Predicting and Explaining Parking Space Sharing Behaviors Using LightGBM and SHAP with Individual Heterogeneity Considered

Ange Wang[a, b], Jiyao Wang[a], Xiao Wen[a], Dengbo He[a, b, c*], Ran Tu[d]

[a] Thrust of Intelligent Transportation, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
[b] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR
[d] HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen
[d] School of Transportation, Southeast University, Nanjing, China
* Corresponding author: dengbohe@hkust-gz.edu.cn

Abstract:
Shared parking has emerged recently and plays a vital role in alleviating parking problems. Several behavioral theoretical frameworks have been adopted to explore influential factors of potential suppliers' sharing intention. However, the heterogeneity of the population has rarely been considered. Thus, in this study, an interpretable Machine Learning (ML) framework has been adopted to investigate parking space sharing behaviors with individual heterogeneity considered. First, a survey study using a stated preference experiment design, questionnaire data from 383 respondents was collected from mainland China. A Latent Class Model (LCM) was used to categorize potential parking space suppliers and three latent classes were formed, consisting of 41.2%, 33% and 25.8% of the respondents. Next, a Light Gradient Boosting Machine (LightGBM) method, which was tested to outperform other commonly used ML models, was employed to quantify the effects of factors that can affect sharing behaviors. The SHapley Additive exPlanation (SHAP) approach was introduced to interpret the LightGBM outputs, and the result shows that the influential factors of parking space-sharing behaviors vary across different latent classes. The findings from our research can guide operators of shared parking to make informed and targeted decisions to facilitate shared parking behaviors of potential parking space suppliers.

Keywords:
Shared parking, Sharing behavior, Heterogeneity, Machine learning, Interpretability

1. Introduction

With the development of society and economy, the number of automobiles has increased rapidly, and the gap between parking spaces and parking demand has become a concern for many growing cities. For example, as of July 2020, the shortage of parking spaces in China has reached 80 million (Pang et al., 2021), which leads to increases in tail gas emission and aggravates environmental pollution (Shoup, 2006). Increasing parking space may be a solution to the shortage of parking spaces. However, it is costly because parking spaces are not storable in terms of time and space, which means the land reserved for parking space is usually determined in certain periods of the day. Fortunately, a closer scrutinization of the temporal and spatial distribution of the parking demand indicates that at the city level, the shortage of parking spaces is less severe. For example, during the working day, the demand for parking spaces in business areas usually exceeds the supply and private parking spaces in residential areas are often idle; while at weekends, the patterns are reversed. To maximize the usage of the parking space, the concept of shared parking was first proposed in 1984 by Lalani (1984) and was firstly implemented in the city of Portland, USA (Xie et al., 2019), which coordinates the use of the existing parking lots and satisfies the demand for parking (Hensher and King, 2001). The core concept involves parking space owners marketing permits for the unused periods of their parking spaces on an electronic parking platform. Travellers in need of parking can then acquire these permits through the platform. The dynamics of supply and demand in this system can be illustrated in Figure 1.

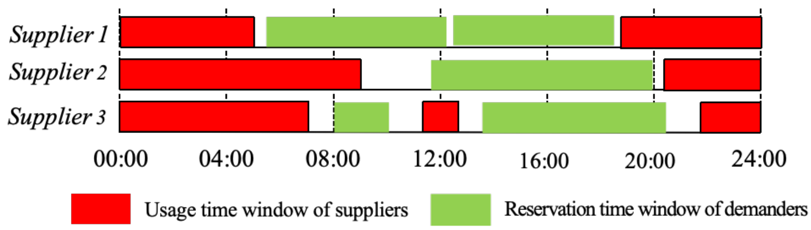


Fig. 1. Illustration of supply and demand in shared parking spaces.

Shared parking can take advantage of the spatial-temporal difference between the parking demands of two or more types of land-use buildings, transforms the problem of unbalanced parking demand into the redistribution of parking spaces and integrates existing parking resources (Kenig and Hocking, 1985). Encouraging shared parking can resolve the shortage of parking spaces, reduce land waste and drivers' parking space searching time, and even alleviate traffic congestion (Litman, 2016).

However, to date, shared parking strategies were mostly implemented in business districts (Zhang et al., 2020), while the penetration rate of shared parking in residential areas is limited, though the residential area takes the largest portion of parking spaces in most cities (2019 Annual Report on Beijing Transportation Development, 2019). For example, 485,000 parking spaces are designated for private use in residential areas in Hong Kong, accounting for nearly 70% of the total number of parking spaces in the city (Yan et al., 2020) and nearly 800,000 private parking spaces in Beijing are idle in daytime of weekdays, mostly in residential areas (2020 Annual Report on Beijing Transportation Development, 2020). The low participation willingness of shared parking in residential area might explain the limited penetration rate of shared parking in residential areas (Ning et al., 2021). Actually, when sharing parking spaces, there is a cooperative relationship between the supplier and the demander of parking spaces - only when both supply and demand sides agree to cooperate can shared parking be possible. Hence, investigating the participation behavior of potential suppliers and demanders and their influential factors are essential for encouraging shared parking.

In previous studies, Most research focuses on the allocation of shared parking spaces (Ji et al., 2022; Xie et al., 2022; Duan et al., 2020), while less attention is given to participation behavior in which most of the research focused on the demand side of shared parking (Ning et al., 2021; Wang et al., 2022; Ardeshiri et al., 2021) and paid little attention to the sharing behavior of suppliers, though suppliers play an equally essential role as compared to demanders (Chase, 2015). Only a few studies focused on sharing intentions of suppliers. For example, Liang et al. (2019) found that potential suppliers' sharing intentions can be explained by the technology acceptance model and theory of planned behavior (TPB) model. Wang et al. (2021) found that the expanded theoretical framework of TPB (i.e., with additional factors such as personal demographic characteristics, travel characteristics and built environment) can better explain the sharing intentions. Yan et al. (2020) further found that part of the socio-demographic characteristics of the suppliers (e.g., age, income) and the income from parking space sharing can affect sharing behavior. Later, (Q. Yan et al., 2022) also confirmed that latent variables such as attitude would impact parking space-sharing behavior. However, these previous studies did not consider the heterogeneity of the population when exploring suppliers' intentions, while the effects of individual heterogeneity are non-negligible (Sfeir et al., 2022; Wali and Khattak, 2022). Besides, while it is undoubtedly important to comprehend the factors that impact sharing behavior, the predictive capacity of models ignored in previous studies is also of utmost significance in enabling shared parking operators to determine potential operational areas based on pertinent factors, including geographical considerations and population characteristics. For example, Ardeshiri et al. (2021) found that considering the heterogeneity of the population can increase the accuracy of motivation prediction of parking space sharing.

Thus, in this study, a stated preference (SP) survey was designed to investigate the factors influencing potential suppliers' behaviors to share their parking spaces. Potential suppliers' socio-demographic characteristics, travel and traffic condition, geographic location of their residential area and the latent psychological factors, as well as behavioral data under different scenarios, were collected in the survey. To account for the heterogeneity of the populations, we employed the latent class model (LCM) to explore the latent classes of respondents. Improving the predictive capability of models is equally important as identifying significant influencing factors and their mechanisms of impact. Therefore, a novel ML method, the light gradient boosting machine (LightGBM) was applied to predict the sharing behaviors of the potential parking space suppliers. It should be noted that, when investigating certain behaviors based on questionnaire survey, most previous research used linear models (e.g., logit model) to explore the influential factors of behavior, which may be problematic if the underlying assumptions of the linear model do not hold. The lightGBM we applied in the study is an emerging ML model in recent years and has been successfully applied in predicting users' intentions or behaviors, such as commuting mode choice (Liu and Xiao, 2022), vehicle lane change behaviors (Ali et al., 2022; Chen et al., 2020), vehicle collisions (Wen et al., 2021), and electric vehicle charging behaviors (Ullah et al., 2022). Finally, to overcome the interpretability issue of ML models (specifically, LightGBM in our study), we adopted the shapley additive explanations (SHAP) method to quantify the nonlinear effect of the influential factors in the LightGBM model (Ali et al., 2022; Chen et al., 2020; Wen et al., 2021). To the best of our knowledge, this is the first time LightGBM and SHAP were used to investigate factors influencing parking space sharing behaviors. The results of this study can provide insights into the optimization of shared parking strategies among populations with different characteristics.

The survey data used in this study has been partially reported in Wang et al. (2021), which identified the factors that significantly affect the behaviors of sharing parking spaces. The contributions of our current paper were threefold. First, it identified the latent classes of potential suppliers with individual heterogeneity considered. Second, it introduced an innovative ML method, the LightGBM, to model the shared parking behaviors of potential suppliers and compared it with other popular ML methods, therefore has extended the application of ML methods in survey-related research. Third, it explored the application of SHAP in quantifying the impact of factors on sharing behaviors, which allows customized policy optimization for different latent groups based on our results.

2 Survey

2.1 Survey Design

Considering the small market penetration rate of shared parking in residential areas, it is difficult to collect revealed preference data regarding parking space sharing behaviors. Thus, an SP survey Hensher, 1994) method was used, in which a wide range of scenarios that may not be currently available were provided to respondents (Jin et al., 2020). In the survey, we considered scenarios defined by three factors, i.e., sharing earning, overtime sharing earning, and parking pressure. First, for parking space suppliers, sharing earning was also identified as an important influencing factor of parking sharing behaviors (Yan et al., 2020). Therefore, in our survey, six hourly sharing earning scenarios are designed: 1 *CNY* / 0.16 *USD*, 2 *CNY* / 0.32 *USD*, 3 *CNY* / 0.48 *USD*, 4 *CNY* / 0.64 *USD*, 5 *CNY* / 0.80 *USD*, 6 *CNY* / 0.96 *USD*. In order to address the concern of owners that their parking demand cannot be satisfied because of the parking sharing, as has been pointed out by previous research (Zhang, 2020), two overtime-sharing earning were used in the survey, i.e., 16 *CNY* / 2.56 *USD* and 24 *CNY* / 3.84 *USD* per hour. At the same time, parking pressure around residential areas may also affect parking space owners' concerns over parking space sharing. Thus, three levels of parking pressure were manipulated in the scenarios, i.e., high, general, and low, qualitatively. The three factors led to 36 different scenarios (sharing earning levels*6, overtime sharing earning levels*2, parking pressure levels*3). Respondents were asked to make a forced choice in the scenarios, i.e., sharing or not sharing. To reduce the number of choice scenarios that each respondent need to answer, a cross-design method (Maclure and Mittleman, 2000) was used and each participant experienced eight (out of 36) different scenarios.

To explore other factors that can be associated with the parking sharing behaviors, the survey also collected participants' demographic information (i.e., gender, age, education and monthly income of the respondents), parking and travel information (i.e., number of cars owned, number of private parking spaces, the use mode of parking spaces, and the frequency of driving travels), and geographic location information (i.e., city, residential location, and land types around residential areas). Lastly, latent variables of the extended Theory of Planned Behavior

(TPB) (i.e., Facilitating Conditions (PC), Subjective Norm (SN), Perceived Behavioral Control (PBC), Perceived Risk (PR), Perceived Benefit (PB), Attitude (ATT)) were also considered (Wang et al., 2021). A five-level Likert scale ranging from 1 (“Strongly disagree”) to 5 (“Strongly agree”) was used for all TPB-related variables. These latent variables and their measurement variables used in LCM are described in Table 1.

Table 1. Description of latent and measurement variables inputted into LCM.

Latent variable	Description	Measurement variable	Reference
Facilitating Conditions (PC)	How much someone believes their organization supports using the system with the necessary technology and resources.	If a credit system for shared parking is established, I would be more willing to offer a shared parking space.	(Thompson et al., 1991; Wang, Guan, Han, et al., 2021)
Subjective Norm (SN)	How much someone feels influenced by others to use the system.	Support and advocacy from the government and media would make me more willing to offer a shared parking space.	(Thompson et al., 1991; Wang, Guan, Han, et al., 2021)
Perceived Behavioral Control (PBC)	How easy or hard it is for someone to use a system.	I believe that I have the capability to offer shared parking spaces through a shared parking system.	(Venkatesh et al., 2003; Wang, Guan, Han, et al., 2021)
Perceived Risk (PR)	How much a person thinks using a system or engaging in an activity might lead to negative outcomes,	If a vehicle parked in my community is involved in an accident, it is very likely to cause trouble for me.	(Im et al., 2008; Wang, Guan, Han, et al., 2021)
Perceived Benefit (PB)	How much someone believes using the system will bring positive outcomes.	Providing shared parking spaces would generate significant economic income for me.	(Lee, 2009; Wang, Guan, Han, et al., 2021)

2.1 Survey Data

Targeting towards people with private parking spaces in residential areas, this survey was conducted online in January-February 2021. Respondents were recruited by email and social medias (e.g., WeChat) in mainland China. Each respondent was paid 2 *CNY / 0.32 USD*. We collected a total of 400 questionnaires. After validation checks and the exclusion of data with logical errors, a total of 383 questionnaires and 3064 choices (each respondent makes choice in eight different scenarios) were included for analysis. The descriptive statistics of the collected variables are summarized in Table 2.

Table 2. Descriptive statistics of the collected variables.

Category	Exogenous variable	Description	Level	Percentage (%)
Socio-demographic characteristics	Age	-	1: 18-25	27.45
			2: 26-30	13.24
			3: 31-40	21.57
			4: 41-50	15.44
			5: 51-60	17.65
			6: > 60	1.47
	Gender	-	1: Male	58.58
			2: Female	41.42
	Education	Education level	1: ≤High school	31.37
			2: Bachelor	38.48
			3: Master	24.51
			4: PhD	5.64
	Income	Monthly income	1: <3000 *CNY* / <416.67 *USD*	25
			2: 3000-5000 *CNY* / 416.67 - 694.44 *USD*	20.83
			3: 5000-7000 *CNY* / 694.44 - 972.22 *USD*	18.14
			4: 7000-9000 *CNY* / 972.22 - 1250.00 *USD*	11.27
			5: 9000-11000 *CNY* / 1250.00 - 1527.78 *USD*	6.62
			6: >11000 *CNY* / >1527.78 *USD* (The exchange rate of 1 USD to 7.20 CNY)	18.14
	Degree_of_ understanding	Understanding of shared parking scheme (The larger the value, the better the understanding)	1	7.6
			2	27.7
			3	44.36
			4	14.71
			5	5.64
Parking and travel factors	Number_of_ parking_spaces	Number of private parking spaces owned	1: One	73.77
			2: Two	19.12
			3: ≥Three	7.11
	Number_of_cars	Number of private cars owned	1: One	11.27
			2: Two	64.95
			3: ≥Three	23.77
	Usage_of_parking_ spaces	Usual use mode of private parking spaces	1: Fixed	32.11
			2: Flexible	67.89
	Driving frequency	Number of trips by car per day	1: Zero	17.89
			2: One	37.99
			3: Two	28.43
			4: ≥Three	15.69
Traffic condition	Convenience of public transportation around the residential area	Convenience of public transportation around the residential area (The larger the value, the more convenient)	1	7.35
			2	8.82
			3	30.15
			4	37.25
			5	16.42
	Traffic pressure around residential areas	Traffic pressure around residential areas (The larger the value, the greater the pressure)	1	5.64
			2	15.93
			3	47.79
			4	8.58
			5	22.06
	Residential parking pressure	-	1: High	25.49
			2: General	52.94
			3: Low	21.57
Geographic location	Location_of_ residential_areas	-	1: Suburb	13.48
			2: Between urban & suburb	19.61

		3: Between suburb & downtown	33.58
		4: Downtown	33.33
City_level	-	1: First-tier city	19.36
		2: Second-tier city	25.49
		3: Third-tier city	25.25
		4: Fourth-tier city	29.9

3 Data Analysis Framework

Fig. 2 illustrates the framework of this study. First, an SP survey was designed, and the data was collected using the "Questionnaire Star" platform in China. Factors that significantly impact the sharing behaviors of potential suppliers were selected using the Multi-Indicator Multi-Causal (MIMIC) model, as reported by Wang et al. (2021). Then, the LCM was applied based on the selected factors to categorize potential suppliers into heterogeneous groups. For readers' convenience, the measurement items of latent variables and the test results of the reliability (Cronbach's alpha, α) and validity (Average Variance Extracted, AVE) of the MIMIC model were shown in Wang et al. (2021). Next, four lightGBM models were trained for all samples as well as latent groups identified in LCM to predict parking space suppliers' sharing behaviors. Finally, the SHAP method was used to calculate the importance and interaction effects of the influential factors. The LCM was established with LatentGOLD Choice 5.0, and the Scikit-learn library (Pedregosa et al., 2011) in Python 3.9 was used to implement LightGBM and SHAP.

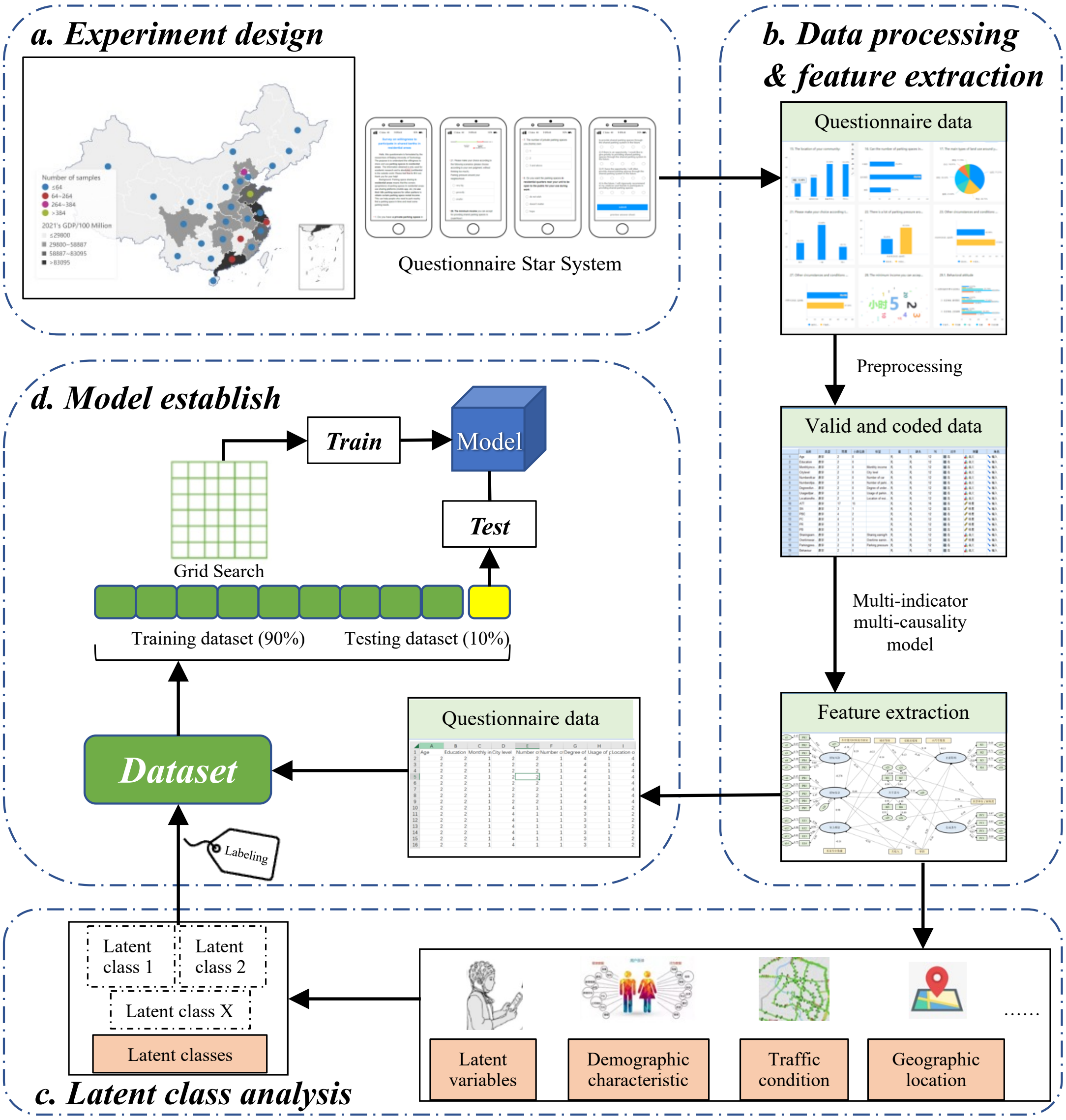


Fig. 2. The proposed framework of the study.

3.1 Latent Class Model (LCM)

LCM is a modeling technique that categorizes the data by assuming a latent grouping variable based on explicit variables (e.g., age, gender, etc., in our study). By estimating parameters using the maximum likelihood method, LCM can more accurately calculate the appropriate number of classes compared to traditional clustering methods, such as K-means clustering (Brusco et al., 2017). In addition, its classification results are sample-independent and thus can be extended to other samples of the same population without re-training the model, which is convenient for practical application (López-Bonilla and López-Bonilla, 2008). The LCM analysis process consists of three steps: probabilistic parameterization, model fitting and parameter estimation, latent classification, and interpretation of results (Wang et al., 2022). The basic LCM can be expressed as follows:

$$P = \sum_{t=1}^{T} (P_t^X \times P_{it}^{age \cdot X} \times P_{jt}^{edu \cdot X} \times P_{kt}^{income \cdot X} \times P_{lt}^{city \cdot X} \times P_{mt}^{PB \cdot X} \times P_{nt}^{PR \cdot X} \times P_{ot}^{PBC \cdot X} \times P_{qt}^{SN \cdot X} \times P_{rt}^{PC \cdot X}), \quad (3)$$

where P denotes the joint probability (the sum of the probabilities of each latent class) of a LCM; P_t^X denotes the probability that the observed data belongs to a particular latent class X, i.e., $P(X = t), t = 1,2, \dots, T$. T is the total number of respondents; *i, j, k, l, m, n, o, q* and *r* are level of the varibles. $age, edu, income, city, PB, PR, PBC, SN$ and PC represent explicit variables. Thus, $P_{it}^{age \cdot X}$ is the conditional probability of the respondent *t* belongs to the ith level of the varible age.

3.2 Light Gradient Boosting Machine (LightGBM)

LightGBM combines two innovative techniques: Gradient-based Single-Side Sampling (GOSS) and Exclusive Factor Bundling (EFB) (Ullah et al., 2022; Wen et al., 2021) and thus is expected to outperform other gradient-boosting methods in terms of training speed and accuracy. The general idea of LightGBM can be illustrated in Equation (4):

$$F_M(x) = \sum_{m=1}^{T} \gamma_m h_m(x), \tag{4}$$

where M represents the maximum number of iterations and T denotes the number of decision trees. The predicted value of GBDT ($F(x)$) is the sum of the outputs of a set of decision tree models ($h(x)$). Interested readers are referred to Ke et al. (2017) for more details.

3.3 Shapley Additive Explanations (SHAP)

For survey-related research, the interpretability of the model is equally important as model accuracy, as only by clarifying the internal mechanism of influencing factors on human behavior can relevant policies be optimized. To guide shared parking practices, it is also vital to quantify the impact of various key factors and their joint impact on parking-sharing behaviors. The SHAP, proposed by Lundberg and Lee (2017), is a novel model interpretation method for post-hoc interpretation of ML methods. The SHAP can reflect the influence of the factors in each sample and calculate the positive and negative effects. In Equation (5), the impact of a factor i on the final outcome is the Shapley value $\emptyset_i$. For example, for a factor subset $S \subset N$ (where N is the set of all factors), to extract the effects of factor i, two models are trained; the first model $f(S \cup \{i\})$ is trained with factor i and the second model $f(S)$ is trained without it, while $S \cup \{i\}$ is the value of input factors. For a more detailed introduction to SHAP, refer to Wen et al. (2022).

$$\emptyset_i = \frac{1}{|N|!} \sum_{S \subseteq N\{i\}} \frac{|S|!(|N|-|S|-1)!}{N} [f(S \cup \{i\}) - f(S)]. \tag{5}$$

4 Results

4.1 Clustering Based on LCM

According to the results of the MIMIC model, nine variables among variables that significantly affect parking space sharing behaviors were used in LCM, including *age, income, education, city_level*, and five latent variables (i.e., *FC, PBC, SN, PB, PR*) (Wang, Guan, Han, et al., 2021). Consistent Akaike Information Criterion (CAIC) and Bayesian Information Criterion (BIC) were used to determine the number of classes (Pacifico and Yoo, 2013). The smaller the values of CAIC and BIC, the higher the model fitting degree. Since the number of potential categories is initially unknown, it is necessary to pre-specify the number of latent categories. Therefore, assuming the samples can be divided into 1 to 10 classes, LatentGOLD

5.0 software was adopted to fit the LCM and estimate parameters, resulting in CAIC and BIC values for different numbers of classes. The CAIC and BIC values under different class numbers were obtained shown in Fig. 3, it can be found that 3 classes led to the best performance (with a classification error of 2.49%), hence potential suppliers were categorized into three groups, i.e., *C1*(41.2% suppliers), *C2* (33.0% suppliers) and *C3* (25.8% suppliers). Besides, the parameter estimation results are shown in Table 3.

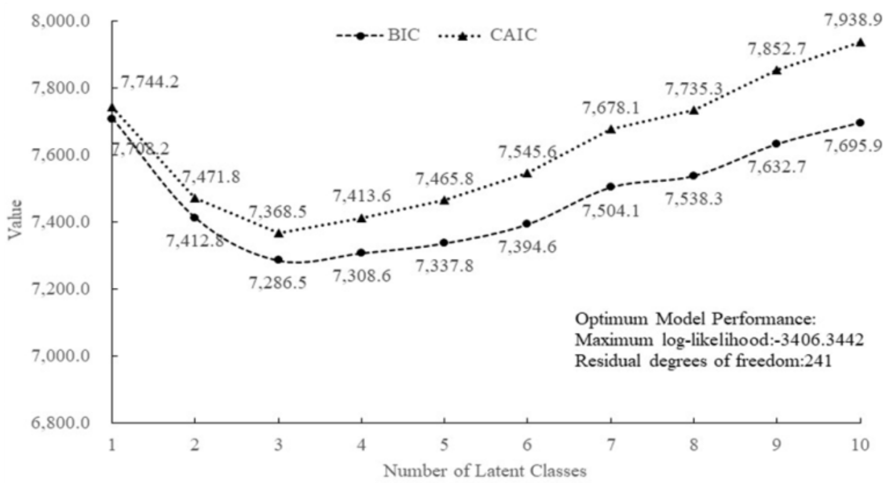


Fig. 3. Optimal number of latent classes for LCM.

Table 3. LCM parameter estimation results (C1, C2, C3 represent 3 latent classes)

Variable	Observable variable	Level	Description	C1	C2	C3
Psychological factor	SN	1	Strongly disagree	0.0000	0.0375	0.0000
		2	Disagree	0.0002	0.0932	0.0004
		3	General consent	0.0743	0.7189	0.1010
		4	Agree	0.7529	0.1497	0.7669
		5	Strongly agree	0.1726	0.0007	0.1317
			Average value	4.0978	2.9830	4.0300
	PBC	1	Strongly disagree	0.0000	0.0094	0.0000
		2	Disagree	0.0011	0.2419	0.0005
		3	General consent	0.1677	0.6922	0.1181
		4	Agree	0.7530	0.0564	0.7663
		5	Strongly agree	0.0782	0.0001	0.1151
			Average value	3.9083	2.7959	3.9959
	PC	1	Strongly disagree	0.0000	0.0094	0.0000
		2	Disagree	0.0000	0.1218	0.0000
		3	General consent	0.0140	0.7249	0.0335
		4	Agree	0.7487	0.1437	0.8413
		5	Strongly agree	0.2373	0.0002	0.1251
			Average value	4.2233	3.0036	4.0915
	PB	1	Strongly disagree	0.0132	0.0670	0.0010

		2	Disagree	0.0951	0.2218	0.0202
		3	General consent	0.4729	0.5074	0.2757
		4	Agree	0.4073	0.2012	0.6527
		5	Strongly agree	0.0114	0.0026	0.0504
			Average value	3.3086	2.8505	3.7313
	PR	1	Very disagree	0.0000	0.0000	0.0000
		2	disagree	0.0278	0.1396	0.0770
		3	General consent	0.2269	0.4199	0.3484
		4	agree	0.5896	0.4014	0.5013
		5	Very much agree	0.1558	0.0390	0.0733
			Average value	3.8734	3.3399	3.5710
	Age	1	<18	0.0001	0.0385	0.0826
		2	18-25	0.3505	0.3762	0.0028
		3	26-30	0.2115	0.1389	0.0006
		4	31-40	0.2990	0.1792	0.1333
		5	41-50	0.1158	0.1500	0.1632
		6	51-60	0.0080	0.0891	0.6174
		7	>60	0.0150	0.0281	0.0001
Personal attribute	City_level	1	1	0.2874	0.1943	0.0965
		2	2	0.3683	0.2191	0.1157
		3	3	0.2318	0.1926	0.2956
		4	4	0.1125	0.3940	0.4922
	Education level	1	High school and below	0.0012	0.2981	0.8649
		2	Undergraduate	0.4940	0.4045	0.1336
		3	Master	0.4124	0.2441	0.0013
		4	Doctor	0.0924	0.0533	0.0002
	Income	1	<3000 *CNY* / <416.67 *USD*	0.2070	0.2782	0.2616
		2	3000-5000 *CNY* / 416.67 - 694.44 *USD*	0.1583	0.1646	0.3888
		3	5000-7000 *CNY* / 694.44 - 972.22 *USD*	0.1388	0.2156	0.1745
		4	7000-9000 *CNY* / 972.22 - 1250.00 *USD*	0.1064	0.1218	0.0943
		5	9000-11000 *CNY* / 1250.00 - 1527.78 *USD*	0.0996	0.0532	0.0489
		6	>11000 *CNY* / >1527.78 *USD*	0.2899	0.1666	0.0319
Latent class probabilities				0.4118	0.3303	0.2579

As shown in Fig. 4, the left vertical axis represents latent classes, while the right vertical axis represents different levels of variables. The bandwidth of the connections between the latent classes and the different levels of variables reflects the magnitude of data flow (or proportions).

Regarding age distribution, individuals in *C1* and *C2* are generally younger compared to those in *C3*. Specifically, in *C1* and *C2*, those under the age of 40 constituting over 85% and 70% of these classes, respectively. Conversely, *C3* demonstrates a higher proportion of elderly individuals, with those aged 50 and above accounting for over 60% of the class.

As for education level, individuals in *C1* predominantly hold bachelor's degrees, followed by master's degrees. In *C2*, bachelor's degrees are most prevalent, followed by high school diplomas or less. In *C3*, over 86% have a high school education or less, followed by 13% holding bachelor's degrees, which corresponds with relatively lower educational levels. Across all three classes, the proportion of individuals with doctoral degrees is the smallest, with *C1* having a relatively higher percentage at 9%.

Concerning income, *C1* and *C2* has a higher proportion of high-income individuals. In *C2*, the distribution across various income groups is relatively even, though the number of high-income individuals is slightly lower than that of low-income individuals. *C3* has a notably higher proportion of low-income individuals.

In terms of city level, individuals in *C1* predominantly reside in second-tier cities, followed by first-tier and third-tier cities, with the lowest proportion residing in fourth-tier cities. In *C2*, the highest proportion resides in fourth-tier cities, with similar proportions in first, second, and third-tier cities. For C3, the highest proportion resides in fourth-tier cities.

Regarding PB, nearly 40% of respondents in *C1* agree that offering shared parking spaces would generate significant economic income, although about 47% maintain a neutral stance. In *C2*, over 50% are neutral, with the proportions of negative and positive attitudes being roughly equal. Over 65% of *C3* believe shared parking contributes significantly to their income, with a very low percentage holding negative views.

Concerning PR, about 60% of *C1* respondents believe that accidents within their community caused by shared parking users would pose problems, while over 22% are neutral. In *C2*, both neutral and positive attitudes exceed 40%. Over half of *C3* respondents are neutral, with close to 35% disagreeing that such incidents would cause them trouble.

In terms of PC, nearly 98% of *C1* respondents indicate a willingness to offer shared parking if a credit system is implemented, aligning with *C3* where over 96% exhibit a positive attitude. In contrast, the majority in *C2* are neutral, with equal proportions of opposing and supporting.

Regarding PBC, over 82% of *C1* respondents believe they are capable of offering shared parking spaces through a system, with *C3* even higher at over 88%, indicating a greater confidence in managing new technologies. About 70% of *C2* respondents are neutral, with close to 25% holding negative views, indicating lower confidence in adopting new technologies.

Finally, regarding SN, over 75% of respondents in C1 and C3 indicate that support and advocacy from the government and media would make them more willing to offer a shared parking space, with an additional 17% in C1 strongly agree with this, and very few holding negative views. In C2, over 70% are neutral, indicating a lower sensitivity to governmental and media influences.

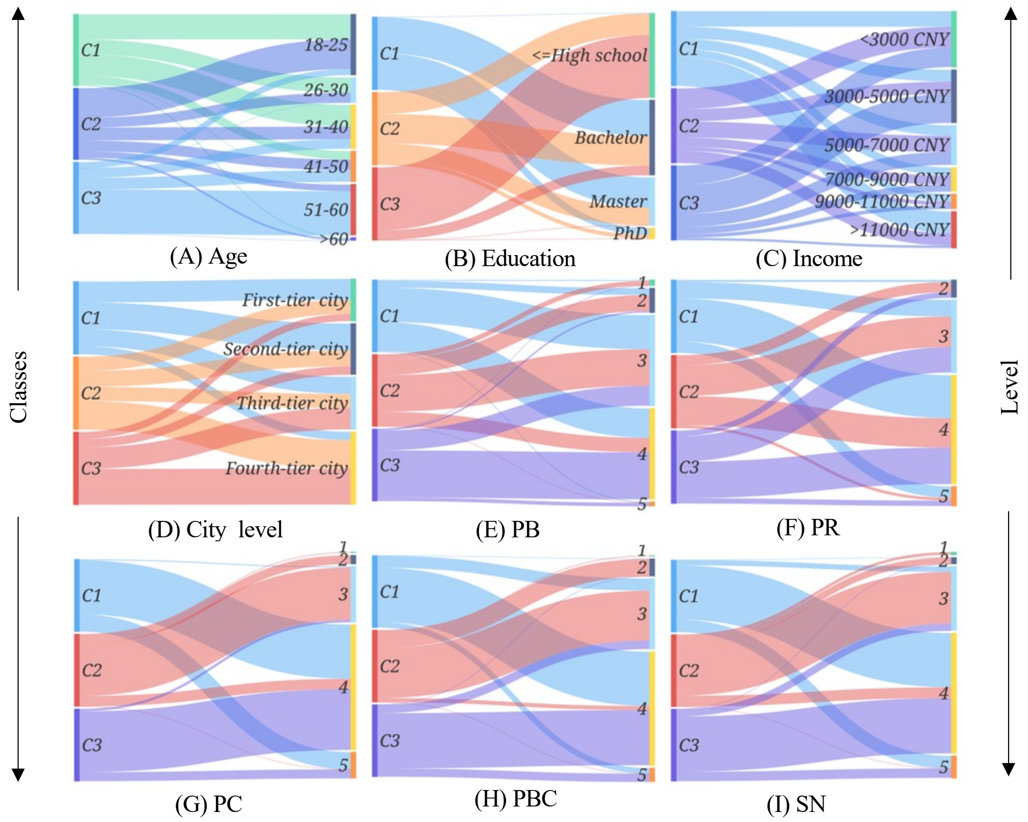


Fig. 4. Distribution of explicit variables by latent classes.

4.2 Sharing Behavior Prediction Based on LightGBM

4.2.1 Hyperparameters

As shown in Table 4, in the LightGBM model, five hyperparameters were tuned. The grid search was implemented to search for the optimal hyperparameter combinations through the scikit-learn _GridSearchCV () _ function in Python. This study developed separate models for the datasets including all samples (*AS*) and the datasets containing each of the three latent groups (*C1, C2,* and *C3*). The optimized hyperparameters on each dataset are listed in Table 2.

Table 4. Hyperparameters of LightGBM model.

Parameter	Description	*AS*	*C1*	*C2*	*C3*
n_estimators	The number of boosting iterations	100	100	100	300
Learning_rate	The model learning rate	0.5	0.1	0.2	0.1
max_depth	The maximum number of splits for base learners	UL	12	3	UL
num_leaves	Maximum number of leaves in one tree	10	10	10	10
min_child_samples	Minimum number of leaf node samples	1	1	5	20

Note: UL means unlimited.

4.2.2 Model evaluation

The performance of the LightGBM was compared to several commonly used ML models, including the logistic regression algorithm (LR), decision tree algorithm (DT), random forest algorithm (RF), neural network algorithm (NN), AdaBoost, and XGBoost. In order to compare the performance of these models, we selected several commonly used metrics (see Table 5).

Table 5. Performance evaluation metrics.

Metric	Description
Accuracy	The accuracy refers to the ratio of correctly predicted samples to the total sample size: $Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$, where true positive ($TP$) is the positive example that is correctly predicted; true negative (TN) refers to a correctly predicted counter example; false positive (FP) is a negative example being falsely predicted as a positive example; false negative (FN) is a positive example being falsely predicted as a negative example (Wei et al., 2022). The higher the accuracy, the better the classification performance of the model.
F1-score	F1 score refers to the weighted average of precision and recall. It is more effective measure of accuracy when the dataset has an imbalanced class distribution. $F_1 - score = \frac{2TP}{2TP+(FP+FN)}$. Again, the higher the F1-score, the better the model.
AUC	The Area Under Curve (AUC) is defined as the area under the Receiver Operating Characteristic (ROC) curve. AUC can measure the quality of sample sorting. The larger the AUC, the better the classification performance of the model.

For all ML models explored in this study, grid search was used to tune their hyperparameters. A 10-fold cross-validation was adopted, with 90% of samples being used as training and 10% of samples for testing. Table 6 summarizes the evaluation metrics of the best-tuned models on test datasets in different datasets. In general, LightGBM yielded better than other ML methods (i.e., Logistic Regression, Decision Tree, Random Forest, AdaBoost, Neural Network, and XGBoost) explored in this study in most datasets. Thus, in the following part of the work, we will analyze the influence of each variable based on the LightGBM results.

Table 6. Comparisons of model performance.

		LR	DT	RF	NN	AdaBoost	XGBoost	**LightGBM**
AS	Accuracy	0.67	0.78	0.85	0.73	0.72	0.82	**0.85**
	F1-score	0.75	0.82	0.88	0.8	0.78	0.85	**0.89**
	AUC	0.72	0.78	0.91	0.82	0.78	0.88	**0.91**
C1	Accuracy	0.61	**0.84**	0.83	0.66	0.66	0.82	0.81
	F1-score	0.68	**0.86**	0.85	0.68	0.71	0.84	0.84
	AUC	0.66	0.84	**0.91**	0.74	0.76	0.90	0.88
C2	Accuracy	0.62	0.82	0.85	0.74	0.77	0.84	**0.87**
	F1-score	0.72	0.86	0.88	0.81	0.83	0.87	**0.90**
	AUC	0.70	0.81	**0.89**	0.76	0.79	0.88	**0.89**
C3	Accuracy	0.62	0.82	0.82	0.71	0.75	0.80	**0.90**
	F1-score	0.71	0.85	0.87	0.78	0.81	0.85	**0.92**
	AUC	0.62	0.83	0.89	0.74	0.76	0.89	**0.94**

4.3 SHAP Results Based on LightGBM

4.3.1 Factor Importance and Discussions

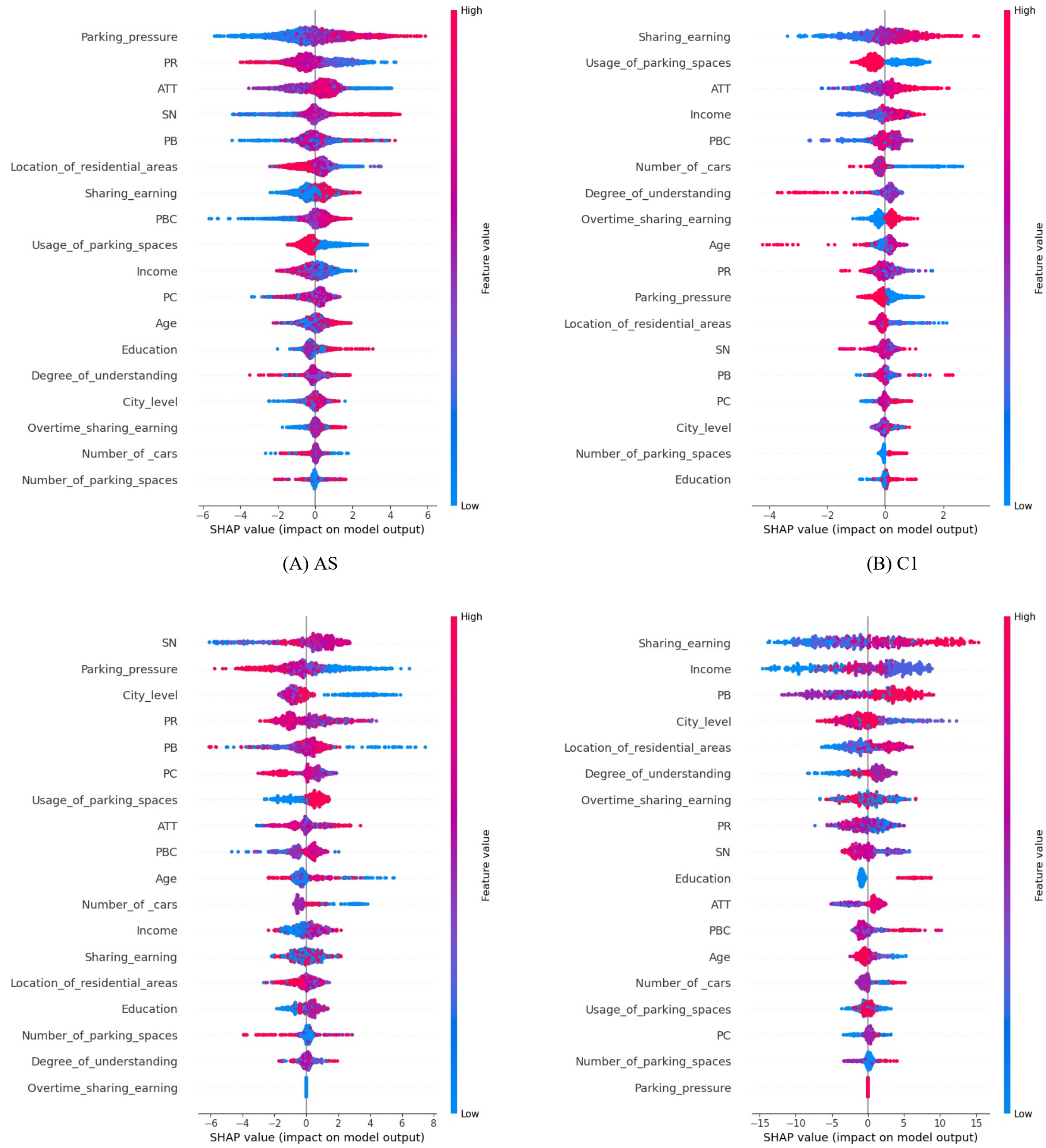
To illustrate how each factor influences the parking space sharing behaviors, the SHAP summary plots for LightGBM models based on *AS, C1, C2* and *C3* groups are shown in Fig. 5. In the SHAP summary plots, the y-axis represents the input factor, and the x-axis represents the SHAP value. Each data point refers to a choice collected by the questionnaire. The value of a factor is displayed in color, with smaller values in blue and larger values in red. The absolute SHAP value represents the effect of a factor on the model output (i.e., the probability of sharing).

A negative (positive) SHAP value represents an inverse (direct) relationship between the input factors and the sharing probability, and the larger the value, the greater the possibility of sharing, and vice versa.

As can be observed in Fig. 5, for some factors, their influence was homogeneous in different groups. For example, in all groups, the larger the *PR* (perceived risk), the smaller the SHAP value, representing the reduced sharing probability (particularly for *AS* and *C2* groups). This might be because residents may have concerns about the over-time occupation of their parking space and increased liability of allowing strangers to enter their neighbourhood or building (Liang et al., 2019) as a result of shared parking. Policy designers as well as the community should propose strategies to alleviate these concerns to facilitate shared parking in residents' areas.

At the same time, we also observed that the influence of some factors was heterogenous in different latent groups. Specifically, as shown in Fig. 5A, in general (in *AS*), *Parking_pressure* was the most influential factor in sharing behavior, and higher *Parking_pressure* values were associated with greater SHAP values (see the x-axis), representing the increased probability of sharing. It is likely that the greater the parking pressure around the residential area, the more the residents will perceive the potential parking demand and the benefits gained from sharing their parking spaces: a similar conclusion was also mentioned by Yan et al. (2019). However, *Parking_pressure* becomes less important in *C1* and *C3* groups. A potential explanation is that the residents in these two groups were from larger cities and had relatively high family incomes. Thus, they felt more pressure to reserve parking spaces for themselves and cared less about the income from sharing their parking space. Instead, in *C1* and *C3* groups, the most influential factor was *Sharing_earning*, and as can be expected, with the increase of *Sharing_earning*, the SHAP value, as well as the probability of providing shared parking spaces, also increased. While the sharing behaviors of the *C2* group are mostly influenced by *SN* (i.e., the influence of people around them, media, etc.). It is likely that respondents in *C2* are in general younger than those in *C1* and *C3* and thus care more about the suggestions from the people around them, such as family members (Yan et al., 2020). It should also be noted that the heterogeneity of factors across groups does not mean the factors would have opposite effects on sharing behaviors. For example, although *PB* (perceived benefit) has the same direction of influence on sharing behaviors in the *C1* and *C3* groups, the degree of influence was larger in the *C3* group, potentially because the residents in the *C3* group had lower family income compared to residents in the *C1* group.

Further, even in the same latent group, the influence of a factor may vary for different individuals. For example, with the increase in *Age*, some respondents in *C1* (Fig. 5B) would have a higher probability (i.e., SHAP value > 0) to share their parking spaces while some may have lower probability (i.e., SHAP value < 0) in sharing behavior. This indicates that there might be sub-groups in the latent groups obtained in our study, potentially because of other factors we have not considered in this study, which should be explored in future studies.



(A) AS (B) C1

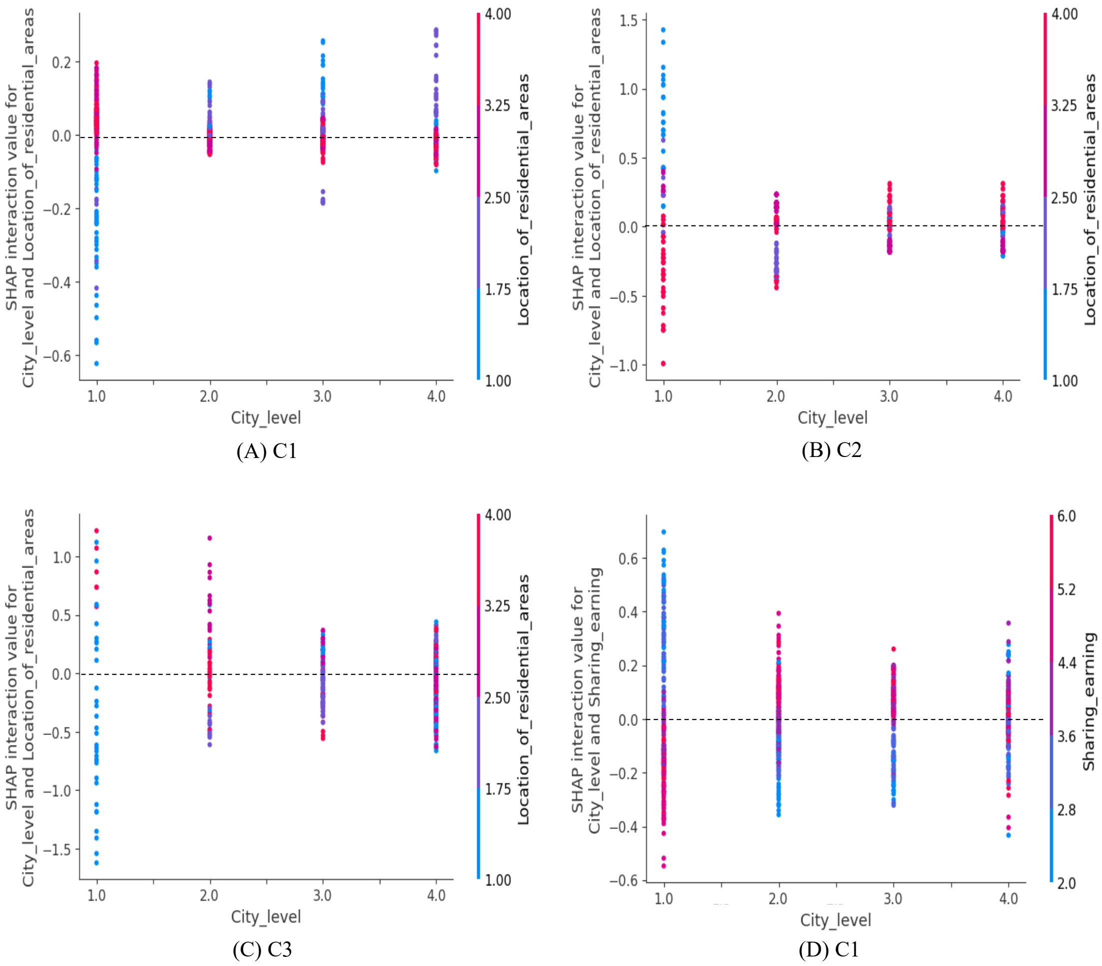
(C) C2 (D) C3

Fig. 5. SHAP summary plots.

4.3.2 Analyses of Factor Dependence Effects and Discussions

To further examine how residents' sharing behaviors are affected by different levels of the factor, the factor dependence plots in SHAP were created, which show the factor dependence effects between factors (similar to the interaction effect in statistical models). In Fig. 6 and 7, the vertical axis is the SHAP interaction values of one factor (secondary factor) when the other factor (main factor) is fixed at a specific level on the horizontal axis. The color of the secondary factor indicates its value, with smaller values in blue and larger values in red. In this paper, factor dependence plots show the dependence effects between the most influential factors that are meaningful to policy design in *C1*, *C2,* and *C3* groups including *City_level,* *Location_of_resdential_areas, Sharing_earning,* and *Income*.

Fig. 6 presents the influence of other factors when the *City_level* is fixed, while Fig. 7 shows the influence of other factors when the *Location_of_resdential_areas* is fixed. We chose *City_level* and *Location_of_resdential_areas* as the main effects for analyses as they are highly related to local policy design. As shown in Fig. 6 and Fig. 7, it is found that the sharing behaviors varied across city development (*City_level*), in different residential areas (*Location_of_resdential_areas*), in families with different family incomes (*Income*), and when they can earn differently from shared parking (*Sharing_earning*). For example, in first-tier cities, residents living closer to downtown (level 4 of *Location_of_resdential_areas*) in *C1* and *C3* were more inclined to share parking spaces, while residents in the suburb (level 1 of *Location_of_resdential_areas*) in *C2* were more willing to share parking spaces. In contrast, in second-tier cities (level 2 of *City_level*), residents living closer to downtown (level 3 of *City_level*) in *C2* and *C3* were more likely to share parking spaces, while residents in the suburbs in *C1* were more willing to share parking spaces. Further, it was found that residents of second-tier and third-tier cities (level 3 of *City_level*) were only willing to share parking spaces when the sharing earning was over 3 *CNY / 0.48 USD* (SHAP interaction value > 0). As another example, as shown in Fig. 7A and Fig. 7C), the threshold for sharing parking spaces among residents located downtown in *C1* was 5 *CNY / 0.8 USD*, while that of the *C3* group was 4 *CNY / 0.64 USD*. These sharing-earning-related findings suggest that the price strategies should be designed accordingly in order to encourage shared parking in different populations.

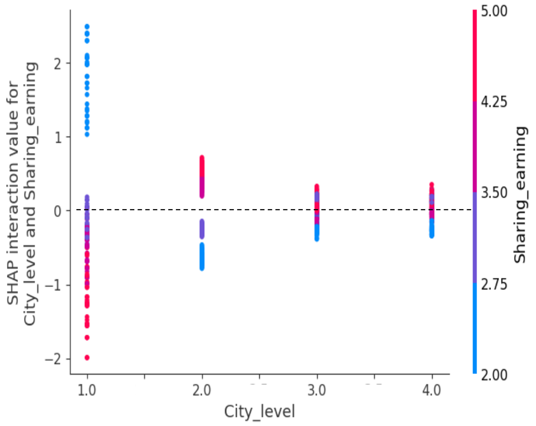


(A) C1

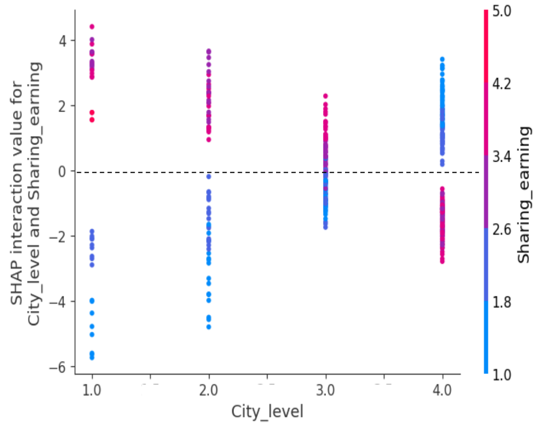
(B) C2

(C) C3

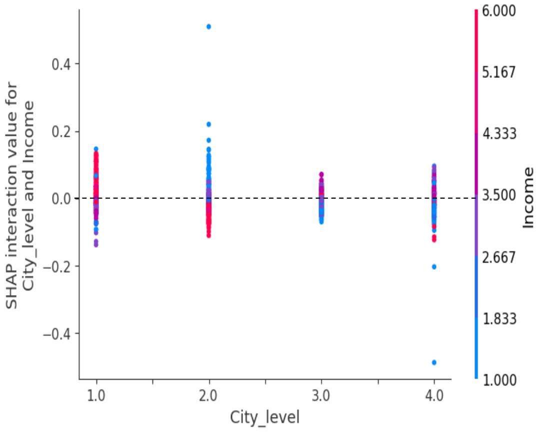
(D) C1



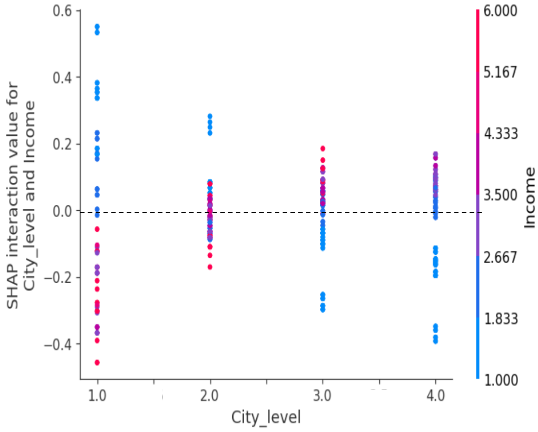
(E) C2



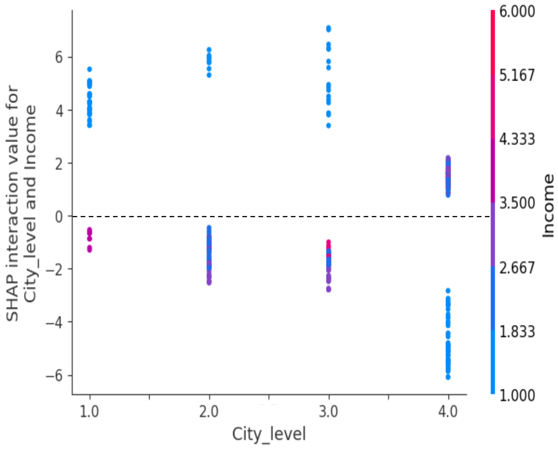
(F) C3



(G) C1

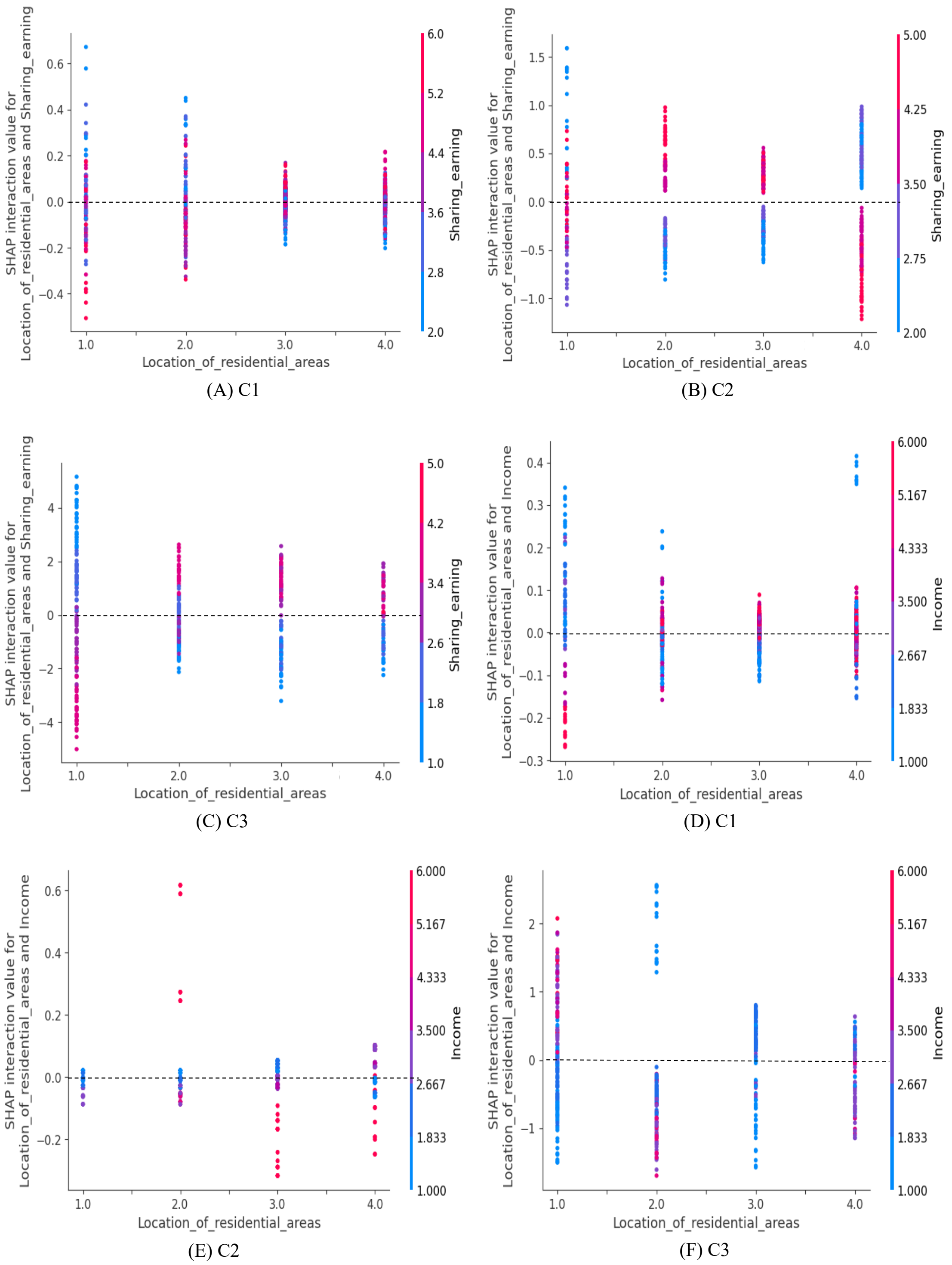


(H) C2



(I) C3

Fig. 6. SHAP interaction effects plots of *City_level*.



(A) C1 (B) C2

(C) C3 (D) C1

(E) C2 (F) C3

Fig. 7. SHAP interaction effect plots of *Location_of_resdential_areas.*

5 General Discussion and Conclusions

In this study, we explored the heterogeneity of parking space sharing behaviors among potential shared parking space suppliers by categorizing them into different latent classes (i.e., latent group of respondents) through LCM and quantified the nonlinear impact of factors on parking space sharing behaviors in different latent classes through the combination of LightGBM and SHAP analysis.

The results show that potential shared parking space suppliers can be divided into three latent groups: *C1* (41.2%), *C2* (33%) and *C3* (25.8%). In general, respondents in *C1* are mostly from first-tier or second-tier cities and are relatively young. Most of them also have bachelor's or master's degrees. Respondents in *C2* have an evenly distributed demographic background but report high *PC*, *PBC,* and *SN*. Respondents in *C3* are mostly elderly people with high school degrees and from fourth-tier cities. The LightGBM model, which was first been applied in parking space sharing behavior prediction, has also been found to outperform other commonly used ML models. The modeling results show that LightGBM has the best predictive performance in terms of accuracy, F1-score, and AUC overall. More accurate models can usually better capture the underlying relationship between parking space-sharing behavior and factors. The introduced LightGBM provides another competitive alternative for modeling parking space-sharing behavior and survey-based research. Based on the optimal LightGBM model, the SHAP analyses were conducted, and the results show the heterogeneity of the influential factors of parking space-sharing behaviors in different latent groups. In other words, the parking space-sharing behaviors in different latent groups are mainly affected by different contributing factors. Specifically, in the *C1* and *C3* groups, *sharing_earning* was the most influential factor in parking space sharing behaviors, which suggested the parking space sharing behaviors of these two latent groups were mainly driven by sharing earning, which means advantageous pricing strategies may greatly promote shared parking in these populations. In contrast, in *C2, sharing_earning* and *overtime_sharing_earning* had a limited impact on the sharing behaviors. Instead, *SN* became the most important factor affecting sharing behaviors, indicating that the behavior decisions of the *C2* group were more likely to be influenced by people around them (e.g., family members). An intuitive explanation was that respondents in the *C2* group are relatively younger and thus they may care more about the suggestions from their family members. Future research may investigate parking space sharing behavior from a group decision-making perspective of view.

In summary, the heterogeneity of the factors as well as the associated behaviors observed in our study indicated that a single policy might not be enough to facilitate parking space-sharing behaviors for different individuals with different characteristics. Thus, simply modeling the sharing behavior without considering the heterogeneity of the respondents may provide limited guidance on the design of the strategies that encourage parking space sharing. When designing shared parking policies, the government or operators should take the heterogeneity of the influential factors into consideration in order to propose customized strategies for different potential contributors in different regions.

One of the limitations of our study relates to the representativeness of the age and gender distributions of our respondents compared to the actual population of private parking space owners. While our research primarily focused on this specific demographic in urban and adjacent urban areas, we acknowledge the absence of comprehensive demographic data for this group. This gap in data limits our ability to affirm that our sample accurately reflects the broader population. Further, incorporating a more balanced gender and age sample could help in drawing more generalized conclusions applicable to the entire population of private parking space owners. Besides, future work should fine-tune the models built in our paper with a larger sample size and calibrate the model based on the data from naturalistic studies or observational studies. Further, additional factors may need to be considered (e.g., weather and individual personality) to further improve the understanding of parking space-sharing behaviors.(Wang, Guan, Qin, et al., 2021)

CRediT authorship contribution statement discussion

Ange Wang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft. **Jiyao Wang**: Methodology, Software, Validation. **Xiao Wen:** Software, Writing – review & editing. **Dengbo He**: Formal analysis, Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 52202425) and partially by the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083), the Guangzhou Municipal Science and Technology Project (No. 2023A03J0011), the Guangzhou Science and Technology Program City-University Joint Funding Project (No. 2023A03J0001).

References

Pang, Y., M. Huang, B. Wu, X. Pan, and F. Liang. 2021. "Research on intelligent composite model for public parking lot matching." In *2021 International Conference on Computer, Blockchain and Financial Development (CBFD)*, Nanjing, April. 107-111. doi:10.1109/cbfd52659.2021.00028.

Shoup, D. 2006. "Cruising for parking." *Transport Policy* 13 (6): 479–486. doi:10.1016/j.tranpol.2006.05.005.

Hensher, D., and J. King. 2001. "Parking demand and responsiveness to supply, pricing and location in the Sydney central business district." *Transportation Research Part A: Policy and Practice* 35 (3): 177–196. doi:10.1016/s0965-8564(99)00054-3.

Lalani, N. 1984. "Evaluating shared parking for new developments." *Public Works,* 115 (2). http://worldcat.org/oclc/1606878.

Xie, J., X. Ye, Z. Yang, X. Yan, L. Lu, Z. Yang, and T. Wang. 2019. "Impact of Risk and Benefit on the Suppliers' and Managers' Intention of Shared Parking in Residential Areas." *Sustainability* 12 (1): 1-17. doi:10.3390/su12010268.

Kenig, N. S., and R. J. Hocking. 1985. *Parking efficiency through shared parking concepts*. New York: American Society of Civil Engineers.

Todd, L. 2016. *Parking Management: Strategies, Evaluation and Planning*. Victoria: Victoria Transport Policy Institute.

Zhang, W, F. Gao, S. Sun, Q. Yu, J. Tang, and B. Liu. 2020. "A Distribution Model for Shared Parking in Residential Zones that Considers the Utilization Rate and the Walking Distance." *Journal of Advanced Transportation* 2020: 1–11. doi:10.1155/2020/6147974.

Beijing Transportation Development Research Institute. 2019. *2019 Annual Report on Beijing Transportation Development*. Beijing: Beijing Transportation Development Research Institute.

Yan, Q., T. Feng, and H. Timmermans. 2020. "Investigating private parking space owners' propensity to engage in shared parking schemes under conditions of uncertainty using a hybrid random-parameter logit-cumulative prospect theoretic model." *Transportation Research Part C: Emerging Technologies* 120(2020): 1-18. doi:10.1016/j.trc.2020.102776.

Yan, Q., T. Feng, and H. Timmermans. 2022. "Attitudes, personality traits and private parking space owners' willingness to engage in shared parking schemes: a hybrid prospect theoretic model." *Transportmetrica A: Transport Science*, 1–29. doi:10.1080/23249935.2022.2138628.

Beijing Transportation Development Research Institute. 2020. *2020 Annual Report on Beijing Transportation Development*. Beijing: Beijing Transportation Development Research Institute.

Ning, Y., M. Yan, S. Xu, Y. Li, and L. Li. 2021. "Shared parking acceptance under perceived network externality and risks: Theory and evidence." *Transportation Research Part A: Policy and Practice* 150 (2021): 1–15. doi:10.1016/j.tra.2021.05.009.

Ji, Y., J. Dong, Z. Lai, and Q. Feng. 2022. "Optimal allocation of shared parking spaces for hospital parkers considering parking choice behavior under bounded rationality." *Transportation Letters* 15(3):1-12. doi:10.1080/19427867.2022.2048226.

Xie, M., S. Lin, Z. Wu, X. Zhang, and Y. Wang. 2022. "Optimal allocation and adjustment mechanism of shared parking slots considering combined parking resources." *Transportation Letters* 15(2):1-12. doi: 10.1080/19427867.2022.2082767.

Duan, M., D. Wu, and H. Liu. 2020. "Bi-level programming model for resource-shared parking lots allocation." *Transportation Letters* 12(7):501-511. doi: 10.1080/19427867.2019.1631596.

Wang, A., H. Guan, J. Guo, Y. Han, and H. Bian. 2022. "An Empirical Study on the Segmentation of Potential Users of Shared Parking Spaces considering Individual Heterogeneity." *Journal of Advanced Transportation* 2022: 1–18. doi:10.1155/2022/2445693.

Ardeshiri, A., F. Safarighouzhdi, and T. H. Rashidi. 2021. "Measuring willingness to pay for shared parking." *Transportation Research Part A: Policy and Practice* 152 (2021): 186–202. doi:10.1016/j.tra.2021.08.014.

Chase, Robin. 2015. "How people and platforms are inventing the collaborative economy and reinventing capitalism." *Public Affairs*, December.

Liang, J. K., T. Eccarius, and C. C. Lu. 2019. "Investigating factors that affect the intention to use shared parking: A case study of Taipei City." *Transportation Research Part A: Policy and Practice* 130 (2019): 799–812. doi:10.1016/j.tra.2019.10.006.

Wang, A., H. Guan, Y. Han, and Y. Cao. 2021. "Private Parking Space Sharing Intention in China: An Empirical Study Based on the MIMIC Model." *Discrete Dynamics in Nature and Society* 2021: 1–13. doi:10.1155/2021/9283686.

Sfeir, G., F. Rodrigues, and M. Abou-Zeid. 2022. "Gaussian process latent class choice models." *Transportation Research Part C: Emerging Technologies* 136 (2022): 1-21. doi:10.1016/j.trc.2022.103552.

Wali, B., and A. J. Khattak. 2022. "A joint behavioral choice model for adoption of automated vehicle ride sourcing and carsharing technologies: Role of built environment & sustainable travel behaviors." *Transportation Research Part C: Emerging Technologies* 136 (2022): 1-25. doi:10.1016/j.trc.2022.103557.

Tang, J., J. Liang, C. Han, Z. Li, and H. Huang. 2019. "Crash injury severity analysis using a two-layer Stacking framework." *Accident Analysis & Prevention* 122 (2019): 226–238. doi:10.1016/j.aap.2018.10.016.

Hiroe, S., A. Oyama, S. Kitora, T. Hiroshi, Y. Ryohei, K. Jun'ichi, and H. Akihiro, et al. 2022. "Does Machine Learning Make More Reliable Diabetes Prediction Model Using Big Data?: Comparison of Gradient Boosting Decision Tree and Logistic Regression." *Research Square.* doi:10.21203/rs.3.rs-1480578/v1.

Liu, J., and L. Xiao. 2022. "Socioeconomic differences in effect size: predicting commuting mode choice of migrants and locals using a light gradient boosting approach." *Transportation*. doi:10.1007/s11116-022-10317-5.

Chun, P. J., S. Izumi, and T. Yamane. 2020. "Automatic detection method of cracks from concrete surface imagery using two-step light gradient boosting machine." *Computer-Aided Civil and Infrastructure Engineering* 36 (1): 61–72. doi:10.1111/mice.12564.

Chen, T., X. Shi, Y. D. Wong, and X. Yu. 2020. "Predicting lane-changing risk level based on vehicles' space-series features: A pre-emptive learning approach." *Transportation Research Part C: Emerging Technologies* 116 (July): 102646. doi:10.1016/j.trc.2020.102646.

Ali, Yasir, Fizza Hussain, Michiel C. J. Bliemer, Zuduo Zheng, and Md Mazharul Haque. 2022. "Predicting and explaining lane-changing behaviour using machine learning: A comparative study." *Transportation Research Part C: Emerging Technologies* 145 (2022): 1-20. doi:10.1016/j.trc.2022.103931.

Wen, X., Y. Xie, L. Wu, and L. Jiang. 2021. "Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP." *Accident Analysis & Prevention* 159 (2021): 1-11. doi:10.1016/j.aap.2021.106261.

Ullah, I., K. Liu, T. Yamamoto, M. Zahid, and A. Jamal. 2022. "Prediction of electric vehicle charging duration time using ensemble machine learning algorithm and Shapley additive explanations." *International Journal of Energy Research*. doi:10.1002/er.8219.

Hensher, D. A. 1994. "Stated preference analysis of travel choices: the state of practice." *Transportation* 21 (2): 107–133. doi:10.1007/bf01098788.

Jin, F., E. Yao, and K. An. 2020. "Analysis of the potential demand for battery electric vehicle sharing: Mode share and spatiotemporal distribution." *Journal of Transport Geography* 82 (2020): 1-12. doi:10.1016/j.jtrangeo.2019.102630.

Zhang, B. 2020. "suggestions on the Implementation of Shared Parking Policy in Shanghai." *Traffic & Transportation* 36 (3): 93–96.

Maclure, M., and M. A. Mittleman. 2000. "Should We Use a Case-Crossover Design?" *Annual Review of Public Health* 21 (1): 193–221. doi:10.1146/annurev.publhealth.21.1.193.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and M. Blondel. 2011. "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research* 12 (2011):2825–2830.

Brusco, M. J., E. Shireman, and D. Steinley. 2017. "A comparison of latent class, K-means, and K-median methods for clustering dichotomous data." *Psychological Methods* 22 (3): 563–580. doi:10.1037/met0000095.

López B., L. Miguel, and J. M. López-Bonilla. 2008. "Postmodernism and Heterogeneity of Leisure Tourist Behavior Patterns." *Leisure Sciences* 31 (1): 68–83. doi:10.1080/01490400802558210.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu. 2017. *Lightgbm: A highly efficient gradient boosting decision tree*. Advances in Neural Information Processing systems.

Lundberg, S. M., and S. I. Lee. 2017. *A Unified Approach to Interpreting Model Predictions*. Advances in Neural Information Processing systems.

Wen, X., Y. Xie, L. Jiang, Y. Li, and T. Ge. 2022. "On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development." *Accident Analysis & Prevention* 168 (2022): 1-15. doi:10.1016/j.aap.2022.106617.

Pacifico, D., and H. I. Yoo. 2013. "Lclogit: A Stata Command for Fitting Latent-Class Conditional Logit Models via the Expectation-Maximization Algorithm." *Stata Journal: Promoting communications on statistics and Stata* 13 (3): 625–639. doi:10.1177/1536867x1301300312.

Wei, T., T. Zhu, and H. Liu. 2022. "Risk prediction and factor analysis of rider's injury severity in passenger car and E-bike accidents based on interpretable machine learning." *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* 2022:1-15. doi:10.1177/09544070221121864.

Yan, X., J. Levine, and R. Marans. 2019. "The effectiveness of parking policies to reduce parking demand pressure and car use." *Transport Policy* 73 (2019): 41–50. doi:10.1016/j.tranpol.2018.10.009.

Appendix

Scenario Choice Part (assuming that you are a shared parking space provider)

Concept description:

Public parking space: refers to the charging commercial parking space built according to urban planning and supporting public buildings for social vehicles.

Overtime-sharing earning: that is, the default fee charged by the parking space users who are still using the shared parking space after the reserved parking period. To compensate for the loss of shared parking providers returning to no parking space available.

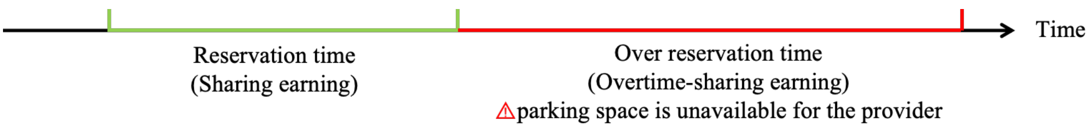


Figure 1. Description of shared parking time window

Description of Figure 1: Suppose that the green period in Figure 1 is the free time of your own parking space, and at this time, someone wants to pay to use your parking space. When the reservation time is over and you need to continue to use your parking space, the demander should leave in time. If the demander does not leave as reserved, he/she will need to pay the overtime-sharing fee to make up for your loss.

1. Please make your choice according to the following scenarios (please choose according to your judgment without thinking too much).

Parking pressure around your community: [single choice] *

○ High (The supply of public parking space is far less than the demand)

○ Moderate (The supply of public parking Spaces basically meets the demand)

○ Low (The supply of public parking space far exceeds the demand)

The following information is not showed to the participant: If the option "high" is selected, then proceed to questions 2 and 3; if the option "moderate" is selected, then proceed to questions 4 and 5; if the option "low" is selected, then proceed to questions 6 and 7.

2. The parking pressure around your community is high, so it is difficult to find the public parking space. The supply of public parking space is far less than the demand, the overtime sharing earning for providing shared parking space is **16 *RMB* / hour** (to make up for the loss of overtime uasge). What choice will you make? [Matrix single-choice] *

Sharing earning (RMB/h)	Provide	Don't provide
1	○	○
2	○	○
3	○	○
4	○	○
5	○	○
6	○	○

3. Other circumstances and conditions are the same as above, if the overtime sharing earning for the shared parking space is 24 *RMB* / hour, what choice will you make?[Matrix single-choice] *

Sharing earning (RMB/h)	Provide	Don't provide
1	○	○
2	○	○
3	○	○
4	○	○
5	○	○
6	○	○

4. The parking pressure around your community is moderate, and the supply of public parking spaces meets the demand. the overtime sharing earning for providing shared parking space is 16 *RMB* / hour. What choice will you make?[Matrix single-choice] *

Sharing earning (RMB/h)	Provide	Don't provide
1	○	○

2	○	○
3	○	○
4	○	○
5	○	○
6	○	○

5. Other circumstances and conditions are the same as above, if the overtime sharing earning of the shared parking space is 24 *RMB* / hour, what choice will you make?[Matrix single-choice] *

Sharing earning (RMB/h)	Provide	Don't provide
1	○	○
2	○	○
3	○	○
4	○	○
5	○	○
6	○	○

6. The parking pressure around your community is low. The supply of public parking space is far greater than the demand. The overtime sharing eanring for providing shared parking space is 16 *RMB* / hour. What choice will you make?[Matrix single-choice] *

Sharing earning (RMB/h)	Provide	Don't provide
1	○	○
2	○	○
3	○	○
4	○	○

5	○	○
6	○	○

7. Other circumstances and conditions are the same as above, if the overtime sharing earning for the shared parking space is 24 *RMB* / hour, what choice will you make?[Matrix single-choice] *

Sharing earning (RMB/h)	Provide	Don't provide
1	○	○
2	○	○
3	○	○
4	○	○
5	○	○
6	○	○