

Article

# The Classification of Application Users Supporting and Facilitating Travel Mobility Using Two-Step Cluster Analysis

Jaroslav Mašek<sup>1</sup>, Vladimíra Štefancová<sup>1,\*</sup> , Jaroslav Mazanec<sup>2</sup>  and Petra Juránková<sup>3</sup>

<sup>1</sup> Department of Railway Transport, Faculty of Operation and Economics of Transport and Communications, University of Zilina, Univerzitna 1, 01026 Zilina, Slovakia

<sup>2</sup> Department of Quantitative Methods and Economic Informatics, Faculty of Operation and Economics of Transport and Communications, University of Zilina, Univerzitna 1, 01026 Zilina, Slovakia

<sup>3</sup> OLTIS Group a. s., Dr. Milady Horákové 1200/27a, 779 00 Olomouc, Czech Republic

\* Correspondence: vladimira.stefancova@fpedas.uniza.sk

**Abstract:** There is a significant and supported trend toward the achievement of ensuring continuous door-to-door travel in the pan-European transport network. Many innovative programs are dedicated to this topic through assigned projects. This paper is based on the concrete partial results of the H2020 project Shift2Rail IP4 to support the deployment of mobility as a service (IP4MaaS). Attitudes towards travel for demonstration sites were assessed based on the outputs of a sample of respondents from two countries. Cooperation in working on the IP4MaaS project was also provided by a partner from Slovakia (UNIZA) and the Czech Republic (OLTIS). Mathematical statistical tools were used to evaluate the available data to find a connection with promoting mobility as a service. This paper aims to identify differences in travelers' needs with a focus on using applications using two-step cluster analysis. The research resulted in the identification of differences in traffic behavior within MaaS activities when comparing different clusters reflecting preferences for using a website or mobile application.

**Keywords:** mobility; cluster analysis; MaaS activities; H2020

**MSC:** 37N40; 62H30; 91C20



**Citation:** Mašek, J.; Štefancová, V.; Mazanec, J.; Juránková, P. The Classification of Application Users Supporting and Facilitating Travel Mobility Using Two-Step Cluster Analysis. *Mathematics* **2023**, *11*, 2192. <https://doi.org/10.3390/math11092192>

Academic Editors: Hongzhi Wang and Ye Chen

Received: 2 March 2023

Revised: 26 April 2023

Accepted: 5 May 2023

Published: 6 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The door-to-door strategy is gradually coming to the fore, even with the idea of providing continuous transport services connecting several operators within one digital platform. Promoting door-to-door mobility is a very current topic that is being studied in numerous studies. Research in Dresden focused on identifying the needs and preferences of public transport users for mobile applications to support mobility [1]. As part of transport planning, the decision of transport routes and the inclusion of private door-to-door services to ensure transport service is taken as a key element [2]. In Singapore, a public transport routing approach was assessed, including door-to-door shared services [3]. The option of continuous traffic service also applies to the reduced-mobility population. Research in Barcelona highlighted the importance of door-to-door transportation services even for disadvantaged road users, emphasizing the influence of road characteristics and the need for priority-based trip allocation [4]. Research in Turkey examined the impact of the availability of transport infrastructure and the attractiveness of the region on regional mobility [5].

Digitization in general provides a wide range of available data, which makes it possible to understand and link user behavior. Processing data from multiple communication systems and interconnected subsystems helps in the development of smart cities [6]. In the context of the design of a smart city, the interaction between mobile applications and transport itself is especially important. When setting up a functional system, a process

consisting of a sufficient amount of traffic data, map documents, and outputs from real-time sensors or transport operators is considered [7]. The identification of factors influencing urban tourists' receptivity as well as citizens traveling for work or school is an important input when assessing mobility [8]. The effective operation of public transport systems is closely related to a detailed understanding of the behavior of its participants. In Hong Kong, research focused on evaluating spatial, temporal, modal, and targeted user group parameters based on data from smart cards [9]. Knowing the social demographics of passengers is the basis for the future design of transport services. This topic was addressed by a study that used data extracted from traffic cards to understand and assess the travel habits of different population groups [10]. A study in Canada focused on assessing the impact of transport mobility benefits on older non-working citizens [11]. Business trips have a significant impact on the increased demand for transport services. A Swedish study tested the interaction of the application of mobility services with the reservation and implementation of business trips in the context of public transport as well [12]. Part of the provision of transport is also the offer of shared mobility services in urban areas. Understanding the decision-making process when choosing public transport or alternative shared means plays a significant role in setting up a mobility plan [13]. A French study evaluated mobility with lower density, and the results showed that it is in such cases that carpooling and walking in conjunction with mobile apps are most applied [14]. Shifting traffic in cities to public transport services ensuring mobility is also closely related to a properly set parking policy, especially in city centers [15]. The choice of transport mode is also connected with the assessment of safety. Travel users perceive the provision of safe driving very sensitively and have high expectations for intelligent transport systems [16,17].

Cluster analysis is widely used in various areas, including the transport sector. In a Portuguese urban area, cluster analysis was used to investigate the segmentation of public transport users and their perception of satisfaction with transport services, revealing four user segments [18]. London's public transport network was an area of research that considered the heterogeneity of commuters and assessed the diversity of urban residents as well as temporal attributes within days and sequences of activities [19]. Another English study focused on the approach to the analysis of travel workflows based on geodemographic classification [20]. The recognition of the important predictor was identified through cluster analysis in traffic congestion, while the factors influencing the management, flow, and functioning of traffic were outlined [21]. The impact of transport on carbon dioxide emissions in Chinese cities was also examined using a two-stage cluster analysis, which was divided into various categories based on the degree of impact of each driver [22]. Another study focused on traffic from the point of view of motorcyclists, and segmented motorcycle crashes into homogeneous clusters through cluster analysis [23].

Mobility as a service (MaaS) can be expressed as a tool to achieve sustainable mobility and increase the share of trips by public transport [24]. It is based on the idea of access to a centralized platform for planning, payment, and travel management along with the combination of several types of transport [25]. From another point of view, MaaS emphasizes the need to focus on finding the most acceptable way of moving and deciding to make that move [26]. The term MaaS represents a tool to support mobility. Numerous studies are devoted to this topic. The London study pointed to the potential of the mobility as a service package from the point of view of supporting shared modes. More than half of the respondents confirmed that they would be interested in trying new modes of transport and, thus, supporting travel by shared modes [27]. MaaS is defined as a user-oriented service concept providing people with door-to-door mobility solutions. The importance of future planned bus transport in the context of MaaS appears to be an essential element in strengthening public transport [28]. The possibility of unified search, reservation, and payment through a digital platform helps to promote mobility. Even though it focuses more on passenger transport, the idea of integration of freight transport was also investigated, as the means of freight transport affects the capacity of road traffic and, thus, affects mobility [29]. The MaaS scheme operates in various industries, and

it is described in many studies. Another London study focused on examining newly existing mobility services such as car sharing or bike sharing and their impact on urban mobility [30]. The legislative framework within the digital market in the context of MaaS is addressed by research that pointed to the need to harmonize the legal framework for personal multimodal transport [31]. A Dutch study also addressed the topic of MaaS implementation and its impact on passengers' transportation, in which five different clusters were identified concerning individuals' inclinations [32]. An Australian study assessed consumer preferences, where willingness to use MaaS was shown to be dependent on age and life stage [33]. A German study addressed MaaS from the perspective of identifying key motivational determinants and their interrelationships [34].

The MaaS initiative provides a comprehensive approach to solving urban mobility through a single interface, with technical support and a journey planner as desirable elements [35]. From the results of research on the use of MaaS in Metro Manila, it was concluded that the respondents highlighted the reliability and cost savings, and about 80% of them would use the MaaS application [36]. A Belgian pilot study looked at the possibilities of replacing the car with the effects of MaaS, with findings indicating the impact of MaaS on car ownership and use [37]. Another study on MaaS focused on mobility as a service preferences in the context of understanding potential demand from the perspective of different subscription options [38]. The international Delphi study gathered the opinions of experts on the future implementation of MaaS, in which their attitudes and reactions to vulnerabilities as well as opportunities were considered [39]. The impact of shared mobility on MaaS was addressed in a case study in Madrid, where emerging shared mobility operators and their provided transport services were investigated [40]. Travel behavior and the connection to MaaS were addressed in another study, where the preference for public transport from the perspective of passengers with a motor vehicle was compared [41]. The need for innovation and the provision of public benefits to the traveling public was highlighted as a significant prerequisite for the future preference for MaaS [42].

Solving key social trends such as reducing greenhouse gas emissions or solving problems with congested roads are positive effects that are expected. Shift2Rail IP4 supports MaaS routing to offer transport services and specific route settings directly with individual operators through a digital platform. The IP4MaaS project aimed to support the deployment of MaaS schemes by testing technologies developed under IP4 within Shift2Rail through complementary IP4MaaS projects through demonstration implementation. This project involved 26 participants from eight countries (Belgium, Spain, Greece, Italy, Poland, Croatia, the Czech Republic, and Slovakia) to achieve jointly set goals. Project activities between partners were strengthened by mutual co-creation such as brainstorming, workshops, and imaginative activities with the aim of harmonizing opinions and achieving set goals through project demonstrations [43].

## 2. Materials and Methods

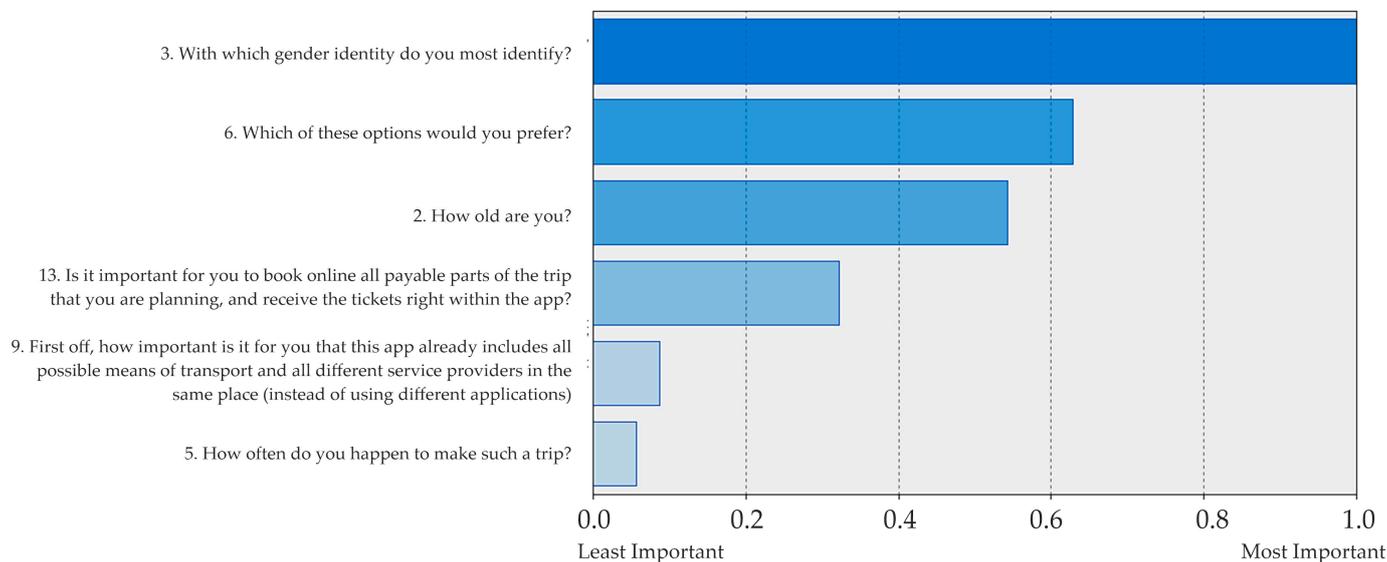
The data from our questionnaire survey led to us obtaining the needs and expectations of potential users within MaaS activities. We used a conversational survey using the Coney tool, while the questions were asked via chat. It was easy to answer the list of questions in the form of an online interview through this platform. The original English version was translated into several languages, while our partial survey used the Czech language. The survey was anonymous and took place from 1 March 2022 to the end of April [43]. The questionnaire was answered by respondents over the age of 18. The questionnaire was divided into two sections: a socio-demographic section (age, gender, social status) and an application section focusing on functionality and user-friendliness. We segmented users using two-step cluster analysis in the statistical analytical program SPSS 26. This program was used to calculate the relative importance of the predictor (see Figure 1) based on the formula below.

$$VI_i = \frac{-\log_{10}(\text{sig}_i)}{\max_{j \in \Omega} (-\log_{10}(\text{sig}_j))} \quad (1)$$

where:

$\Omega$  is a set of predictor and evaluation fields;

$sig_i$  is a p-value computed by applying a certain test. If  $sig_i$  equals 0, set  $sig_i = \text{MinDouble}$ , where  $\text{MinDouble}$  is the minimal double value [44].



**Figure 1.** Predictor importance of all inputs.

Clustering is a known way of segmentation in many areas, including k-means, hierarchical clusters, and two-step clusters. Two-step cluster analysis is a different algorithm from traditional clustering techniques for handling categorical and continuous variables, determining the automatic selection of the number of clusters, and scalability. In other words, the method uses categorical and continuous variables [45]. Two-step cluster analysis is useful when the dataset has a complex structure or when there is no clear prior knowledge about the number of clusters. This approach provides a flexible and effective way to uncover the underlying structure in the data and generate meaningful insights.

Distance measures explain how the similarity between two clusters is computed using log-likelihood and Euclidean distance. Euclidean measures only use continuous variables, in contrast to log-likelihood distance. This method assumes that all variables are independent. Moreover, continuous variables should be normally distributed, and categorical variables should be multinomial.

### 2.1. Log-Likelihood Distance

The log-likelihood distance (also known as the log-likelihood ratio) is a measure of the difference between two probability distributions. It is commonly used in information theory and machine learning to compare the fit of two models to a given set of data. The log-likelihood distance is defined as the difference between the log-likelihood of the observed data using one model and the log-likelihood of the same data using a second model. A smaller log-likelihood distance between two models indicates that the two models are more similar in their predictions for the observed data, while a larger log-likelihood distance indicates that the models are less similar. In other words, a model with a higher log-likelihood is considered to be a better fit for the data, and the log-likelihood distance can be used to compare the goodness-of-fit of different models. Finally, using log-likelihood distance in two-step cluster analysis can provide a more robust clustering method compared to methods that only use Euclidean distance or other measures of similarity. The distance between clusters  $i$  and  $j$  is defined as

$$d_{(i)(j)} = \xi_i + \xi_j - \xi_{(ij)} \tag{2}$$

$$\xi_s = -N_s \left( \left( \sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{sk}^2) \right) + \left( \sum_{k=1}^{K^B} \hat{E}_{sk} \right) \right) \tag{3}$$

$$\hat{E}_{sk} = - \sum_{l=1}^{L_k} \left( \frac{N_{skl}}{N_s} * \log \left( \frac{N_{skl}}{N_s} \right) \right) \tag{4}$$

where:

$K^A$  is the total number of continuous variables;

$K^B$  is the total number of categorical variables;

$L_k$  is the number of categories for the  $k$ th categorical variable;

$N_s$  is the total number of data records in cluster  $s$ ;

$N_{skl}$  is the number of records in cluster  $s$  whose categorical variable  $k$  takes the  $l$  category;

$\hat{\sigma}_k^2$  is the estimated variance of the continuous variable  $k$ ;

$\hat{\sigma}_{sk}^2$  is the estimated variance of the continuous variables  $k$  in cluster  $j$ ;

$d_{(i)(j)}$  is the distance between the  $i$  and the  $j$ ;

$(i, j)$  is the index representing cluster formed by combining clusters  $i$  and  $j$  [45].

### 2.2. Optimal Cluster Number

The optimal number of clusters is identified by the maximum value of the ratio of distance measures according to [46,47] in the statistical analytical program IBM SPSS 26. BIC and AIC are calculated for each number of clusters within a specific range. These indicators identify the optimal number of clusters. AIC (Akaike information criterion) and BIC (Bayesian information criterion) are both measures of the goodness-of-fit of a statistical model used to compare different models and select the best one. Both AIC and BIC balance the model’s goodness-of-fit with its complexity, as a model with too many parameters can easily fit the data too well, but at the cost of overfitting. In general, AIC and BIC provide similar information, but BIC tends to favor simpler models, while AIC is more balanced between fit and complexity. The choice of which measure to use depends on the specific problem and the trade-off between model fit and complexity that is desired. BIC statistics for a partition with  $R$  clusters are calculated as:

$$BIC_R = -2 \sum_{r=1}^R \xi_r + m_r \log(N) \tag{5}$$

with

$$m_r = R \left\{ 2K^A + \sum_{k=1}^{K^B} (L_k - 1) \right\} \tag{6}$$

where:

$BIC_R$  is the Bayesian information criterion;

$m_r$  is the ratio in the  $r$ -cluster developed during the hierarchical clustering stage;

$L_k$  is the number of groups in  $k$  categorical variables [47,48].

In addition, we also monitored the ratio of BIC changes and the ratio of distance measures. However, the statistical analytical program automatically determines the optimal number of clusters without the author’s decision [49].

### 2.3. Cluster Quality

The silhouette coefficient is a measure of the quality of a clustering solution in unsupervised machine learning. It provides a way to assess the similarity of the data points within a cluster and the dissimilarity between different clusters. The silhouette coefficient ranges from  $-1$  to  $1$ , where a value close to  $1$  indicates that the data points in a cluster are well separated and similar to each other, while a value close to  $-1$  indicates that the data points in a cluster are dissimilar and assigned to the wrong cluster. A value close to

0 indicates that the data points are indifferently similar to their cluster and the neighboring clusters. In other words, the silhouette value identifies poor classification (from  $-1.0$  to  $0.2$ ), fair classification (from  $0.2$  to  $0.5$ ), and good classification (from  $0.5$  to  $1.0$ ) [50]. This coefficient is computed for each data point, and the average silhouette score for all data points provides an overall measure of the quality of the clustering solution. The silhouette score can be used to compare different clustering solutions and to determine the optimal number of clusters for a given dataset. The silhouette value is calculated as

$$SW_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (7)$$

$$a_i = \frac{\sum_{j \in C_g (j \neq i)} D_{ij}}{n_g - 1} \quad (8)$$

$$b_i = \frac{\sum_{j \in C_h} D_{ij}}{n_h} \quad (9)$$

$$SC = \frac{\sum_{i=1}^n SW_i}{n} \quad (10)$$

where:

$SW_i$  is the silhouette coefficient for the  $i$ th object;

$SC$  is the average silhouette coefficient;

$a_i$  is the average of the minimum distance between the  $i$ th objects in the same cluster (average intra-cluster distance);

$b_i$  is the average of the minimum distance between the  $i$ th objects in a different cluster (average inter-cluster distance);

$C_g, C_h$  are cluster elements;

$D_{ij}$  is the distance;

$n_g, n_h$  are the number of objects in the  $g$ th ( $h$ th) cluster;

$n$  is the total number of observations [50,51].

It is a widely used metric for evaluating clustering solutions and is particularly useful for datasets with defined clusters or for datasets with a large number of clusters.

### 3. Results

**Optimal cluster number.** Table 1 reveals important metrics such as BIC, BIC change, ratio of BIC, and ratio of distance measures determining the optimal number of clusters. The maximum value of the ratio of distance measures identifies four clusters as the optimal number.

**Cluster quality.** The silhouette measure is higher than  $0.2$ . This metric represents the fair zone. These results demonstrate that the behaviors are significantly different from each other, but respondents in individual groups have similar features and preferences.

**Cluster structure.** Table 2 shows that the total number of respondents is 350, but all four clusters consist of 261 respondents. Other respondents are excluded due to missing data. As can be seen, clusters consist of different numbers of respondents. We find that the third cluster consists of more than 80 respondents (31.40%). On the other hand, the first cluster includes fewer than 50 respondents (18.00%).

Moreover, Figure 1 shows that the most significant predictors are gender (100%), user preference focusing on using a website, mobile application, or both (63%), and age category (54%). Other predictors are less significant. Predictor importance explains the relative importance of each predictor.

**Table 1.** Optimal cluster number based on ratio of distance measures.

Number of Clusters	Schwarz’s Bayesian Criterion (BIC)	BIC Change <sup>a</sup>	Ratio of BIC Changes <sup>b</sup>	Ratio of Distance Measures <sup>c</sup>
1	2794.695			
2	2603.075	−191.62	1.000	1.220
3	2459.052	−144.023	0.752	1.127
4	2339.487	−119.565	0.624	1.648
5	2295.396	−44.091	0.230	1.058
6	2257.667	−37.728	0.197	1.066
7	2226.747	−30.92	0.161	1.061
8	2201.740	−25.007	0.131	1.325
9	2200.588	−1.152	0.006	1.027
10	2201.336	0.749	−0.004	1.088
11	2207.889	6.552	−0.034	1.052
12	2217.721	9.832	−0.051	1.031
13	2229.430	11.709	−0.061	1.020
14	2242.326	12.896	−0.067	1.201
15	2265.164	22.838	−0.119	1.042

<sup>a</sup> The changes are from the previous number of clusters in the table. <sup>b</sup> The ratios of changes are relative to the change for the two-cluster solution. <sup>c</sup> The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

**Table 2.** The total sample divided into four clusters.

	N	% of Combined	% of Total
Cluster	1	47	18.00%
	2	70	26.80%
	3	82	31.40%
	4	62	23.80%
	Combined	261	100.00%
Excluded Cases	89		25.40
Total	350		100.00%

Figure 2 shows that the respondents are divided into four clusters. These clusters are made up of respondents with similar demographic characteristics and preferences for using applications. As can be seen, these clusters are created using six input categorical variables such as gender, age category, trip frequency, and user preferences. The color scale distinguishes predictors by within-cluster importance (see the legend).

Table 3 demonstrates that the first cluster consists exclusively of respondents aged 18 to 24. Other age groups are not represented in this cluster. On the other hand, the fourth cluster consists of respondents from 25 to 64 years old; both age categories from 25 to 44 and 45 to 64 have equal representation. Finally, the second and third clusters have diverse representations in all age categories. We find that the second cluster consists primarily of respondents from 18 to 24 years old, unlike the third cluster. This cluster consists of respondents from 25 to 44 years old (44.80% of this age group).

**Table 3.** The age of the respondents and their division into clusters.

		18–24 Years		25–44 Years		45–64 Years	
		Frequency	%	Frequency	%	Frequency	%
Cluster	1	47	42.70	0	0.00	0	0.00
	2	32	29.10	17	19.50	21	32.80
	3	31	28.20	39	44.80	12	18.80
	4	0	0.00	31	35.60	31	48.40
Combined		110	100.00	87	100.00	64	100.00

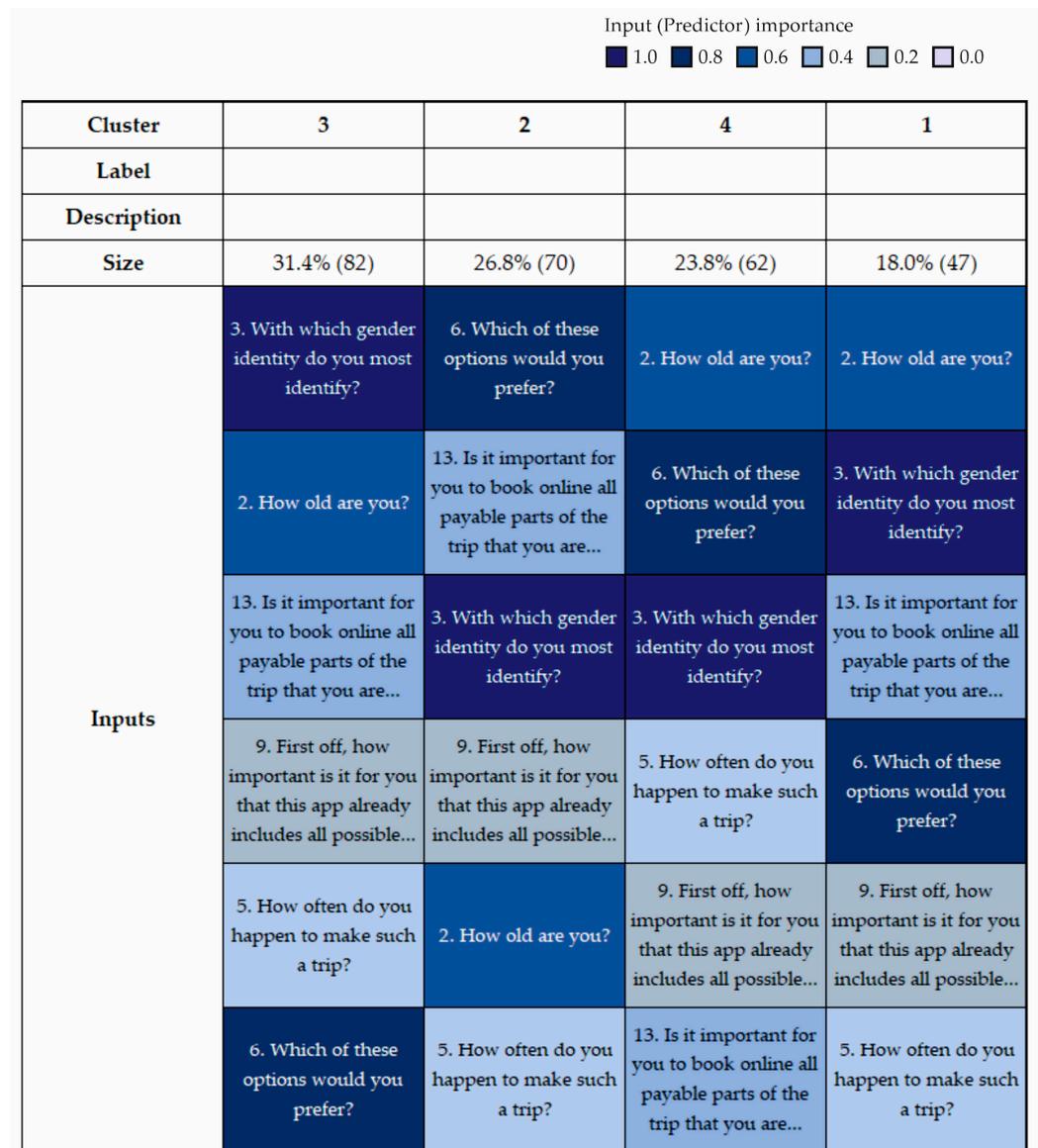


Figure 2. The four user clusters (inputs sorted by within-cluster importance).

Table 4 reveals that the first cluster consists only of men, unlike the third group. On the other hand, the second and fourth groups are made up of both sexes, but the majority are men.

Table 4. The gender of the respondents and their division into clusters.

		Female		Male	
		Frequency	%	Frequency	%
Cluster	1	0	0.00	47	28.80
	2	3	3.10	67	41.10
	3	82	83.70	0	0.00
	4	13	13.30	49	30.10
Combined		98	100.00	163	100.00

Table 5 demonstrates that respondents in the second and third clusters often go on trips, unlike the other groups.

**Table 5.** The frequency of trips and their division into clusters.

		Very Often		Frequently		Rarely		Sometimes		Never	
		Frequency	%	Frequency	%	Frequency	%	Frequency	%	Frequency	%
Cluster	1	6	10.70	13	18.30	11	21.60	16	23.20	1	7.10
	2	20	35.70	21	29.60	7	13.70	18	26.10	4	28.60
	3	16	28.60	27	38.00	21	41.20	18	26.10	0	0.00
	4	14	25.00	10	14.10	12	23.50	17	24.60	9	64.30
Combined		56	100.00	71	100.00	51	100.00	69	100.00	14	100.00

Table 6 shows that mobile applications are popular for three of the four clusters; the exception is the second cluster. This cluster prefers a web application as opposed to a mobile application.

**Table 6.** The user preferences and their division into clusters.

		Web App		Mobile App		I Do Not Care	
		Frequency	%	Frequency	%	Frequency	%
Cluster	1	0	0.00	47	24.50	0	0.00
	2	29	76.30	15	7.80	26	83.90
	3	9	23.70	68	35.40	5	16.10
	4	0	0.00	62	32.30	0	0.00
Combined		38	100.00	192	100.00	31	100.00

Table 7 demonstrates that access to information about all possible means of transport in the same place is extremely and very important for most users.

**Table 7.** The degrees of importance of having traffic schedules and services in one place and their distribution into clusters.

		Extremely Important/Very Important		Moderately Important		Not Important at All/Slightly Important	
		Frequency	%	Frequency	%	Frequency	%
Cluster	1	45	19.40	1	5.60	1	9.10
	2	55	23.70	6	33.30	9	81.80
	3	72	31.00	10	55.60	0	0.00
	4	60	25.90	1	5.60	1	9.10
Combined		232	100.00	18	100.00	11	100.00

Table 8 reveals that respondents substantially prefer online booking and quick access to tickets in the app in all clusters except the second cluster.

**Table 8.** The degrees of importance of having access to all tickets in one app and their distribution into clusters.

		Extremely Important/Very Important		Moderately Important		Not Important at All/Slightly Important	
		Frequency	%	Frequency	%	Frequency	%
Cluster	1	47	25.10	0	0.00	0	0.00
	2	24	12.80	27	58.70	19	67.90
	3	70	37.40	6	13.00	6	21.40
	4	46	24.60	13	28.30	3	10.70
Combined		187	100.00	46	100.00	28	100.00

Moreover, Appendix A contains the tables dividing respondents into clusters according to selected variables.

Cluster comparison. Figure 3 compares all four clusters. We find that the third (red) cluster consists only of women who prefer mobile applications to web applications (almost 83%). Most women are between 25 and 45 years old (47.6%). These women often go on trips, and online booking of all parts of the trip and quick access to transport information in one application is very or extremely important. The other groups include mainly men. Moreover, the first (green) cluster comprises only men. Men prefer the mobile app in all clusters except the second cluster. The results show that most of the men are between 18 and 24 years old. Finally, frequent travel is typical for the second (light blue) cluster, with a lower preference for online booking of all parts of the trip as the only cluster. However, information about all possible means of transport is very or extremely important. Even though the first (green) and fourth (dark blue) clusters travel less often than the other groups, both groups prefer one mobile application with all important information about means of transport and online booking with easy access to tickets.

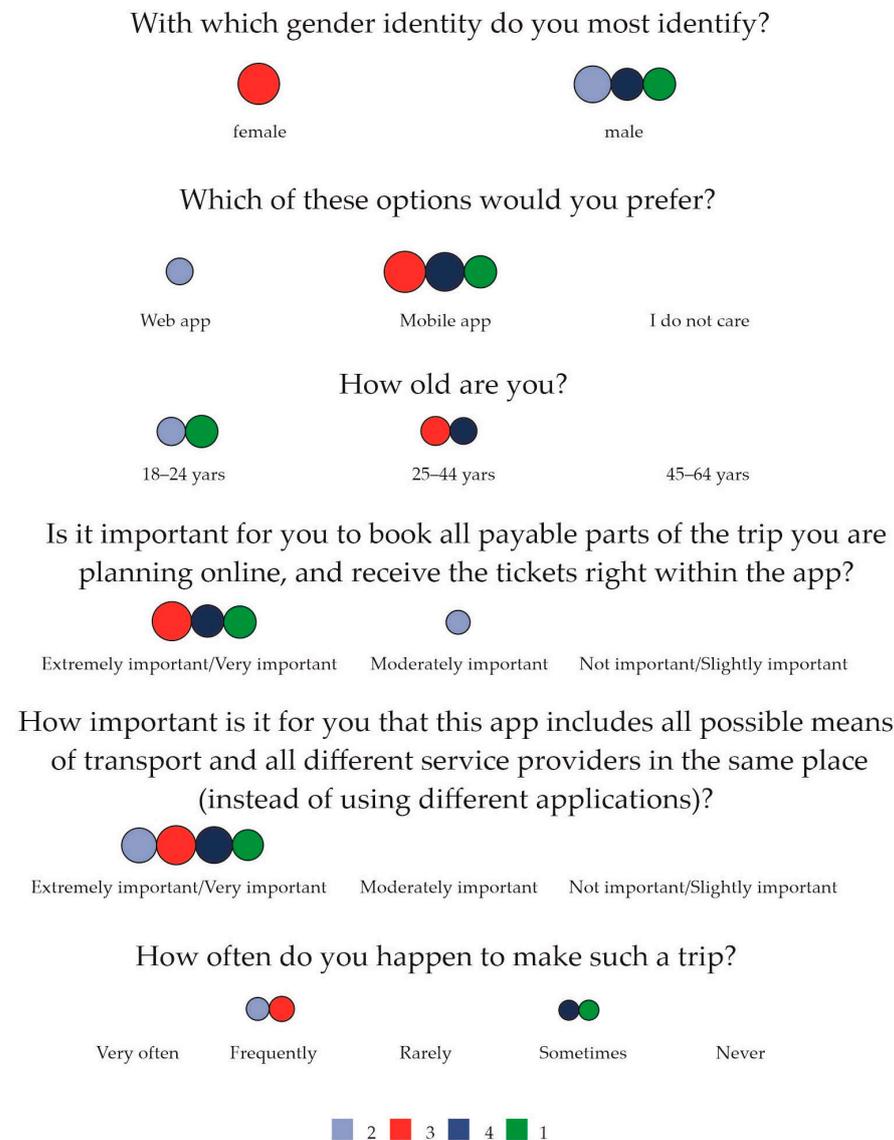


Figure 3. Cluster comparison.

#### 4. Discussion

The field of transport offers a lot of data, which with proper analysis and interpretation can bring a lot of benefits. The following studies performed cluster analysis for sequencing enormous amounts of traffic data to categorize them according to related characteristics and understand mutual associations [52,53]. The cluster segmentation technique made it possible to sort travel attendees based on travel distance and frequency of attendance to present a proposal for market applications [54]. In our case, the respondents were divided according to similar demographic characteristics as well as preferences for using a web or mobile application. Two-step cluster analysis was used in the selection of transport simulation model parameter values for the rapid processing of enormous amounts of data with both continuous and categorical variables [55]. Our dataset was also redistributed using a two-step clustering method based on the significance of the variables. Cluster analysis was applied to create homogeneous transport markets, where markets of comparable sizes were combined into one group [56]. The assignment of our sample resulted from grouping into homogeneous groups according to travel behavior. In another study, passenger demand for public transport services was predicted through the spatial clustering of applications [57]. Our research focused on comparing the preferences of users using a web or a mobile application. A study in the United States assessed the interaction between travelers and online travel systems using cluster analysis [58]. Support for spatial planning of transport in cities is related to the efficient processing of traffic data. Cluster analysis evaluated travel data in terms of the interaction of travel flows and spatial structure [59]. The impact of fares on the status of public transport was investigated by a study in Italy that used cluster analysis to segment passengers [60]. In our article, we examined the preference for online reservations with easy access to tickets. The prediction of the flow of passenger transport was conducted using cluster analysis, which divided the stations into six categories concerning their patterns of passenger transport [61].

Our paper focused on identifying differences in passenger needs through a statistical analytical tool. Two-step cluster analysis was applied to a wide range of available data from a survey reflecting traffic behavior within MaaS activities. The sample consisted of approximately 300 respondents divided into four clusters. The most influential predictors were gender, user preferences for using a website or mobile application, and age composition. The third cluster was made up of working-age women who prefer mobile applications to web applications, for whom quick access to online booking information is essential. The other groups were dominated by men, while the first cluster was composed exclusively of men. Even though the second cluster in terms of frequent travel was not extremely interested in online booking, the opposite was true in the other clusters. In the first, third, and fourth clusters, respondents preferred reservations through a mobile application, while they rated the availability of traffic information as extremely important.

The limited sample that was available for analytical purposes can be considered a limitation of our research. As part of further research activities, we would be interested in expanding awareness of MaaS activities among the public as well.

#### 5. Conclusions

Traffic movement prediction is determined by the demand for transport services. Identifying real demand is an essential aspect that is facilitated by constantly advancing and improving intelligent transport systems with designed mobility applications. Profiling public transport users is an important prerequisite for understanding the perception and behavior of passengers using public transport services. Achieving the segmentation of public transport participants with the same or at least related features is possible through cluster analysis.

In recent years, the MaaS initiative has been developed, which focuses on understanding travel behavior and combines the offer of transport services of operators on one global platform. The intention of this direction is the possibility of providing online reservations,

payments, and the availability of travel information for business but also private trips by supporting all modes of transport within the framework of public transport.

In this paper, the partial results of the demonstration activities of Shift2Rail IP4 for the Slovak and Czech sides were evaluated. Data reflecting travel attitudes were obtained from the online conversational inquiry tool. Cluster analysis was used to evaluate the sample, in which an optimal number of four clusters was identified. Individual clusters consisted of different numbers of respondents with similar characteristics.

This article aimed to identify the differences in the needs of passengers to determine the direction of the use of the application with a focus on the future development of mobility as a service. The importance of individual predictors was determined based on the attitudes towards the input variables. The three most important predictors were considered in descending order, namely gender, web or mobile application preference, and age category. The results of our research indicated that respondents are interested in timely traffic information and prefer a mobile application for online booking. The benefit of this paper was to highlight the partial demographic results affecting travel activities within MaaS for the Czech and Slovak sides. The results of this research for the Czech and Slovak participation are useful for supplementing the overall evaluation of the perception of MaaS in the world.

**Author Contributions:** Conceptualization, J.M. (Jaroslav Mašek), V.Š., J.M. (Jaroslav Mazanec) and P.J.; methodology, J.M. (Jaroslav Mazanec) and V.Š.; software, J.M. (Jaroslav Mazanec); validation, J.M. (Jaroslav Mašek) and P.J.; formal analysis, J.M. (Jaroslav Mazanec) and V.Š.; investigation, V.Š. and J.M. (Jaroslav Mazanec); resources, J.M. (Jaroslav Mašek); data curation, J.M. (Jaroslav Mašek) and P.J.; writing—original draft preparation, V.Š. and J.M. (Jaroslav Mazanec); writing—review and editing, J.M. (Jaroslav Mašek), V.Š., J.M. (Jaroslav Mazanec) and P.J.; visualization, J.M. (Jaroslav Mazanec) and V.Š.; supervision, J.M. (Jaroslav Mašek) and P.J.; project administration, J.M. (Jaroslav Mašek), V.Š. and J.M. (Jaroslav Mazanec); funding acquisition, J.M. (Jaroslav Mašek). All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper was supported by the Shift2Rail Joint Undertaking under the European Union’s Horizon 2020 research and innovation program under grant agreement no. 101015492, Shift2Rail IP4 to support the deployment of Mobility as a Service (IP4MaaS).

**Data Availability Statement:** All data used here are available on request from the authors.

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

## Appendix A

**Table A1.** Age structure.

Age	Cluster							
	1	2	3	4	1	2	3	4
	abs.				%			
18–24 years	47	32	31	0	100.00	45.71	37.80	0.00
25–44 years	0	17	39	31	0.00	24.29	47.56	50.00
45–64 years	0	21	12	31	0.00	30.00	14.63	50.00
Total	47	70	82	62	100.00	100.00	100.00	100.00

**Table A2.** Gender structure.

Gender	Cluster							
	1	2	3	4	1	2	3	4
	abs.				%			
Female	0	3	82	13	0.00	4.29	100.00	20.97
Male	47	67	0	49	100.00	95.71	0.00	79.03
Total	47	70	82	62	100.00	100.00	100.00	100.00

**Table A3.** Frequency of trip.

Frequency of Trip	Cluster							
	1	2	3	4	1	2	3	4
	abs.				%			
Very often	6	20	16	14	12.77	28.57	19.51	22.58
Frequently	13	21	27	10	27.66	30.00	32.93	16.13
Rarely	11	7	21	12	23.40	10.00	25.61	19.35
Sometimes	16	18	18	17	34.04	25.71	21.95	27.42
Never	1	4	0	9	2.13	5.71	0.00	14.52
Total	47	70	82	62	100.00	100.00	100.00	100.00

**Table A4.** User preference.

Preference	Cluster							
	1	2	3	4	1	2	3	4
	abs.				%			
Web app	0	29	9	0	0.00	41.43	10.98	0.00
Mobile app	47	15	68	62	100.00	21.43	82.93	100.00
I do not care	0	26	5	0	0.00	37.14	6.10	0.00
Total	47	70	82	62	100.00	100.00	100.00	100.00

**Table A5.** Degree of importance of having access to all possible means of transport and all different service providers in the same place.

Importance	Cluster							
	1	2	3	4	1	2	3	4
	abs.				%			
Extremely important/Very important	45	55	72	60	95.74	78.57	87.80	96.77
Moderately important	1	6	10	1	2.13	8.57	12.20	1.61
Not important at all/Slightly important	1	9	0	1	2.13	12.86	0.00	1.61
Total	47	70	82	62	100.00	100.00	100.00	100.00

**Table A6.** Degree of importance of all payable parts of the trip.

Importance	Cluster							
	1	2	3	4	1	2	3	4
	abs.				%			
Extremely important/Very important	47	24	70	46	100.00	34.29	85.37	74.19
Moderately important	0	27	6	13	0.00	38.57	7.32	20.97
Not important at all/Slightly important	0	19	6	3	0.00	27.14	7.32	4.84
Total	47	70	82	62	100.00	100.00	100.00	100.00

## References

1. Stopka, U. Identification of User Requirements for Mobile Applications to Support Door-to-Door Mobility in Public Transport. In *Human-Computer Interaction. Applications and Services, Proceedings of the 16th International Conference, HCI International 2014, Heraklion, Crete, Greece, 22–27 June 2014*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 513–524. [\[CrossRef\]](#)
2. Ceder, A. Urban mobility and public transport: Future perspectives and review. *Int. J. Urban Sci.* **2021**, *25*, 455–479. [\[CrossRef\]](#)
3. Medina, S.A.O.; Wang, B. Public transport routing including fixed schedule, shared on-demand and door-to-door services. *Procedia Comput. Sci.* **2019**, *151*, 846–851. [\[CrossRef\]](#)
4. Portell, L.; Morera, S.; Ramalhinho, H. Door-to-Door Transportation Services for Reduced Mobility Population: A Descriptive Analytics of the City of Barcelona. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4536. [\[CrossRef\]](#)
5. Mraih, R.; Anis, R. Impact of transport infrastructure on regional mobility: Application of social equity approach. In Proceedings of the 2013 International Conference on Advanced Logistics and Transport, Sousse, Tunisia, 29–31 May 2013; pp. 284–289. [\[CrossRef\]](#)
6. Zou, X.; Cao, J.; Sun, W.; Guo, Q.; Wen, T. Flow data processing paradigm and its application in smart city using a cluster analysis approach. *Clust. Comput.* **2019**, *22*, 435–444. [\[CrossRef\]](#)
7. Bellini, P.; Bilotta, S.; Cenni, D.; Nesi, P.; Paolucci, M.; Soderi, M. Knowledge Modeling and Management for Mobility and Transport Applications. In Proceedings of the IEEE 4th International Conference on Collaboration and Internet Computing (CIC), Philadelphia, PA, USA, 18–20 October 2018; pp. 364–371. [\[CrossRef\]](#)
8. Souza, V.S.; Marques, S.R.B.D.V. Factors influencing urban tourists' receptivity to ecogamified applications: A study on transports and mobility. *Int. J. Tour. Cities* **2022**, *8*, 820–843. [\[CrossRef\]](#)
9. Halvorsen, A.; Koutsopoulos, H.N.; Ma, Z.; Zhao, J. Demand management of congested public transport systems: A conceptual framework and application using smart card data. *Transportation* **2020**, *47*, 2337–2365. [\[CrossRef\]](#)
10. Abbasi, S.; Ko, J.; Min, J. Measuring destination-based segregation through mobility patterns: Application of transport card data. *J. Transp. Geogr.* **2021**, *92*, 103025. [\[CrossRef\]](#)
11. Spinney, J.E.; Scott, D.; Newbold, B. Transport mobility benefits and quality of life: A time-use perspective of elderly Canadians. *Transp. Policy* **2009**, *16*, 1–11. [\[CrossRef\]](#)
12. Andersson, A.; Hiselius, L.W.; Berg, J.; Forward, S.; Arnfalk, P. Evaluating a Mobility Service Application for Business Travel: Lessons Learnt from a Demonstration Project. *Sustainability* **2020**, *12*, 783. [\[CrossRef\]](#)
13. Vinayak, P.; Dias, F.F.; Astroza, S.; Bhat, C.R.; Pendyala, R.M.; Garikapati, V.M. Accounting for multi-dimensional dependencies among decision-makers within a generalized model framework: An application to understanding shared mobility service usage levels. *Transp. Policy* **2018**, *72*, 129–137. [\[CrossRef\]](#)
14. Le Boennec, R.; Nicolai, I.; Da Costa, P. Assessing 50 innovative mobility offers in low-density areas: A French application using a two-step decision-aid method. *Transp. Policy* **2019**, *83*, 13–25. [\[CrossRef\]](#)
15. Cirianni, F.M.M.; Leonardi, G. Analysis of transport modes in the urban environment: An application for a sustainable mobility system. *Sustain. City IV Urban Regen. Sustain.* **2006**, *93*, 637–645. [\[CrossRef\]](#)
16. Malone, K.; Silla, A.; Johanssen, C.; Bell, D. Safety, mobility and comfort assessment methodologies of intelligent transport systems for vulnerable road users. *Eur. Transp. Res. Rev.* **2017**, *9*, 1–16. [\[CrossRef\]](#)
17. Manasseh, C.; Sengupta, R. Middleware to enhance mobile communications for road safety and traffic mobility applications. *IET Intell. Transp. Syst.* **2010**, *4*, 24–36. [\[CrossRef\]](#)
18. Vicente, P.; Reis, E. Profiling public transport users through perceptions about public transport providers and satisfaction with the public transport service. *Public Transp.* **2016**, *8*, 387–403. [\[CrossRef\]](#)
19. Goulet-Langlois, G.; Koutsopoulos, H.N.; Zhao, J. Inferring patterns in the multi-week activity sequences of public transport users. *Transp. Res. Part C Emerg. Technol.* **2016**, *64*, 1–16. [\[CrossRef\]](#)
20. Martin, D.; Gale, C.; Cockings, S.; Harfoot, A. Origin-destination geodemographics for analysis of travel to work flows. *Comput. Environ. Urban Syst.* **2018**, *67*, 68–79. [\[CrossRef\]](#)
21. Yaduvanshi, R.; Bansal, S.; Kumar, A. Factors Affecting Traffic Management using Two Step Cluster. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 1184–1189. [\[CrossRef\]](#)
22. Qin, H.; Huang, Q.; Zhang, Z.; Lu, Y.; Li, M.; Xu, L.; Chen, Z. Carbon dioxide emission driving factors analysis and policy implications of Chinese cities: Combining geographically weighted regression with two-step cluster. *Sci. Total Environ.* **2019**, *684*, 413–424. [\[CrossRef\]](#)
23. Chang, F.; Xu, P.; Zhou, H.; Chan, A.H.; Huang, H. Investigating injury severities of motorcycle riders: A two-step method integrating latent class cluster analysis and random parameters logit model. *Accid. Anal. Prev.* **2019**, *131*, 316–326. [\[CrossRef\]](#)
24. Santos, G.; Nikolaev, N. Mobility as a Service and Public Transport: A Rapid Literature Review and the Case of Moovit. *Sustainability* **2021**, *13*, 3666. [\[CrossRef\]](#)
25. Tomaino, G.; Teow, J.; Carmon, Z.; Lee, L.; Ben-Akiva, M.; Chen, C.; Leong, W.Y.; Li, S.; Yang, N.; Zhao, J. Mobility as a service (MaaS): The importance of transportation psychology. *Mark. Lett.* **2020**, *31*, 419–428. [\[CrossRef\]](#)
26. Hensher, D.A.; Xi, H. Mobility as a service (MaaS): Are effort and seamlessness the keys to MaaS uptake? *Transp. Rev.* **2022**, *42*, 269–272. [\[CrossRef\]](#)
27. Matyas, M.; Kamargianni, M. The potential of mobility as a service bundles as a mobility management tool. *Transportation* **2019**, *46*, 1951–1968. [\[CrossRef\]](#)

28. Hensher, D.A. Future bus transport contracts under a mobility as a service (MaaS) regime in the digital age: Are they likely to change? *Transp. Res. Part A Policy Pract.* **2017**, *98*, 86–96. [CrossRef]
29. Le Pira, M.; Tavasszy, L.A.; Correia, G.H.D.A.; Ignaccolo, M.; Inturri, G. Opportunities for integration between Mobility as a Service (MaaS) and freight transport: A conceptual model. *Sustain. Cities Soc.* **2021**, *74*, 103212. [CrossRef]
30. Kamargianni, M.; Li, W.; Matyas, M.; Schäfer, A. A Critical Review of New Mobility Services for Urban Transport. *Transp. Res. Procedia* **2016**, *14*, 3294–3303. [CrossRef]
31. Murati, E. Mobility-as-a-service (MaaS) digital marketplace impact on EU passengers' rights. *Eur. Transp. Res. Rev.* **2020**, *12*, 62. [CrossRef]
32. Alonso-González, M.J.; Hoogendoorn-Lanser, S.; van Oort, N.; Cats, O.; Hoogendoorn, S. Drivers and barriers in adopting Mobility as a Service (MaaS)—A latent class cluster analysis of attitudes. *Transp. Res. Part A Policy Pract.* **2020**, *132*, 378–401. [CrossRef]
33. Vij, A.; Ryan, S.; Sampson, S.; Harris, S. Consumer preferences for Mobility-as-a-Service (MaaS) in Australia. *Transp. Res. Part C Emerg. Technol.* **2020**, *117*, 102699. [CrossRef]
34. Schikofsky, J.; Dannewald, T.; Kowald, M. Exploring motivational mechanisms behind the intention to adopt mobility as a service (MaaS): Insights from Germany. *Transp. Res. Part A Policy Pract.* **2020**, *131*, 296–312. [CrossRef]
35. Georgakis, P.; Almohammad, A.; Bothos, E.; Magoutas, B.; Arnaoutaki, K.; Mentzas, G. Heuristic-Based Journey Planner for Mobility as a Service (MaaS). *Sustainability* **2020**, *12*, 10140. [CrossRef]
36. Hasselwander, M.; Bigotte, J.F.; Antunes, A.P.; Sigua, R.G. Towards sustainable transport in developing countries: Preliminary findings on the demand for mobility-as-a-service (MaaS) in Metro Manila. *Transp. Res. Part A Policy Pract.* **2022**, *155*, 501–518. [CrossRef]
37. Storme, T.; De Vos, J.; De Paepe, L.; Witlox, F. Limitations to the car-substitution effect of MaaS. Findings from a Belgian pilot study. *Transp. Res. Part A Policy Pract.* **2020**, *131*, 196–205. [CrossRef]
38. Ho, C.Q.; Mulley, C.; Hensher, D.A. Public preferences for mobility as a service: Insights from stated preference surveys. *Transp. Res. Part A Policy Pract.* **2020**, *131*, 70–90. [CrossRef]
39. Jittrapirom, P.; Marchau, V.; van der Heijden, R.; Meurs, H. Future implementation of mobility as a service (MaaS): Results of an international Delphi study. *Travel Behav. Soc.* **2020**, *21*, 281–294. [CrossRef]
40. Arias-Molinares, D.; García-Palomares, J.C. Shared mobility development as key for prompting mobility as a service (MaaS) in urban areas: The case of Madrid. *Case Stud. Transp. Policy* **2020**, *8*, 846–859. [CrossRef]
41. Alyavina, E.; Nikitas, A.; Njoya, E.T. Mobility as a service and sustainable travel behaviour: A thematic analysis study. *Transp. Res. Part F Traffic Psychol. Behav.* **2020**, *73*, 362–381. [CrossRef]
42. Smith, G.; Sochor, J.; Karlsson, I.M. Mobility as a Service: Development scenarios and implications for public transport. *Res. Transp. Econ.* **2018**, *69*, 592–599. [CrossRef]
43. Project Description-IP4MaaS D2.3 Demonstration Requirements and Scenarios F-REL\_Final. Available online: [https://www.ip4maas.eu/wp-content/uploads/2022/07/IP4M-D2.3-Demonstration-requirements-and-scenarios-F-REL\\_final.pdf](https://www.ip4maas.eu/wp-content/uploads/2022/07/IP4M-D2.3-Demonstration-requirements-and-scenarios-F-REL_final.pdf) (accessed on 15 January 2023).
44. IBM SPSS Statistics 22 Algorithms. Available online: [https://www.sussex.ac.uk/its/pdfs/SPSS\\_Statistics\\_Algorithms\\_22.pdf](https://www.sussex.ac.uk/its/pdfs/SPSS_Statistics_Algorithms_22.pdf) (accessed on 15 January 2023).
45. IBM SPSS Statistics Base 28. Available online: [https://www.ibm.com/docs/en/SSLVMB\\_28.0.0/pdf/IBM\\_SPSS\\_Statistics\\_Base.pdf](https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Statistics_Base.pdf) (accessed on 15 January 2023).
46. Şchiopu, D. Applying TwoStep cluster analysis for identifying bank customers' profile. *Buletinul* **2010**, *62*, 66–75.
47. Tevdovski, D. Twostep cluster analysis: Segmentation of the largest companies in Macedonia. In Proceedings of the International Scientific Conference: Challenges for Analysis of the Economy, the Businesses and Social Progress, Szeged, Hungary, 19–21 November 2009.
48. Astuti, A.B.; Fernandes, A.A.R.; Amaliana, L.; Yanti, I.; Isaskar, R. Step Cluster Analysis for Tourist Segmentation Coastal Object for Green Marketing Strategy. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *239*, 012019. [CrossRef]
49. Radenović, Ž.; Boshkov, T. Economic Effects of Congress Tourism: Two-Step Cluster Approach. *Chall. Tour. Bus. Logist. 21st Century* **2022**, *5*, 185–192. [CrossRef]
50. Supandi, A.; Saefuddin, A.; Sulvianti, I.D. Two step Cluster Application to Classify Villages in Kabupaten Madiun Based on Village Potential Data. *Xplore J. Stat.* **2021**, *10*, 12–26. [CrossRef]
51. Goyal, N.; Gupta, K.; Kumar, N. Clustering-based hierarchical framework for multiclass classification of leaf images. *IEEE Trans. Ind. Appl.* **2022**, *58*, 4076–4085. [CrossRef]
52. Řezanková, H. Different approaches to the silhouette coefficient calculation in cluster evaluation. In Proceedings of the 21st International Scientific Conference AMSE Applications of Mathematics and Statistics in Economics, Kutná Hora, Czech Republic, 29 August–2 September 2018; pp. 1–10.
53. Arongna, A.; Sulonggaowa, S. Application of travel service recommendation algorithm based on cloud computing. *J. Phys. Conf. Ser.* **2021**, *1852*, 42090. [CrossRef]
54. Warnick, R.B.; Bojanic, D.C.; Mathur, A.; Ninan, D. Segmenting Event Attendees Based on Travel Distance, Frequency of Attendance, and Involvement Measures: A Cluster Segmentation Technique. *Event Manag.* **2011**, *15*, 77–90. [CrossRef]
55. Zenina, N.; Romanovs, A.; Merkurjev, Y. Transport Simulation Model Calibration with Two-Step Cluster Analysis Procedure. *Inf. Technol. Manag. Sci.* **2015**, *18*, 49–56. [CrossRef]
56. Gremm, C. The effect of intermodal competition on the pricing behaviour of a railway company: Evidence from the German case. *Res. Transp. Econ.* **2018**, *72*, 49–64. [CrossRef]

57. Thiagarajan, R.; Prakashkumar, S. Identification of Passenger Demand in Public Transport Using Machine Learning. *Webology* **2021**, *18*, 223–236. [[CrossRef](#)]
58. Park, S.; Kim, D.-Y. Assessing language discrepancies between travelers and online travel recommendation systems: Application of the Jaccard distance score to web data mining. *Technol. Forecast. Soc. Chang.* **2017**, *123*, 381–388. [[CrossRef](#)]
59. Akin, D.; Alasalvar, S. Estimate Urban Growth and Expansion by Modeling Urban Spatial Structure Using Hierarchical Cluster Analyses of Interzonal Travel Data. *Int. J. Syst. Dyn. Appl.* **2016**, *5*, 16–41. [[CrossRef](#)]
60. Salis, S.; Barabino, B.; Useli, B. Segmenting Fare Evader Groups by Factor and Cluster Analysis. *WIT Trans. Built Environ.* **2017**, *176*, 503–515. [[CrossRef](#)]
61. Park, Y.; Choi, Y.; Kim, K.; Yoo, J.K. Machine learning approach for study on subway passenger flow. *Sci. Rep.* **2022**, *12*, 2754. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.