Can Bike-Sharing Reduce Car Use in Alexandroupolis? An Exploration through the Comparison of Discrete Choice and Machine Learning Models

Santhanakrishnan Narayanan [1,*], Nikita Makarov [1], Evripidis Magkos [2], Josep Maria Salanova Grau [2], Georgia Aifadopoulou [2] and Constantinos Antoniou [1]

1   TUM School of Engineering and Design, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany
2   Hellenic Institute for Transport, Centre for Research & Technology Hellas, 6th km Charilaou-Thermis Road, 57001 Thessaloniki, Greece
*   Correspondence: santhanakrishnan.narayanan@tum.de

Abstract: The implementation of bike-sharing systems (BSSs) is expected to lead to modifications in the travel habits of transport users, one of which is the choice of travel mode. Therefore, this research focuses on the identification of factors influencing the shift of private car users to BSSs based on stated preference survey data from the city of Alexandroupolis, Greece. A binary logit model is employed for this purpose. The estimation results indicate the impacts of gender, income, travel time, travel cost and safety-related aspects on the mode shift, through which behavioural insights are derived. For example, car users are found to be twice as sensitive to the cost of BSSs than to that of car. Similarly, they are highly sensitive to BSS travel time. Based on the behavioural findings, policy measures are suggested under the following categories: (i) finance, (ii) regulation, (iii) infrastructure, (iv) campaigns and (v) customer targeting. In addition, a secondary objective of this research is to obtain insights from the comparison of the specified logit model with a machine learning approach, as the latter is slowly gaining prominence in the field of transport. For the comparison, a random forest classifier is also developed. This comparison shows a coherence between the two approaches, although a discrepancy in the feature importance for gender and travel time is observed. A deeper exploration of this discrepancy highlights the hurdles that often occur when using mathematically more powerful models, such as the random forest classifier.

Keywords: bike sharing; shared mobility; sustainable urban mobility; smart cities; mode choice; mode shift; machine learning

1. Introduction

Bike sharing is defined as the shared use of a bicycle, in which a user accesses a fleet of bicycles offered on public space [1]. Bike-sharing systems (BSSs) have a long history, with the first known scheme launched in 1965 in Amsterdam [2]. Thanks to advances in information and communications technology (ICT), modern BSSs are characterised by wireless pick-up, drop-off, and real-time GPS tracking of the bicycles [3]. Furthermore, today, around 2000 BSS stations are found around the world [4]. The growing popularity of BSSs is supported by the associated social, economic, and environmental benefits, such as decrease in private car ownership, reduction in energy consumption and emissions, lowering of travel costs, lessening of congestion, improvement in public health and creation of environmental awareness, thus helping the cities to move towards sustainable mobility and become smarter [5–7]. Furthermore, BSSs act as an efficient choice for first/last-mile trips [8,9], thereby increasing transit ridership.

Despite their growing popularity, not all BSSs deployed around the world are successful [10]. Hence, for effective planning, there is a need to understand the conditions under

which a BSS will be used, especially with a focus on the influence of socio-demographic characteristics of local citizens. In addition, several studies have hypothesized that the shift towards BSSs comes mainly from sustainable modes of transport rather than private cars [11,12], although reduction in car ownership is found to be an impact of BSS in the literature [13]. Therefore, the perceptions and attitudes of private car users towards BSSs have to be further studied. Thus, the primary research objective of this paper is to identify the factors that influence the shift of car users towards bike sharing, based on a discrete choice model that focuses on the implementation of a new service in the city of Alexandroupolis (No bike-sharing service existed in the city, when this research was conducted. Nevertheless, as on March 2023, a new service is being operated at a small-scale, with around 22 shared bikes and 500 registered users) in Greece A secondary objective of this research is to compare the discrete choice model with a machine learning approach, namely the random forest classifier. This comparison is performed as machine learning models are slowly gaining prominence in the field of transport, with a view that they have better predictive power than the traditional models. The contributions of this research work are the following:

- Identification of factors affecting the shift of car users towards bike sharing;
- Comparison of a discrete choice model with a machine learning approach;
- Derivation of behavioural, policy and other relevant insights

The identified factors can support the design of BSSs, and comparison can help to find the similarities and differences between the two modelling approaches as well as the related implications. Finally, the insights can aid the shift of car users towards a BSS. The remainder of this paper is structured as follows: A summary of the existing literature is presented in Section 2. Subsequently, the methodology (i.e., data collection and model estimation) is presented in Section 3, followed by Section 4 with the data collection and the descriptive statistics of the collected research sample. Then, the modelling results are consolidated in Section 5, along with Section 6, in which the implications of the modelling results are discusses. Finally, the conclusions are elucidated in Section 7.

2. State of the Art

In this section, to support the primary research objective, the literature focusing on the factors influencing the demand for BSSs is initially examined. Subsequent to that, in view of the secondary research objective, the findings from studies comparing logit models with random forest approaches, with a focus on mode choice modelling, are summarised. The studies were collected by querying the Scopus database using an open source Python script (https://github.com/nsanthanakrishnan/Scopus-Query, accessed on 7 March 2023) from Narayanan and Antoniou [14].

2.1. Factors Influencing the Demand for BSSs

In the pertinent literature, the influencing factors of the demand for BSSs are explored using both revealed and stated preference data. For example, based on a stated preference survey, travel time, trip purpose and income are shown by Politis et al. [15] to influence the mode choice between private car and BSS in Thessaloniki. Exploring the mode choice between walking, private car, E-Bikes, car sharing and bike sharing, Li and Kamargianni [16] found out that travel time and travel cost are significant factors for choosing a BSS. In a subsequent study, Li and Kamargianni [17] estimated nested logit models to evaluate modal substitution patterns, and they observed that the conventional modes do not have a nesting effect with the shared mobility services. Furthermore, the coefficients in the mode choice model are observed to differ according to trip distance, which is distinguished as low, medium and long distances.

Narayanan and Antoniou [18] also identified three different distance segments (i.e., <2 km, 2 to 5 km and 5 to 15 km) for the shared mobility services. They used a household survey from the city of Madrid and distinguished the characteristics of the shared mobility services from that of the conventional modes using a discrete choice model. Based on

a review of multiple studies related to BSSs, Fishman et al. [12] identify convenience and cost to be major factors for using BSSs. Cost is also found to be a significant factor in Ma et al. [19], whose conclusion is based on a binary logit model, with the dependent variable being whether or not a survey respondent shifts his/her commuting mode to a BSS.

With regards to socio-demographics, based on a survey of 3000 individuals, Raux et al. [20] observed that the majority of BSS users in Lyon (France) are male and hold higher social positions when compared with the general population. Additionally, the authors found that proximity to the nearest BSS station is a major factor for the use of BSSs. However, based on a stated preference survey conducted in Beijing (China), Campbell et al. [11] conclude that BSSs will draw users from across the social spectrum. Based on an analysis of the trip records of a BSS in Oslo (Norway), Böcker et al. [21] state that gender and age play a significant role for the use of BSSs. The system is found to be less used by women and older age groups.

Focusing on the difference between Millennials', Gen Xers' and Baby Boomers' bike-sharing ridership, Wang et al. [22] also found that most of the bike share trips are made by older Millennials (born between 1979 and 1988). Furthermore, weather factors are shown to have less of an impact on younger Millennials' BSS use. These findings are based on zero-inflated negative binomial models, which were developed using New York's Citi Bike system data. Lee et al. [23] also utilise a zero-inflated negative binomial regression model to analyse the frequency of use of shared mobility services. Variables, such as travel distance, age and gender, are found to influence the use frequency.

Tran et al. [24] use linear regression models to predict station-level BSS demand in Lyon (France). The focus is on both long term and short term subscribers. The authors conclude that long term subscribers use the BSS often together with trains for commuting trips, whilst short term subscribers' trip purposes are more varied. Furthermore, students are determined to be an important group among BSS users. Other factors that influence the demand for BSSs are traffic safety concerns and limitations in the existing cycling infrastructure [5]. The positive influence of supportive cycling infrastructure is also concluded in Shen et al. [25].

To summarise, findings from the literature show that the demand for BSSs is affected by socio-demographic variables (age, income and gender), trip characteristics (trip distance, travel time, travel cost and trip purpose), and attitudes of the individuals (safety concerns). Therefore, for effectively planning the implementation of a BSS in the city of Alexandroupolis (Greece), this research will explore these variables for the mode shift from private car towards BSS.

2.2. Existing Comparisons between Logit Models and Random Forest Classifiers

Looking at the existing literature on the comparison between the two approaches, Zhao et al. [26] conclude that random forest models perform significantly better (in terms of predictive accuracy) than multinomial and mixed logit models. Nevertheless, both approaches agree on several aspects of the behavioural outputs (i.e., variable importance and the direction of association between independent variables and mode choice). However, the random forest model is found to produce behaviourally unreasonable arc elasticities (i.e., the elasticity of the dependent variable with respect to an independent variable between two given points). Thus, there exists a trade-off between predictive accuracy and behavioural soundness when choosing between the two approaches. Hagenauer and Helbich [27] compare seven different classifiers using a two-year Dutch travel diary dataset, and state that the random forest classifier performs better than other models, including neural networks and multinomial logit models. For estimating feature importance, the authors use the permutation importance method described in Altmann et al. [28] and observe that the importance of several variables varies according to the classifier chosen.

On the other hand, based on an exploration of mode choice between car-sharing system and traditional modes, Ceccato et al. [29] conclude that a binary logit model results

in a more reliable prediction than a random forest classifier. Nevertheless, the significant variables in both models are found to be same. Liang et al. [30] perform a comparison of a logit model, a support vector machine model and a random forest classifier using large-scale household mobility survey data collected in Milan. The authors observe that the support vector machine model and the random forest classifier perform marginally better than the multinomial logit model on the full dataset of over 20,000 samples, while the results fluctuate significantly when the sample size is smaller.

Summarising the existing literature on the comparison between logit and random forest approaches, one can clearly see that prediction performance is highly dependent on the analysis context (i.e., the dataset and the type of classification problem). Therefore, it is impossible to formulate a general conclusion that one approach is superior to the other. Given the above finding and that the application of machine learning models, such as the random forest classifier, in the field of transportation is being increasingly observed [31], this research will explore how a binary logit model will perform against a random forest approach for the modelling of mode choice between private car and BSS.

3. Methodology

The overall methodological framework is outlined in Figure 1 and described in the following subsections. Data collection through a stated preference survey is initially performed, followed by a descriptive analysis to understand the research sample. Then, a binary logit model is estimated, followed by the development of a random forest classifier. Subsequently, a comparison is performed between the binary logit model and the random forest classifier.
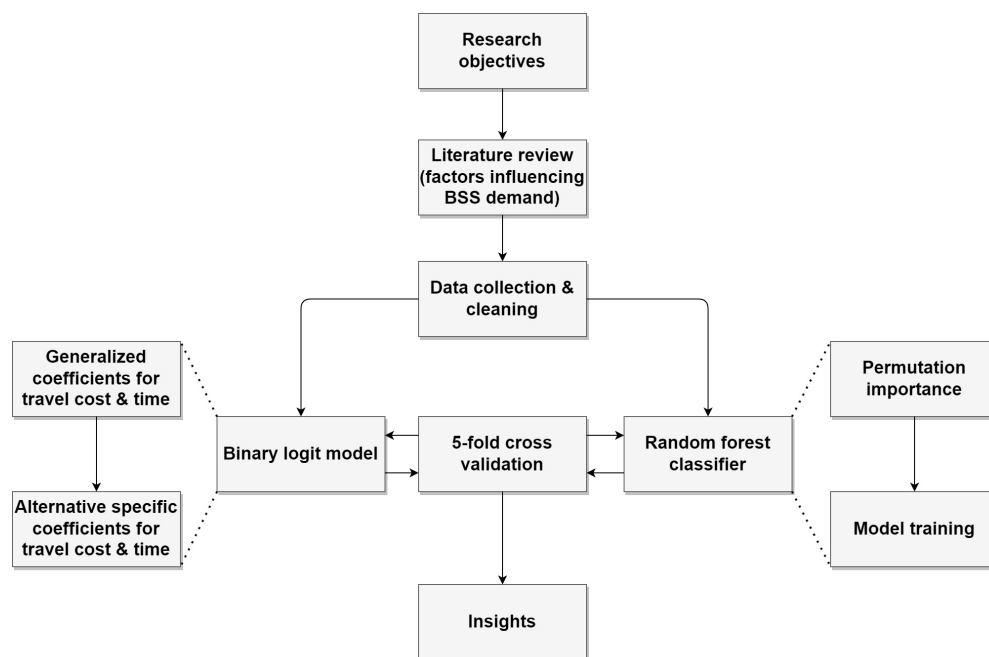

Figure 1. Methodological framework.

3.1. Discrete Choice Model

A binary logit model is selected due to its mathematical simplicity and widespread use in mode choice modelling. Such a model is based on the random utility theory. The decision makers are assumed to exhibit rational economic behaviour, and the choice set does not affect the decision taking process. The attractiveness of each alternative can be represented by a utility that the decision maker will enjoy if they choose it. The (Gumbel) distributions of the error terms associated with the utility function of each alternative in a binary logit model are the same (i.e., the means are zero and the variances are equal) and are independent from each other (i.e., not correlated).

The estimation is carried out using the "mlogit" package [32] in R statistical computing software [33]. The model specification is developed in a stepwise fashion. The decision to keep an independent variable is based on the *p*-value (significance level of 0.10) of the corresponding variable, log-likelihood test and the statistical parameters "AIC" and "BIC". In addition, the selection of variables is based on 5-fold cross-validation [34], with classification accuracy [35] being used as the scoring metric. This implies that, each time, the model is estimated on 80% of the data (train data), and the accuracy is computed on the remaining 20% test set. The procedure is repeated five times to ensure that each data point is present at least once in the test set. The accuracy is computed as an average of all five runs.

At first, a binary logit model with generalised coefficients for travel cost and time is estimated, and then another model with alternative specific coefficients is estimated. This is performed to ascertain whether the individuals perceive the cost and time differently for car and BSS. Following the estimation of a discrete choice model, a machine learning model is developed using the same data.

3.2. Machine Learning Model

A random forest classifier is selected, which utilises many decision trees as an ensemble method, along with a bagging approach to ensure that the individual trees have low correlation [36]. The result is that the approach is relatively stable with respect to noise. Since a random forest model is a non-statistical approach, it makes no formal assumptions of the training data and the prediction residuals. Therefore, it can handle skewed and multimodal data as well as categorical data that are ordinal or non-ordinal. Furthermore, it is capable of handling complex interaction effects between the input variables. Nevertheless, the predictions from each tree must have very low correlations.

For training the random forest classifier, SciKit-Learn [37] is used in a Python environment. Based on cross-validation, the number of trees is selected to be 100 and the maximum depth of each tree is set to 5 to avoid overfitting. All other parameters are kept to their default values, as changing them does not improve the prediction accuracy. The selection of input variables is based on the importance of the variables for the given classification problem. To extract the importance of the input variables, also called feature importance, the permutation importance method [28] is employed. To make the results more stable and obtain less biased results for each feature, this method is run ten times on a test dataset. It is to be noted that, although the random forest classifier does not suffer from correlation issues, the permutation importance method may split importance between two variables if they are correlated. Hence, if the highly correlated variables are found to be important, only one of them is included, and for this, the accuracy based on cross-validation is used.

3.3. Model Comparison and Validation

Finally, a comparison is made between the binary logit model and the random forest classifier to understand the similarities and differences. This comparison is interesting, as the random forest classifier is usually capable of handling more complex relationships than the logit model. For the comparison, the classification accuracy of the final logit and random forest models, based on 5-fold cross-validation, is used. Thus, the cross-validation technique is used for both variable selection and the comparison of the two approaches.

4. Data Collection and Results

4.1. Data Collection

As mentioned in Section 1, the primary objective of this study is to identify the factors influencing the shift of car users towards BSSs. To achieve this objective, a stated preference survey was designed and disseminated to residents of the city of Alexandroupolis in Northern Greece. The survey is divided mainly into two parts. The first part of the survey targeted respondents' socio-demographics and travel behaviour, focusing on the following three different activities: (i) most frequent destination, (ii) leisure and (iii) sports. The

second part of the survey intended to capture the choice between the existing mode choice and the BSS.

The survey was conducted online in the Greek language through Google Forms in the month of September 2020. It was launched under the title "Installing a bike sharing system in the city of Alexandroupolis". Although COVID restrictions were lowered during the time of the survey, the respondents were reluctant to participate in face-to-face questionnaires due to the fear of disease transmission. Each survey respondent was provided with six or seven different scenarios (with different travel times and costs) to quantify their intention to switch from their existing mode to a BSS. It is to be noted that collection of multiple responses from the participants of stated preference surveys is standard practice in the field of transport, which is performed to increase the number of data points for model estimations.

According to the objectives of this study, a subsample of the survey (i.e., the private car users) is used in this study. A preliminary descriptive analysis was performed to understand the sample distribution and the willingness to shift from the existing mode to a BSS. Then, a discrete choice model, with the dependent variable being the choice between private car and BSS, is estimated.

4.2. Descriptive Statistics

After cleaning the dataset for missing and inappropriate values, the research sample consisting of 385 observations for the mode choice between private car and BSS remained. Based on the rule of thumb proposed by McFadden [38] and Pearmain [39], the statement made by Lancsar and Louviere [40] according to their empirical experience, and the equation for minimum sample size suggested in Orme [41], the sample size of 385 observations can be deemed to be sufficient for the current study.

Table 1 summarises the descriptive statistics of the research sample. In this research sample, 44% of the respondents are female and 56% are male, with 7% having a monthly family income between 0 and 400 Euros, 11% between 401 and 800 Euros, 27% between 801 and 1200 Euros, 25% between 1201 and 1600 Euros, 11% between 1601 and 2000 Euros and 19% over 2000 Euros per month. The distribution of the participants' age is as follows: 2% of the participants belong to the age group 18–24, 45% belong to the group 25–39, 44% belong to the group 40–54, 7% belong to the group 55–64 and 2% are over 64 years old. Regarding the employment status, 31% of the participants are self-employed, 24% are state employees, 31% are private employees, 5% are unemployed, 2% are students, and 7% belong to the category "other". Overall, a good representation of gender and income is observed. However, students and individuals aged below 25 and above 64 are not adequately represented. The under-representation of older people can be attributed to the use of an online survey, which was necessitated by the COVID situation.

It is believed that BSSs are used (comparatively) more for leisure trips. Hence, the frequency of performing leisure trips was enquired in the survey. Around 11% of the participants responded that they perform leisure trips every day, 58% two or three times a week, 22% once a week, 7% rarely and 2% never. When asked about how safe the participants think a bike ride is in Alexandroupolis, 18% of the survey respondents stated that it is a high risk, 22% stated that it is risky, 31% were neutral, 24% stated safe and 5% stated that it is very safe. The participants were also asked about how likely they might use a BSS for commuting trips. To this question, 22% of the participants answered extremely unlikely, 18% stated unlikely, 20% were neutral, 9% stated likely and 13% answered extremely likely. A similar distribution is observed for sports and leisure activities, as well as for the trip to the local market.

A further question is related to the implementation of public bicycles in Alexandroupolis, to which 9% of the participants strongly oppose, 9% somewhat oppose, 13% being neutral, 29% somewhat favour and 40% strongly favour. Additionally, the survey participants were asked whether a public bicycle system will help in promoting sustainable urban mobility, and 82% of the respondents answered yes whilst 18% answered no to the question.

The attitude of the research sample shows that the majority of the respondents generally favour the implementation of BSSs.

Table 1. Summary of descriptive statistics.

Gender	Female	44%		18–24	2%
	Male	56%		25–39	45%
Household income (Euros/month)	0–400	7%	Age	40–54	44%
	401–800	11%		55–65	7%
	801–1200	27%		>65	2%
	1201–1600	25%	Bike safety perception	Very safe	5%
	1601–2000	11%		Safe	24%
	>2000	19%		Neutral	31%
Employment status	Self-employed	31%		Risky	22%
	State employees	24%		High Risk	18%
	Private employees	31%	No. of leisure trips per week	Less than 2	31%
	Other	7%		At least 2	69%
	Unemployed	5%	Mode choice	Car	46%
	Student	2%		BSS	54%

5. Estimation Results

In this section, the estimation results for the logit and the random forest models are presented. First, the binary logit model results are described, and the results of the random forest classifier are then elucidated.

5.1. Binary Logit Models

As mentioned in the Methodology section, a binary logit model with generalised coefficients for travel cost and time is first estimated, the results of which are shown on the left side of Table 2. On the right side of the table, the estimation results for a model with alternative specific coefficients are shown. The likelihood ratio test shows that the latter model is better than the former. This result conveys that the individuals perceive cost and time differently for car and BSS. Furthermore, although the time variable for the choice "car" has a negative coefficient, it is found to be insignificant and was hence removed from the final model specification. This conveys the fact that, when comparing a car with a BSS, the car users are not significantly influenced by the time that they would require to travel by car but are, however, sensitive to BSS travel time. The model with alternative specific coefficients is considered as the best model (utility specification shown in Equation (1)), and further exploration of the coefficients will be based on that.

$$
\begin{aligned}
Utility(Car) &= Cost : Car \\
Utility(BSS) &= Intercept + Cost : BSS + Time : BSS + Perception(Bike\ is\ not\ safe) + \\
&\quad Perception(Bike\ is\ safe) + Leisure\ trips \geq\ 2/week + \\
&\quad Household\ income < \text{€}\,1200/month + Employed\ by\ state + Gender(Male)
\end{aligned}
\tag{1}
$$

Table 2. Estimation results for logit models.

Generalised Coefficients for Cost & Time				Alternative Specific Coefficients for Cost and Time			
Variable	Est.	Std. Err.	*p*-Value	Variable	Est.	Std. Err.	*p*-Value
Intercept	−1.249	0.402	0.002 (**)	Intercept	1.677	0.966	0.082 (.)
Cost	−0.026	0.004	<0.001 (***)	Cost:BSS	−0.032	0.005	<0.001 (***)
				Cost:car	−0.018	0.006	0.002 (**)
Time	−0.035	0.017	0.042 (*)	Time:BSS	−0.184	0.060	0.002 (**)
Perception—bike is not safe	−0.707	0.292	0.015 (*)	Perception—bike is not safe	−0.726	0.296	0.014 (*)
Perception—bike is safe	1.093	0.343	0.002 (**)	Perception—bike is safe	1.111	0.346	0.002 (**)
Leisure trips ≥2/week	1.250	0.285	<0.001 (***)	Leisure trips ≥2/week	1.277	0.289	<0.001 (***)
Household income < EUR 1200/month	−0.653	0.260	0.012 (*)	Household income < EUR 1200/month	−0.665	0.263	0.011 (*)
Employed by state	0.965	0.318	0.002 (**)	Employed by state	0.985	0.320	0.002 (**)
Gender—male	0.533	0.254	0.035 (*)	Gender—male	0.544	0.256	0.033 (*)
Cross-validation accuracy	72.1%			Cross-validation accuracy	72.4%		
R^2	0.23			R^2	0.24		
Log-likelihood	−204.74			Log-likelihood	−201.87		
AIC	427.48			AIC	423.75		
BIC	463.06			BIC	463.27		

Note: Car is the base alternative. For the perception variable, neutral is kept as the base category. Cost is expressed in terms of cents and time in terms of minutes. "Time:car" has a negative coefficient; however, it is found to be insignificant. Different categories of the income variable were tested and only the aforementioned dummy, which corresponds to the low income group, resulted in a significant effect. Age is found to be insignificant, which makes sense as the majority of the participants (96%) belong to the age group of usual shared mobility user segment. (.) $p < 0.1$; (*) $p < 0.05$; (**) $p < 0.01$; (***) $p < 0.001$.

Based on the coefficients shown in Table 2, it can be concluded that the respondents of the survey are almost twice as sensitive to the cost of the BSS as they are to the car. Similarly, respondents are sensitive to the BSS travel time but not to the car travel time. It can also be seen that the negative perception of bike safety reduces the probability of using a BSS, whilst the opposite is true if an individual feels that biking is safe, although the former has less impact. On the positive side, individuals who make leisure trips for at least twice a week have significantly higher odds of using the bike-sharing system. Likewise, state employed car users are more likely to shift to a BSS when the system is introduced. However, having a household income of less than 1200 Euros per month reduces the odds of using the BSS. Looking at the gender aspect, males are more probable to use the BSS.

5.2. Random Forest Model

The final model with the highest accuracy uses identical variables to the logit model. The average accuracy is 74.8%, i.e., 2.4% higher than the logit model. The improved accuracy could be attributed to the fact that the random forest classifier is capable of modelling non-linear effects in the dataset. However, considering that the accuracy only improves by 2.4%, it is possible that the non-linear effects are not substantial.

In Table 3, the permutation feature importance of the variables in the best random forest classifier is shown, together with the standard deviation across the permutation runs. The higher the permuted mean in accuracy, the more important the feature is. On the other hand, the higher the standard deviation, the lower the confidence regarding that feature's importance.

Table 3. Results for the random forest model.

Variable	If Permuted, Mean Loss in Accuracy (in Percentage Points)—Relative Importance	Standard Deviation Around Mean (in Percentage Points) —Uncertainty of Importance
cost:BSS	8.3%	3.2%
cost:car	1.8%	1.8%
time:BSS	0.7%	1.2%
Perception—bike is not safe	3.6%	2.8%
Perception—bike is safe	3.3%	1.1%
Leisure trips $\geq$ 2/week	4.9%	3.7%
Household income <EUR 1200/month	2.5%	2.0%
Employed by state	2.7%	2.0%
Gender—male	0.6%	0.5%
Cross-validation accuracy	74.8%	

Comparing the relative importance of variables in the random forest classifier and the logit model, one can observe that the cost of the BSS and making leisure trips at least twice per week are highly important variables in both the cases. Bike safety is also seen to be an important feature in both models, however, the negative perception has a higher importance in the case of the random forest model. The cost associated with a car trip and having a monthly income of less than 1200 Euros per month are also important in both models.

In contrast to the logit model, the random forest model shows a low importance for BSS travel time. Investigating this further, it was found that the random forest classifier can approximate BSS travel time using the variables "cost:car" and "cost:BSS". As "time:BSS", "cost:BSS" and "cost:car" only have a few unique data points, the random forest classifier overfits to "cost:car" and "cost:BSS" and does not require "time:BSS". Thus, the importance of the feature "time:BSS" is very low. Although we are not able to provide a certain explanation at this point, we postulate that the issue could be an artefact of the survey design. The stated preference surveys are usually designed for discrete choice models and not with a focus on machine learning models, such as random forest classifier, and hence the survey design might have caused the aforementioned issue. A similar approximation issue is also found for the gender variable. A separate random forest model with gender as the output variable was trained, which shows that income, bike safety perception and the dummy for conducting at least two leisure trips per week can be used to classify gender with a 69% accuracy.

6. Discussion

6.1. Insights for Policymakers

No bike-sharing service existed in Alexandroupolis when this research was performed. The discussion in this section can support the preparation of a roadmap for the implementation of a new service in Alexandroupolis (as well as in other similar cities) and also serve as a starting point to enhance the current cycling infrastructure in the city.

Based on the coefficients of the logit model, which is supported by the results of the random forest classifier, it can be concluded that individuals are more sensitive towards the cost of the BSS than the cost of car trips. This implies that for successful implementation of a BSS, the user cost should be kept low. Based on Figure 2, in case of a trip of 15 min costing 50 cents to use a private car, the cost of the bike-sharing system should be less than 28 cents for a women (who is not a low income individual, not a state employee and feels biking is safe) to achieve a probability of greater than 0.5 to shift towards the service. Within this regard, a subsidization scheme, similar to the ones implemented for public transport (PT) in many cities around the world, would be beneficial. In particular, given that a BSS is shown to increase transit ridership in the existing literature, the development of a subsidization scheme for PT pass holders is suggested, and the two systems have to be integrated to attract more users. Nevertheless, with a higher penetration, it might also be possible to lower the cost without subsidies in the future. On the other hand, the cost for car trips could be increased by the implementation of road pricing schemes and higher parking fees.
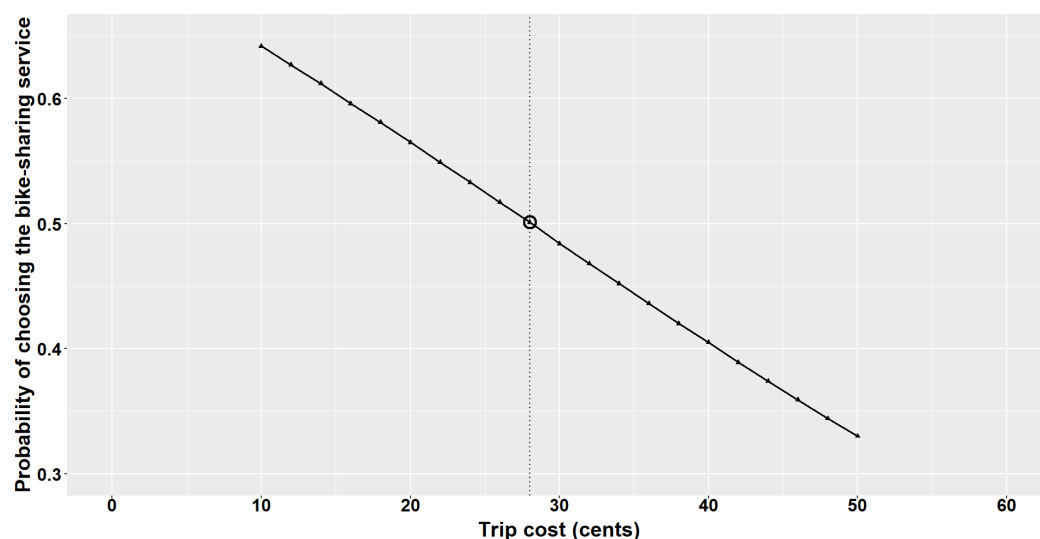
Note: Example considered is a women who is not a low income individual, not a state employee and feels biking is safe

Figure 2. Change in the probability to shift towards the bike-sharing service according to the cost of the service for a trip of 15 min costing 50 cents to use a private car.

Concerning travel time, the estimation results show that the car users are highly sensitive to BSS travel time. Hence, a substantial travel time difference between car and BSS (BSS being quicker) is a necessity. This means that a BSS will be preferred in areas where there are access restrictions for cars (i.e., car users have to detour significantly) and the parking places are few (requiring substantial cruising to find a parking spot). Nevertheless, such measures should be carefully planned and introduced. If not, such regulations may lead to diminished turnover to the local shop and restaurant owners. On a different note, the competitiveness of BSSs can be improved by the introduction of shortcuts for bicycles, e.g., cycle paths through parks. The positive impact of shortcuts for bicycles on travel time difference between cars and cycles has already been shown in the literature [42].

Perception of bike safety is an important factor for the mode choice between car and BSS. Therefore, there is a need to establish a positive perception of bike safety among the citizens. Alexandroupolis is almost flat and, therefore, easily bikeable. Nevertheless, the current dedicated cycling network in the city extends for only about 15 km, covering the main points of interest in the city (in particular, the seaside city, providing a pleasant and unique experience for the riders). Thus, suggested measures include the improvement of cycling infrastructure (e.g., implementation of more dedicated cycle lanes and extension

of the current cycling network) and creation of bike safety campaigns. Techniques from the field of growth hacking [43,44] can be helpful. Moreover, provision of additional accessories, such as helmets, is also important to increase the safety perception [45].

Given that state employees and individuals who frequently perform leisure trips are more likely to use a BSS, they can be targeted for early adoption, reducing the initial reservation that usually exists for the introduction of any new system.

Since the individuals from households with monthly income below EUR 1200 are less likely to switch to a BSS, a financial motivation scheme has to be designed to attract such individuals. An example of such a scheme is to get paid for using a BSS to commute to the office. This scheme could be introduced by companies, and in fact, similar schemes (i.e., getting paid for cycling to companies) already exist [46]. Such financial schemes may also nudge more females to use a BSS. Nevertheless, awareness campaigns may have to be carried out to disseminate the benefits of cycling and attract more companies to initiate such schemes. Similarly, local authorities can initiate educational campaigns targeting the citizens to share information about what is on offer and how to use it [47].

To allow a better overview of the insights discussed in this section, the influencing factors identified in the previous section have been grouped and summarised in Table 4 according to policy measures.

Table 4. Policy measures based on identified factors.

Policy Measure	Relevant Factors	Recommendation
Finance	cost:BSS	BSS subsidization scheme, similar to the ones implemented for PT, especially for PT pass holders.
	Household income <EUR 1200/month, Gender—male	Financial motivation scheme, e.g., monetary incentive for using BSS to commute to the office.
Regulation	cost:car	Implementation of road pricing schemes for car use.
	time:BSS	Access restrictions for cars, thereby making BSS more competitive in terms of travel time.
Infrastructure	Perception of bike safety	Improvement of cycle infrastructure, e.g., implementation of dedicated cycle lanes.
	time:BSS	Introduction of shortcuts for cycles, e.g., cycle paths through parks.
Campaigns	Perception of bike safety	Creation of bike safety campaigns. Techniques from the field of growth hacking can be helpful.
	Household income <EUR 1200/month, Gender—male	Awareness campaigns targeted at companies, to disseminate the benefits of cycling, thereby stimulating them to provide financial incentives to their employees for using BSS.
Customer targeting	Leisure trips ≥ 2/week, employed by state	Target state employees and individuals who frequently perform leisure trips for early adoption.

6.2. Insights for Modellers

It is common to use a generalised coefficient for cost when estimating a logit model for mode choice. The justification for such a use is the following: a Euro is perceived to be a Euro, be it for one mode or the other. However, the results presented in Table 2 clearly show that the cost is perceived differently for car and BSS. Hence, care should be taken by future researchers when modelling mode choice.

The comparison of the cross-validation accuracy of the logit and the random forest models show that the latter performs only slightly better, which is also observed

in Ceccato et al. [29]. Hence, the presence of non-linear interactions is very low or nil, supporting the underlying linear assumptions of the binary logit model. The logit model provides the key interpretable results, whereas the random forest classifier validates that no complex effect that influence the mode choice is missed. While the application of machine learning models, such as the random forest classifier, are increasingly observed in the literature [31], the results from this research show that simple interpretable models, such as logit models, are still effective for certain cases.

The discrepancy in the feature importance, for both gender and travel time, highlights the different hurdles that often occur when using mathematically more powerful models, such as the random forest classifier. Stated preference surveys are usually designed with few levels of factors, aiming at the development of discrete choice models, and thus one should be careful when using random forest classifier for such a dataset.

7. Conclusions

Bike-sharing systems (BSSs), when properly implemented, can lead to positive impacts in smart cities and improve living conditions. While the popularity of BSSs is growing, not all BSSs deployed around the world are successful. The existing literature points out that the choice of using BSSs is influenced by socio-demographic variables (e.g., income and gender), trip characteristics (e.g., travel time, travel cost and trip purpose), and attitudes of the individuals (e.g., safety concerns). Furthermore, even though a reduction in car ownership is found to be an impact of BSSs in the literature, there are also studies hypothesizing that the shift towards BSSs is mainly from sustainable modes of transport rather than from private cars. Therefore, the perceptions and the attitudes of private car users towards BSSs have to be studied. Hence, this research identifies the factors that influence the shift of car users to BSSs based on a discrete choice model (binary logit model) that focuses on the implementation of a new service in the city of Alexandroupolis (Greece). In particular, the results from this study reflect the citizens' choices subsequent to a major pandemic lockdown. In addition, this research also compares the discrete choice model with a machine learning approach, namely the random forest classifier, to find the similarities and differences and to ascertain the implications.

Assessing the significant factors from the logit model, it can be concluded that the cost of BSSs plays a major role in the (reluctance to) shift from private cars. Car users are found to be twice as sensitive to the cost of BSSs than to that of cars. Therefore, subsidies for BSS use can be effective in increasing the penetration of the system, which is a pull measure. Nevertheless, with higher penetration levels, it might also be possible in future to lower the cost without subsidies. On the other hand, a push measure could be to introduce road pricing schemes for car use. In addition to the cost of BSSs, the travel time also plays a major role, and car users are highly sensitive to BSS travel time. Hence, shortcuts for bicycles (e.g., cycle paths through parks) to decrease the travel time can be helpful. The positive impact of shortcuts for bicycles has already been shown in the literature [42]. In order to modify users' habit of using private cars, vehicle access restrictions can be implemented, thereby causing the car users to detour significantly and experience higher travel times.

Apart from the cost and travel time factors, the shift of car users towards BSSs also depends on the perception of bike safety. Suggested measures in this regard include improvement of cycle infrastructure (e.g., implementation of dedicated cycle lanes) and creation of bike safety campaigns. To reduce the reservation that may exist during the initial stages of a BSS, state employees and individuals who frequently perform leisure trips can be targeted as early adopters. To attract individuals from households with monthly income below EUR 1200, a financial motivation scheme may have to be designed. An example of such a scheme is to provide financial benefits for people using the BSS to commute to the office, which is already showed to be advantageous [46].

This study includes a comparison between the estimated logit model and the random forest classifier, since machine learning models are slowly gaining prominence in the field of transport, with a view that they have better predictive power than the traditional models.

The comparison shows that the random forest classifier performs only slightly better, which shows that the presence of non-linear interactions is very low, supporting the underlying linear assumption of the binary logit model. This demonstrates that simple interpretable models, such as logit models, are still efficient for specific use cases. Nevertheless, a discrepancy in the feature importance between the two models for gender and travel time highlight the different hurdles that often occur when using mathematically more powerful models, such as the random forest classifier. Stated preference surveys are usually designed with few levels of factors, aiming at the development of discrete choice models, and one should be careful when using random forest classifiers for such datasets.

Future research should include the exploration of impacts of the suggested measures on the penetration of BSSs. The examination of the influence of students is limited in this study, because of their lower representation. Thus, further investigation is suggested. Another interesting direction for future research is the comparison of logit models with other machine learning approaches. Besides, the existence of a distinction in the survey design requirements between discrete choice and machine learning models needs to be further analysed. Though the sample size of 385 responses is sufficient for satisfactory analysis, a larger sample would have led to a richer analysis.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike information criterion
BIC	Bayesian information criterion
BSS	bike-sharing systems
ICT	information and communications technology

References

1. Büttner, J.; Petersen, T. Optimising Bike Sharing in European Cities: A Handbook. 2011. Available online: http://mobility-workspace.eu/wp-content/uploads/OBIS_Handbook_EN.pdf (accessed on 7 March 2023).
2. Shaheen, S.A.; Guzman, S.; Zhang, H. Bikesharing in Europe, the Americas, and Asia. *Transp. Res. Rec. J. Transp. Res. Board* **2010**, *2143*, 159–167. https://doi.org/10.3141/2143-20.
3. Shaheen, S.A.; Cohen, A.P.; Martin, E.W. Public Bikesharing in North America. *Transp. Res. Rec. J. Transp. Res. Board* **2013**, *2387*, 83–92. https://doi.org/10.3141/2387-10.

4.   Meddin, R.; DeMaio, P.; O'Brien, O.; Rabello, R.; Yu, C.; Seamon, J.; Benicchio, T.; Han, D.; Mason, J. The Bike-Sharing World Map. 2022. Available online: https://bikesharingworldmap.com (accessed on 7 March 2023).

5.   Bakogiannis, E.; Siti, M.; Tsigdinos, S.; Vassi, A.; Nikitas, A. Monitoring the first dockless bike sharing system in Greece: Understanding user perceptions, usage patterns and adoption barriers. *Res. Transp. Bus. Manag.* **2019**, *33*, 100432. https://doi.org/10.1016/j.rtbm.2020.100432.

6.   Fontes, T.; Arantes, M.; Figueiredo, P.V.; Novais, P. A Cluster-Based Approach Using Smartphone Data for Bike-Sharing Docking Stations Identification: Lisbon Case Study. *Smart Cities* **2022**, *5*, 251–275. https://doi.org/10.3390/smartcities5010016.

7.   Zhang, Y.; Mi, Z. Environmental benefits of bike sharing: A big data-based analysis. *Appl. Energy* **2018**, *220*, 296–301. https://doi.org/10.1016/j.apenergy.2018.03.101.

8.   Chen, Z.; van Lierop, D.; Ettema, D. Dockless bike-sharing systems: what are the implications? *Transp. Rev.* **2020**, *40*, 333–353. https://doi.org/10.1080/01441647.2019.1710306.

9.   Lu, M.; Hsu, S.C.; Chen, P.C.; Lee, W.Y. Improving the sustainability of integrated transportation system with bike-sharing: A spatial agent-based approach. *Sustain. Cities Soc.* **2018**, *41*, 44–51. https://doi.org/10.1016/j.scs.2018.05.023.

10. Hamann, T.K.; Güldenberg, S.; Renzl, B. Overshare and collapse: How sustainable are profit-oriented company-to-peer bike-sharing systems? *Die Unternehm.* **2019**, *73*, 345–373. https://doi.org/10.5771/0042-059X-2019-4-345.

11. Campbell, A.A.; Cherry, C.R.; Ryerson, M.S.; Yang, X. Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. *Transp. Res. Part Emerg. Technol.* **2016**, *67*, 399–414. https://doi.org/10.1016/j.trc.2016.03.004.

12. Fishman, E.; Washington, S.; Haworth, N. Bike Share: A Synthesis of the Literature. *Transp. Rev.* **2013**, *33*, 148–165. https://doi.org/10.1080/01441647.2013.775612.

13. Basu, R.; Ferreira, J. Planning car-lite neighborhoods: Does bikesharing reduce auto-dependence? *Transp. Res. Part Transp. Environ.* **2021**, *92*, 102721. https://doi.org/10.1016/j.trd.2021.102721.

14. Narayanan, S.; Antoniou, C. Electric cargo cycles—A comprehensive review. *Transp. Policy* **2022**, *116*, 278–303. https://doi.org/10.1016/j.tranpol.2021.12.011.

15. Politis, I.; Fyrogenis, I.; Papadopoulos, E.; Nikolaidou, A.; Verani, E. Shifting to Shared Wheels: Factors Affecting Dockless Bike-Sharing Choice for Short and Long Trips. *Sustainability* **2020**, *12*, 8205. https://doi.org/10.3390/su12198205.

16. Li, W.; Kamargianni, M. Providing quantified evidence to policy makers for promoting bike-sharing in heavily air-polluted cities: A mode choice model and policy simulation for Taiyuan-China. *Transp. Res. Part Policy Pract.* **2018**, *111*, 277–291. https://doi.org/10.1016/j.tra.2018.01.019.

17. Li, W.; Kamargianni, M. Steering short-term demand for car-sharing: A mode choice and policy impact analysis by trip distance. *Transportation* **2020**, *47*, 2233–2265. https://doi.org/10.1007/s11116-019-10010-0.

18. Narayanan, S.; Antoniou, C. Shared mobility services towards Mobility as a Service (MaaS): What, who and when? *Transp. Res. Part Policy Pract.* **2023**, *168*. https://doi.org/10.1016/j.tra.2023.103581.

19. Ma, X.; Yuan, Y.; van Oort, N.; Hoogendoorn, S. Bike-sharing systems' impact on modal shift: A case study in Delft, the Netherlands. *J. Clean. Prod.* **2020**, *259*, 120846. https://doi.org/10.1016/j.jclepro.2020.120846.

20. Raux, C.; Zoubir, A.; Geyik, M. Who are bike sharing schemes members and do they travel differently? The case of Lyon's "Velo'v" scheme. *Transp. Res. Part Policy Pract.* **2017**, *106*, 350–363. https://doi.org/10.1016/j.tra.2017.10.010.

21. Böcker, L.; Anderson, E.; Uteng, T.P.; Throndsen, T. Bike sharing use in conjunction to public transport: Exploring spatiotemporal, age and gender dimensions in Oslo, Norway. *Transp. Res. Part Policy Pract.* **2020**, *138*, 389–401. https://doi.org/10.1016/j.tra.2020.06.009.

22. Wang, K.; Akar, G.; Chen, Y.J. Bike sharing differences among Millennials, Gen Xers, and Baby Boomers: Lessons learnt from New York City's bike share. *Transp. Res. Part Policy Pract.* **2018**, *116*, 1–14. https://doi.org/10.1016/j.tra.2018.06.001.

23. Lee, S.; Smart, M.J.; Golub, A. Difference in travel behavior between immigrants in the US and US born residents: The immigrant effect for car-sharing, ride-sharing, and bike-sharing services. *Transp. Res. Interdiscip. Perspects* **2021**, *9*, 100296. https://doi.org/10.1016/j.trip.2020.100296.

24. Tran, T.D.; Ovtracht, N.; d'Arcier, B.F. Modeling Bike Sharing System using Built Environment Factors. *Procedia CIRP* **2015**, *30*, 293–298. https://doi.org/10.1016/j.procir.2015.02.156.

25. Shen, Y.; Zhang, X.; Zhao, J. Understanding the usage of dockless bike sharing in Singapore. *Int. J. Sustain. Transp.* **2018**, *12*, 686–700. https://doi.org/10.1080/15568318.2018.1429696.

26. Zhao, X.; Yan, X.; Yu, A.; Van Hentenryck, P. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behav. Soc.* **2020**, *20*, 22–35. https://doi.org/10.1016/j.tbs.2020.02.003.

27. Hagenauer, J.; Helbich, M. A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Syst. Appl.* **2017**, *78*, 273–282. https://doi.org/10.1016/j.eswa.2017.01.057.

28. Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340–1347. https://doi.org/10.1093/bioinformatics/btq134.

29. Ceccato, R.; Chicco, A.; Diana, M. Evaluating car-sharing switching rates from traditional transport means through logit models and Random Forest classifiers. *Transp. Plan. Technol.* **2021**, *44*, 160–175. https://doi.org/10.1080/03081060.2020.1868084.

30. Liang, L.; Xu, M.; Grant-Muller, S.; Mussone, L. Household travel mode choice estimation with large-scale data—An empirical analysis based on mobility data in Milan. *Int. J. Sustain. Transp.* **2019**, *15*, 70–85. https://doi.org/10.1080/15568318.2019.1686782.

31. Hillel, T.; Bierlaire, M.; Elshafie, M.Z.; Jin, Y. A systematic review of machine learning classification methodologies for modelling passenger mode choice. *J. Choice Model.* **2021**, *38*, 100221. https://doi.org/10.1016/j.jocm.2020.100221.

32. Croissant, Y. Estimation of Random Utility Models in R: The mlogit Package. *J. Stat. Softw.* **2020**, *95*, 1–41. https://doi.org/10.18637/jss.v095.i11.

33. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.

34. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer Texts in Statistics, Springer: New York, NY, USA, 2013; Volume 103. https://doi.org/10.1007/978-1-4614-7138-7.

35. Hossin, M.; Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations. *Int. J. Data Min. Knowl. Manag. Process.* **2015**, *5*, 1–11. https://doi.org/10.5121/ijdkp.2015.5201.

36. Dietterich, T.G. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Mach. Learn.* **2000**, *40*, 139–157. https://doi.org/10.1023/A:1007607513941.

37. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

38. McFadden, D.L. Econometric analysis of qualitative response models. In *Handbook of Econometrics*; Elsevier: Amsterdam, The Netherlands, 1984; pp. 1395–1457. https://doi.org/10.1016/S1573-4412(84)02016-X.

39. Pearmain, D. *Stated Preference Techniques: A Guide to Practice*, 2nd ed.; Steer Davies Gleave: London, UK, 1991.

40. Lancsar, E.; Louviere, J. Conducting discrete choice experiments to inform healthcare decision making: A user's guide. *Pharm. Econ.* **2008**, *26*, 661–677. https://doi.org/10.2165/00019053-200826080-00004.

41. Orme, B. Sample size issues for conjoint analysis studies. In *Sequim: Sawtooth Software Technical Paper*. LLC: Manhattan Beach, CA, USA, 1998.

42. Gruber, J.; Narayanan, S. Travel Time Differences between Cargo Cycles and Cars in Commercial Transport Operations. *Transp. Res. Rec. J. Transp. Res. Board* **2019**, *2673*, 623–637. https://doi.org/10.1177/0361198119843088.

43. Bohnsack, R.; Liesner, M.M. What the hack? A growth hacking taxonomy and practical applications for firms. *Bus. Horizons* **2019**, *62*, 799–818. https://doi.org/10.1016/j.bushor.2019.09.001.

44. Herttua, T.; Jakob, E.; Nave, S.; Gupta, R.; Zylka, M.P. Growth Hacking: Exploring the Meaning of an Internet-Born Digital Marketing Buzzword. In *Designing Networks for Innovation and Improvisation*; Zylka, M.P., Ed.; Springer Proceedings in Complexity, Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 151–161. https://doi.org/10.1007/978-3-319-42697-6_15.

45. Turoń, K. Complaints Analysis as an Opportunity to Counteract Social Transport Exclusion in Shared Mobility Systems. *Smart Cities* **2022**, *5*, 875–888. https://doi.org/10.3390/smartcities5030044.

46. Hu, E. By paying staff to cycle to work, companies benefit from more savings, less sick leave and improved satisfaction. *Bus. Insid.* **2018**. Available online: https://www.businessinsider.com/heres-how-companies-paying-staff-to-cycle-to-work-end-up-saving-money-2018-7 (accessed on 7 March 2023).

47. Schaefer, C.; Stelter, A.; Holl-Supra, S.; Weber, S.; Niehaves, B. The Acceptance and Use Behavior of Shared Mobility Services in a Rural Municipality. *Smart Cities* **2022**, *5*, 1229–1240. https://doi.org/10.3390/smartcities5040062.