therefore, the factors found in these studies may not enhance multimodal travel. Although the real multimodal travel situation can be restored to some extent through traditional questionnaire survey methods (Olafsson et al., 2016), the surveyed population is small, and it is not easy to restore the commuting situation of the entire city scale.

More importantly, when measuring the efficiency and fairness of bike-sharing multimodal travel, most studies are based on the perspective of individual travel (Wang et al., 2022), that is, measuring the efficiency improvement of a single travel trajectory, rather than embedding it into the entire city commute for analysis, it is easy to get the conclusion that the efficiency improvement at the individual level can be ignored at the city-wide level. In addition, the use of bike-sharing is related to the built environment and is also affected by a series of socio-economic factors such as the income level of the region, population density, and the number of jobs (Ricci, 2015; Chen et al., 2020). And Fotheringham et al. (2017) found that different factors only have an impact on specific scales. For example, Liu et al. (2023) found that different distance thresholds from public transportation stations have different impacts on transportation mode choices, and it is not a linear relationship. This means that there are non-linear relationships and threshold effects of factors affecting multi-modal travel at different scales (Smart and Klein, 2018; Wu et al., 2019b). If we ignore the non-linear effects, the impact of variables will be misestimated. These non-linear effects are often unestimable by traditional spatial statistical methods. Traditional regression models are global models that can provide important information such as variables being positively or negatively correlated and specific coefficients, but it is difficult to provide non-linear and threshold effect information like machine learning models, and this information is often of practical significance for policymakers and stakeholders. Therefore, machine learning models have been applied in the field of urban transportation and built environment research (Tao and Cao, 2023; Yin et al., 2023). Scholars use GBDT and XGBoost and other machine learning algorithms to replace traditional regression algorithms, studying the non-linear threshold effects of the built environment and socio-economic factors on public transportation (Tang et al., 2020; Xiao et al., 2021; Liu et al., 2023).

In summary, the efficiency and equality issues of bike-sharing-public transit multimodal travel still need further exploration and integration. From the built environment and mixed land use to the spatial effects of jobs-housing commuting, to the non-linear threshold effects of socio-economic factors, these are all issues we must face when considering the combination of bike-sharing and public transportation. These research gaps provide us with further research directions. In the following sections, we will elaborate on the research approach of this study based on the content of the above literature review.

As a supplement to the existing literature, in terms of data, this study will use three months of mobile phone location data integrated jobs-housing commuting OD data and four months of bike-sharing data. Through strict and cautious data processing methods (see sections 4.1-4.2), these two independent data sets will be integrated to analyze the efficiency and equality issues of bike-sharing multimodal travel in the entire city. Since the data set spans a long time, the combination of data from the two sources can not only identify the relatively stable commuting and travel patterns in the city but also filter out the effects of sudden events such as holidays, weather, and traffic regulations on travel patterns. In terms of methods, considering the unequal spatial dependence of multimodal travel (e.g., areas with a large number of commuters and many public transportation stops will have more people adopting multi-modal travel), we introduced the spatial Gini index to evaluate the efficiency and equality of the city's global scale and further identified the local scale communities of bike-sharing multimodal travel through community detection algorithms (see section 4.3). These local areas represent communities with high frequencies of using bike-sharing multimodal travel, trying to exclude endogenous spatial non-stationarity interference as much as possible. Then we use the Taylor index to decompose the inequality within and between communities, thereby gaining a multiscale understanding of how bike-sharing affects the urban public transportation system. Finally, we employ the spatial entropy of POI as a metric for land use diversity, capturing nuances of the built environment. By incorporating socio-economic indicators such as housing prices and urban village areas, we leverage gradient boosting decision trees (GBDT) and interpretable machine learning techniques to discern the non-linear threshold effects of these factors on bike-sharing multimodal travel.

3. Datasets

In this study, we harness multiple datasets to unravel the intricate role of biKe-sharing within the public transit system of Shenzhen. At the core of this analysis are two mobility datasets: one capturing the jobs-housing commuting patterns derived from mobile phone app traces and the other detailing dockless bike-sharing trips sourced from a governmental API. These datasets, representative of Shenzhen's urban dynamics for the year 2021, are further enriched by integrating Point of Interest (POI) data, offering insights into the city's built environment and land use patterns. Additionally, we incorporate socioeconomic indicators, such as house prices and urban village areas, to provide a holistic understanding of the factors influencing commuting choices. Together, these datasets not only shed light on the current state of urban mobility but also pave the way for informed urban planning and transportation strategies.

3.1. Jobs-housing commuting data

We collected mobile phone location data for residents of Shenzhen from GeTui from April 1 to June 30, 2021 (second quarter of 2021). GeTui is a data company aggregating anonymous location data from mobile phone apps (GeTui, 2022), offering services similar to SafeGraph in the USA. Such datasets have been utilized in population mobility research and provide a reliable representation of urban mobility (Chen et al., 2023). The dataset was created based on high-resolution (100 m) Software Development Kit (SDK) location data from users of more than 100 smartphone apps. It encapsulates the jobs-housing relationships of millions of smartphone users in a Geohash6 (1.2KM*0.6KM grid) scale origin-destination (OD) matrix format.

To address potential concerns about user privacy, it's essential to note that this dataset is derived from three months of user historical data to calculate the city's OD commuting volume. While we cannot report the exact number of mobile users in the area due to privacy concerns, the dataset allows us to analyze the population whose home or workplace is in the target area on the analysis day. By tracing back the historical data of this population for 24 cycles (6 months), if more than half of the data falls within the target area, they are considered as permanent residents of the area. We identified the resident population of Shenzhen in 2019 as 10,218,569, in 2020 as 11,101,839, and in 2021 as 11,459,449. The latest census reported the permanent population of Shenzhen in 2020 as 17,560,061, implying that this dataset covers between 58.19% and 65.26% of Shenzhen's permanent population. Although the data does not cover 100% of the resident population, this is understandable. According to the 2020 census, the population aged 0-14 in Shenzhen was 2.6534 million (15.11%), and those aged 60 and above were 940,700 (5.36%). These groups are less likely to use smartphones, and even if they do, they might not be involved in work commuting and thus might be excluded from the jobs-housing data identification. Moreover, the sample size of this dataset is significantly larger than traditional survey data (typically $< 5\%$), indicating its strong representativeness for the region.

As illustrated in Fig. A.1, the specific data aggregation method is as follows:

1) Time-segmented location reporting frequency was used to calculate weights for the working hours (weekdays 10:00-17:00) and non-working hours (weekdays 21:00 to the next day 6:00 and non-working day periods). Based on the frequency of location reports, weights were assigned to different hours. The top three grids with the highest weights for both periods were identified daily.

2) The results from the past 12 weeks were aggregated. Based on the weights from the past 12 weeks, the weight values of each reported location were obtained. The locations with the highest weight values during working and

non-working hours were selected as potential workplaces and residences, respectively. A higher score indicates a higher reliability of the identified residence and workplace.

Through the data aggregation process described above, we have successfully compiled commuting origin-destination (OD) data for 2613 jobs-housing grids in Shenzhen, totaling 769,164 entries. Each entry contains details about the number of commuters and the latitude-longitude coordinates of the OD pair. Fig. 3 a visualizes the jobs-housing OD network, while Fig. A.2 a and Fig. A.2 b depict the visualization of the origin and destination grids, respectively. The color gradient of the grids provides insights into the number of commuters starting from and arriving at each grid, offering a comprehensive view of the spatial distribution of commuting populations throughout the city. This dataset offers a detailed perspective on Shenzhen's commuting patterns and serves as a valuable resource for further analysis and research.

3.2. Socio-economic data

Building upon the jobs-housing commuting data, it's essential to understand the socio-economic factors that influence these commuting patterns. Commuters' transportation mode choices are not only influenced by the spatial distribution of jobs and housing but also significantly by travel costs (DeSalvo and Huq, 1996). This choice is further moderated by individual economic status and the availability of alternative transportation options (Fearnley et al., 2018).

While many studies on bike-sharing have delved into urban travel costs and the built environment, there's a noticeable gap in research focusing on the socio-economic determinants. In this context, we introduce house prices and the area of urban villages within a jobs-housing grid as proxies for an area's socioeconomic status. House prices serve as an indicator of the economic affluence and purchasing power of residents in a given area. Conversely, urban villages in Chinese cities, characterized by their underdeveloped infrastructure and non-modernized built environments, mirror the city's residential and social divisions (Guo et al., 2021). Given the dispersed nature of urban villages in Shenzhen, their area within a grid offers insights into the socioeconomic development level of that specific grid, further enriching the understanding of the jobs-housing commuting patterns.

3.3. Dockless bike-sharing data

Dockless shared bicycle data was obtained by calling the API of the Shenzhen government data open platform [`https://opendata.sz.gov.cn/`], and we acquired 141,404,316 rows of data. Each row of data records the origin and destination (OD) of a shared bicycle trip, along with the start and end times and duration. The dataset spans 122 days, from March 1 to July 1, 2021. Each row of bicycle trip information includes bicycle ID, starting point coordinates, user ID, departure time, destination coordinates, and arrival time.

Considering the primary focus of this study on the "first mile" and "last mile" of bike-sharing trips, we believe that trips that are too short or too long cannot be considered as part of multimodal travel. Instead, they are more likely to represent single-mode trips using shared bicycles. This rationale led us to filter out trips based on certain criteria. Specifically, drawing on previous studies (Wu et al., 2019a; Guo and He, 2020), we removed data for trip distances up to 100 m and over 5 km. The exclusion of these trips is grounded in the understanding that they do not typically represent the "first mile" or "last mile" of a multimodal journey.

Furthermore, the duration of shared bicycle trips is generally short. Studies on public bicycle systems in Melbourne, Brisbane, Washington D.C., Minnesota, and London have shown that trip durations typically range between 16 to 22 minutes (Chen et al., 2020). This further justifies the decision to exclude trips with durations longer than 30 minutes. Additionally, research on the integration of shared bicycles with public transportation in China has found that most connections occur within a range of 500-2000 m from public transport stops (Guo et al., 2021). Furthermore, the decision to focus on trips between 100 meters and 5 kilometers is based on the typical commuting distances for shared bicycles. Trips shorter than 100 meters are often too brief to represent meaningful commuting, while those longer than 5 kilometers exceed the typical "last mile" distance and may not align with the primary use case of shared bicycles for short-distance, multimodal commuting.

3.4. POI data

We used Point of Interest (POI) data to measure the city's built environment and land use utilization. The POIs of Shenzhen were collected from the Gaode Map platform [`https://lbs.amap.com/`] for the year 2021. While there are established methods like Corine Land Cover to measure land use mix, the choice of POI data was driven by its timeliness, capturing the most recent urban changes that align with the bike-sharing dataset in 2021. Additionally, the Gaode Map platform provides a nuanced view of mixed-use with 17 distinct categories, offering a more detailed representation than broader classifications found in other datasets. This high-resolution POI data seamlessly complements other datasets, ensuring consistency in the analysis. Furthermore, our decision is underpinned by prior research (Yue et al., 2017; Xia et al., 2021; Im and Choi, 2019), which have effectively employed POI data in similar contexts. The information entropy formula we adopted offers a robust metric for assessing land use diversity, where higher entropy values indicate a richer mix of city functions. The formula is

$$\text{poi\_entropy}_{\text{grid}} = -\sum_{i=1}^{N} P_i \times \log_2 P_i \tag{1}$$

where N is the type of POIs and the value is 17, the number of POIs of each type is A_1, $A_2 \ldots A_N$ and the total number is A. The probability of each type of function is $P_i = A_i/A$. The level of information entropy can reflect the degree of mixing of city functions, and the higher the entropy value, the more types of functions and the higher the degree of mixing. This study also calculated the number of business and residential POIs within each jobs-housing grid. Previous studies found that the above explanatory variables correlate with human travel behavior (Liu et al., 2018), so the variable adoption is more reasonable.

4. Methodology

This section delineates the comprehensive methodology employed in our study, as illustrated in Fig. 2 We commence by harnessing bike-sharing and mobile phone data to discern multimodal travel trips. By analyzing extensive data over an extended timeframe, we identify the nexus between bike-sharing and public transport stations, thereby uncovering the underlying multimodal travel dynamics. Subsequently, we introduce the Gini and Thayer indices, elucidating how they facilitate the measurement of equality in our study. Efficiency, another pivotal metric, is gauged by computing the time saved through multimodal travel. To further refine our understanding, we deploy a community detection algorithm, enabling us to discern the local scale of multimodal travel and thereby assess efficiency and equality at both global and local scales. Concluding this section, we delve into the realm of interpretable machine learning, leveraging the prowess of LightGBM and SHAP. This approach aids us in discerning the factors influencing the adoption of multimodal travel, with a particular emphasis on socioeconomic determinants.

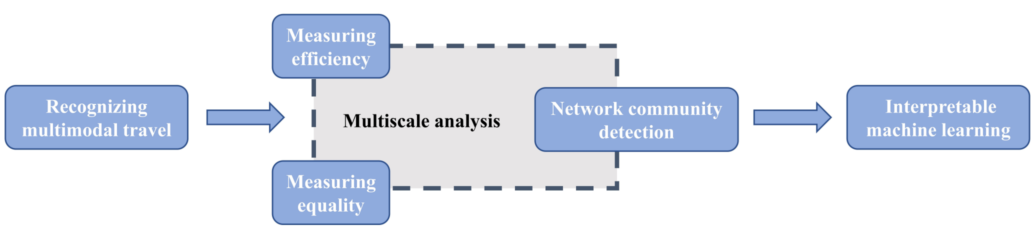

Figure 2: Research Methodology Workflow.

4.1. Identifying bike sharing - public transit multimodal travel trips

Public transit primarily serves the commuting needs of urban residents, forming long-term and stable travel patterns. This study aims to explore the supplementary role of bike-sharing in these patterns, especially in the "first mile" and "last mile" of public transit commuting. A pivotal hypothesis of this research is that bike-sharing trips arriving at or departing from public transit stations(within 100m) and having undergone rigorous data filtering, are indicative of a bike-sharing-public transit multimodal travel. Jin et al. (2019) first regarded Uber passengers who got off within 100 meters of a public transit station as multimodal connections. Wu et al. (2019a) pioneered the concept of considering rides within a 100 meters radius of subway stations as bike-sharing-subway transfer trips. Similar methodologies were employed by Guo and He (2020), Wang et al. (2020), and Guo et al. (2021) to investigate the interplay between bike-sharing and public transit. Furthermore, a survey by Guo et al. (2021) involving 1,167 participants from 22 subway stations in Shenzhen, China revealed that over 95% of commuters either parked or initiated their shared bicycle rides within 100 meters of a subway station. Numerous studies have underscored the spatial correlation between bike-sharing usage and public transit stations, highlighting their synergistic relationship (Nair et al., 2013; Schimohr and Scheiner, 2021; Saltykova et al., 2022).

Previous research directly categorized shared bicycle trips within 100 meters of public transit stations as multimodal transport. In contrast, this research's approach, supported by a larger dataset and strict data processing measures, makes the hypothesis more convincing. Admittedly, despite utilizing a vast dataset spanning 3 months of mobile phone data and 4 months of bike-sharing data, we must admit that not all trips are multimodal. In reality, many trips, such as those to nearby eateries or shops around transit stations, might not encompass multiple modes of transport. It has always been challenging to record the complete travel modes of humans. Most scholars use questionnaires or sensors carried by humans. The former is limited in its ability to conduct large-scale surveys, especially across an entire city, while the latter involves cost and privacy issues, making it difficult to apply in many studies. One advantage of big data, despite its inherent biases and inability to definitively record multimodal trips, is its capacity to offer insights through meticulous data processing and conditional constraints.

Therefore, to ensure that the bike-sharing data aligns with this research theme and hypothesis, we conducted meticulous and cautious data processing. Firstly, the bike-sharing dataset spans 122 days, which we believe is sufficient to identify stable travel patterns and filter out random trips and unexpected events such as holidays or adverse weather. The analysis is limited to the morning peak (7:00 a.m. to 9:00 a.m.) and evening peak (5:30 p.m. to 7:30 p.m.), the two main commuting periods, which matches well with the commuting data (aggregating from mobile phone data between 9:00 p.m. and 6:00 a.m. and between 10:00 a.m. and 5:00 p.m.). Secondly, when constructing a 100-meter buffer around public transit stations (Fig. 3 b), we assume that a shared bicycle completes a multimodal trip when it arrives or departs from the 100-meter buffer. Drawing on previous research on bike-sharing and public transit connections, we excluded trips that are less than 100 meters or more than 5 kilometers, as well as trips lasting more than 30 minutes (Wu et al., 2019a; Guo and He, 2020). Notably, the bike-sharing dataset contains user IDs, allowing us to distinguish consistent travel patterns of individual users. During the data cleaning process, we only considered users who used bike-sharing services more than 20 times during peak hours in a month, as there are an average of 20 working days in a month. After rigorous data processing, we obtained 20,568,361 bike-sharing multimodal trips, accounting for approximately 15.55% of the 141,404,316 original trips.

From the jobs-housing origin grid to the public transit stations, we recorded the first mile of multimodal travel. From the public transit stations to the jobs-housing destination grid, we recorded the last mile of multimodal travel. Public transit stations are divided into bus stations and subway stations, so we can ultimately identify eight types of multimodal travel, namely the first mile of travel during the morning and evening peaks (bike-sharing-bus, bike-sharing-metro) and the last mile of travel during the morning and evening peaks (bus-bike-sharing, metro-bike-sharing).

4.2. Measuring efficiency

Leveraging Baidu Maps, one of China's premier map navigation platforms [`https://map.baidu.com`], we planned the route for multimodal travel. This allowed us to determine the time cost for a jobs-housing OD using both the walking-public transportation mode and the bike-sharing-public transportation mode. By comparing the two time costs, we calculated the time savings (TS) for each jobs-housing grid, representing the efficiency gain from using bike-sharing for multimodal commutes. The following equations depict the calculation of the first mile multimodal trip efficiency gain for each jobs-housing grid:

$$\mathrm{TS}_i = \frac{\sum_{j=1}^{\mathrm{N}_i} (\mathrm{TCW}_j - \mathrm{TCB}_j)}{\mathrm{N}_i} \tag{2}$$

$$\mathrm{MLR}_i = \frac{\sum_{\varepsilon=1}^{4} \mathrm{count}_\varepsilon}{\mathrm{N}_i} \tag{3}$$

$$\mathrm{TTS}_i = \mathrm{TS}_i \times \mathrm{MLR}_i \tag{4}$$

Here, TS_i represents the average time saved for all commuting trajectories originating from grid i by adopting the bike-sharing-public transit multimodal travel compared to the walking-public transit travel. N_i is the total number of job-housing commuting trajectories originating from grid i. j represents a specific job-housing OD trajectory, originating precisely within grid i, hence the total number of j equals N_i. TCW_j and TCB_j respectively represent the commuting time for trajectory j using the walking-public transportation mode and the bike-sharing-public transportation mode. In Equation (3), ε represents the type of the first-mile multimodal travel, which are the morning peak: bike-sharing-bus, bike-sharing-metro and the evening peak: bike-sharing-bus, bike-sharing-metro, thus totaling 4 types. This formula calculates the probability of adopting multimodal bike-sharing travel for the first mile originating from grid i (MLR_i). Finally, we obtain the total travel time efficiency gain for grid i (TTS_i) by multiplying TS_i by MLR_i.

In this study, it's important to note that the bike-sharing data and mobile phone data are sourced from two distinct datasets. We employed mobile location data spanning from April 1 to June 30, 2021, to discern the consistent commuting patterns within each jobs-housing grid, each approximately sized at 1.2KM*0.6KM. By comparing the number of multimodal trips by bike-sharing in each grid to the overall commuting volume, we were able to determine the proportion of such multimodal trips for each grid. This metric serves as an indicator, suggesting the likelihood of residents in a particular grid opting for bike-sharing as part of their multimodal commuting routine over an extended period. The insights derived from these two comprehensive datasets offer a representative snapshot of bike-sharing's role in multimodal urban travel. Notably, this data-driven approach presents a more cost-effective alternative to traditional methods like questionnaire surveys or sensor-based tracking, especially when scaled to larger urban areas. Additionally, it sidesteps potential concerns related to privacy and research ethics.

4.3. Measuring equality

To assess the equality of bike-sharing multimodal trips, we employ two widely recognized indices: the Gini index and the Thiel index.

In this study, we first calculated the Gini index of efficiency gains brought about by multimodal trips on a global scale for each grid. This aimed to analyze the equality of bike-sharing multimodal trips throughout the city scale. The Gini index is given by:

$$G = \frac{\sum_i \sum_j |y_i - y_j|}{2n^2\bar{y}} \tag{5}$$

Here, $\bar{y}$ is the average efficiency gains in multimodal trips for the grid. y_i and y_j represent the efficiency gains of the i_{th} and j_{th} grids, respectively. The term $2n^2\bar{y}$ is a normalization factor, ensuring the Gini index remains between 0 (complete equality) and 1 (complete inequality).

Considering the spatial dependency of inequality in multimodal trips, we introduced the spatial Gini index:

$$G_{spatial} = \frac{\sum_i \sum_j w_{i,j}|y_i - y_j|}{2n^2\bar{x}} + \frac{\sum_i \sum_j (1 - w_{i,j})|y_i - y_j|}{2n^2\bar{x}} \tag{6}$$

In this equation, $w_{i,j}$ from the binary spatial weights matrix indicates the spatial relationship between the i_{th} and j_{th} grids. If two grids are neighbors, $w_{i,j}$ is 1; otherwise, it's 0.

The Thiel index, a measure used to assess spatial inequality, is especially relevant in the fields of Geographic Information Systems (GIS) and spatial economics (Shorrocks and Wan, 2005; Novotnŷ, 2007). After obtaining the local scale of multimodal trips through community detection, the spatial Gini index can only measure the equality of grids located within all communities, making it a global model. The Thiel index, however, can be decomposed into two parts: B (Between-group inequality) and W (Within-group inequality), allowing us to measure fairness between and within communities, analyzing the equality of multimodal trips on a local scale.

$$
\begin{aligned}
T &= \sum_{i=1}^{m}\left(\frac{y_i}{\sum_{i=1}^{m} y_i}\ln\left[m\frac{y_i}{\sum_{i=1}^{m} y_i}\right]\right) \\
&= \left[\sum_{g=1}^{w} s_g \ln\left(\frac{m}{m_g} s_g\right)\right] + \left[\sum_{g=1}^{w} s_g \sum_{i\in g} s_{i,g}\ln\left(m_g s_{i,g}\right)\right] \\
&= B + W
\end{aligned}
\tag{7}
$$

In this formula, m represents the total number of grids, and y_i represents the efficiency gain of the i_{th} grid. The first equation of (7) calculates overall inequality based on the concept of entropy from information theory. In the second equation, the number of grids in community g is m_g, and the total number of multimodal trip communities is w. $s_g = \frac{\sum_{i\in g} y_i}{\sum_i y_i}$ represents the proportion of the total efficiency gain of community g to the overall gain, while $s_{i,g} = \frac{y_i}{\sum_{i\in g} y_i}$ represents the proportion of the efficiency gain of the i_{th} grid in community g to the total gain of that community. In the context of Theil's index, values close to 0 indicate a low level of inequality, whereas values approaching 1 signify a high level of inequality.

4.4. Multimodal travel network community detection

Community detection algorithms in complex network analysis are designed to group nodes in a graph such that connectivity within groups (communities) is maximized relative to cross-community connections. In a spatial graph, nodes associated with locations naturally form regions that can be represented by, for instance, the convex hull, alpha shapes, or simply the centroid of the geometries associated with the nodes. The diameter or area of the convex hull can be used to derive a spatial scale.

The primary motivation behind applying community detection in this study is to gain a multiscale understanding of how bike-sharing affects the urban public transportation system. Before community detection, we analyzed the global scale of bike-sharing multimodal travel throughout the city using the original grid. Given the city's public transportation system's inherent spatial disparities, such as the concentration of public transportation stops in city centers and the prevalence of bike-sharing deployments, we aimed to minimize the influence of these intrinsic factors. Through community detection, our objective was to pinpoint communities characterized by a high frequency of bike-sharing multimodal travel. This method enabled a more nuanced assessment of equity both within and between these communities at a localized scale.

We employed the fast unfolding algorithm (Blondel et al., 2008), which is essentially a specialized clustering algorithm for network (OD) data. The algorithm works by continuously optimizing modularity to discover community structures in the network. Modularity measures how edges in the network are concentrated within specific communities compared to a random placement. By maximizing modularity, the algorithm effectively groups tightly connected nodes into communities while ensuring sparse connections between these communities.

Modularity is a measure used in community detection algorithms to evaluate the quality of a division of a network into communities or clusters normalized to take values in [-1, 1] (Newman, 2006). It is generally considered that modularity between 0.3 and 0.7 indicates a more appropriate result of community division (Newman, 2004). In this study, we computed a range of 0.3

to 0.7 for all travel networks extracted from the data. The modularity was calculated by the formula,

$$\text{Modularity } = \frac{1}{2m}\sum_{i,j}\left[w_{ij}-\frac{k_i k_j}{2m}\right]\delta\left(c_i,c_j\right) \tag{8}$$

Let $\delta(c_i, c_j)$ be defined by the function $f(x)$ such that:

$$f(x) = \begin{cases} 1, & \text{if } c_i = c_j \\ 0, & \text{if } c_j \neq c_i \end{cases}$$

where w_{ij} denotes the edge-weight between nodes v_i and v_j. The term k_i represents the cumulative edge-weights associated with node v_i. The symbol c_i designates the community to which node v_i belongs. Furthermore, $2m$ signifies the aggregate of all edge weights present in the network.

4.5. Interpretable machine learning: LightGBM and SHAP

The application of machine learning, particularly in the realm of urban transportation, has become increasingly prevalent due to its ability to capture complex non-linear relationships and threshold effects that traditional spatial statistical methods might overlook (Tang et al., 2020; Xiao et al., 2021; Liu et al., 2023). In this study, we opted for machine learning over traditional regression models to address potential multicollinearity issues and accommodate missing values, and outliers, which are often challenging for traditional regression models. Understanding why a model makes a certain prediction is crucial to interpret results, explain differences between models, and assess to what extent we understand the phenomenon under analysis. Moreover, the Ethics Guidelines for Trustworthy AI of the EU High-Level Expert Group on AI suggest that the behavior of AI systems should be transparent, explainable, and trustworthy. We use both LightGBM (Ke et al., 2017) and SHAP (Lundberg and Lee, 2017) frameworks to open the black box of machine learning and analyze what factors influence bike sharing multimodal travel.

LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework that is efficient and scalable. Travel behavior is complex, leading to potential multicollinearity issues when introducing many independent variables. LightGBM is particularly adept at handling large datasets and addressing the challenges posed by multicollinearity among independent variables. The GBDT algorithm, an integral part of LightGBM, is an iterative decision tree algorithm that can analyze the non-linear threshold effect of different influencing factors, providing a precise reference for realistic and sustainable traffic planning. This feature is invaluable for policymakers and stakeholders, guiding decisions like deploying varying numbers of shared bikes in areas with different housing prices to prevent resource wastage.

The GBDT algorithm, often referred to as MART (Multiple Additive Regression Trees), is an iterative decision tree algorithm (Freund et al., 1996). It seamlessly integrates the principles of decision trees with gradient boosting techniques. Initially, the data samples are partitioned into various subgroups using a decision tree. Subsequently, the mean of the observations within each subgroup serves as the prediction for those observations. This step results in a prediction error, which the GBDT algorithm utilizes to recalibrate the weights of each independent variable for the subsequent rounds of classification and prediction. An illustrative example of the GBDT algorithm is provided below:

Algorithm 1 Gradient Boosting Decision Tree (GBDT)

Input: Training data $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, number of trees T, learning rate η
Output: Final model $F(x)$
Initialize model with a constant prediction: $F_0(x) = \arg\min_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$
for $t = 1$ to T **do**
    Compute the negative gradient (pseudo-residuals):
    for $i = 1$ to n **do**
        $r_{it} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F=F_{t-1}}$
    end for
    Fit a decision tree to the pseudo-residuals, resulting in leaves $\{R_{t1}, R_{t2}, \ldots, R_{tJ}\}$
    for $j = 1$ to J **do**
        $\gamma_{tj} = \arg\min_\gamma \sum_{x_i \in R_{tj}} L(y_i, F_{t-1}(x_i) + \gamma)$
    end for
    Update the model:
    $F_t(x) = F_{t-1}(x) + \eta \sum_{j=1}^{J} \gamma_{tj} I(x \in R_{tj})$
end for
return $F(x) = F_T(x)$

Moreover, GBDT can generate partial dependence plots, which can be used to assess the nonlinear relationships between variables through multivariate analysis. It also allows for the assessment of the interaction effects between two or more independent variables. By providing the relative importance of the independent variables, GBDT offers insights into the significance of each variable in planning practice.

We use SHapley Additive exPlanations (SHAP) to understand how input features determine the output of GBDT. SHAP is based on game theory and estimates the contribution of each feature based on the best Shapley value(Mangalathu et al., 2020), indicating how the presence or absence of the feature changes the model prediction for a particular instance compared to the average predictive value of the dataset.

5. Results

In this section, we begin by assessing the efficiency enhancements achieved by incorporating bike-sharing into urban multimodal travel, shedding light on both the improvements and the spatial inequities at the city scale. Transitioning to a more granular perspective, we identify and analyze specific communities at the local scale, emphasizing areas with high bike-sharing usage and the disparities inherent within them. In our in-depth analysis, we utilize advanced machine learning to uncover the nonlinear threshold effects of key determinants, including commuting costs, socio-economic factors, and built environment attributes, on the adoption of bike-sharing in multimodal travel.

5.1. Efficiency and equity at the city scale

When evaluating the efficiency improvements brought about by integrating bike-sharing into multimodal travel, it's crucial to consider the broader, city-wide perspective. By calculating the first and last mile total travel time efficiency gain (TTS_i) for each jobs-housing grid, we can gauge the citywide efficiency enhancement of multimodal travel with bike-sharing. Fig. 3 c, d illustrates the distribution of these efficiency gains for the first and last mile, respectively.

From a global scale, the efficiency gains introduced by bike-sharing to the existing public transportation system appear relatively modest. These gains are primarily concentrated in the city center, particularly in the mid-western regions, where they can save about 1-10 minutes per multi-modal journey. Notably, the first mile sees a more pronounced efficiency improvement than the last mile. However, in areas farther from the city center, bike-sharing doesn't seem to enhance commuting efficiency significantly. This observation underscores a broader issue of equity at the citywide scale. While certain central areas benefit from the integration of bike-sharing, outlying regions remain relatively underserved, highlighting an inherent spatial inequality in the distribution of these efficiency gains.

Building on the observations of spatial inequality in the efficiency gains brought about by bike-sharing, it's essential to delve deeper into the metrics that quantify this disparity. The spatial distribution of these gains, as previously discussed, is not uniform across the city, suggesting a pronounced spatial inequality in the benefits of bike-sharing for urban public transportation.

To quantify this spatial inequality, we computed the spatial Gini index for both the first and last mile TTS. The results are telling: a spatial Gini index of 0.8996 for the first mile and 0.8750 for the last mile. Conventionally, a Gini index exceeding 0.5 is indicative of high inequality. These values, therefore, underscore a pronounced inequality in the efficiency improvements

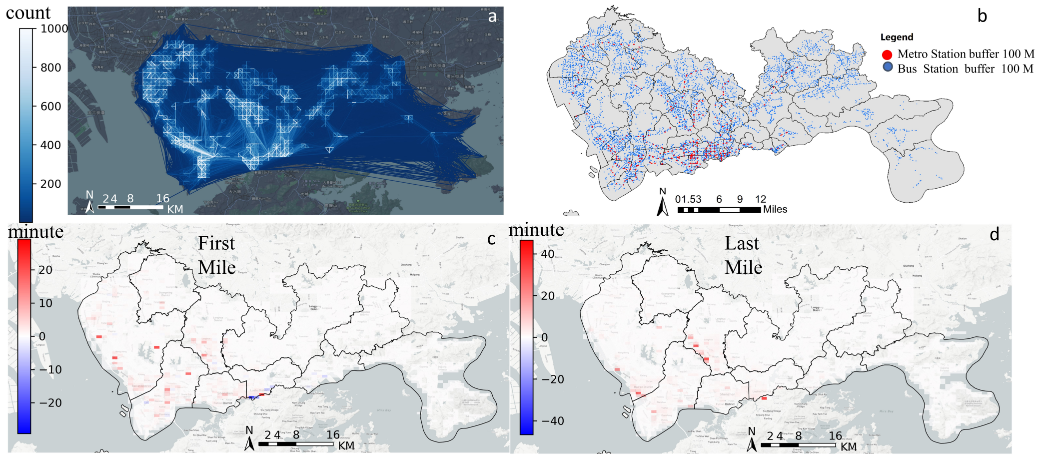

Figure 3: Commuting OD networks, public transit station buffer and multimodal travel efficiency gains. (a) Jobs-housing commuting OD network, where 2,631 jobs-housing grids serve as the nodes of the network, and 769,164 commuting origin-destination pairs act as the connecting edges. The color gradient of the edges, transitioning from dark to light, indicates the number of commuters (count), with lighter shades representing higher commuter counts. (b) 100-meter buffer zone of bus stations and metro stations in Shenzhen. (c, d) Each grid's total travel time efficiency gain, TTS (minutes), (c) denotes the TTS from the grid to the public transit station (first mile), and (d) represents the TTS from the public transit station to the grid (last mile).

brought about by bike-sharing on a citywide scale. This inequality is further visualized by the near-vertical Lorenz curve depicted in Fig. 4 c, d.

Further insights into this inequality can be gleaned from the kernel density distribution of the TTS, as shown in Fig. 4 a, b. Here, a significant clustering of values around 0 is evident, with only a sparse distribution above 10. Notably, while the first mile values display a broader spread, the last mile efficiency improvements register higher values. This distribution pattern might be attributed to the strategic decisions of bike-sharing service providers. By concentrating bike-sharing resources near public transportation stations, they potentially reduce management costs. This strategy is further evidenced by Appendi Fig. A.3, where it's clear that the last mile usage of bike-sharing significantly outpaces that of the first mile.

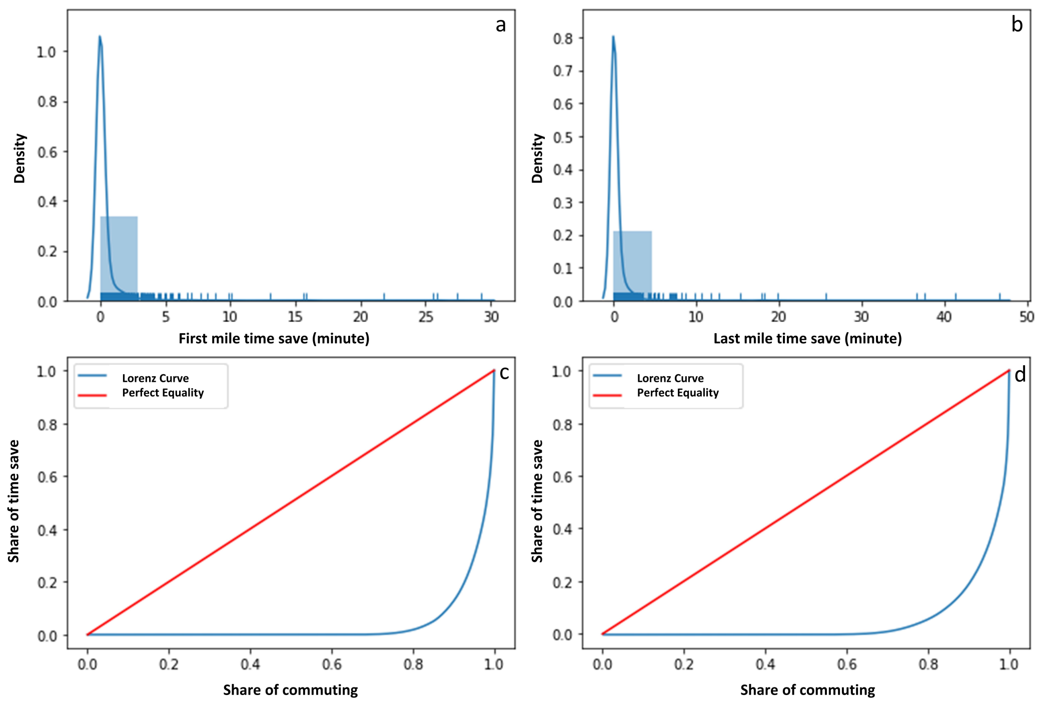

Figure 4: Kernel density distribution plot (a, b) with Lorenz curve(c, d) for efficiency gain of bike-sharing multimodal travel. (a, c) illustrate bike-sharing multimodal travel for the first mile and (b, d) illustrate multimodal travel for the last mile.

In the city-scale analysis, we observed pronounced spatial inequalities in the efficiency improvements brought about by bike-sharing in urban public transportation. However, attributing these disparities solely to bike-sharing might be an oversimplification. External factors, such as the uneven distribution of transportation infrastructure, can also play a significant role. For instance, the less dense arrangement of metro and bus stops in eastern Shenzhen, as depicted in Fig. 3 b, could inherently lead to lower efficiency gains in that region. Such disparities in transportation infrastructure can significantly skew the overall picture of inequality at the city-wide level.

While the global perspective provides a broad understanding, it's essential to delve deeper to discern the nuances of these disparities. Specifically, we need to investigate whether similar patterns of inequality persist in localized areas where bike-sharing is frequently integrated with other modes of transportation. By focusing on these high-frequency areas, we can better isolate the impact of bike-sharing from other potential confounding factors and gain a clearer understanding of its role in shaping transportation equity.

In the following subsection, we will explore equity at the local scale, employing community detection algorithms to identify and analyze regions with high bike-sharing multimodal travel usage. This granular approach will allow us to determine if bike-sharing inherently contributes to transportation inequalities or if other factors are predominantly at play.

5.2. Equity at the local scale

Through the application of the community detection algorithm from complex network theory, we identified mesoscale multimodal travel communities, as illustrated in Fig. 5 . The modularity of these communities surpasses 0.85, suggesting robust intra-community connections and validating the appropriateness of our community identification. The spatial distribution of these communities aligns well with regions of high commuting demand, as depicted in Fig. 3 a.

Notably, the spatial extent of bike-sharing-bus travel communities is considerably more expansive than that of bike-sharing-metro communities. This trend is likely influenced by the denser distribution of bus stations compared to metro stations. As observed in Fig. A.3, the coverage and quantity of bike-sharing-bus multimodal travel surpass that of bike-sharing-metro. The former spans the southwestern and central regions of Shenzhen, while the latter is primarily confined to stations in the southwest. Additionally, there's a pronounced central-peripheral structure within these travel communities: the city's core is characterized by larger, more cohesive communities, whereas its outskirts are dotted with smaller, fragmented ones.

Given that each identified travel community represents a city region with a high frequency of multimodal trips, they serve as ideal units for assessing local-scale inequalities in bike-sharing-public transit integration. To quantify these disparities, we employed Theil's index to compute both within-group (Theil(W)) and between-group (Theil(B)) inequalities for various multimodal travel communities. The summation of these components provides a comprehensive view of local equity through Theil's total index.

Table 1 displays the intra-community and inter-community inequalities for the eight types of bike-sharing-public transit multimodal travel. From the spatial Gini coefficient ($G_{spatial}$), we discern that, for both the first and last miles, the inequality in bike-sharing-metro multimodal travel is lower than that of bike-sharing-bus. This observation is also captured by the Theil index. Crucially, the inequality between different travel communities is typically less pronounced than the inequality within individual communities. Moreover, the inter-community inequality constitutes a minor fraction of the overall inequality, as indicated by $\text{Theil(B)}_{\text{share}}$.

Table 1: Within-group inequality and between-group inequality for eight types of multimodal travel communities.

	Type	$G_{spatial}$	Theil(B)	Theil(W)	Theil	$\text{Theil(B)}_{\text{share}}$
First Mile	a B-B-M	0.77	0.399	0.783	1.182	33.8%
	b B-M-M	0.669	0.193	0.626	0.819	23.5%
	c B-B-E	0.783	0.384	0.856	1.24	31%
	d B-M-E	0.684	0.233	0.632	0.865	27%
Last Mile	e B-B-M	0.713	0.387	0.572	0.959	40.4%
	f M-B-M	0.601	0.237	0.395	0.632	37.5%
	g B-B-E	0.735	0.415	0.624	1.0387	39.9%
	h M-B-E	0.598	0.214	0.41	0.625	34.3%

Note: Types a-h correspond to the multimodal travel communities a-h in Figure 5. For the first mile: B-B-M is bike-sharing-bus in the morning, B-M-M is bike-sharing-metro in the morning, B-B-E is bike-sharing-bus in the evening, and B-M-E is bike-sharing-metro in the evening. For the last mile: B-B-M is bus-bike-sharing in the morning, M-B-M is metro-bike-sharing in the morning, B-B-E is bus-bike-sharing in the evening, and M-B-E is metro-bike-sharing in the evening. G_{spatial} is the overall spatial Gini index. Theil(B) and Theil(W) indicate between and within community differences, respectively. Theil represents the total index, and $\text{Theil(B)}_{\text{share}}$ is the proportion of Theil(B) in the total Theil index.

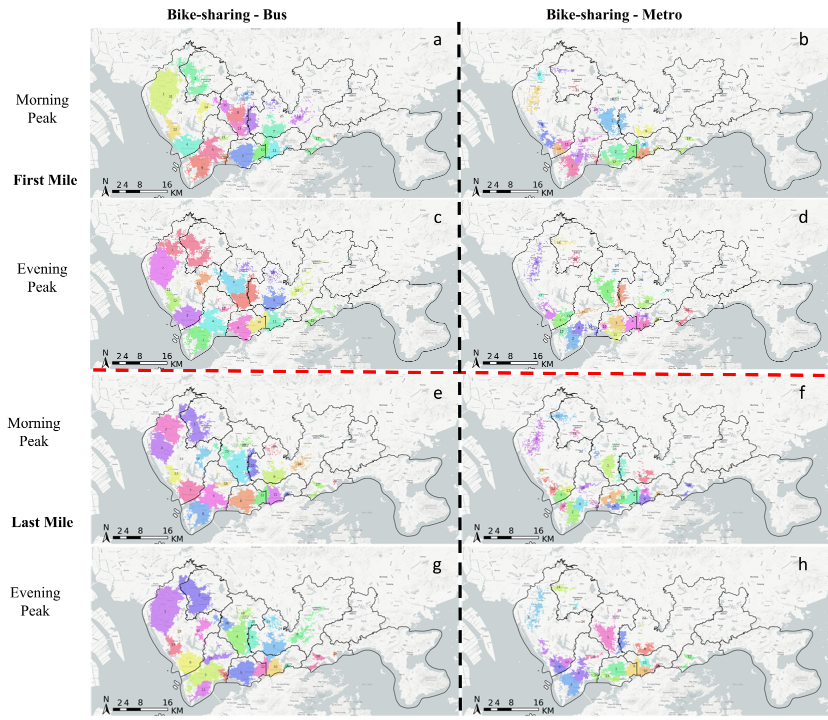

Figure 5: Eight types of multimodal travel communities. Above the red dashed line: first mile; below: last mile. Left of the black dashed line: bike-sharing-bus; right: bike-sharing-metro. (a, c) Morning and evening peak for first mile bike-sharing-bus. (b, d) Morning and evening peak for first mile bike-sharing-metro. (e, g) Morning and evening peak for last mile bus-bike-sharing. (f, h) Morning and evening peak for last mile metro-bike-sharing.

On one hand, this suggests that the communities identified through our community detection algorithm exhibit significant spatial homogeneity and comparability in multimodal travel patterns. On the other hand, it underscores that even within communities with high multimodal travel frequencies, substantial inequalities persist. An exception to this trend is observed in the last-mile metro-bike-sharing travel groups (Types f and h), where the inequality is less than 0.5. This indicates that bike-sharing deployments at metro stations effectively cater to commuters' needs during peak hours, facilitating the last mile of their journeys. However, for all other multimodal travel types, the inequality remains pronounced. This highlights the limited effectiveness of bike-sharing in enhancing the overall efficiency of the urban public transportation system and underscores the emergence of new disparities, evident at both local and global scales.

The root causes of these disparities might be attributed to the inherent spatial imbalances of the existing public transportation system, the commuting needs of urban residents in terms of jobs and housing, and the socio-economic development across different regions. To delve deeper into the underlying reasons for this observed inefficiency and inequality, a robust machine learning approach is warranted, setting the stage for our subsequent analysis.

5.3. Nonlinear threshold effects of influencing factors

Building upon the insights from our literature review, which highlighted the multifaceted roles of bike-sharing in urban commuting, the significance of jobs-housing commuting, and the concerns surrounding the efficiency and equality of multimodal travel, we sought to delve deeper into the determinants of bike-sharing usage. Recognizing the potential non-linear relationships and threshold effects emphasized in prior studies (Tang et al., 2020; Xiao et al., 2021; Liu et al., 2023), we employed the Gradient Boosting Decision Trees (GBDT) model, implemented through the LightGBM framework.

We formulated two distinct models to predict bike-sharing usage for multimodal travel within the jobs-housing grid: one for the first mile and another for the last mile.

1. **Commuting Cost**: Drawing from the literature that emphasizes the importance of commuting time and distance on transportation mode choices (Redmond and Mokhtarian, 2001; Guo et al., 2021), we included variables such as the average commuting time by public transportation ("commu_duration"), average commuting distance ("commu_distance"), and the average distance and duration for the first/last mile by bicycle ("first/last_distance" and "first/last_duration"). Additionally, we considered the time saved when opting for car commuting over public transportation ("cartime_save") and the time saved by using a shared bike for the first/last mile compared to walking ("savetime").
2. **Built Environment**: Prior research has consistently shown that the built environment, especially mixed land use, plays a pivotal role in influencing bike-sharing usage (Caulfield et al., 2017; Guo and He, 2020; Chen et al., 2020). Thus, we incorporated variables like mixed land use calculated using POI data ("poi_entropy"), the number of company POIs ("company_poi"), and the number of commercial residential POIs ("residence_poi").
3. **Socio-economic Status**: Recognizing the socio-economic disparities in bike-sharing usage as highlighted in the literature (Ricci, 2015; Chen et al., 2020; Eren and Uz, 2020), we included variables representing the working population ("work_pop"), living population ("home_pop"), urban village areas ("village_area"), and house prices within the jobs-housing grid ("house_price").

To ensure the robustness of our model, we employed a grid search to determine the optimal combination of hyperparameters. The final objective function was set to Poisson, with the evaluation function being the mean squared error (MAE). The model's configuration included a leaf number of 59, a maximum decision tree depth of 8 iterations, and a learning rate of 0.01. The models achieved commendable performance with the lowest MAEs of 22.56 and 31.32, and pseudo-R^2 values of 0.41 and 0.56, respectively.

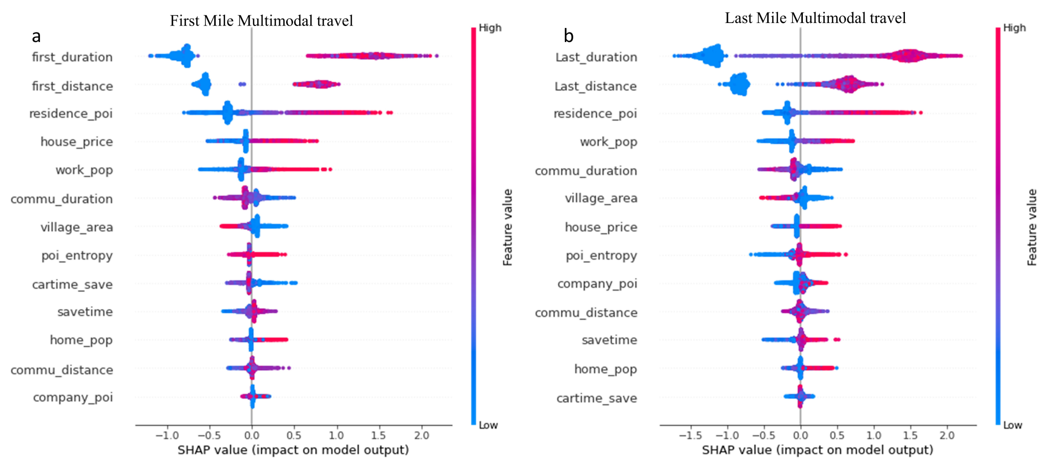

Figure 6: Distribution of Shapley values for features in first mile (a) and last mile (b) multimodal travel. Features are ranked by importance from top to bottom. Dots represent instances of bike sharing-public transit travel, with color indicating feature value (blue: low, red: high). Horizontal position shows the Shapley value, indicating the feature's contribution to multimodal travel likelihood. For clarity, dots are jittered vertically to prevent overlap.

Fig. 6 depicts the distribution of Shapley values for the influencing factors of multimodal travel during the first mile and the last mile. The duration and distance of both the first and last mile travel exhibit pronounced SHAP values, marking them as the two most pivotal predictive features. A longer distance and time for the first/last mile segment of a commute indicate a higher propensity for commuters within that jobs-housing grid to opt for bike-sharing as part of their public transportation trips. Furthermore, influencing the choice of bike-sharing during the first mile, the number of commercial residential POIs emerges as the third most influential feature, followed by housing prices in the fourth position. In contrast, for the last mile, housing prices rank seventh, with the number of working population taking the fourth spot. This suggests that while there are similarities in the primary influencing factors for multimodal travel during the first and last miles, there are also distinct differences.

The interaction dependence plot for the first mile travel feature, as shown in Fig. 7 a, c, reveals a distinct nonlinear threshold effect for both the duration and distance of the first mile travel. Notably, when the duration of the first mile exceeds 200 seconds and the travel distance surpasses 250 meters, this nonlinear effect becomes pronounced. Beyond these thresholds, these two features significantly influence the choice of bike-sharing as a mode of transportation. This nonlinear effect is also evident in the last mile travel, as illustrated in Fig. 8 c, d. However, a key difference is observed: the distance threshold for the last mile is around 500 meters. This indicates that in multimodal travel decisions, people's sensitivity to distance varies between the first and last miles.

The house price demonstrates a nonlinear threshold effect on the adoption of first-mile multimodal trips. As depicted in Fig.7 b, when the house price exceeds approximately RMB 40,000, a trend emerges: higher house prices are associated with a higher probability of first-mile multimodal trip adoption. This effect is also evident for last-mile trips, as indicated in Fig.A.7

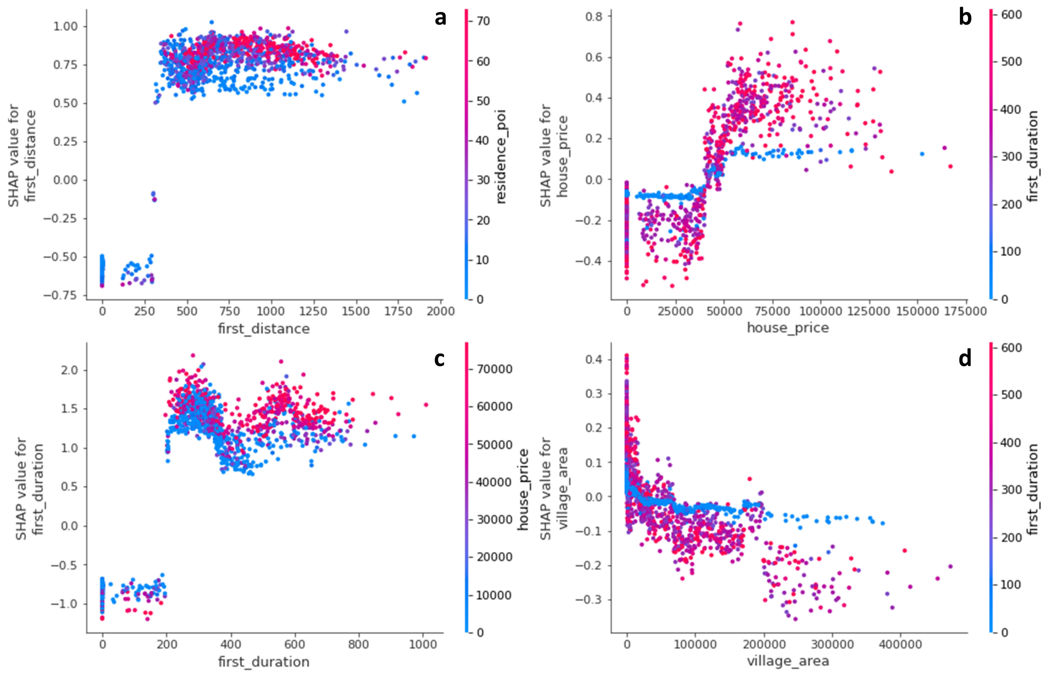

Figure 7: First mile travel feature interaction dependence plot. (a) The average distance and duration for the first mile. (b) Housing prices (in yuan per square meter). (c) The average duration for the first mile. (d) Urban village areas (in square meters).

a. Moreover, the size of urban villages has an influence on the adoption of both first and last-mile multimodal trips. Specifically, larger urban villages are associated with a decreased likelihood of adopting these trips, as shown in Fig.7 d for the first mile and Fig.A.7 c for the last mile.

These observed effects, especially concerning house prices and urban village sizes, suggest that, compared to groups with higher socioeconomic status, those with lower socioeconomic status are less likely to improve their commuting efficiency through bike-sharing. Furthermore, bike-sharing services seem to be more prevalent in affluent urban areas, such as city centers.

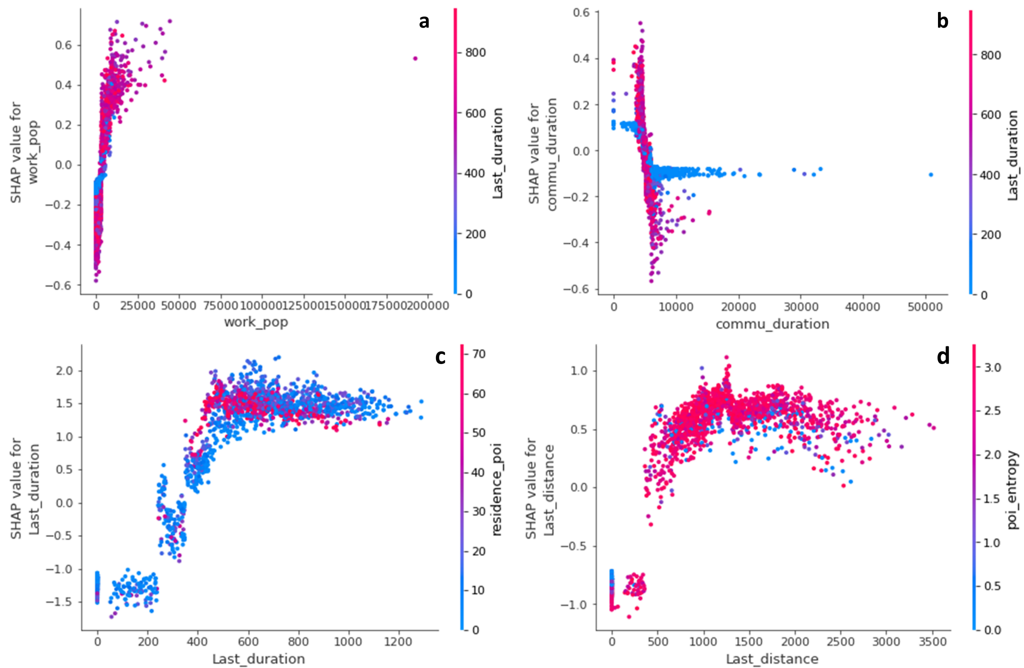

Figure 8: Last mile travel feature interaction dependence plot. (a) Number of working population. (b) The average commuting time by public transportation. (c) The average duration for the last mile. (d) The average distance and duration for the first mile.

In the last mile model, while the house price stands out as the 4_{th} most influential feature in the first mile, the number of the working population takes precedence as a significant predictor. A plausible interpretation for this is that areas with a larger working population naturally have more amenities or infrastructure supporting bike-sharing, leading to a higher likelihood of people using bike-sharing for their last mile trips. However, this observation should be considered in conjunction with other factors that might influence this trend. It's essential to understand that these factors often interact with each other. As depicted in Fig.8 a, the interaction dependence plot between the working population and the average duration for the last mile reveals that regions with longer last mile commuting time and a larger working population have a higher probability of bike-sharing adoption.

Moreover, as shown in Fig.8 b, as the commuting time to a specific location lengthens, the likelihood of opting for multimodal travel diminishes. This indicates that those with extended commutes might not find the amalgamation of public transportation and bike-sharing as beneficial as those with shorter journeys. This trend is consistent for both the first and last mile, as seen in Fig.A.6 b. Notably, when the commuting duration surpasses a threshold of 6000 seconds (approximately 1.7 hours), the interactive effect between commuting time and the duration of the first/last mile vanishes. This suggests that the previously observed nonlinear threshold effect for the first/last mile duration doesn't encourage those with exceedingly long commutes to adopt bike-sharing. Given that individuals residing on the city's outskirts face longer commuting durations, this might further underscore that bike-sharing predominantly benefits those in urban centers.

Additionally, a similar nonlinear effect is observed with land use mix. When this mix nears a value of 3, areas with a higher mix show a greater propensity for bike-sharing. Whether considering commute costs, socio-economic factors, or built environment characteristics, the machine learning models reveal these nonlinear thresholds and interaction effects, providing valuable insights for city planners and stakeholders.

6. Conclusion and Discussion

This research, harnessing mobile phone and bike-sharing big data, pinpointed eight types of bike-sharing-public transit multimodal travel. The findings indicate that, although bike-sharing does augment the existing public transit system, its potency in amplifying the efficiency of multimodal travel is largely confined to a narrow scope of areas. Noteworthy inequalities emerge in this multimodal travel both on global and local urban scales. The results of this study challenge the homogeneous assumptions of the existing literature (Wang et al., 2020; Guo and He, 2020; Yu et al., 2021), which presupposes that bike-sharing universally enhances the efficiency of the existing public transit system and that this enhancement is spatially homogeneous. Such assumptions likely overstate the impact of bike-sharing on urban public transportation.

Within city centers and in specific configurations, such as the metro-bike-sharing for last-mile solutions, bike-sharing can enhance both efficiency and equity. However, it doesn't seem to benefit those from lower socioeconomic backgrounds or those residing in the city's outskirts. Concurrent research from North America, Europe, and Australia reinforces this assertion, suggesting that contemporary bike-sharing systems predominantly serve the transportation needs of an increasingly privileged demographic entrenched within the urban center (Ricci, 2015; Fishman et al., 2014; de Chardon, 2019).

The primary factors influencing bike-sharing-public transit multimodal travel are the first/last mile distance and duration. This suggests that the decision to use public transportation for multimodal travel is tightly linked to how accessible transit stations are within an individual's neighborhood. Existing research has also recognized the significance of station proximity and layout in influencing bike-sharing usage (Guo and He, 2020; Guo et al., 2021; Willberg et al., 2021). Additionally, housing prices reflecting socioeconomic status and the area of urban villages also play pivotal roles in determining the utilization of bike-sharing in multimodal travel. Studies by Guo et al. (2021) and others support the finding that the likelihood of using bike-sharing is lower in urban villages. Insights derived from machine learning models reveal nonlinear threshold effects of commuting costs, socioeconomic factors, and the built environment on multimodal travel. Such insights can guide planners and operators to give particular consideration to connections with existing public transit systems during placement and allocation.

This research is characterized by two primary limitations. Firstly, the foundational data stems from China, which prompts the question: To what extent can the conclusions drawn here be extrapolated to other nations with bike-sharing systems? The bike-sharing landscape in China is distinct, marked by its substantial private investments and intense competition. Introduced in 2016, the bike-sharing phenomenon in China witnessed an explosive growth, with the number of shared bicycles escalating from 2 million in 2017 to a staggering 23 million. By 2020, the market was inundated with over 30 competing bike-sharing brands (Hu and Creutzig, 2022). This rapid expansion, coupled with lenient public policies and subpar coordination, precipitated a significant resource wastage, infamously known as the "bike-sharing graveyard." Furthermore, it raised questions about the actual efficiency of urban transportation (Wang and Sun, 2022). Drawing parallels, case studies from

Europe and North America reflect a pattern of low bike-sharing utilization rates, with a tendency to cater predominantly to a privileged demographic, thereby fostering social exclusion (de Chardon, 2019). Reacting to these challenges, local governments in China have now pivoted to enforcing restrictive measures to temper the unchecked competition among bike-sharing enterprises. This shift underscores the necessity for in-depth research: How does a service, largely backed by private capital, interface with and influence established public transportation systems? It's worth noting that the aim of this paper isn't to castigate this nascent transportation modality but to furnish valuable insights for urban planners and stakeholders, and in doing so, pave the way for subsequent research endeavors.

The second limitation revolves around our methodology. We anchored our analysis on two disparate big data sets to determine the probability of multimodal travel. This approach, while expansive, inherently challenges the verifiability of the resultant multimodal travel data. To mitigate this, we delved deep, sifting through extensive timeframes and billions of data points, fortified by a rigorous and thorough data processing protocol. Our analytical lens, however, was primarily trained on urban demographics exhibiting stable job-housing commute patterns. This meant sidelining segments without regular commuting behaviors, like those without fixed jobs. Such a focus inadvertently limits the breadth of our assessment, restricting our understanding of multimodal commuting efficiency and equity across all urban strata. Yet, a silver lining emerges. Preliminary findings, underpinned by machine learning algorithms, resonate with extant literature. This congruence provides a modicum of assurance about the credibility of the multimodal travel data we've unearthed.

In conclusion, the cornerstone of sustainable urban development hinges on amplifying the equity and accessibility of public transportation. The insights gleaned from our study on bike-sharing's multimodal travel inequality shed invaluable light on future sustainable urban planning trajectories. Governments stand to enhance the spatial equity of bike-sharing through judicious policies, ensuring this transport modality penetrates even the more marginalized urban pockets and extends its reach to the economically disadvantaged. Moreover, the integration of bike-sharing initiatives should harmoniously dovetail with prevailing public transport systems. Blind, unchecked investments not only risk monumental resource wastage but also imperil the optimization of current transit operations, potentially spawning myriad urban challenges. As we champion novel transport alternatives, it's paramount to strike a judicious balance between efficiency and equity. A holistic approach, steered by the synergy of diverse stakeholder groups, trumps isolated government or supplier-driven initiatives, preventing potential skews in this delicate equilibrium.

7. Code and data availability

The bike-sharing data can be accessed from the Shenzhen Municipal Government's Open Data Platform at `https://opendata.sz.gov.cn/`. Additional related codes and datasets are available at `https://github.com/Liu-Zhihang/bike-sharing`. Due to privacy concerns, mobile phone data is not available for distribution. For a more in-depth understanding of this research, we have also constructed an interactive visualization website, which can be accessed at `https://zhihangliu.cn/projects/Sharingbike/Morning_bike_sharing.html`.

References

Audikana, A., Ravalet, E., Baranger, V., Kaufmann, V., 2017. Implementing bikesharing systems in small cities: Evidence from the swiss experience. Transport Policy 55, 18–28.

Bauman, A., Crane, M., Drayton, B.A., Titze, S., 2017. The unrealised potential of bike share schemes to influence population physical activity levels–a narrative review. Preventive medicine 103, S7–S14.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment 2008, P10008.

Caggiani, L., Camporeale, R., Marinelli, M., Ottomanelli, M., 2019. User satisfaction based model for resource allocation in bike-sharing systems. Transport Policy 80, 117–126.

Caggiani, L., Colovic, A., Ottomanelli, M., 2020. An equality-based model for bike-sharing stations location in bicycle-public transport multimodal mobility. Transportation Research Part A: Policy and Practice 140, 251–265.

Castillo-Manzano, J.I., Castro-Nuño, M., López-Valpuesta, L., 2015. Analyzing the transition from a public bicycle system to bicycle ownership:

A complex relationship. Transportation Research Part D: Transport and Environment 38, 15–26.

Caulfield, B., O'Mahony, M., Brazil, W., Weldon, P., 2017. Examining usage patterns of a bike-sharing scheme in a medium sized city. Transportation research part A: policy and practice 100, 152–161.

Cerutti, P.S., Martins, R.D., Macke, J., Sarate, J.A.R., 2019. "green, but not as green as that": An analysis of a brazilian bike-sharing system. Journal of cleaner production 217, 185–193.

de Chardon, C.M., 2019. The contradictions of bike-share benefits, purposes and outcomes. Transportation research part A: policy and practice 121, 401–419.

de Chardon, C.M., Caruso, G., Thomas, I., 2017. Bicycle sharing system 'success' determinants. Transportation research part A: policy and practice 100, 202–214.

Chen, R., Zhang, M., Zhou, J., 2023. Jobs-housing relationships before and amid covid-19: An excess-commuting approach. Journal of Transport Geography 106, 103507.

Chen, W., Liu, X., Chen, X., Cheng, L., Wang, K., Chen, J., 2022. Exploring year-to-year changes in station-based bike sharing commuter behaviors with smart card data. Travel Behaviour and Society 28, 75–89.

Chen, Z., van Lierop, D., Ettema, D., 2020. Dockless bike-sharing systems: what are the implications? Transport Reviews 40, 333–353.

Cheng, L., Mi, Z., Coffman, D., Meng, J., Liu, D., Chang, D., 2021. The role of bike sharing in promoting transport resilience. Networks and spatial economics , 1–19.

Cheng, P., Xu, C., Lebreton, P., Yang, Z., Chen, J., 2019. Terp: Time-event-dependent route planning in stochastic multimodal transportation networks with bike sharing system. IEEE Internet of Things Journal 6, 4991–5000.

De Chardon, C.M., Caruso, G., Thomas, I., 2016. Bike-share rebalancing strategies, patterns, and purpose. Journal of transport geography 55, 22–39.

DeSalvo, J.S., Huq, M., 1996. Income, residential location, and mode choice. Journal of urban economics 40, 84–99.

Duran, A.C., Anaya-Boig, E., Shake, J.D., Garcia, L.M.T., de Rezende, L.F.M., de Sá, T.H., 2018. Bicycle-sharing system socio-spatial inequalities in brazil. Journal of transport & health 8, 262–270.

Eren, E., Uz, V.E., 2020. A review on bike-sharing: The factors affecting bike-sharing demand. Sustainable cities and society 54, 101882.

Faghih-Imani, A., Eluru, N., El-Geneidy, A.M., Rabbat, M., Haq, U., 2014. How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (bixi) in montreal. Journal of transport geography 41, 306–314.

Faghih-Imani, A., Hampshire, R., Marla, L., Eluru, N., 2017. An empirical analysis of bike sharing usage and rebalancing: Evidence from barcelona and seville. Transportation Research Part A: Policy and Practice 97, 177–191.

Fearnley, N., Currie, G., Flügel, S., Gregersen, F.A., Killi, M., Toner, J., Wardman, M., 2018. Competition and substitution between public transport modes. Research in Transportation Economics 69, 51–58.

Fishman, E., Washington, S., Haworth, N., 2014. Bike share's impact on car use: Evidence from the united states, great britain, and australia. Transportation research Part D: transport and environment 31, 13–20.

Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale geographically weighted regression (mgwr). Annals of the American Association of Geographers 107, 1247–1265.

Freund, Y., Schapire, R.E., et al., 1996. Experiments with a new boosting algorithm, in: icml, Citeseer. pp. 148–156.

Fu, C., Huang, Z., Scheuer, B., Lin, J., Zhang, Y., 2023. Integration of dockless bike-sharing and metro: Prediction and explanation at origin-destination level. Sustainable Cities and Society 99, 104906.

Gallotti, R., Barthelemy, M., 2014. Anatomy and efficiency of urban multimodal mobility. Scientific reports 4, 6911.

Gao, K., Yang, Y., Li, A., Qu, X., 2021. Spatial heterogeneity in distance decay of using bike sharing: An empirical large-scale analysis in shanghai. Transportation research part D: transport and environment 94, 102814.

GeTui, 2022. Anonymous travel dataset. Available at: `https://getui.com/` (Accessed on 11 December 2022).

Gu, T., Kim, I., Currie, G., 2019. To be or not to be dockless: Empirical analysis of dockless bikeshare development in china. Transportation Research Part A: Policy and Practice 119, 122–147.

Guo, Y., He, S.Y., 2020. Built environment effects on the integration of dockless bike-sharing and the metro. Transportation Research Part D: Transport and Environment 83, 102335.

Guo, Y., Yang, L., Lu, Y., Zhao, R., 2021. Dockless bike-sharing as a feeder mode of metro commute? the role of the feeder-related built environment: Analytical framework and empirical evidence. Sustainable Cities and Society 65, 102594.

Hoffmann, M.L., 2016. Bike lanes are white lanes: Bicycle advocacy and urban planning. U of Nebraska Press.

Hosford, K., Winters, M., 2018. Who are public bicycle share programs serving? an evaluation of the equity of spatial access to bicycle share service areas in canadian cities. Transportation research record 2672, 42–50.

Hu, B., Zhong, Z., Zhang, Y., Sun, Y., Jiang, L., Dong, X., Sun, H., 2022. Understanding the influencing factors of bicycle-sharing demand based on residents' trips. Physica A: Statistical Mechanics and its Applications 586, 126472.

Hu, J.W., Creutzig, F., 2022. A systematic review on shared mobility in china. International journal of sustainable transportation 16, 374–389.

Hu, Y., Wang, F., 2016. Temporal trends of intraurban commuting in baton rouge, 1990–2010. Annals of the American Association of Geographers 106, 470–479.

IDTP, 2020. The impact of the covid-19 pandemic on urban transportation and future trends. Available: `http://www.itdp-china.org/news/?newid=155&lang=0`.

Im, H.N., Choi, C.G., 2019. The hidden side of the entropy-based land-use mix index: Clarifying the relationship between pedestrian volume and land-use mix. Urban studies 56, 1865–1881.

Jäppinen, S., Toivonen, T., Salonen, M., 2013. Modelling the potential effect of shared bicycles on public transport travel times in greater helsinki: An open data approach. Applied geography 43, 13–24.

Jia, Y., Fu, H., 2019. Association between innovative dockless bicycle sharing programs and adopting cycling in commuting and non-commuting trips. Transportation research part A: policy and practice 121, 12–21.

Jin, S.T., Kong, H., Sui, D.Z., 2019. Uber, public transit, and urban transportation equity: A case study in new york city. The Professional Geographer 71, 315–330.

Kapuku, C., Kho, S.Y., Kim, D.K., Cho, S.H., 2021. Assessing and predicting mobility improvement of integrating bike-sharing into multimodal public transport systems. Transportation research record 2675, 204–213.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems 30.

Koglin, T., Mukhtar-Landgren, D., 2021. Contested values in bike-sharing mobilities–a case study from sweden. Journal of transport geography 92, 103026.

Li, A., Zhao, P., Huang, Y., Gao, K., Axhausen, K.W., 2020. An empirical analysis of dockless bike-sharing utilization and its explanatory factors: case study from shanghai, china. Journal of Transport Geography 88, 102828.

Lin, J.J., Wang, N.L., Feng, C.M., 2017. Public bike system pricing and usage in taipei. International Journal of sustainable transportation 11, 633–641.

Liu, L., Kong, H., Liu, T., Ma, X., 2022. Mode choice between bus and bike-sharing for the last-mile connection to urban rail transit. Journal of Transportation Engineering, Part A: Systems 148, 04022017.

Liu, M., Liu, Y., Ye, Y., 2023. Nonlinear effects of built environment features on metro ridership: An integrated exploration with machine learning considering spatial heterogeneity. Sustainable Cities and Society 95, 104613.

Liu, Z., Shen, Y., Zhu, Y., 2018. Inferring dockless shared bike distribution in new cities, in: Proceedings of the eleventh ACM international conference on web search and data mining, pp. 378–386.

Lu, M., Hsu, S.C., Chen, P.C., Lee, W.Y., 2018. Improving the sustainability of integrated transportation system with bike-sharing: A spatial agent-based approach. Sustainable cities and society 41, 44–51.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems 30.

Ma, X., Zhang, X., Li, X., Wang, X., Zhao, X., 2019. Impacts of free-floating bikesharing system on public transit ridership. Transportation research part D: transport and environment 76, 100–110.

Ma, Y., Lan, J., Thornton, T., Mangalagiu, D., Zhu, D., 2018. Challenges of collaborative governance in the sharing economy: The case of free-floating bike sharing in shanghai. Journal of cleaner production 197, 356–365.

Mangalathu, S., Hwang, S.H., Jeon, J.S., 2020. Failure mode and effects analysis of rc members based on machine-learning-based shapley additive explanations (shap) approach. Engineering Structures 219, 110927.

Nair, R., Miller-Hooks, E., Hampshire, R.C., Bušić, A., 2013. Large-scale vehicle sharing systems: analysis of vélib'. International Journal of Sustainable Transportation 7, 85–106.

Newman, M.E., 2004. Fast algorithm for detecting community structure in networks. Physical review E 69, 066133.

Newman, M.E., 2006. Modularity and community structure in networks. Proceedings of the national academy of sciences 103, 8577–8582.

Novotnỳ, J., 2007. On the measurement of regional inequality: does spatial dimension of income inequality matter? The Annals of Regional Science 41, 563–580.

Olafsson, A.S., Nielsen, T.S., Carstensen, T.A., 2016. Cycling in multimodal transport behaviours: Exploring modality styles in the danish population. Journal of transport geography 52, 123–130.

Otero, I., Nieuwenhuijsen, M.J., Rojas-Rueda, D., 2018. Health impacts of bike sharing systems in europe. Environment international 115, 387–394.

Park, C., Sohn, S.Y., 2017. An optimization approach for the placement of bicycle-sharing stations to reduce short car trips: An application to the city of seoul. Transportation Research Part A: Policy and Practice 105, 154–166.

Pellicer-Chenoll, M., Pans, M., Seifert, R., López-Cañada, E., García-Massó, X., Devís-Devís, J., González, L.M., 2021. Gender differences in bicycle sharing system usage in the city of valencia. Sustainable Cities and Society 65, 102556.

Radzimski, A., Dzięcielski, M., 2021. Exploring the relationship between bike-sharing and public transport in poznań, poland. Transportation research part A: policy and practice 145, 189–202.

Redmond, L.S., Mokhtarian, P.L., 2001. The positive utility of the commute: modeling ideal commute time and relative desired commute amount. Transportation 28, 179–205.

Ricci, M., 2015. Bike sharing: A review of evidence on impacts and processes of implementation and operation. Research in Transportation Business & Management 15, 28–38.

Saltykova, K., Ma, X., Yao, L., Kong, H., 2022. Environmental impact assessment of bike-sharing considering the modal shift from public transit. Transportation Research Part D: Transport and Environment 105, 103238.

Schimohr, K., Scheiner, J., 2021. Spatial and temporal analysis of bike-sharing use in cologne taking into account a public transit disruption. Journal of Transport Geography 92, 103017. URL: `https://www.sciencedirect.com/science/article/pii/S0966692321000703`, doi:`https://doi.org/10.1016/j.jtrangeo.2021.103017`.

Shaheen, S., Chan, N., 2016. Mobility and the sharing economy: Potential to facilitate the first-and last-mile public transit connections. Built Environment 42, 573–588.

Shorrocks, A., Wan, G., 2005. Spatial decomposition of inequality. Journal of Economic Geography 5, 59–81.

Smart, M.J., Klein, N.J., 2018. Remembrance of cars and buses past: How prior life experiences influence travel. Journal of Planning Education and Research 38, 139–151.

Sun, Q., Chen, H., Long, R., Yang, J., 2023. Who will pay for the "bicycle cemetery"? evolutionary game analysis of recycling abandoned shared bicycles under dynamic reward and punishment. European Journal of Operational Research 305, 917–929.

Tang, T., Liu, R., Choudhury, C., 2020. Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data. Sustainable Cities and Society 53, 101927.

Tao, T., Cao, J., 2023. Exploring nonlinear and collective influences of regional and local built environment characteristics on travel distances by mode. Journal of Transport Geography 109, 103599.

Wang, J., Kwan, M.P., Cao, W., Gong, Y., Guo, L., Liu, Y., 2022. Assessing changes in job accessibility and commuting time under bike-sharing scenarios. Transportmetrica A: Transport Science , 1–17.

Wang, R., Lu, Y., Wu, X., Liu, Y., Yao, Y., 2020. Relationship between eye-level greenness and cycling frequency around metro stations in shenzhen, china: A big data approach. Sustainable Cities and Society 59, 102201.

Wang, Y., Sun, S., 2022. Does large scale free-floating bike sharing really improve the sustainability of urban transportation? empirical evidence from beijing. Sustainable Cities and Society 76, 103533.

Willberg, E., Salonen, M., Toivonen, T., 2021. What do trip data reveal about bike-sharing system users? Journal of Transport Geography 91, 102971.

Wu, C., Kim, I., 2020. Analyzing the structural properties of bike-sharing networks: Evidence from the united states, canada, and china. Transportation Research Part A: Policy and Practice 140, 52–71.

Wu, W., Hong, J., 2017. Does public transit improvement affect commuting behavior in beijing, china? a spatial multilevel approach. Transportation Research Part D: Transport and Environment 52, 471–479.

Wu, X., Lu, Y., Lin, Y., Yang, Y., 2019a. Measuring the destination accessibility of cycling transfer trips in metro station areas: A big data approach. International journal of environmental research and public health 16, 2641.

Wu, X., Tao, T., Cao, J., Fan, Y., Ramaswami, A., 2019b. Examining threshold effects of built environment elements on travel-related carbon-dioxide emissions. Transportation Research Part D: Transport and Environment 75, 1–12.

Xia, X., Lin, K., Ding, Y., Dong, X., Sun, H., Hu, B., 2021. Research on the coupling coordination relationships between urban function mixing degree and urbanization development level based on information entropy. International Journal of Environmental Research and Public Health 18, 242.

Xiao, L., Lo, S., Liu, J., Zhou, J., Li, Q., 2021. Nonlinear and synergistic effects of tod on urban vibrancy: Applying local explanations for gradient boosting decision tree. Sustainable cities and society 72, 103063.

Yang, J., Cao, J., Zhou, Y., 2021. Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in shenzhen. Transportation Research Part A: Policy and Practice 144, 74–88.

Yang, X.H., Cheng, Z., Chen, G., Wang, L., Ruan, Z.Y., Zheng, Y.J., 2018. The impact of a public bicycle-sharing system on urban public transport networks. Transportation research part A: policy and practice 107, 246–256.

Yao, Z., Kim, C., 2022. Analyzing the multiscale patterns of jobs-housing balance and employment self-containment by different income groups using lehd data: A case study in cincinnati metropolitan area. Computers, Environment and Urban Systems 96, 101851.

Yin, C., Cao, J., Sun, B., Liu, J., 2023. Exploring built environment correlates of walking for different purposes: Evidence for substitution. Journal of transport geography 106, 103505.

Yu, S., Liu, G., Yin, C., 2021. Understanding spatial-temporal travel demand of free-floating bike sharing connecting with metro stations. Sustainable Cities and Society 74, 103162.

Yue, Y., Zhuang, Y., Yeh, A.G., Xie, J.Y., Ma, C.L., Li, Q.Q., 2017. Measurements of poi-based mixed use and their relationships with neighbourhood vibrancy. International Journal of Geographical Information Science 31, 658–675.

Zhang, L., Zhang, J., Duan, Z.y., Bryde, D., 2015. Sustainable bike-sharing systems: characteristics and commonalities across cases in urban china. Journal of cleaner production 97, 124–133.

Zhang, Y., Lin, D., Mi, Z., 2019. Electric fence planning for dockless bike-sharing services. Journal of cleaner production 206, 383–393.

Appendix A.

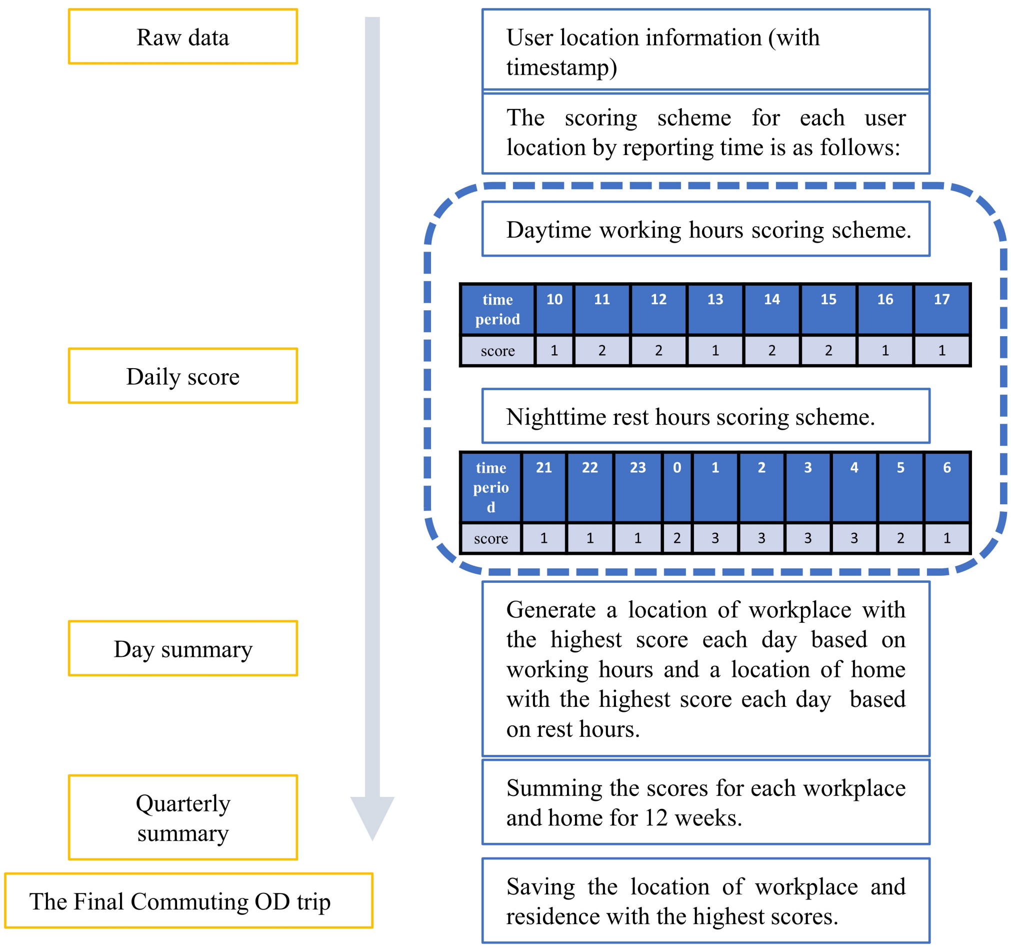

Figure A.1: Location data aggregation process.

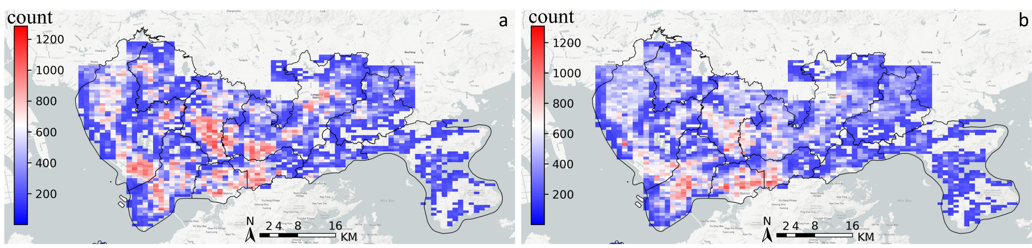

Figure A.2: Jobs-housing commuting origin and destination grids. The color gradient from blue to red indicates the number of commuters within the grid. (a) Origin grid. (b) Destination grid.

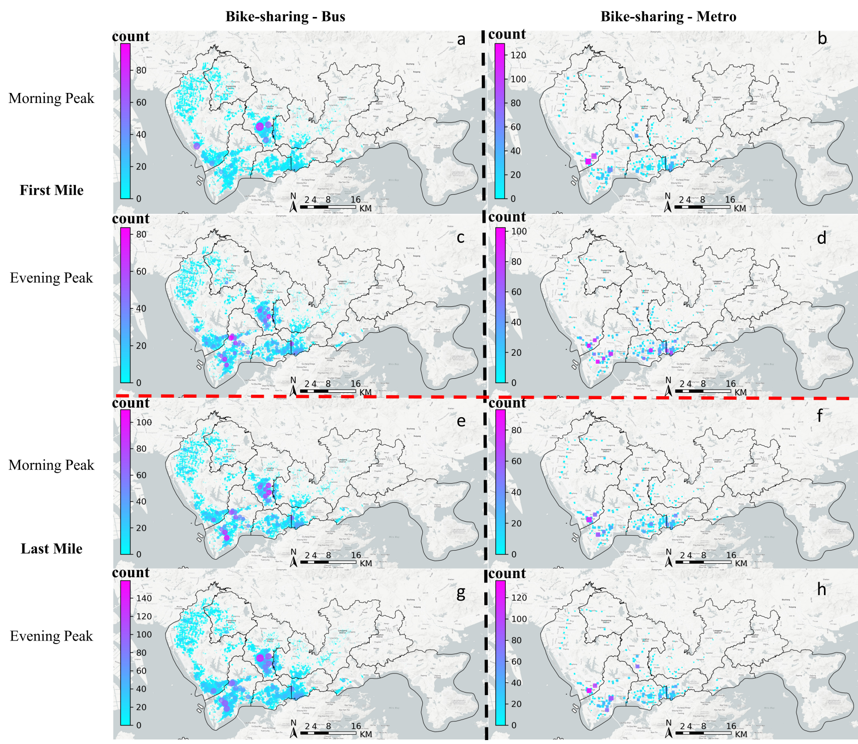

Figure A.3: Eight types of bike-sharing multimodal travel volume at each station. Above the red dashed line: first mile; below: last mile. Left of the black dashed line: bike-sharing-bus; right: bike-sharing-metro. (a, c) Morning and evening peak for first mile bike-sharing-bus. (b, d) Morning and evening peak for first mile bike-sharing-metro. (e, g) Morning and evening peak for last mile bus-bike-sharing. (f, h) Morning and evening peak for last mile metro-bike-sharing.

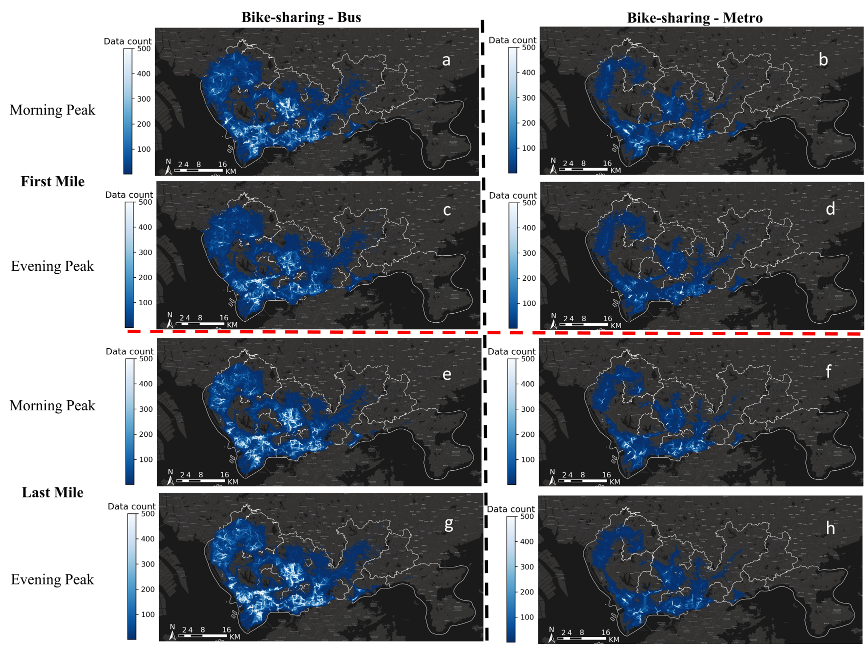

Figure A.4: Eight types of bike-sharing multimodal travel OD networks. Above the red dashed line: first mile; below: last mile. Left of the black dashed line: bike-sharing-bus; right: bike-sharing-metro. (a, c) Morning and evening peak for first mile bike-sharing-bus. (b, d) Morning and evening peak for first mile bike-sharing-metro. (e, g) Morning and evening peak for last mile bus-bike-sharing. (f, h) Morning and evening peak for last mile metro-bike-sharing.

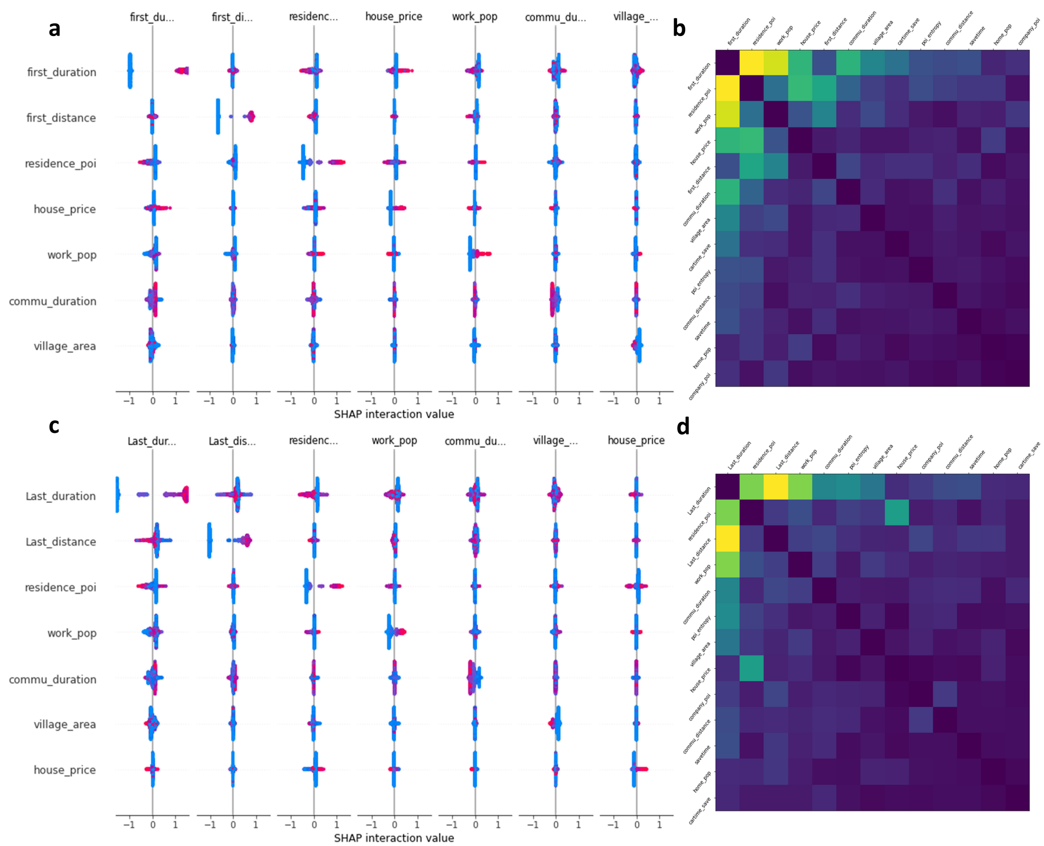

Figure A.5: First and last mile GBDT model feature variable interaction summary plots.

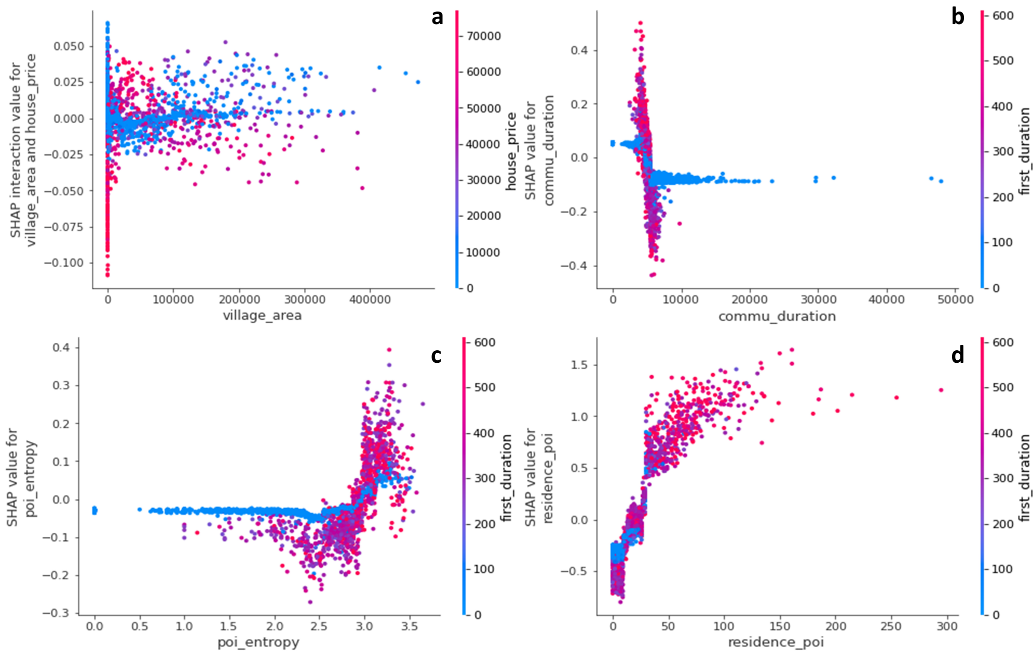

Figure A.6: First mile travel feature interaction dependence plot. (a) Urban village areas (in square meters). (b) The average commuting time by public transportation. (c) POI entropy representing mixed land use. (d) Number of commercial residential POIs.

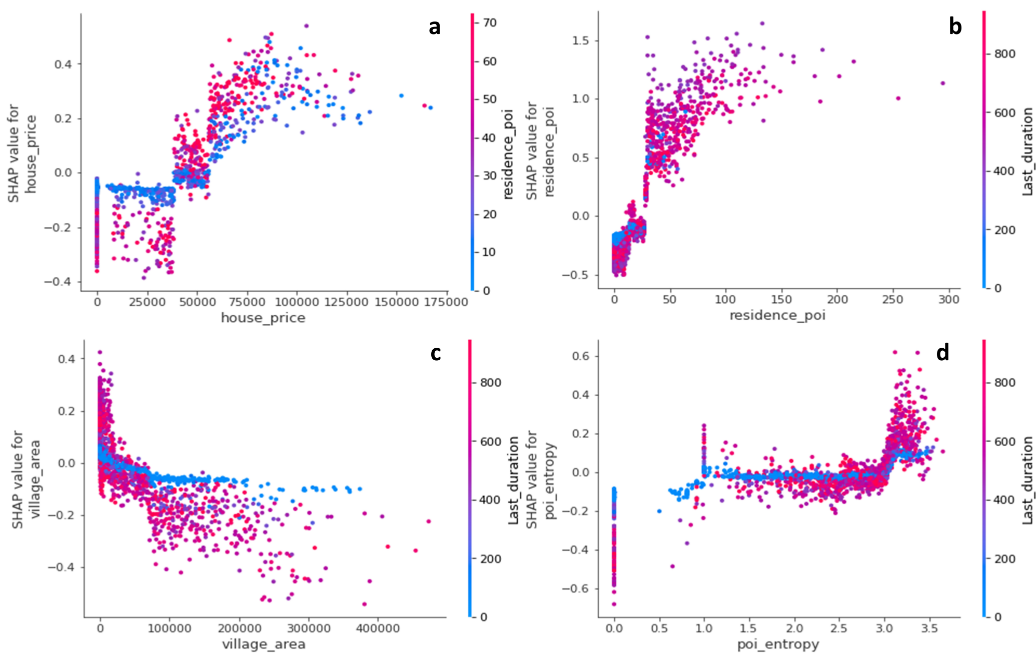

Figure A.7: Last mile travel feature interaction dependence plot. (a) Housing prices (in yuan per square meter). (b) Number of commercial residential POIs. (c) Urban village areas (in square meters). (d) POI entropy representing mixed land use.