Preprints.org

Article

The Nonlinear and Threshold Effect of Built Environment on Ride-Hailing Travel Demand

Jiexiang Yin , Feiyan Zhao [*] , Wenyun Tang , Jianxiao Ma [*]

Article

The Nonlinear and Threshold Effect of Built Environment on Ride-Hailing Travel Demand

Jiexiang Yin, Feiyan Zhao *, Wenyun Tang and Jianxiao Ma *

College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing 210037, China; yinjiexiang@126.com (J.Y.); tangwy@njfu.edu.cn (W.T.)

* Correspondence: zfy@njfu.edu.cn (F.Z.); majx@njfu.edu.cn (J.M.)

Abstract: While numerous studies have explored the correlation between the built environment and ride-hailing demand, few have assessed their nonlinear interplay. Utilizing ride-hailing order data and multi-source built environment data from Nanjing, China, this paper uses the machine learning method, eXtreme Gradient Boosting (XGBoost), combined with SHapley Additive exPlanations (SHAP) and Partial Dependence Plots (PDP) to investigate the impact of built environment factors on ride-hailing travel demand, including their non-linear and threshold effects. The findings reveal that dining facilities have the most significant impact, with a contribution rate of 30.75%, on predicting ride-hailing travel demand. Additionally, financial, corporate, and medical facilities also exert considerable influence. The built environment factors need to reach a certain threshold or within a certain range to maximize the impact of ride-hailing travel demand. population density, land use mix and distance to the subway station collectively influence ride-hailing demand. The results are helpful for TNCs to allocate network ride-hailing resources reasonably and effectively.

Keywords: ride-hailing; built environment; XGBoost; SHAP; nonlinearity

1. Introduction

Transportation Network Companies (TNCs) such as Uber, Lyft and DiDi build service platforms based on Internet technology to match vehicles online and provide travel services for travelers in real-time, also known as ride-hailing services. Amid the swift progress of information technology and the popularity of smartphones, ride-hailing as a shared travel mode has reduced passenger waiting time and improved travel efficiency, and is gradually becoming one significant mode of daily urban transportation [1]. The scale of ride-hailing users in China continues to expand, the ride-hailing user base has surged to 453 million at the end of 2021. Although ride-hailing provides passengers with high accessibility services, researches have shown that there is an alternative or complementary relationship between ride-hailing and transit [2,3]. Some researchers have found that ride-hailing has replaced traditional modes of transportation, leading to various issues including traffic congestion, increased energy consumption, and higher vehicle mileage, which in turn contribute to additional traffic pressure [4].

As a microscopic manifestation of urban spatial form, the built environment significantly influences travel patterns [5–7]. A multitude of scholars have scrutinized the nexus between the built environment and traffic travel, and studied different models, including metro [8,9], taxi [10–12] and dockless bike-sharing [13,14]. As a rising mode of travel, ride-hailing is gradually affecting the change in residents' travel demand. The built environment factors also wield a significant influence in this phenomenon. In the pursuit of promoting green, low-carbon, and sustainable urban development, delving into the impact mechanism between the built environment and ride-hailing becomes imperative.

Researchers have discussed the relationship between the built environment and the demand for ride-hailing travel. Most of them start from the characteristics of spatial attributes, including spatial autocorrelation and spatial heterogeneity. Spatial autocorrelation effects are usually explained by spatial econometric models [15]. Geographically Weighted Regression (GWR) enhances regression accuracy by creating a local spatial weight matrix to estimate spatial variations [16]. The GWR and

its variations are frequently utilized to investigate the spatial diversity in how the built environment influences ride-hailing travel demand [17–19]. On the one hand, these models assume that there is a linear or generalized linear relationship between the two, and the conclusions may have some deviations. On the other hand, the machine learning method possesses the advantage of not adhering to the assumption of linear relationships between variables in multivariate fitting, allowing for more flexible modeling of complex relationships, which helps find complex nonlinear relationships[20]. Based on this, contemporary research has started to explore the nonlinear impact of the built environment on travel behavior and the threshold effect of the built environment can also be observed [9,21,22]. However, there are few studies on the non-linear correlation and threshold effect amid the built environment and ride-hailing travel demand.

To address this gap, the study utilizes weekday ride-hailing travel data from Nanjing, China, and applies machine learning methods, XGBoost, to explore the complex relationship between the built environment and the ride-hailing travel demand. Using the XGBoost model to identify the important built environment characteristics influencing ride-hailing travel demand. The SHAP summary plot and the traditional PDP are used to explain the threshold effect between the built environment and the ride-hailing service. It furnishes a foundation for TNCs to enhance their current ride-hailing service, and has policy implications for how future urban land use and transportation strategies can effectively affect the ride-hailing service.

The remainder of this paper is structured as follows. Section 2 provides a review of research concerning the built environment's influence on ride-hailing, with a particular focus on exploring the nonlinear relationship between influencing factors and travel behavior. Section 3 delineates the study area and outlines the pertinent variables involved. Section 4 introduces the modeling approach adopted in this study. Section 5 introduces the model estimation findings and analyzes them. Section 6 provides an overview of the primary research outcomes and suggests future avenues for investigation.

2. Literature Review

2.1. Literature Review

In recent years, ride-hailing services are becoming increasingly pivotal in the overall transportation system. Research on the key factors influencing ride-hailing trips has received widespread attention, including social demographics [23,24], user attitudes and preferences [25,26] , built environment characteristics [17–19], and other factors such as weather conditions [23]. Among these, the urban built environment, as a crucial factor affecting residents' travel behavior, is closely associated with ride-hailing trips. Built environment characteristics are commonly described using the "5Ds": density, diversity, design, destination accessibility, and distance to transit [5,6]. This framework has now been expanded to "7Ds," encompassing demand management and demographic factors.

Various aspects of the built environment, such as population density, land use patterns, and transportation infrastructure, are frequently studied to assess their correlation with ride-hailing trips. For instance, in exploring the linear relationship between the built environment and ride-hailing travel demand, Alemi et al [27] utilized binary logit models to explore the joint and separate impacts of multiple factors on Uber/Lyft usage in California. The findings suggest that a diverse mix of land use and improved regional auto accessibility positively influence the utilization of ride-hailing services. Similarly, Zhang et al. [3] investigated the correlation between the intensity of ride-hailing trips and the density of Points of Interest (POI) using ride-hailing data from Chengdu, China, using ordered logistic regression models. They found that different types of POIs had varying ramifications for ride-hailing trip intensity, with transportation facility density having the greatest impact, followed by scenic spot density, while company density had no significant influence.

To examine the spatial correlation between the built environment and ride-hailing travel demand, Sabouri et al [28] explored the impact of the built environment on ride-hailing travel demand through multilevel modeling, utilizing Uber travel data from 24 distinct regions across the United States. The results indicated a positive correlation between Uber demand and land use mix and bus station density, suggesting that ride-hailing services complement the first/last-mile connectivity of public transit, a conclusion also reached by Ghaffar et al [23]. Dean and Kockelman

[15] utilized Spatial Autoregressive (SAR) models and Structural Equation Models (SEM) to reflect spatial autocorrelation effects in census tracts. They compared the results with Ordinary Least Squares (OLS) models to explore how population and land use variables affect ride-hailing travel demand. The findings showed that an increase in job opportunities in retail and entertainment sectors promotes ride-hailing travel demand, while areas with pedestrian infrastructure and greater distance from public transit stations reduce ride-hailing travel demand.

To capture the spatial heterogeneity of the built environment's influence on ride-hailing trips, Wang and Noland [18] utilized DiDi travel data from Chengdu, China, estimating both global models and Geographically Weighted Regression (GWR) models. They examined the spatial variation of factors influencing ride-hailing trips during morning peak, evening peak, and late-night periods. The results indicated that high land use mix can create environments conducive to walking and encourage non-motorized travel [5], which differs from the conclusions of Sabouri et al [28] and Ghaffar et al [23]. Liu et al [17] utilized ride-hailing order data from Shenzhen, China, to develop a Geographically Weighted Quantile Regression (GWQR) model, exploring the temporal, spatial, and trip count variations of built environment factors. They discovered that ride-hailing services diminish the appeal of buses, subways, bicycles, and taxis, a trend particularly noticeable during weekdays. High land use mix and dense commercial areas attract ride-hailing trips, although the correlation between land use mix and ride-hailing trip counts exhibited no consistent pattern of growth. Yu and Peng [19] utilized ride-hailing trip data from Texas and applied the Geographically Weighted Panel Regression (GWPR) model to examine the spatial relationship between the built environment and ride-hailing travel demand. They discovered substantial geographical variations in the influence of built environment factors, which were comprehensive in nature. Land use mix positively impacted ride-hailing travel demand throughout the study area.

Previous research has employed diverse models to investigate the interplay between the built environment and ride-hailing travel demand, encompassing linear relationships and spatial effects. However, they have neglected the non-linear and threshold effects between these factors.

2.2. Nonlinear Effects between Traffic Travel and Influencing Factors

Analyzing the connection between influencing factors and travel behavior can be approached using various methods. In recent years, machine learning-based techniques such as Random Forests, Gradient Boosting Decision Trees (GBDT), and XGBoost have been extensively utilized to investigate such matters. For example, in investigating key factors influencing mode choice, Ding et al [29] utilized a gradient-boosting logistic regression model to incorporate built environment factors at both residential and workplace locations. They investigated the impact of the built environment and commuting patterns on commuting mode selection, unveiling that built environment variables at the workplace hold more significance than those at the residence. Moreover, most built environment variables display non-linear relationships with the choice of commuting mode. He et al [30] through a survey of car users, defined 4km as the threshold for short-distance travel and employed Random Forests to examine the essential factors influencing transportation mode selection for short-distance travelers. They analyzed the complex relationships and found significant threshold effects for key influencing factors on mode choice. Specifically, they pinpointed 1.2 kilometers as the tipping point for car and active mode selection, with a notable surge in the likelihood of car usage beyond this threshold.

Regarding the study of the nonlinear relationship between travel distance and influencing factors, Ji et al [20] utilized bicycle travel survey data from Xi'an, China, utilizing the XGBoost model to investigate the nonlinear and interplay effects between the built environment and cycling distance. They utilized SHAP for interpretation. The results indicated that the road network structure pattern contributed the most, and bicycle lane infrastructure also played an important role. They also analyzed the interaction effects of key variables in the road network structure, such as average geodesic distance, and other variables on cycling distance. Tao and Cao [21] constructed their analysis using regional travel data from the Twin Cities, US, and utilized GBDT to examine the non-linear correlation between the built environment and travel distance across driving, public transit, and active modes. The results revealed nonlinear and threshold effects between the built environment and travel distance, with different transportation modes exhibiting different impact effects.

Numerous scholars have also focused on the nonlinear impacts of influencing factors on travel demand. For instance, Tu et al [22] segmented the ride-hailing trip data into single/shared trips based

on a carpooling identification algorithm, and utilized the GBDT model to investigate the relationship between the built environment and carpooling origin-destination points. The findings revealed that proximity to the city center, land use diversity, and road density are pivotal factors influencing travel behavior, and threshold effects were analyzed using partial dependence plots. Du et al [8] aimed at identifying determinants of subway ridership, applied the GBRT model to explore the nonlinear impact that accessibility has on subway ridership from a spatiotemporal perspective. They found that accessibility indicators collectively contributed over 60% to predicting subway ridership at different times. Peng et al [9] utilized LightGBM to study the nonlinear, threshold, and synergistic effects of last-mile facilities on subway ridership. The results showed that last-mile facilities made the largest contribution to predicting subway ridership and needed to reach a attain threshold to maximize subway ridership. Through 2D-PDP analysis of their synergistic effects, they found that the influence of last-mile facilities on subway ridership may vary depending on the provision of public transit and built environment factors.

Furthermore, scholars have delved into the intricate interplay between ride-hailing and alternative transportation modes. For instance, Jin et al [31] employed the GBDT model to distinguish between weekdays and weekends, studying the nonlinear correlation between the built environment and the integrated utilization of subway and ride-hailing services. They found different patterns of built environment influence between subway-originated and subway-destinated trips. Zhang et al [3] based on the National Household Travel Survey data from San Diego, utilized the Hierarchical Negative Binomial Generalized Additive Model (HNBGAM) to explore the nonlinear association between ride-hailing trips and public transit utilization. They discovered that the frequency of ride-hailing trips plays a complementary and substitutive role in public transit use. Ride-hailing has gradually become an integral part of urban transportation. Unraveling the intricate dynamics between the built environment and ride-hailing travel demand can enable Transportation Network Companies (TNCs) to strategically allocate ride-hailing resources, enhance service levels, and mitigate overutilization of ride-hailing services. However, the current body of research on the nonlinear relationship between travel behavior and influencing factors has scarcely delved into its implications for ride-hailing travel demand.

3. Data

3.1. Study Area

Nanjing, serving as the capital of Jiangsu Province, holds significant importance as a central city in eastern China. The permanent population of Nanjing is 9.49 million. By the end of 2022, Nanjing's urban population has reached 8.26 million, with an urbanization rate of 87.01% [32]. Nanjing provided online ride-hailing services as early as 2013 [31]. At present, about 13,000 ride-hailing vehicles are legally operated in Nanjing. This paper focuses on the main urban area of Nanjing, including Qinhuai District, Xuanwu District, Jianye District, Gulou District, Yuhuatai District, and Qixia District, with a total area of 787.45 km².

Many researches have refined the analysis of data distribution and the influence of the built environment on travel behavior by dividing rectangular grids. Therefore, the study area is divided into 500m*500m rectangular grids. It is generally believed that the grid with a travel demand of 0 or less does not have strong urban functions [33]. Therefore, the grid with several ride-hailing points greater than or equal to 10 is retained. After screening, a total of 1555 grids are obtained as research units, as shown in Figure 1.
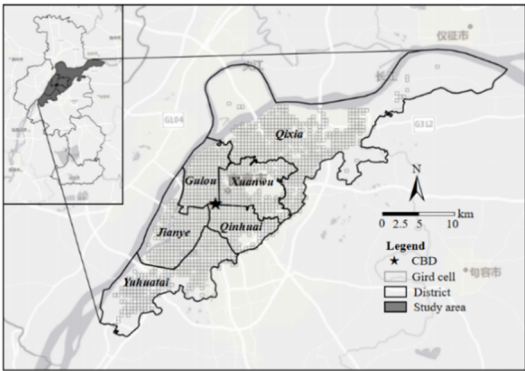
Figure 1. Study area.

3.2. Variables

The study utilized ride-hailing order data from Nanjing City for five working days, from April 11th to April 15th, 2022 (Monday to Friday). The dataset includes essential fields as shown in Table 1, comprising order ID, vehicle ID, pick-up/drop-off time, and vehicle pick-up/drop-off latitude and longitude information. The data underwent screening and cleaning procedures, such as removing irrelevant fields, excluding orders with missing pick-up/drop-off latitude and longitude or time information, and eliminating orders with trip durations less than 2 minutes or exceeding 2 hours. After these processes, a total of 835,370 valid records were obtained. The dependent variable is the count of ride-hailing pick-up points within each grid.

Table 1. Examples of the ride-hailing orders in the dataset.

Order ID	Vehicle ID	Pick-up time	Pick-off time	Pick-up location (LON, LAT)	Pick-off Location (LON, LAT)
TS120220411012600XXX	SADXXX87	2022/4/11 01:33:59	2022/4/11 01:50:56	(118.746650, 32.021847)	(118.787378, 32.048361)
15fb8a0ea4422477fXXX	SA1XXXY	2022/4/12 09:59:20	2022/4/12 10:21:39	(118.823082, 31.964883)	(118.637144, 31.930987)
TS120220413155000XXX	SADXXX19	2022/4/13 15:54:21	2022/4/13 16:20:28	(118.787103, 32.069792)	(118.734399, 32.127606)
TS120220414151003XXX	SADXXX53	2022/4/14 15:14:28	2022/4/14 15:33:21	(118.816308, 32.066452)	(118.763203, 32.009617)
17753565138XXX	SA8XXXC	2022/4/15 12:34:17	2022/4/15 12:44:36	(118.779821, 32.029380)	(118.787625, 32.043140)

The explanatory variables use "5Ds" to describe the built environment, "density" includes population density and POI density. Using the population grid data of WorldPop 100m accuracy (https://www.worldpop.org/), the Population density is calculated by dividing the total population in the grid by the grid area. POI is a kind of point data in an electronic map, including name, address, coordinate and category. It provides important data support in the research of urban built environment [20]. We selected 10 types of POI facilities for the study, the density value is obtained by dividing the number of each type of POI facility point by the grid area. The "design" adopts the description of urban road length, slow road length and intersection, and extracts the road network data of Nanjing based on OpenStreetMap (https://www. openstreetmap.org/). The length of the urban road is obtained by calculating the sum of the length of the urban expressway, main road, secondary road, and branch road. The length of a slow road is obtained by calculating the sum of the length of the sidewalk and bicycle lane. The data of road intersections are extracted and the number of road intersections in each grid is calculated. The "diversity" of land use mixing degree expression was gauged by computing the entropy index of 10 types of POI.

$$EI = \frac{-\sum A_{is} \ln A_{is}}{\ln S_i} \tag{1}$$

Where A_{is} denotes the share of land utilization category s within grid i, while S_i represents the count of distinct land types integrated within grid i. Bus and subway are important components of urban public transportation. The euclidean distance measured from the centroid point to the closest bus or subway station in each grid represents the "distance to transit". "Destination accessibility" is articulated through the euclidean distance from a given location to the CBD. The descriptive explanation of variables is shown in Table 2.

Table 2. Descriptive statistics of variables.

Variable	Description	Mean	S.D.
Dependent variables			
ride-hailing travel demand	Number of ride-hailing trips divided by the grid area (count/km²)	536.22	712.48
Independent variables			
Density			
Population density	Population size divided by the grid area (person/km²)	12641.95	19460.06
Dining facility	Number of dining facilities divided by the grid area (count/km²)	69.09	130.90
Company	Number of companies divided by the grid area (count/km²)	37.79	67.99
Shopping facility	Number of shopping facilities divided by the grid area (count/km²)	99.77	228.37
Financial facility	Number of financial facilities divided by the grid area (count/km²)	6.41	16.57
Accommodation service	Number of accommodation services divided by the grid area (count/km²)	11.26	32.99
Science & Education & Culture	Number of Science & Education & Culture facilities divided by the grid area (count/km²)	22.99	44.82
Scenic spot	Number of scenic spots divided by the grid area (count/km²)	5.23	18.76
Commercial residence	Number of commercial residences divided by the grid area (count/km²)	17.79	24.02
Leisure service	Number of leisure services divided by the grid area (count/km²)	5.32	12.40
Medical facility	Number of medical facilities divided by the grid area (count/km²)	12.87	23.21
Design			
Road length	The length of roads divided by the grid area (km /km²)	1.97	1.15
Non-motorized road length	The length of non-motorized roads divided by the grid area (km /km²)	0.21	0.51

Number of Road nodes	Number of road nodes divided by the grid area (count/km²)	7.07	9.31
Diversity			
Land use mix	The entropy value of thirteen categories of POIs	0.71	0.28
Destination accessibility			
Distance to CBD	Distance from the grid centroid to CBD (km)	10.16	5.73
Distance to transit			
Distance to bus stop	Distance from the grid centroid to the nearest bus stop (km)	0.31	0.21

4. Methodology

4.1. XGBoost

XGBoost proposed by Chen and Guestrin [34] based on GBDT shows good advantages in prediction performance and prevention of overfitting. Therefore, this study uses XGBoost to investigate the nonlinear relationship between built environment variables and ride-hailing travel demand.

This method belongs to a forward iterative model. The iterative process will train multiple trees. Finally, the aggregated prediction from each tree within the sample serves as the predicted value for the entire sample in the model:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) \tag{2}$$

Where is the predicted value of ride-hailing travel demand in the grid i after the iteration t, $f_k(x_i)$ is the predicted value of the tree k, and x_i is the built environment in the grid i.

Overfitting is one of the defects that hinder the accuracy and performance of the model in machine learning [20], XGBoost prevents model overfitting by adding regularization terms:

$$L(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{t} \Omega(f_k) \tag{3}$$

$$\Omega(f_k) = \gamma T_k + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{4}$$

Where n represents the total number of grids within the study area, $l(y_i, \hat{y}_i)$ represents the loss function calculated between the predicted value and the true value, $\Omega(f_k)$ is the regularization term, T_k is the number of leaf nodes of the tree k, w_j^2 is the weight assigned to the tree node, γ and λ are penalty factors respectively.

We employ 5-fold cross-validation to determine the optimal parameter configureurations, when XGBoost has the best fitting effect, the hyperparameters are set as follows: "n_estimults:100, gamma=0.01, learning_rate:0.05, alpha=1, max_depth:5".

4.2. Model explanation

SHAP is a method for interpreting machine learning models, originally extended from the Shapley value concept in game theory [35]. Lundberg and Lee [36] proposed a unified predictive interpretation framework, which enables the quantification of each feature's contribution within the model and utilizes the Shapley value to illustrate the influence distribution of each feature on the model output [9,37]:

$$\phi_j = \sum_{z' \subseteq x'} \frac{|z'|!(P - |z'| - 1)!}{P!}(f_x(z') - f_x(z'/j)) \tag{5}$$

where ϕ_j denotes the contribution of feature j, P is the number of features, $|z'|$ represents the number of non-zero entries in z', and $z' \subseteq x'$ represents all z' vectors where the non-zero entries form a subset of those in x'.

Partial dependence plots (PDP) can effectively reveal the nonlinear relationship between input features and the target variable, enabling a deeper understanding of how the model utilizes these features to make decisions, defined as:

$$F_s(x_s) = E_{x_c}[F(x_s, x_c)] \tag{6}$$

$$\overline{F}_s(x_s) = \frac{1}{n}\sum_{i=1}^{n} F(x_s, x_c) \tag{7}$$

Where x_s is the built environment variable to be analyzed, x_c is the other built environment variable except x_s, $F_s(x_s)$ represents the forecasted value of the ride-hailing travel demand when taking the mean value and taking different values. The partial dependence function defined in equations (6) and (7) is derived from the average level of other built environment variables x_c to analyze the influence of the built environment variable x_s on the ride-hailing travel demand.

5. Results and Discussion

5.1. Relative Importance Analysis

Table 3 illustrates the relative significance of built environment variables in influencing ride-hailing travel demand. A higher importance value indicates a greater impact of the built environment variable on ride-hailing travel demand. The sum of importance values for all variables totals 100%. Due to significant differences in the quantity of built environment features across different dimensions, average relative importance is employed to assess the significance of the five dimensions of the built environment. Among these dimensions, density has the highest average relative importance (7.89%), with destination accessibility (3.05%) following closely behind. Meanwhile, the average relative importance of design, diversity, and distance to transit are 1.86%, 1.48%, and 1.53%, respectively. In the ranking of built environment variables, the top five in relative importance are all POI facilities, namely dining facilities, financial facilities, medical facilities, companies, and shopping facilities. This indicates that POI facilities significantly impact ride-hailing travel demand, consistent with previous research [3], where restaurants contribute the most to ride-hailing travel demand at 30.75%.

Table 3. The relative importance of independent variables.

Variables	Relative importance	Ranking
Density (Average of relative importance: 7.89%)		
Population density	1.36%	15
Dining facility	30.75%	1
Company	5.94%	4
Shopping facility	4.86%	5
Financial facility	26.42%	2
Accommodation service	3.19%	7
Science & Education & Culture	1.34%	16
Scenic spot	0.59%	18
Commercial residence	3.56%	6
Leisure service	2.55%	9
Medical facility	6.28%	3
Design (Average of relative importance: 1.86%)		
Road length	2.07%	10
Non-motorized road length	1.88%	11
Number of Road nodes	1.63%	13
Diversity (Average of relative importance: 1.48%)		
Land use mix	1.48%	14

Destination accessibility (Sum of relative importance: 3.05%)		
Distance to CBD	3.05%	8
Distance to transit (Average of relative importance: 1.53%)		
Distance to bus stop	1.23%	17
Distance to subway station	1.82%	12

5.2. SHAP Summary Plot of Independent Variables

The SHAP summary plot reveals the order of various built environment features and how they positively or negatively influence the target variable. In the plot, each colored point corresponds to a sample, with the color indicating the magnitude of the built environment feature values. Higher feature values are depicted in red, while lower feature values are depicted in blue. The x-axis illustrates the SHAP values corresponding to each built environment feature.

As shown in Figure 2, based on the SHAP feature ranking, it is found that the top 4 ranked built environment features are similar to the relative importance ranking, all of which are points of POI facilities, namely dining facilities, financial facilities, companies, and medical facilities. Among them, dining facilities are considered the most important, with the highest contribution to relative importance ranking as well. However, the road length is ranked 5th in the SHAP feature ranking, which differs from the relative importance ranking.

Samples with fewer dining facilities tend to have negative SHAP values, while samples with larger dining facilities tend to have positive values, indicating a positive correlation between dining facilities and SHAP values. This is because areas with higher-density dining facilities often coincide with shopping malls or densely populated areas, resulting in higher travel demand. Similarly, POI facilities such as finance, company, medical, shopping, accommodation, and leisure also exhibit similar feature effects, while tourist attraction POI show negative effects. The distance to CBD and subway accessibility are negatively correlated with SHAP values, indicating that closer proximity to the CBD results in higher travel demand. Similarly, other built environment variables that exhibit significant negative correlations with SHAP values include the distance to subway station. Road length and the number of road nodes show positive correlations with SHAP values. Conversely, the non-motorized road length shows a negative correlation. This is because road nodes enhance the connectivity of urban roads, and a more complete and smooth road network structure attracts more ride-hailing travel demand [19]. Conversely, a focus on non-motorized transportation development will reduce motorized travel and suppress the demand for ride-hailing travel demand.
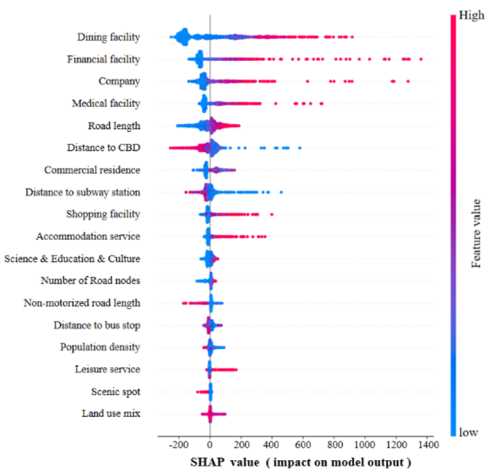

Figure 2. SHAP summary plot of independent variables.
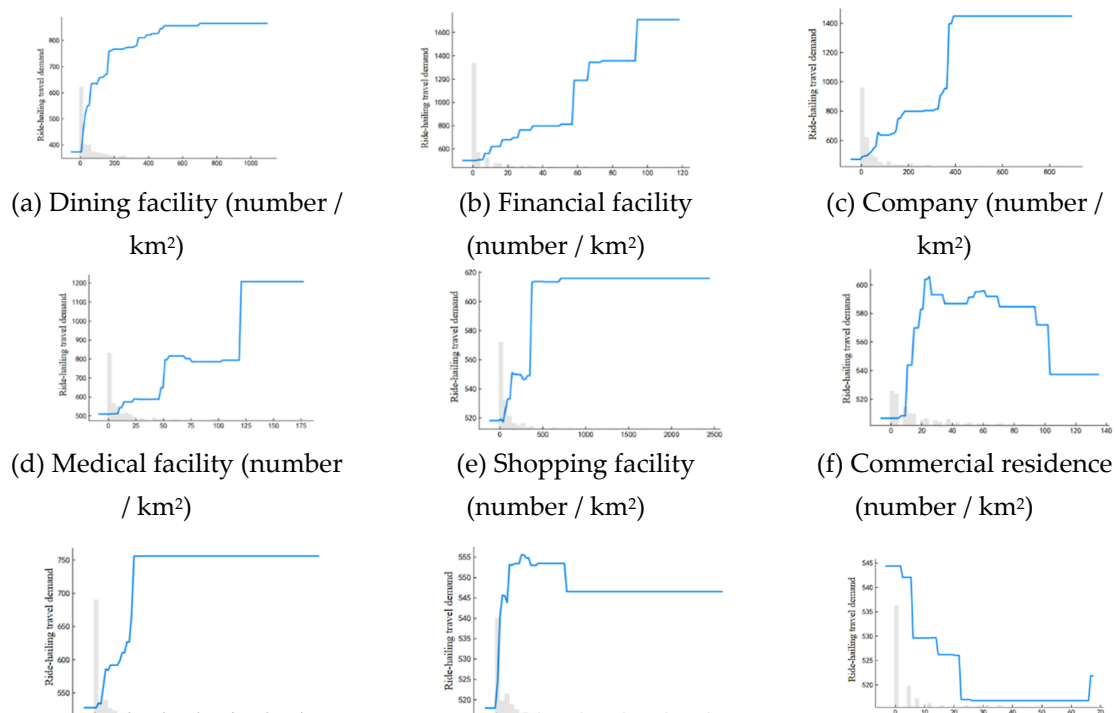
5.3. Marginal Effects Analysis

The above results help determine the importance and positive/negative feedback of each built environment variable in influencing ride-hailing travel demand. To further clarify the nonlinear

relationship between individual built environment variables and the demand for ride-hailing travel, partial dependence plots can be utilized to visualize the marginal impacts of built environment features on model predictions [38]. This allows capturing the effective scope of influence and threshold impacts on ride-hailing travel demand.

Figure 3 illustrates the non-linear relationship between density variables and ride-hailing travel demand. Dining, finance, company, medical, shopping, accommodation, and leisure POI facilities positively influence ride-hailing travel demand. As the density of POI facilities rises, so does the demand for ride-hailing. However, the range of influence and thresholds vary among different types of POI facilities. For example, the influence of dining facility density on ride-hailing travel demand gradually increases between 0 and 700, and stabilizes after exceeding 700. Similarly, for financial facilities, the impact stabilizes after surpassing 100.

Scenic spots have a negative impact on ride-hailing travel demand, as illustrated in Figure 3(i). A noticeable decrease in demand occurs within the range of 0 to 20. This could be attributed to the fact that scenic spots are typically located near public transportation hubs, making it convenient for tourists to travel without relying on ride-hailing services. Additionally, these scenic spots may often be situated in suburban areas, where the cost of utilizing ride-hailing services is relatively higher. The influence of commercial residence areas and Science & Education & Culture areas on ride-hailing travel demand exhibits a trend of rising and then declining, as shown in Figure 3(f) and 3(h). This indicates that there are certain advantageous intervals for influencing ride-hailing travel demand. Commercial residence areas are most attractive to ride-hailing travel demand within the range of 20 to 90, while Science & Education & Culture areas exhibit peak influence between 50 and 150.

As shown in Figure 3 (k), population density generally impacts ride-hailing travel demand positively. Within the range of 0 to 20,000 person/km², there is significant fluctuation in ride-hailing travel demand. The lowest demand for ride-hailing occurs when the population density is 5000 person/km². This suggests that areas with either sparse or excessively dense populations exhibit higher ride-hailing travel demand. This could be attributed to several factors: in regions with low population density, public transportation accessibility may be poor while road connectivity is high, making ride-hailing a preferred mode of transportation due to the lack of viable alternatives. Additionally, in densely populated areas, where the separation between residential and commercial areas is pronounced, ride-hailing becomes more favorable.
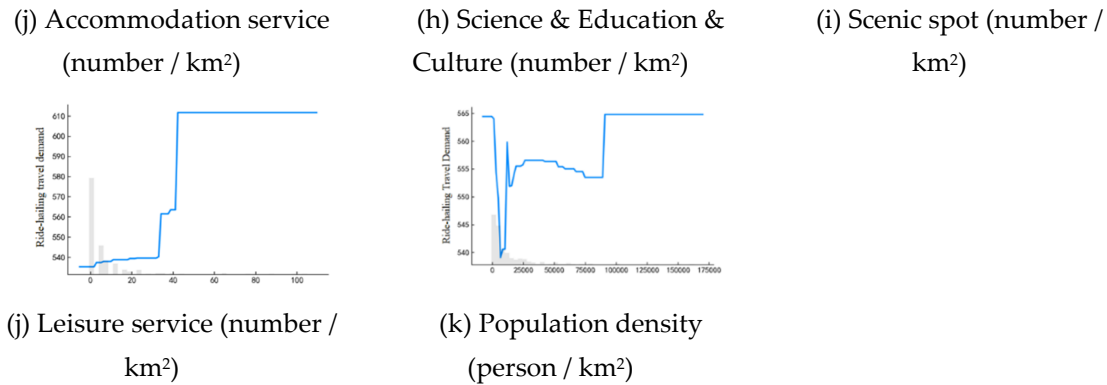

(a) Dining facility (number / km²)

(b) Financial facility (number / km²)

(c) Company (number / km²)

(d) Medical facility (number / km²)

(e) Shopping facility (number / km²)

(f) Commercial residence (number / km²)

(j) Accommodation service (number / km²)

(h) Science & Education & Culture (number / km²)

(i) Scenic spot (number / km²)

(j) Leisure service (number / km²)

(k) Population density (person / km²)

Figure 3. Nonlinear effect of density variables on ride-hailing travel demand.

Figure 4(a) illustrates a noticeable positive impact of road length on ride-hailing travel demand within the grid, particularly within the range of 0 to 4. As road length increases, there is a clear rise in ride-hailing travel demand, which stabilizes once the road length surpasses this range. In Figure 4(b), when the length of slow roads within the grid reaches 1 km, there is a sharp decrease of approximately 500 ride-hailing travel trips. Figure 4(c) illustrates that the influence of the number of road nodes on ride-hailing travel demand is most significant within the range of 0 to 3. As the number of road nodes increases from 3 to around 40, there is a gradual increase in demand. However, this impact stabilizes once the number of road nodes exceeds approximately 40 per grid.
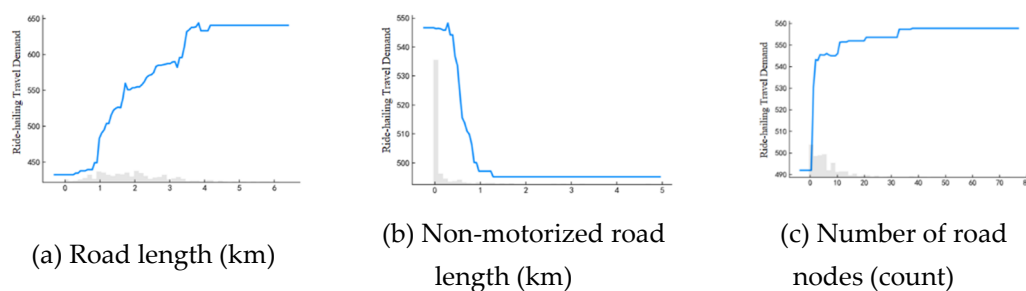
(a) Road length (km)

(b) Non-motorized road length (km)

(c) Number of road nodes (count)

Figure 4. Nonlinear effect of design variables on ride-hailing travel demand.

As illustrated in Figure 5, the demand for ride-hailing remains constant when the land use mix is between 0 and 0.4. However, the demand for ride-hailing reaches its peak when the land use mix approaches around 0.6. This finding aligns with previous research [23,27,28], which suggests that higher land use mix attracts ride-hailing services to provide transportation options. Conversely, between 0.6 and 0.8 in the land use mix, there is a notable decrease in demand for ride-hailing. This indicates that when the land use mix attains a particular level of completeness in terms of POI facilities within the regional grid, it is more conducive to walking trips [5,18].
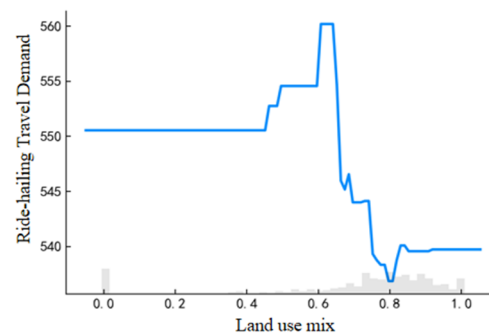
Figure 5. Nonlinear effect of diversity variables on ride-hailing travel demand.

As shown in Figure 6, the demand for ride-hailing increases as the distance to the CBD decreases, which corresponds with the pattern of higher demand closer to the CBD. Ride-hailing drivers often prefer to operate in the city center to meet the higher demand. When the distance from CBD is between 0 and 20 km, there is a noticeable decrease in ride-hailing travel demand, particularly within the range of 0 to 1 km, where the negative impact on ride-hailing travel demand is most significant. Ride-hailing travel demand sharply decreases from nearly 900 to 600. Beyond a distance of 20 km from CBD, the impact tends to stabilize.
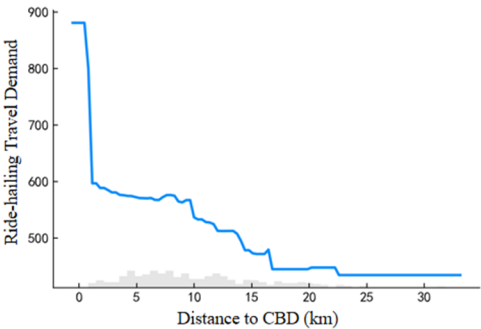

Figure 6. Nonlinear effect of destination accessibility on ride-hailing travel demand.

As shown in Figure 7(a), When the distance to the bus station falls within the range of approximately 0 to 0.5 km, there is a significant decline in ride-hailing travel demand. Beyond a distance of 3 km, the impact of ride-hailing tends to stabilize. Similarly, as illustrated in Figure 7(b), the distance to the metro station shows a continuous reduction in ride-hailing travel demand within the range of 0 to 0.35 km, reaching its lowest point at 0.35 km. This phenomenon may be attributed to the integration of ride-hailing with public transportation, aiming to address the "last-mile" problem of commuting [27,28]. We also observed that bus stations have a larger radius of influence. Conversely, the ride-hailing travel demand increases slightly after reaching a distance of 0.4 km from metro stations, and then stabilizes. This may be because ride-hailing services complement the "first-mile" commute to connect with metro stations, contributing to this phenomenon.
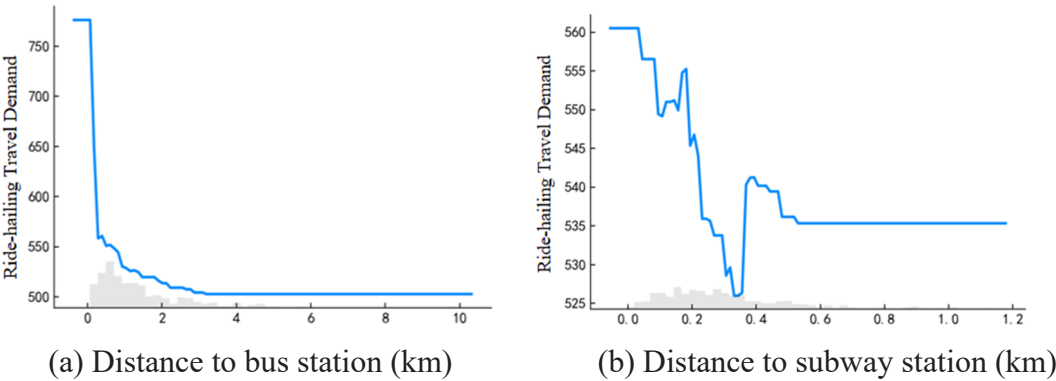

(a) Distance to bus station (km)   (b) Distance to subway station (km)

Figure 7. Nonlinear effect of distance to transit on ride-hailing travel demand.

6. Conclusions

This study utilizes ride-hailing data from Nanjing and integrates multi-source data including population, urban road network, POI and public transportation stations to characterize built environment indicators from the perspective of the "5Ds.". By applying the XGBoost model, the study explores the impact of built environment factors on ride-hailing travel demand. Through the combined analysis of SHAP and PDP, the study elucidates the contribution, nonlinearity, and threshold impacts of built environment features on predicting ride-hailing travel demand. The research findings provide valuable insights for urban development, transportation planning, and ride-hailing resource allocation.

Firstly, the article estimated the average relative importance of the "5Ds" and assessed the relative importance and ranking of various built environment features on ride-hailing travel demand.

The results indicate that the density dimension has the highest average relative importance at 7.89%, and the contribution rate of dining facilities to predicting ride-hailing travel demand is highest at 30.75%. Secondly, through the SHAP summary plot analysis, the article examined how built environment features positively or negatively affect the target variable. Based on the SHAP feature ranking, it was found that the top 4 built environment features are dining, finance, company, and medical facilities, all of which are POI facilities. TNCs should prioritize the supply and allocation of ride-hailing vehicles around these built environment features. Finally, through the PDP plot visualization, the article illustrated the nonlinearity and threshold effects between built environment features and ride-hailing travel demand. The results indicate that built environment features need to reach a certain threshold or be within a certain range to exert the maximum impact on ride-hailing travel demand., with different thresholds and ranges for different built environment features. Analyzing the marginal impacts of population density, land use mix, and distance to subway stations revealed their comprehensive impact on ride-hailing travel demand. Understanding the nonlinearity and threshold effects between built environment variables and ride-hailing travel demand can help TNCs allocate ride-hailing resources reasonably and prevent overuse.

The article has certain limitations, and future research could explore these aspects:(1) The study focused on ride-hailing travel demand over a typical workweek without analyzing specific periods. However, ride-hailing travel demand typically exhibits peak and off-peak periods during the day, and demand characteristics may vary from weekdays to weekends [33]. Clarifying the nonlinear impacts of factors influencing ride-hailing travel demand during different periods is important. Future research could investigate how factors such as built environment features influence ride-hailing travel demand during peak and off-peak hours, both on weekdays and weekends. (2) Both built environment features and ride-hailing travel demand possess spatial attributes. Future research could analyze spatial effects in machine learning models. It could investigate whether interpretable machine learning models incorporating spatial effects outperform traditional geographically weighted models or exhibit similar or improved performance. This comparative examination would shed light on the effectiveness of different modeling approaches in capturing the spatial dynamics of ride-hailing travel demand affected by built environment features.

Author Contributions: Conceptualization, J.M., F.Z., W.T., and J.Y.; methodology, J.M. and W.T.; software, F.Z.; validation, J.Y. and F.Z.; writing—original draft preparation, J.Y. and F.Z.; writing—review and editing, J.M., F.Z., W.T., and J.Y.; visualization, F.Z. and J.Y.; supervision, J.M., W.T. and F.Z.; funding acquisition, F.Z. and J.Y All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data belongs to Nanjing regulatory platform. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rayle, L.; Dai, D.; Chan, N.; Cervero, R.; Shaheen, S. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transp. Policy*. **2016**, *45*, 168-178. [CrossRef]
2. Young, M.; Allen, J.; Farber, S. Measuring when Uber behaves as a substitute or supplement to transit: An examination of travel-time differences in Toronto. *J. Transp. Geogr*. **2020**, *82*, 102629. [CrossRef]
3. Zhang, B.; Chen, S.; Ma, Y.; Li, T.; Tang, K. Analysis on spatiotemporal urban mobility based on online car-hailing data. *J. Transp. Geogr*. **2020**, *82*, 102568. [CrossRef]
4. Shi, K.; Shao, R.; De Vos, J.; Cheng, L.; Witlox, F. The influence of ride-hailing on travel frequency and mode choice. *Transport. Res. Part D-Transport. Environ*. **2021**, *101*, 103125. [CrossRef]
5. Ewing, R.; Cervero, R. Travel and the Built Environment. *J. Am. Plan. Assoc*. **2010**, *76*, 265-294. [CrossRef]
6. Ewing, R.; Cervero, R. Travel and the built environment - A synthesis. *Transp. Res. Record*. **2001**, *1780(1)*, 87-114.[CrossRef]

7. Sun, B.; Ermagun, A.; Dan, B. Built environmental impacts on commuting mode choice and distance: Evidence from Shanghai. *Transport. Res. Part D-Transport. Environ*. **2017**, *52*, 441-453. [CrossRef]

8. Du, Q.; Zhou, Y.; Huang, Y.; Wang, Y.; Bai, L. Spatiotemporal exploration of the non-linear impacts of accessibility on metro ridership. *J. Transp. Geogr*. **2022**, *102*, 103380. [CrossRef]

9. Peng, B.; Zhang, Y.; Li, C.; Wang, T.; Yuan, S. Nonlinear, threshold and synergistic effects of first/last-mile facilities on metro ridership. *Transport. Res. Part D-Transport. Environ*. **2023**, *121*, 103856. [CrossRef]

10. Chen, C.; Feng, T.; Ding, C.; Yu, B.; Yao, B. Examining the spatial-temporal relationship between urban built environment and taxi ridership: Results of a semi-parametric GWPR model. *J. Transp. Geogr*. **2021**, *96*, 103172. [CrossRef]

11. Qian, X.; Ukkusuri, S.V. Spatial variation of the urban taxi ridership using GPS data. *Appl. Geogr*. **2015**, *59*, 31-42. [CrossRef]

12. Zhu, P.; Li, J.; Wang, K.; Huang, J. Exploring spatial heterogeneity in the impact of built environment on taxi ridership using multiscale geographically weighted regression. *Transportation*. **2023**, 1-35. [CrossRef]

13. Ma, X.; Ji, Y.; Yuan, Y.; Oort, N.V.; Jin, Y.; Hoogendoorn, S. A comparison in travel patterns and determinants of user demand between docked and dockless bike-sharing systems using multi-sourced data. *Transp. Res. Pt. A-Policy Pract*. **2020**, *139*, 148-173. [CrossRef]

14. Wang, Y.; Li, J.; Su, D.; Zhou, H. Spatial-temporal heterogeneity and built environment nonlinearity in inconsiderate parking of dockless bike-sharing. *Transp. Res. Pt. A-Policy Pract*. **2023**, *175*, 103789. [CrossRef]

15. Dean, M.D.; Kockelman, K.M. Spatial variation in shared ride-hail trip demand and factors contributing to sharing: Lessons from Chicago. *J. Transp. Geogr*. **2021**, *91*, 102944. [CrossRef]

16. Fotheringham, A.S.; Charlton, M.E.; Brunsdon, C. Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environ. Plan. A*. **1998**, *30*, 1905-1927. [CrossRef]

17. Liu, F.; Gao, F.; Yang, L.; Han, C.; Hao, W.; Tang, J. Exploring the spatially heterogeneous effect of the built environment on ride-hailing travel demand: A geographically weighted quantile regression model. *Travel Behav. Soc*. **2022**, *29*, 22-33. [CrossRef]

18. Wang, S.; Noland, R.B. Variation in ride-hailing trips in Chengdu, China. *Transport. Res. Part D-Transport. Environ*. **2021**, *90*, 102596. [CrossRef]

19. Yu, H.; Peng, Z.-R. Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. *J. Transp. Geogr*. **2019**, *75*, 147-163. [CrossRef]

20. Ji, S.; Wang, X.; Lyu, T.; Liu, X.; Wang, Y.; Heinen, E.; Sun, Z. Understanding cycling distance according to the prediction of the XGBoost and the interpretation of SHAP: A non-linear and interaction effect analysis. *J. Transp. Geogr*. **2022**, *103*, 103414. [CrossRef]

21. Tao, T.; Cao, J. Exploring nonlinear and collective influences of regional and local built environment characteristics on travel distances by mode. *J. Transp. Geogr*. **2023**, *109*, 103599. [CrossRef]

22. Tu, M.; Li, W.; Orfila, O.; Li, Y.; Gruyer, D. Exploring nonlinear effects of the built environment on ridesplitting: Evidence from Chengdu. *Transport. Res. Part D-Transport. Environ*. **2021**, *93*, 102776. [CrossRef]

23. Ghaffar, A.; Mitra, S.; Hyland, M. Modeling determinants of ridesourcing usage: A census tract-level analysis of Chicago. *Transp. Res. Pt. C-Emerg. Technol*. **2020**, *119*, 102769. [CrossRef]

24. Gomez, J.; Aguilera-García, Á.; Dias, F.F.; Bhat, C.R.; Vassallo, J.M. Adoption and frequency of use of ride-hailing services in a European city: The case of Madrid. *Transp. Res. Pt. C-Emerg. Technol*. **2021**, *131*, 103359. [CrossRef]

25. Dong, X. Trade Uber for the Bus? An Investigation of Individual Willingness to Use Ride-Hail Versus Transit. *J. Am. Plan. Assoc*. **2020**, *86*, 222-235. [CrossRef]

26. Loa, P.; Habib, K.N. Examining the influence of attitudinal factors on the use of ride-hailing services in Toronto. *Transp. Res. Pt. A-Policy Pract*. **2021**, *146*, 13-28. [CrossRef]

27. Alemi, F.; Circella, G.; Handy, S.; Mokhtarian, P. What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California. *Travel Behav. Soc*. **2018**, *13*, 88-104. [CrossRef]

28. Sabouri, S.; Park, K.; Smith, A.; Tian, G.; Ewing, R. Exploring the influence of built environment on Uber demand. *Transport. Res. Part D-Transport. Environ*. **2020**, *81*, 102296. [CrossRef]

29. Ding, C.; Cao, X.; Wang, Y. Synergistic effects of the built environment and commuting programs on commute mode choice. *Transp. Res. Pt. A-Policy Pract*. **2018**, *118*, 104-118. [CrossRef]

30. He, M.; Pu, L.; Liu, Y.; Shi, Z.; He, C.; Lei, J. Research on Nonlinear Associations and Interactions for Short-Distance Travel Mode Choice of Car Users. *J. Adv. Transp*. **2022**, *2022*, 8598320. [CrossRef]

31. Jin, T.; Cheng, L.; Zhang, X.; Cao, J.; Qian, X.; Witlox, F. Nonlinear effects of the built environment on metro-integrated ridesourcing usage. *Transport. Res. Part D-Transport. Environ*. **2022**, *110*, 103426. [CrossRef]

32. Statistical Communiqué on National Economic and Social Development of Nanjing in 2022. Available online: [CrossRef] (accessed on 12 Oct 2023).

33. He, Z. Portraying ride-hailing mobility using multi-day trip order data: A case study of Beijing, China. *Transp. Res. Pt. A-Policy Pract*. **2021**, *146*, 152-169. [CrossRef]

34. Chen, T.; Guestrin, C.; Assoc Comp, M. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, 2016, Aug 13-17, 2016; pp. 785-794. [CrossRef]

35. Shapley, L.S. *A Value for N-Person Games*; RAND Corporation: Santa Monica, CA, 1952. [CrossRef]

36. Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Arxiv*. **2017**, *30*, 07874. [CrossRef]

37. Štrumbelj, E.; Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Systs*. **2014**, *41*, 647-665. [CrossRef]

38. Li, Z. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Comput. Environ. Urban Syst*. **2022**, *96*, 101845. [CrossRef]