Scientific Journal of Silesian University of Technology. Series Transport
Zeszyty Naukowe Politechniki Śląskiej. Seria Transport

Volume 123 **2024**

p-ISSN: 0209-3324

e-ISSN: 2450-1549

DOI: https://doi.org/10.20858/sjsutst.2024.123.2

Journal homepage: http://sjsutst.polsl.pl

Article citation information:
Jazi, M., Gazder, U., Arsalan, M., Mehdi, M.R. Forecasting future public transport mode choice behaviour of commuters in Bahrain using Logit and classification tree models: a comparative study. *Scientific Journal of Silesian University of Technology. Series Transport*. 2024, **123**, 23-55. ISSN: 0209-3324.
DOI: https://doi.org/10.20858/sjsutst.2024.123.2.

Marwa JAZI[1], Uneb GAZDER[2], Mudassar ARSALAN[3], Mohammed Raza MEHDI[4]

FORECASTING FUTURE PUBLIC TRANSPORT MODE CHOICE BEHAVIOUR OF COMMUTERS IN BAHRAIN USING LOGIT AND CLASSIFICATION TREE MODELS: A COMPARATIVE STUDY

Summary. Global trade and social relationships are greatly facilitated by transportation. However, the majority of nations, including Bahrain, face substantial challenges with their transportation systems. For the development of technical solutions that can promote the progress of these transport systems, it is now crucial to have a complete understanding of travel demands and driver's characteristics. This paper aims to explore the influential factors concerning travel mode choice in Bahrain and utilize mode choice models to forecast the probable utilities of various future public transport modes. The study utilizes diverse, 3864 data records extracted from previous surveys as well as a recent one conducted within this research. Subsequently, using Minitab software, two types of mode choice models were built, namely the logit model and the classification tree model, focusing on modelling the future transportation system, considering potential

[1] Department of Civil Engineering, University of Bahrain, Sakhir, Bahrain. Email: marwah.jazi@gmail.com. ORCID: https://orcid.org/0009-0000-2671-8086
[2] Department of Civil Engineering, University of Bahrain, Sakhir, Bahrain. Email: ugazder@uob.edu.bh. ORCID: https://orcid.org/0000-0002-9445-9570.
[3] School of Computer, Data and Mathematical Sciences, Western Sydney University, Sydney, Australia. Email: mharsalan@gmail.com. ORCID: https://orcid.org/0000-0001-9622-5930.
[4] Department of Civil Engineering, DHA Suffa University, Karachi, Pakistan. Email: mohammedrazamehdi@dsu.edu.pk. ORCID: https://orcid.org/0009-0009-3783-3456

public transport modes (Public Bus, Metro, and Tram). The analysis of the data identified trip cost as the top predictor, moreover, direct, and quick travel, accessibility, and convenience were also found to significantly influence the choice of travel mode in Bahrain. Additionally, the findings indicate that the metro is the preferred choice for future public transport, with a strong preference observed for a combination of metro and tram. The research also suggests, in terms of model performance, that when capturing more complex patterns, as in this study, the classification tree outperforms the multinomial Logit model. Overall, the research provides valuable insights into mode choice in Bahrain and highlights the important factors influencing commuting decisions. The results of this study can support the development of an efficient public transportation system that would satisfy the needs and preferences of commuters in Bahrain and ultimately lead to a sustainable and accessible transportation infrastructure in the country.

Keywords: public transport, Bahrain, mode choice, Logit model, decision tree

1. INTRODUCTION

In the last century, the trends of urbanization and resulting population expansion were going at an alarmingly high rate [1]. These trends give a boost to the increase in car ownership with consequences of frequent congestion delays [2], increase in environmental damage [3], loss of life and property with road crashes [4] and consumption of valuable resources including fuel, land and national budget [5]. Realizing these issues, public transportation implementation and promotion of public transportation modes became a top priority for governments around the world [6]. Therefore, a number of options have been planned and implemented within the domain of public transportation including bus rapid transit, light rail systems, trams, etc. [7].

Urbanization and population expansion in the Kingdom of Bahrain eventually led to a significant increase in automobile ownership levels, which worsened traffic congestion on the Kingdom's road system. According to the General Directorate of Traffic, the number of automobiles in Bahrain increased dramatically from about 400,000 in 2009 to almost 700,000 by the end of 2019. This and the absence of a reliable transit system made it more difficult to travel and resulted in delays of up to several hours during times of high traffic. Consequently, designing and launching an effective public transport service, that would eventually contribute to minimizing congestion, delay and traffic accidents, became a necessity. However, the question that remains is: what public transport mode is best fit to the mode choice behaviour of commuters in Bahrain? Therefore, a thorough mode choice analysis is needed to understand the mode choice preferences and explore the factors that have the greatest influence on the traveller's decision-making process.

This research aims to explore the influential factors concerning travel mode choice in Bahrain, identify the driver's characteristics affecting his/her mode choice, and utilize mode choice models like the Logit model and Classification tree model to forecast the probable utility of various future public transport modes, including bus, metro, and tram. Such systems have already been adopted in other countries, including those neighbouring the study area (Bahrain), including Saudi Arabia [8], Dubai [9], Qatar [10], etc. However, it should be noted that the implementation and success of these modern public transport systems depends heavily on demand and coverage area [11]. Hence, Bahrain could be a peculiar case from these aspects as it has a relatively smaller area with a population of less than 2 million people [12]. The above factors justify the carrying out of current research, and it is expected that it will reveal unique

and important features for transportation planners and researchers for their work related to modern public transportation systems. As stated above, the scope of this research focuses on future transportation modes for Bahrain, which are non-existent currently, so stated preference data was utilized as this was the only option. The study utilizes two different discrete choice models and compares between their performance and insights for prediction of mode choice. The results of this study are expected to provide significant policy recommendations for countries like Bahrain who intend to progress forward in sustainable development of the transportation sector.

2. LITERATURE REVIEW

Travel forecasting models are the core of the transportation planning process, and they are considered as a measure to detect travel needs of cities [13]. These models usually employ mathematical equations and algorithms to simulate travel patterns and behaviours [14]. In practice, it was only after World War II, in the 1960s, that travel-modelling applications commenced, and the classic four-stage model was gradually constructed. This model breaks down the research area into homogenous traffic analysis zones, in which the data required for building the model is collected [15]. The four-stage model process can be divided into two main phases. The first phase focuses on collecting, evaluating, calibrating and validating data to determine the travel demand, while the second phase loads this demand onto the network to formulate equilibration of route choice. These two phases can be further divided into four distinct stages known as trip generation, trip distribution, modal split and traffic assignment [16].

Mode choice or modal split is the third stage of the four-stage model. This stage focuses on the traveller's behaviour in relation to the selection of travel mode [13]. The traveller's decision is mainly influenced by demand variables, such as income, vehicle ownership, household size and location, as well as supply variables, which include travel time, travel cost and transfer time [17].

Discrete choice models, specifically the Logit model, have been a common tool in transportation planning since the 1960s for predicting traveller mode selection [18]. These models assume that each travel mode, such as car, bus, and train, has a specific level of utility, based on the variables mentioned above, which subsequently affects the traveller's decision and preferences [19]. According to the Logit model, the utility function of a mode, and the probability that a traveller chooses it, is expressed by the following equations (1 and 2), respectively:

$$U_m = C + A_1X_1 + A_2X_2 \dots\dots + A_iX_i \qquad (1)$$

Where:
U_m = Utility function of mode 'm'
C = Constant
A_i = Coefficients (weight of each attribute based on survey's data)
X_i = Independent Variables correlated with the mode choice [20]

$$P_m^i = \frac{e^{U_{mi}}}{\sum_j e^{U_{mi}}} \qquad (2)$$

Where:
P_m^i = Probability of trip maker i choosing mode m out of j alternatives
U_{mi} = utility of alternative m for trip maker i [21]

Mode choice behaviour can be formed and explained using various theoretical forms, such as planned behaviour, habit formation, norm activation, and others. One of the most prominent frameworks among them is the rational choice theory, which presents the mode choice as a result of compromise between costs and benefits of each mode choice [22]. Logit models are suitable tools to apply the framework due to its utility functions, as mentioned above.

In 2014, a study by Ratrout et al. [21] investigated the present travel trends and the changes in border mode choice behaviour anticipated due to the introduction of probable future train and ferry services between Dammam-Khobar metropolitan area of the Kingdom of Saudi Arabia and the Kingdom of Bahrain. The researchers collected data through questionnaire-based interviews and used it in the development of Logit models. It was recognized that most of the travellers favoured the car over airplane from the existing modes and generally travelled through this route for recreation, transit flights and social visits. When the potential train and ferry services were presented to the travellers, it was observed that the train service was considered more attractive than the ferry service, while the latter being more appealing to single travellers compared to the car.

The study by Abdullah et al. [24] provides yet another instance of the Logit model in use. Because of the outbreak of the Coronavirus (COVID-19) pandemic, public transport experienced a severe decline in the number of its users worldwide. This issue inspired Abdullah et al. [24] to investigate and analyse travellers' behaviour regards choosing a travel mode under COVID-19 conditions. The study was conducted in Lahore, Pakistan in which, 1516 responses were collected through a questionnaire. To accomplish the aim of the research, a binary Logit model was developed involving private and public transport. It was concluded that gender, ownership, income, trip frequency, education, profession, and safety had a significant influence on people's choices. Adding to that, it was observed that females are more likely to choose public transport over private modes, compared to males.

Despite relying solely on discrete models in the past, transport planners started to investigate more sophisticated alternative machine learning approaches after the remarkable advancement in machine learning research that showed numerous successes of its application [18]. Decision trees are a popular machine learning technique. Using a structure similar to a flowchart or a tree, decision trees, such as Classification trees and Regression trees, categorize data into groups and effectively illustrates the connections between characteristics and potential outputs [19]. Decision trees with binary splits are the most often used types of decision trees. To calculate each split, the data is first categorized according to the required features, and for each feature, potential binary split sites are then examined [18].

Oral and Tecim [25] conducted a study to forecast mode choice of trips in district Buca in Izmir, Turkey using decision tree method. The model was built based on data extracted from a household survey done in 2007. According to the results obtained, travel time, purpose of trip, driving licence, number of vehicles, house ownership, age, occupation, and public transport card ownership play a significant role in determining the mode choice of an individual living in Buca. However, the origin and destination of trips had no major influence on mode choice.

Another study done in São Paulo Metropolitan Area, Brazil by Lindner et al. [26] compared the prediction efficiency of Artificial Neural Network (ANN) and Classification Tree models with that of a binary Logit model. The models were utilized to forecast motorized mode choice, using data from an Origin-Destination survey conducted in 2007. Adding to that, 70% of

the data was used for model building while 30% of it was used for testing and validation purposes. As a result, classification Trees proved to have the best prediction efficiency (80% match rate) followed by ANN (79%) and, finally, by the binary Logit model (74%). Similar to the above-mentioned study, other researchers have also utilized Logit and tree models for analyzing the mode choice behaviour. Among them [27] and [28] have used tree models while [29] have used Logit models along with other models. This provides evidence about the suitability of both these models for application in the field of transportation mode choice prediction. The above-mentioned studies found a number of aspects of in which the Logit and tree models correspond to each other in terms of modelling the travel behaviour. At the same time, the Logit utility equations could provide valuable insights regarding the marginal impacts and elasticities of variables in a convenient manner. On the other hand, tree models could be useful in preparing a rule-base which could aid in developing policy guidelines.

The review of above literature shows a lack of studies related to Gulf Cooperation Council (GCC) countries, especially Bahrain. This being said, there have been studies which have emphasized the problem of traffic congestion in the country and highlighted the lack of practical steps being taken to mitigate this issue [30]. One of the major issues in this regard is the dependent on private vehicles, resulting in high car ownership rates [31]. The current study, by focusing on Bahrain, could provide a unique aspect to the literature because Bahrain has a smaller area and does not have any of the contemporary public transportation modes, which are applied elsewhere in the region. Furthermore, the above studies have mostly compared between existing mode choices, or between a future mode and other existing modes. There have not been any studies found which compare multiple future modes, especially those related to public transportation. This study is an attempt to fill these gaps.

3. DATA COLLECTION

This section provides a comprehensive overview of the data sources and collection methods used to obtain the necessary information for modelling mode choice. This includes a thorough description of the two primary data sources used in this research: data extracted from previous surveys and data collected from a questionnaire conducted as part of the current study.

3.1. Data Extracted from Previous Surveys

To gain a better understanding of the travel behaviour of a specific population, previous surveys can be an invaluable resource. With a focus on attaining the objectives of this research, information was extracted from online surveys carried out by civil engineering students at the university of Bahrain.

By looking at the responses collectively, it was concluded that information concerning gender, age, nationality, occupation, salary, driving licence and car ownership appeared in at least two of the earlier surveys. Additionally, Origin-Destination, current mode, purpose of trip, travel time, total cost of trip and respondents' willingness to use potential future public transport systems (public bus, metro, and tram) were also observed to be common. However, since each survey has a different structure and options' ranges, it was necessary to communize the responses to build a data framework that is applicable for analysis. The details of this process are demonstrated in Table 1.

Tab. 1

Standardization of data items

Data Item	Description
Gender	The letter M is used to refer to males, whereas the letter F is used for females.
Age	Based on the available spectrum of values, age is categorized to be "Under 18", "18 – 25", "26 – 35", "36 – 45" or "Above 45".
Nationality	This data item has two options: Bahraini and Non-Bahraini.
Origin-Destination	To standardize the data, corresponding governorates are used (Capital, Muharraq, Northern and Southern) mentioning the exact locations, if available, between brackets.
Occupation	The realm of options for this item consists of employee, student, retired, unemployed and others.
Average Salary	This item is of a continuous nature with values ranging from under 100 Bahraini Dinar (BD) to above 2000 BD.
Driving Licence	"Yes" is used to refer to respondents who are driving licensed, and "No" for those who are not.
Car Ownership	This item provides information about the number of cars owned by each respondent. The options are 0, 1, 2, 3, 3+.
Purpose of Trip	This item has four options, namely work, education, shopping/leisure and other.
Travel Time	Based on the spectrum of values, travel time (minutes) is categorized to be "0 – 10", "11 – 20", "21 – 30", "31 – 40". "41 – 50", "51 – 60", "61 – 120" and "120+".
Total Cost of Trip	This item is of a continuous nature, with values ranging from 0 BD to 10 BD.
Current Mode	The realm of options for this item consists of private car, sharing car, bus (public bus, school bus and private bus) and non-motorized transportation (walking and cycling).
Future Mode and Feeder	Since each survey focused on obtaining respondents' willingness to use a specific public transport mode, the following assumptions were considered when creating the compiled dataset: ▪ Respondents who currently use bus services will continue to use them in the future. ▪ Respondents who showed willingness to use public transportation in general are considered to have a positive response for all three potential modes. ▪ Feeder modes are derived from the surveys directly or from the details of the study corresponding to it; otherwise it is kept empty.

Following the compiling and standardizing process discussed above, the resulting dataset encompasses a total sample size of 3741 responses. However, the number of data available for each item ranged from, 3722 to 1343 responses with "Average Salary" having the lowest value. Insufficient data undermines the accuracy and reliability of any analysis; therefore, it was necessary to conduct an additional survey to increase the overall data and ultimately lead to more informed and effective decision-making.

3.2. Data Collected from a Recent Survey

To gather comprehensive information, the questionnaire was organised to ensure that all relevant questions were covered and effectively conveyed to the respondents. It consisted of 15 questions that can be categorized as follows:

- Commuter's Information: gender, age, nationality, occupation, driving licence and number of cars owned.
- Trip-related questions: Origin-Destination, purpose of trip, Travel time, Total cost of trip and current travel mode corresponding to the chosen purpose.
- Future public transport modes related questions: respondents were asked to select the public transportation modes (public bus, metro, tram, and none) they are willing to use in the future and how they would select to travel to and from public transport stations.

In order to obtain sufficient and diverse responses, the questionnaire was distributed through various online channels and was used as the basis for interviews done at several locations across the country. Furthermore, it is noteworthy to mention that respondents were provided with a brief description of the presented public transport modes to guarantee credible responses and that all questions were designed to be of a multiple-choice nature, with the exception of trip cost, to facilitate easy collection of data. As a result, 409 complete responses were successfully collected.

3.3. Creating the Final Dataset

Following the diligent process of extracting, compiling, and collecting data necessary to achieve the objectives of the research, a comprehensive dataset of 3864 responses was finally formed. The final dataset is a culmination of responses from multiple surveys that targeted people of different age-groups, gender, and professions, making it a rich source of information for mode choice modelling. Moreover, to ensure that the final dataset is accurate and reliable, responses lacking data concerning current mode and/or more than half of the items in question were filtered out. This step improved data performance by reducing insufficiencies in data and streamlining the process. Tab. 2 and Figs 1-2 highlight the key statistical metrics of the final dataset.

Tab. 2

Key statistical metrics of the final dataset

Data Item	Statistical Metrics
Gender	M = 51% F = 49%
Nationality	Bahraini = 61% Non-Bahraini = 39%
Age	Under 18 = 5% 18 – 25 = 37% 26 – 35 = 29% 36 – 45 = 14% Above 45 = 15%
Current Travel Mode	Car = 71% Sharing Car = 9% Bus = 19% Non-motorized Transportation = 1%

Future Mode and Feeder	Among the 2043 respondents who chose only one future mode, 759 (37%) reported they would prefer to travel by metro, while 675 (33%) and 609 (30%) stated they would choose to travel by tram and public bus respectively. 20% of bus users were also willing to use the metro. 11% of metro users expressed interest in using both tram and public bus. 10% of tram users were willing to use metro too. The most popular feeder for all public transport modes was found to be car, followed by walking and bus.

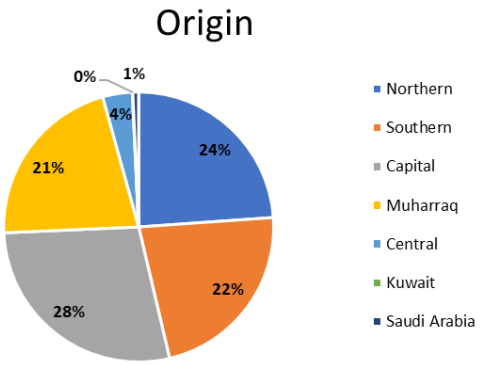

Fig. 1. Origin data percentages

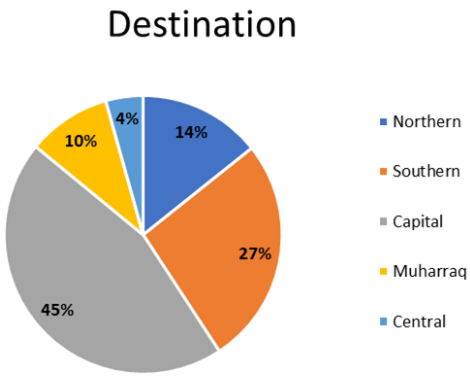

Fig. 2. Destination data percentages

4. MODEL BUILDING

A thorough process of data preparation, model creation, and validation is needed for building mode choice models. This section will discuss the approach used to create mode choice forecasting models for Bahrain, including the generation of logit and classification tree models using "Minitab Statistics" software. The section also outlines the procedure followed to confirm the validity of the dataset and the precision of the built models.

4.1. Logit model

Data preparation is an integral step in model building, as it includes transforming and restructuring raw data to attain optimal results. Accordingly, to achieve an appropriate data structure for constructing a multinomial Logit model, the data variables were classified as either continuous predictors or categorical predictors. Tab. 3 presented below displays a summary of the scales utilized for each category. Following that, data was divided into four categories, each displaying responses associated with a specific mode of transport (Car, Metro, Public Bus and Tram). For instance, to ensure accurate representation of the data, a person who showed willingness to use the tram and metro, their response was classified under both categories.

Tab. 3

Variables for modelling

Category	Variable	Scale
Continuous	Age	Continuous Number
	Travel Time (minutes)	Continuous Number
	Trip Cost (BD)	Continuous Number
	Salary (BD)	Continuous Number
Categorical	Gender	1 for Male, 0 for Female
	Nationality	1 for Bahraini, 0 for non-Bahraini
	Origin	Discrete Numbers (1 - 4) each representing a governorate
	Destination	Discrete Numbers (1 - 4) each representing a governorate
	Future Mode	Car, Tram, Public Bus, Metro
	Trip Purpose	1 for Work, 2 for Education, 3 for Shopping, 4 for Other
	Occupation	1 for Employees, 2 for Students, 3 for Others
	Driving Licence	1 for Yes, 0 for No
	Car Ownership	Discrete Numbers (1 - 4)

To develop a multinomial logit model, several models were built utilizing different variables until the best combination was identified. Then the data utilized by the optimal model was divided randomly into two datasets: training data (70%), which was used to rebuild the model, and testing data (30%) utilized to verify the accuracy of the model. The details of the model, coefficients' analysis, odds ratio and goodness-of-fit tests' results are described below.

4.1.1. Utility Equations

The selected model makes use of a total response of 1796, 1437 of which are used for model training and 359 are used for testing purposes. It employs five variables, two of which are of a continuous nature, namely travel time and age, and three categorical variables: gender, driving licence and car ownership. Adding to that, the Logit model was developed using car as the reference mode. Tabs 4-5 present the counts and percentages of each mode for each dataset.

Tab. 4

Data used for logit model training

Variable	Value	Count	Percentage
Future Mode	Car (Reference Event)	206	14%
	Tram	625	43%
	Public Bus	122	8%
	Metro	484	34%
	Total	1437	100%

Tab. 5

Data used for logit model testing

Variable	Value	Count	Percentage
Future Mode	Car (Reference Event)	51	14%
	Tram	157	44%
	Public Bus	30	8%
	Metro	121	34%
	Total	359	100%

4.1.2. Coefficients Analysis

Based on the coefficients' analysis, results demonstrated below in Tab. 6, it can be said that most of the predictors utilized in the model are statistically significant for one mode or the other. The variables which have a significant impact on all modes include car ownership, and travel time. However, it is interesting to note that specific predictors were deduced to be significant for some modes but insignificant for others. For example:

1. Gender "Male" was not significant for tram and public buses, but was significant for metro.
2. Owning a driving licence is significant for public bus only.
3. Car ownership was significant for public bus and metro, but not tram.
4. Age was significant for tram only.

Some of the variables, which were part of the survey, were not found to be significant in the model. It appears that it could be relevant to the interdependency between the variables. For example, travel time would consider the origin and destination pairs, and salary could be linked with the car-ownership. As mentioned in the literature review, the variables from the survey have been considered by other researchers in the past. More discussion about the variables and their impact is provided in the proceeding section.

Tab. 6

Coefficient analysis for multinomial logit model

Predictor	Coef	SE Coef	Z	P	Significance
Logit 1: (Tram/Car)					
Gender					
Male	-0.199	0.176	-1.130	0.258	No
Age	-0.043	0.009	-4.840	0.000	YES
TT	0.023	0.008	2.990	0.003	YES
Do you have a driving licence					YES
Yes	-0.413	0.265	-1.560	0.119	No
How many cars do you own					
1	-0.144	0.365	-0.390	0.694	No
2	0.726	0.379	1.920	0.055	No
3	0.707	0.383	1.850	0.065	No
4	0.738	0.380	1.940	0.052	No
Logit 2: (Public Bus/Car)					
Gender					
Male	0.051	0.248	0.210	0.837	No
Age	0.008	0.013	0.610	0.545	No
TT (minutes)	0.024	0.009	2.760	0.006	YES
Do you have a driving licence					
Yes	-0.864	0.345	-2.500	0.012	YES
How many cars do you own					
1	-1.221	0.401	-3.050	0.002	YES
2	-1.614	0.458	-3.520	0.000	YES
3	-2.106	0.516	-4.080	0.000	YES
4	-3.644	0.804	-4.530	0.000	YES
Logit 3: (Metro/Car)					
Gender					
Male	0.615	0.182	3.380	0.001	YES
Age	-0.015	0.010	-1.540	0.123	No
TT (minutes)	0.022	0.008	2.950	0.003	YES
Do you have a driving licence					
Yes	0.520	0.307	1.690	0.090	No
How many cars do you own					
1	-0.956	0.342	-2.800	0.005	YES
2	-1.818	0.376	-4.840	0.000	YES
3	-4.898	0.791	-6.190	0.000	YES
4	-3.075	0.428	-7.180	0.000	YES

4.1.3. Odds Ratios:

Tab. 7 displayed below highlights the odds ratios for both continuous and categorical predictors utilized in the multinomial logit model.

Tab. 7

Odds ratio multinomial logit model

Predictor	Odds Ratio	95% CI	
		Lower	Upper
Logit 1: (Tram/Car)			
Gender			
Male	0.82	0.58	1.16
Age	0.96	0.94	0.97
TT (minutes)	1.02	1.01	1.04
Do you have a driving licence			
Yes	0.66	0.39	1.11
How many cars do you own			
1	0.87	0.42	1.77
2	2.07	0.98	4.34
3	2.03	0.96	4.29
4	2.09	0.99	4.40
Logit 2: (Public Bus/Car)			
Gender			
Male	1.05	0.65	1.71
Age	1.01	0.98	1.03
TT (minutes)	1.02	1.01	1.04
Do you have a driving licence			
Yes	0.42	0.21	0.83
How many cars do you own			
1	0.29	0.13	0.65
2	0.20	0.08	0.49
3	0.12	0.04	0.33
4	0.03	0.01	0.13
Logit 3: (Metro/Car)			
Gender			
Male	1.85	1.29	2.64
Age	0.99	0.97	1.00
TT (minutes)	1.02	1.01	1.04
Do you have a driving licence			
Yes	1.68	0.92	3.07
How many cars do you own			
1	0.38	0.20	0.75
2	0.16	0.08	0.34
3	0.01	0.00	0.04
4	0.05	0.02	0.11

4.1.4. Goodness-of-fit

The results of conducting Pearson and Deviance chi-square tests on the model with respect to the training and testing datasets are demonstrated in Tab. 8 below. It is observed that both the Pearson chi-square test (p-value of 0.622) and the deviance chi-square test (p-value of 0.994) for the training data suggest good fits. However, despite the p-value of 0.849 obtained from the deviance chi-square test for the testing data, which indicates a good fit, Pearson chi-square returns a poor fit with a p-value of 0. The multinomial logit model can be concluded to be a reliable approach for forecasting outcomes for the datasets in question, but further analysis (done during validation) is essential to validate the model's accuracy.

Tab. 8

Multinomial logit model data goodness-of-fit

Data	Method	Chi-Square	P
Training	Pearson	743	0.62
	Deviance	660	0.99
Testing	Pearson	583	0.00
	Deviance	393	0.84

4.2. Classification Tree Model

For the classification tree model, no major alteration or adjustment is required to prepare the raw data for analysis. However, as for the Logit model, the data was divided into four categories, each displaying responses associated with a specific mode of transport (Car, Metro, Public Bus, and Tram). Besides, to facilitate the model building process and improve its performance, the car ownership groups (3) and (3+) were merged.

Using Minitab, a multinomial classification tree was developed having 40 terminal nodes, out of which 11 represent Car mode, 15 represent Metro mode, 5 represent Public bus and 9 represent Tram, and a misclassification cost of 0.4487. The model utilizes 4356 responses, 1357 (31.15%), 1147 (26.33%), 947 (21.74%) and 905 (20.78%) of which belong to Car, Metro, Public Bus, and tram users respectively. It consists of 40 decisive nodes and makes use of 11 important predictors, namely trip cost, car ownership, Origin-Destination, age, gender, purpose of trip, travel time, driving licence, occupation, and Salary. Fig. 3 demonstrates the relative variable importance for each predictor with respect to the top predictor, which is trip cost. It was observed that the tree model was able to highlight more complex relationships with its hierarchical form, consequently, incorporating more variables from the available set. The most significant observation from the logit model was related to the trip cost, which was the top predictor in this case, which did not have any significant impact on the logit model. This variable has been identified in a number of studies with a significant impact on the mode choice, such as Feneri et al. [32] in addition to those mentioned in the literature review.

4.3. Split-half Reliability Test

Split-half reliability is a statistical method generally applied to evaluate the internal consistency of data or measures. By randomly dividing the data into two identical parts and comparing the results, the correlation coefficient between them can be computed as a measure

of consistency and as an indicator to the extent to which the data can be generalized for larger populations [33].

To determine the internal consistency of the data used to build the Logit model, The dataset used for this model was divided into two equal samples, each containing 898 responses. The analysis covered six items, namely Future mode, travel time, gender, age, driving licence and car ownership. Using the "Spearman-Brown Formula", the adjusted correlation coefficient was calculated to be 0.71, indicating a good level of internal consistency and suggesting a strong positive relationship between the variables [34].

Similarly, the data used to develop the classification tree model was split into two subsets, each consisting of 2178 responses. The items included in the analysis were future mode, travel time, gender, origin, destination, occupation, car ownership, trip cost, driving licence and salary. The adjusted correlation coefficient was calculated to be 0.68 suggesting a moderate to high level of consistency and reliability within the data.

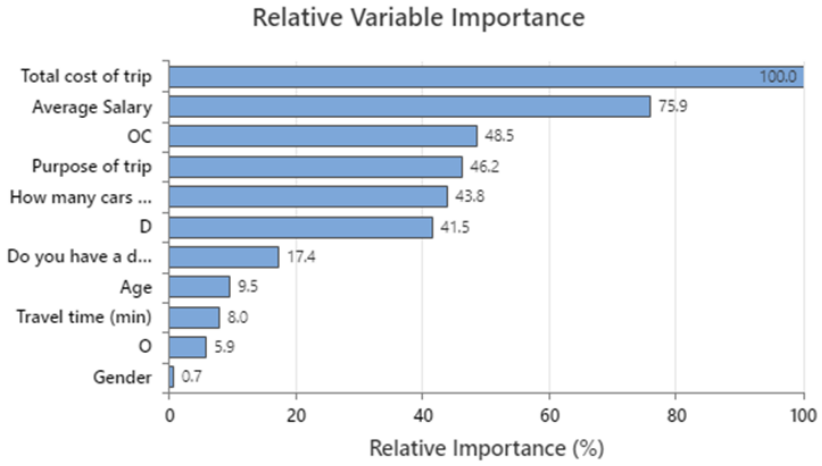

Fig. 3. Relative Variable Importance

4.4. Model Initial Validation

A rigorous procedure was employed to validate the accuracy and effectiveness of the models generated. For the multinomial logit model, the dataset was firstly divided into training and testing datasets as outlined previously, then used to compute the model's goodness-of-fit and finally the predictive performance of the model was evaluated using the AUC ROC method. The classification tree, on the other hand, was validated using the 10-fold cross-validation method.

4.4.1. Logit Model

The probabilities for each mode in both the training and testing datasets were computed and then utilized to generate ROC curves. The ROC curves were developed by calculating the true positive rates and false positive rates at probability thresholds ranging from 0-0.95 [35]. By plotting this information and fitting a mathematical equation to the curves, the AUC values were obtained (Tab. 12). In summary, the findings suggest that the model performed distinctly well in predicting the public transport modes, with high AUC values noted for both training and testing data. However, the AUC values obtained for cars were relatively low. Focusing on

public transport modes, it can be concluded that the multinomial logit model developed within this research has acceptable accuracy and a good predictive power.

4.4.2. Classification Tree

The classification tree model generated consists of 40 terminal nodes, out of which 11 represent Car mode, 15 represent Metro mode, 5 represent Public Bus and 9 represent Tram. A sample view of the detailed tree is presented in Fig. 4. It was not possible to include the entire tree, but the proceeding sections will explain and discuss the node rules of the tree. The overall misclassification error for the training and testing data were noted to be higher than expected compared to some previous studies [36], being 35.9% and 37.9% respectively. It is interesting to mention that, by studying the results, car and metro modes are more likely to be misclassified compared to other modes, with misclassification errors reaching as high as 47% (car) and 41.9% (metro). However, despite the mentioned limitations, the AUC values obtained from the analysis still indicate that the model has a fairly good predictive performance with tram having the highest value of 0.9151, followed by public bus with a value of 0.8832 and finally by car and metro having values 0.8269 and 0.8100 respectively (Tab. 12).

5. VALIDATION SURVEY

To evaluate the competency of the models developed and their accuracy in forecasting current mode choice in Bahrain, a validation survey was conducted. The survey was conducted by asking participants to provide general information about themselves, information about their usual trips and to specify their preferred future public mode of transport in general and in a hypothetical transportation scenario. Ultimately, a total of 49 diverse responses were collected.

5.1 Questionnaire Structure

The validation survey adapted a similar structure to the questionnaire described previously. However, it included an additional question where participants were asked to indicate their preferred public transport system among several hypothetical scenarios presented to them. These scenarios were fashioned based on interpretations of the models created in the second phase of the study and common practices observed around the world. All scenarios presented to the participants included pedestrian-friendly areas with wider side walks and better lighting to encourage walking, bike rental stations for those who prefer to cycle for a small fee and an app to plan and book services provided by the system. The survey included the following hypothetical scenarios:

- Metro: fee is 0.5 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates. Payment is done using a smart card system. The metro system would operate on a regular schedule, with frequent trains arriving and departing every 5-10 minutes. Parking lots near metro stations are available for short-term and long-term parking (0.2 BD per hour, first hour is free, monthly subscription available at discounted rates). Metro runs between the main areas in the country including airport, shopping malls, educational area and universities.
- Tram: fee is 0.3 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates. Payment is done using a smart card system. The trams run on dedicated tracks, making them faster than regular road transport. Parking lots near tram stations are

available for short-term and long-term parking (0.2 BD per hour, first hour is free). There are multiple tram lines in the country, each with several stops along the way. The lines link major landmarks in each city.

- Public Bus: fee is 0.3 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates. Payment is done using a smart card system. The system is designed to serve all areas of the country.
- Metro and Tram: Tram fee is 0.3 BD per trip and Metro fee is 0.5 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates to provide unlimited rides on both trams and metro trains. Payment is done using a smart card system. The trams run on dedicated tracks, making them faster than regular road transport. They act as a connecting link between the residential and commercial areas. The metro trains are the backbone of the system and run between the main districts, including the airport, educational area, and other important locations. The metro and tram system would operate on a regular schedule, with frequent trains arriving and departing every 5–10 minutes.
- Metro and Public Bus: Bus fee is 0.3 BD and Metro fee is 0.5 BD. Weekly, monthly and yearly passes are also available at discounted rates to provide unlimited rides on both public buses and metro trains. Payment is done using a smart card system. The bus routes are designed to connect neighbourhoods and feed the metro. The metro trains cover longer distances and offer quick transportation across the country. The metro and bus system would operate on a regular schedule, with frequent trains arriving and departing every 5-10 minutes.
- Metro, Tram, and Public Bus: fee is 0.4 BD per trip per mode or 1 BD daily for unlimited trips for all modes. Weekly, monthly, and yearly passes are also available at discounted rates to provide unlimited rides on all modes. Payment is done using a smart card system. The trams run on dedicated tracks and connect the downtown areas with surrounding urban neighbourhoods. While buses cover the more remote, residential, and suburban areas. The metro trains are the backbone of the system and run between the main districts, including the airport, educational area, and other important locations. The system would operate on a regular schedule, with frequent trains and buses arriving and departing every 5-10 minutes.

Although it was initially planned to replace the future mode question with the future scenarios question in the questionnaire, it was ultimately added as a separate section. This decision came because of a pilot survey that was conducted to assess whether providing additional information would impact respondents' choices. Four individuals took part in the survey and were requested to answer two separate questions. The first question asked them to state their future mode preferences simply based on minimal information, while the second question required them to choose one of the detailed hypothetical scenarios. Interestingly, the extra information presented in the scenarios altered the respondents' choices, and all four responded differently to the two questions (Tab. 9). Consequently, this led to the decision to include the scenarios as an independent section. This allows the collection of data that matches the earlier data utilized to build the models and evaluate their performance, as well as suggest a more informed transportation system through the scenarios.

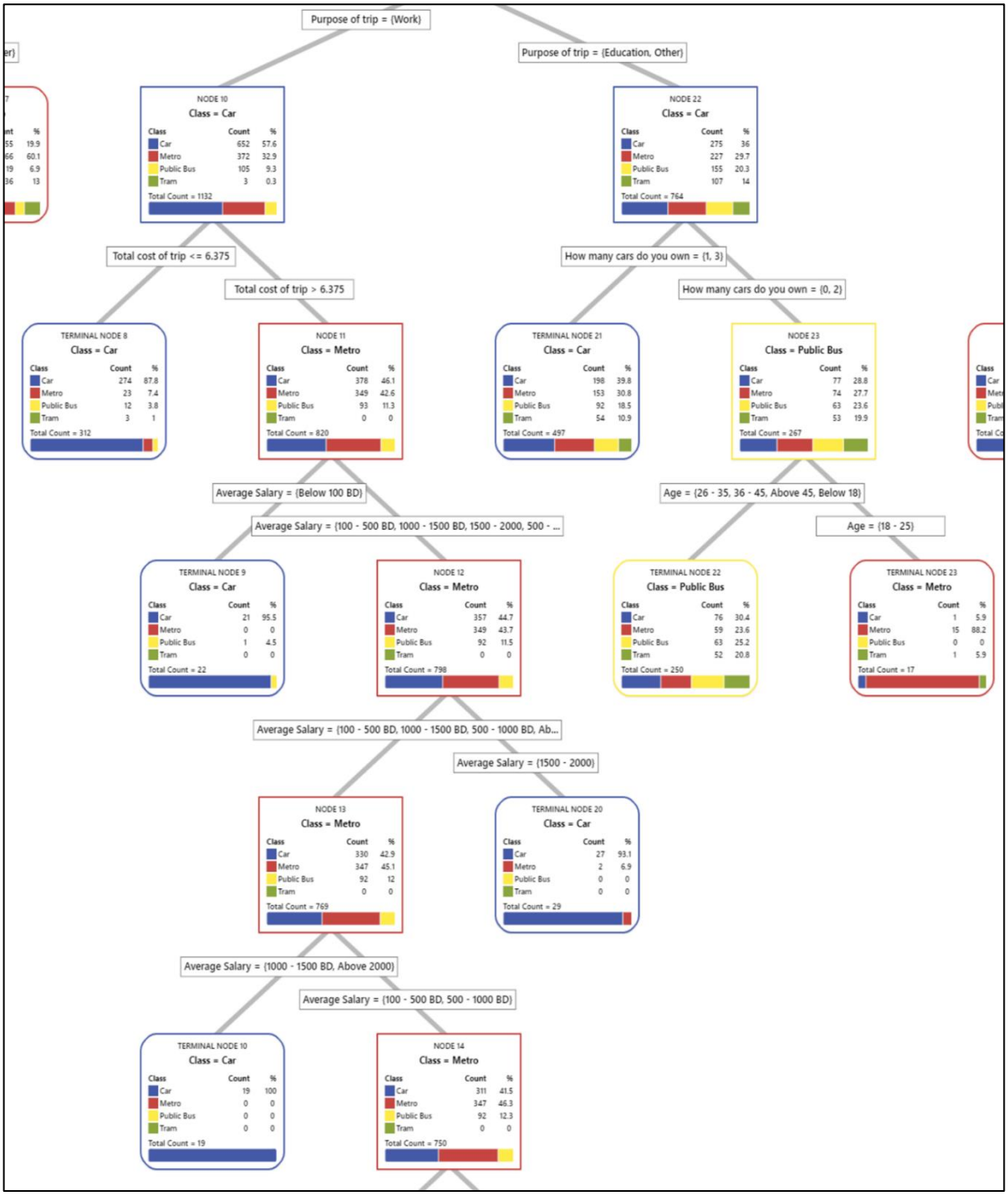

Fig. 4. A sample view of detailed tree

Tab. 9

Pilot survey results

No.	Answer to Question 1	Answer to Question 2
1	Metro and Tram	Metro, Tram and Public Bus
2	Metro	Metro, Tram and Public Bus
3	Public Bus	Metro
4	Metro and Tram	Metro, Tram, and Public Bus

Subsequently, through social media platforms like Instagram and WhatsApp and in-person interviews, a total of 49 responses were collected, 94% of which belonged to car users and 6% to public bus users. Male respondents accounted for 61% of the sample, while females constituted the remaining 39%. Besides, most participants fell into the 36-45 age group, followed by participants over 45 in age, then those aged 26-35 and lastly by those younger than 25.

5.2. Validation Survey Results

Inputting the data collected from the validation survey into the models, the classification tree outperformed the multinomial logit model with accuracy rates of 67% for the classification tree and 55% for the logit model (Tab. 10-11). It is worth noting that the models were tested to accurately predict at least one of the modes the respondents are willing to use in the future.

Tab. 10

Logit model confusion matrix (validation survey)

Actual/Predict	Car	Public Bus	Tram	Metro
Car	0	0	0	2
Public Bus	0	0	1	5
Tram	0	0	1	4
Metro	0	0	10	26
Total Positive	27			
Accuracy	55%			

Tab. 11

Classification tree model confusion matrix

Actual/Predict	Car	Public Bus	Tram	Metro
Car	1	0	0	1
Public Bus	1	3	0	2
Tram	0	0	2	3
Metro	4	5	0	27
Total Positive	33			
Accuracy	67%			

Tab. 12

Logit model vs. classification tree model

Model	AUC	Initial validation	Validation-survey Accuracy
Logit Model	Car = 0.6500 Public Bus = 0.7600 Tram = 0.8000 Metro = 0.8000	74%	55%
Classification Tree	Car = 0.8269 Public Bus = 0.8832 Tram = 0.9151 Metro = 0.8100	62%	67%

From the comparison of the models, it seems that the tree model was more robust since it provided better accuracy on the validation dataset. Moreover, it can also be observed that the logit model had more variation in its accuracies for specific modes, especially car, which is the most dominant mode. On the other hand, tree model accuracies were more consistent in this regard. However, it should be noted that the structure of the tree model was more complex than simple logit utility functions. The utility functions did not include important variables such as cost due to the statistical restrictions, but they provided insights about the impact and elasticity of different variables in a convenient manner. Hence, use of both models could be justified for mode choice studies, as they may highlight different aspects of the mode choice modelling process.

5.3 Future Preferences Scenarios Results

Upon inspecting the percentages shown in Fig. 5, it was revealed that the majority of the respondents prefer using a combination of metro and tram as their mode of transportation, accounting for 33% of the sample. Concurrently, 18% of the participants chose metro only, 17% of the respondents leaned toward a combination of metro, public bus, and tram and 15% favoured a combination of metro and public bus. However, only 11% and 4% of participants selected Tram only and public bus only, respectively. These choices indicate that direct and quick travel, accessibility, and conveniences play a significant role in influencing the choice of transport mode.

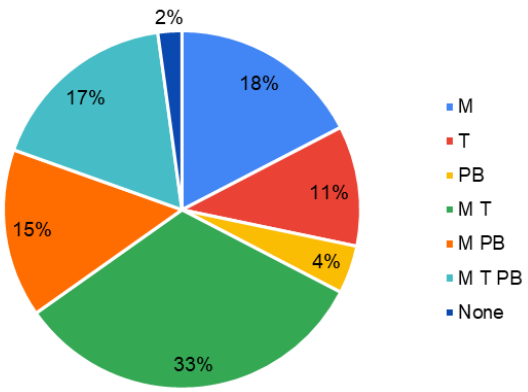

Fig. 5. Hypothetical scenarios results percentages

6. RESULTS AND DISCUSSION

6.1. Logit Model Results

Based on the odds ratios shown in Tab. 6, the following statements can be concluded regarding the variables affecting forecasting public transport mode choice (refer to Tab. 13 for summary of findings):

Tram / Car: It can be concluded that the tram is more likely to be used by younger commuters and for trips with longer travel times, as the odds of choosing it decrease by 4% with age and increase by 2% with travel time. Besides, while there is an evident effect on the odds of selecting tram over car based on gender, driving licence and car ownership, Tab. 5 states that these variables are considered statistically insignificant for tram.

Public Bus / Car: In light of the observations, it can be concluded that car ownership and possession of a driving licence are the main factors that impact deciding public bus over car for individuals in Bahrain. Commuters who own more cars are less likely to travel by public bus, while those who do not have a driving licence are more likely to use it, particularly for longer trips. This trend is also captured by [37] in Malaysia, where more than half of those who do not own a private car are regular bus riders. Moreover, although gender and age appear to have a slight effect on the odds of choosing a public bus over a car, the p-values for these variables shown in Tab. 5 suggest that they are statistically insignificant.

Metro / Car: It can be established that gender, car ownership and travel time play a prominent role in influencing commuters to opt for the metro over car for their trips. In fact, the odds ratios indicate that males, who own no, or fewer cars are more likely to pick metro as their mode of transport, especially for longer trips. Moreover, individuals with a driving licence are observed to be 68% more likely to use the metro, that being said, the possession of a driving licence is a statistically insignificant predictor for metro.

Tab. 13

Multinomial logit model findings summary

Mode	Gender (relative to F)	Age	Travel Time (minutes)	Driving Licecse (relative to no-DL)	Car Ownership (relative to no-car)
Choose Tram over Car	M is 18% less	-4%	+2%	DL is 34% less	1 car = 18% less 2 cars is 107% more 3 cars is 103% more 3+ cars is 109% more
Choose Public Bus over Car	M is 5% more	+1%	+2%	DL is 58% less	1 car = 71% less 2 cars is 80% less 3 cars is 88% less 3+ cars is 97% less
Choose Metro Car	M is 85% more	-1%	+2%	DL is 68% more	1 car = 62% less 2 cars is 84% less 3 cars is 99% less 3+ cars is 95% less

6.2. Classification Tree Model Results

The multinomial classification tree node rules were classified based on trip cost, hence providing a better understanding of its effect on forecasting public transport mode choice, in Bahrain being the top predictor.

Studying the findings, it was observed that:

1. Commuters, who declined the use of public transportation in the future, mainly own 1 car and travel for work purposes.
2. The Metro is more likely to be used by commuters who own 0 or 1 car for work and shopping trips. A study conducted in Riyadh [38] aimed at exploring the factors influencing car users to switch to the metro has indicated a similar finding. The research suggests that households owning more than 1 car are less likely to make the shift to metro usage.
3. Public bus is more probable to be used nearly equally for all purposes.
4. The tram is likely to be used for shopping trips by students or commuters who own 2 or 3 cars.

The obtained node rules from the multinomial classification tree can support the development of future public transport systems that would effectively serve commuters in Bahrain. Furthermore, these findings could also assist in planning and designing the frequency, routes, fare, and other operational parameters of public transport systems, based on the most likely users and purposes shown for each mode by the models in this study. Besides, it would aid in understanding the complex travel patterns and interrelations between modes better and ultimately result in policies and measures that would attract more public transport ridership in the future if implemented.

7. FUTURE POLICY GUIDELINES

The findings from both models can be combined to conclude some important policy measures, which are as follows. Implementation of Metro seems to be appealing for people who have driving licences, and own cars, and these travellers are willing to use this mode for their work and shopping trips. Previous literature clearly shows that car-ownership [39] and work trips [40] are major causes of recurrent congestion. Hence, it is expected that the metro would be highly effective in reducing congestion on the roads, especially during peak hours. The reason for the preference of metro could be related to its implementation and popularity in neighbouring countries, such as Dubai, Qatar, and Saudi Arabia. Hence, another recommendation which could be drawn is that travellers are more willing to shift to modes which are popular in their region.

The tram seemed to be the second most preferred mode, especially for shopping trips. In all cases, public transportation choice is more likely for longer travel times. In the case of Bahrain, which has a relatively smaller area, this could relate to congested commercial areas and the time taken to find the parking space. Hence, public transportation modes are more likely to be adopted for commercial and business-related areas. Another common trend is the lesser probability of choosing public transportation modes by people having a license and/or car, which is also intuitive as these items provide more flexibility to the travellers in terms of schedule of their trip and will not include any walking or waiting time. Hence, stricter policy measures could be required to reduce the tendency for acquiring licences and owning a car.

These measures could include enforcing stricter regulations for licences and vehicle ownership and increasing the monetary requirements of these processes. The later will also affect the travel cost, which is the most influential factor on mode choice found in this study, as well as in other regions [32]. However, it should be noted that these policy measures could prove to create counterproductive without the provision of efficient, and convenient alternative modes of public transportation.

8. CONCLUSIONS AND RECOMMENDATIONS

To conclude, this research has concentrated on putting forward the demanding need for the development of effective forecasting models for public transport mode choice. Through the utilization of two types of models, namely the logit model and the classification tree model, and statistical techniques such as AUC ROC method, this study has investigated the most influential factors that shape commuters' behaviour and choices.

The findings propose critical contributions to the process of constructing solid recommendations for a public transport system that would lead to improved transportation conditions in the Kingdom of Bahrain and other similar situations in the Gulf region and/or with small size countries. This section summarizes the major conclusions obtained from the research, which are believed to be a valuable extension to the existing literature on transportation planning in Bahrain and provide significant information to policymakers and transportation planners worldwide.

The primary outcomes derived from this research regarding mode choice and public transportation in Bahrain as per the available dataset can be distilled in the following points:

1. Trip cost is identified to be the key predictor governing future mode choice in Bahrain.
2. Commuters with high socio-economic status are more likely to travel by car than use bus services.
3. Bus services are recognized to be mainly used by students for education-related trips.
4. Metro is identified to be the top choice of commuters for future public transportation, whether as the sole mode or in connection with tram, public bus or both, with a stronger preference for the metro and tram combination.
5. The tram is established as an immensely renowned mode of transportation for trips undertaken for shopping and leisure purposes.
6. Interestingly, commuters who own multiple cars are more likely to choose tram as their preferred mode of transportation than those who own none or one only.
7. The research indicates that direct and quick travel, accessibility, and conveniences play a significant role in influencing the choice of transport mode.
8. When it comes to capturing more complex relationships and multinomial responses, the classification tree model surpasses the multinomial logit model.

Based upon the findings of this research, it is recommended to prioritize the implementation of public transportation modes which are popular in neighbouring countries, such as the metro. The increase in ridership of such modes could be increased when coupled with policy measures to discourage acquisition of driving licences and car-ownership through the use of regulatory and financial restrictions.

Various potential research avenues could be investigated to enhance our understanding of commuters' mode choice behaviour in Bahrain. Principally, the following recommendations can be considered as possible schemes for future research. Researchers can investigate

the influence of other factors, especially qualitative, on mode choice such as comfort, safety, and network characteristics. Researchers can also explore the impact of transport measures and policies, such as public transport subscriptions, road pricing, parking fees and congestion charges on mode choice. Furthermore, the interactions between transportation and health, climate change and energy can also be investigated. Lastly, the effect of urban design, such as lighting, pedestrian and cycling friendly streets, on the mode choice of commuters in Bahrain could also be explored.

References

1. Davis K. 2015. "The urbanization of the human population". In: *The City Reader*. Chapter: 43-53. Routledge.
2. Opoku O., O. Yeboah, E. Gyamfi, G. Afful. 2022. „Rising car ownership and traffic congestion in the university of cape coast campus". *International Journal of Research in Science & Engineering* 24: 10-21.
3. Ding C., Y. Wang, T. Tang, S. Mishra, C. Liu. 2018. „Joint analysis of the spatial impacts of built environment on car ownership and travel mode choice". *Transportation Research Part D: Transport and Environment* 60: 28-40.
4. Hamed M.M., B.M. Al-Eideh. 2020. „An exploratory analysis of traffic accidents and vehicle ownership decisions using a random parameters logit model with heterogeneity in means". *Analytic Methods in Accident Research* 25: 100116.
5. Kalmykova Y., L. Rosado, J. Patrício. 2016. „Resource consumption drivers and pathways to reduction: economy, policy and lifestyle impact on material flows at the national and urban scale". *Journal of Cleaner Production* 132: 70-80.
6. Hamurcu M., T. Eren. 2020. „Strategic planning based on sustainability for urban transportation: An application to decision-making". *Sustainability* 12(9): 3589.
7. Ojo T.K. 2019. „Quality of public transport service: An integrative review and research agenda". *Transportation Letters* 11(2): 104-116.
8. Alotaibi O., D. Potoglou. 2017. „Perspectives of travel strategies in light of the new metro and bus networks in Riyadh City, Saudi Arabia". *Transportation Planning and Technology* 40(1): 4-27.
9. Mushtaha E.S., R. Nahlé, N. Tahmaz, M. AlKadry. 2019. „Designing guidelines for metro stations in developing countries: The case of Dubai". *International Review for Spatial Planning and Sustainable Development* 7(4): 83-96.
10. AlKhereibi A., M. AlSuwaidi, R. Al-Mohammed, S. Pokharel, M. Ayari. 2021. "An integrated urban-transport smart growth model around metro stations: A case of Qatar". *Transportation Research Interdisciplinary Perspectives* 10: 100392.
11. Al Suleiman S., A. Cortez, A. Monzón, A. Lara. 2023. "How to improve public transport usage in a medium-sized city: key factors for a successful bus system". *European Transport Research Review* 15(1): 47.
12. Alaali F., H. Naser. 2020. "Economic development and environmental sustainability: evidence from Bahrain". *Energy, Ecology and Environment* 5: 211-219.
13. Sowjanya D., D. Tahlyan, C.R. Sekhar. 2014. "Travel demand modelling for a metropolitan city". In: *International Conference on Recent Trends and Challenges in Civil Engineering*: 19-40.
14. Waghmare A., G. Yadav, K. Tiwari. 2022. "Four step travel demand modelling for urban transportation planning". *Sci. Eng. Technol.* 5: 1254.

15. Xintong M. 2021. “A critical overview of four-stage model under the background of the rise of ride-sharing”. In: *E3S Web of Conferences* 253: 03020. EDP Sciences.
16. Nesamani K.S., J.D. Saphores, M.G. McNally, R. Jayakrishnan. 2017. “Estimating impacts of emission specific characteristics on vehicle operation for quantifying air pollutant emissions and energy use”. *Journal of Traffic and Transportation Engineering* 4(3): 215-229.
17. Modi K.B., L.B. Zala, F.S. Umrigar, T.A. Desai. 2011, May. “Transportation planning models: a review”. In: *National Conference on Recent Trends in Engineering and Technology*. Gujarat India.
18. Hillel I., E. Gazit, A. Nieuwboer, L. Avanzino, L. Rochester, A. Cereatti, J.M. Hausdorff. 2019. “Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring”. *European Review of Aging and Physical Activity* 16(1): 1-12.
19. Pineda-Jaramillo J. D. 2019. “A review of Machine Learning (ML) algorithms used for modelling travel mode choice”. *Dyna* 86(211): 32-41.
20. Panter J., E. Heinen, R. Mackett, D. Ogilvie. 2016. “Impact of new transport infrastructure on walking, cycling, and physical activity”. *American Journal of Preventive Medicine* 50(2): e45-e53.
21. Pham B.T., T.V. Phong, H.D. Nguyen, C. Qi, N. Al-Ansari, A. Amini, A., D. Tien Bui. 2020. “A comparative study of kernel logistic regression, radial basis function classifier, multinomial naïve bayes, and logistic model tree for flash flood susceptibility mapping”. *Water* 12(1): 239.
22. Puan O.C., Y.A.H. Hassan, N. Mashros, M.K. Idham, N.A. Hassan, M.N.M. Warid, M.R. Hainin. 2019. “Transportation mode choice binary logit model: A case study for Johor Bahru city”. In: *IOP Conference Series: Materials Science and Engineering* 527(1): 012066. IOP Publishing.
23. Javaid A., F. Creutzig, S. Bamberg. 2020. “Determinants of low-carbon transport mode adoption: systematic review of reviews”. *Environmental Research Letters* 15(10): 103002.
24. Abdullah M., N. Ali, M.A. Javid, C. Dias, T. Campisi. 2021. “Public transport versus solo travel mode choices during the COVID-19 pandemic: Self-reported evidence from a developing country”. *Transportation Engineering* 5: 100078.
25. Oral O.L., V.A.H.A.P. Tecim. 2013. “Using decision trees for estimating mode choice of trips in Buca-Izmir”. *The International Archives of the Photogrammetry,* Remote Sensing and Spatial Information Sciences 40: 139-145.
26. Lindner A., C.S. Pitombo, A.L. Cunha. 2017. “Estimating motorized travel mode choice using classifiers: An application for high-dimensional multicollinear data”. *Travel Behaviour and Society* 6: 100-109.
27. Sekhar C.R., E. Madhu. 2016. “Mode choice analysis using random forest decision trees”. *Transportation Research Procedia* 17: 644-652.
28. Hagenauer J., M. Helbich. 2017. “A comparative study of machine learning classifiers for modelling travel mode choice”. *Expert Systems with Applications* 78: 273-282.
29. Zhao X., X. Yan, A. Yu, P. Van Hentenryck. 2020. “Prediction and behavioural analysis of travel mode choice: A comparison of machine learning and logit models”. *Travel Behaviour and Society* 20: 22-35.
30. Jedidi A., A. Mahdi, A.J. Ali. 2021. “Traffic congestion in the kingdom of Bahrain: social mobile application solution”. In: *3rd Smart Cities Symposium (SCS 2020)* 2021: 337-343. DOI: 10.1049/icp.2021.0883.

31. Kumar M., A.S. Alsharqi, M.A. Al-Romaihi. 2021, December. “Consumers' choice of vehicle parameters and their impact on environment in relation to gasoline price increase in Bahrain”. In: *2021 International Conference on Decision Aid Sciences and Application (DASA)*: 792-796. IEEE.
32. Feneri A.M., S. Rasouli, H.J. Timmermans. 2022. “Modelling the effect of Mobility-as-a-Service on mode choice decisions”. *Transportation Letters* 14(4): 324-331.
33. Steinke A., B. Kopp. 2020. “RELEX: An Excel-based software tool for sampling split-half reliability coefficients”. *Methods in Psychology* 2: 100023.
34. de Vet H.C., L.B. Mokkink, D.G. Mosmuller, C.B. Terwee. 2017. “Spearman-Brown prophecy formula and Cronbach's alpha: different faces of reliability and opportunities for new applications”. *Journal of Clinical Epidemiology* 85: 45-49.
35. Fawcett T. 2008. “PRIE: a system for generating rulelists to maximize ROC performance”. *Data Mining and Knowledge Discovery* 17: 207-224.
36. Xiao Z., Y. Wang, K. Fu, F. Wu. 2017. “Identifying different transportation modes from trajectory data using tree-based ensemble classifiers”. *ISPRS International Journal of Geo-Information* 6(2): 57.
37. Chee W.L., J.L. Fernandez. 2013. “Factors that influence the choice of mode of transport in Penang: A preliminary analysis”. *Procedia-Social and Behavioural Sciences* 91: 120-127.
38. Anwar A.M., A.T. Oakil, A. Muhsen, A. Arora. 2023. “What would it take for the people of Riyadh city to shift from their cars to the proposed metro?”. *Case Studies on Transport Policy* 12: 101008.
39. Rahman M.M., P. Najaf, M.G. Fields, J.C. Thill. 2022. “Traffic congestion and its urban scale factors: Empirical evidence from American urban areas”. *International Journal of Sustainable Transportation* 16(5): 406-421.
40. Samal S.R., P.G. Kumar, J.C. Santhosh, M. Santhakumar. 2020, December. “Analysis of traffic congestion impacts of urban road network under Indian condition”. In: *IOP Conference Series: Materials Science and Engineering* 1006(1): 012002. IOP Publishing.

Appendix: Survey Questionnaire

Forecasting Public Transport Mode Choice in Bahrain

Kindly fill this questionnaire which is conducted by a Civil Engineering student in university of Bahrain, supervised by Dr. Uneb Gazder for a Master's thesis. The obtained results will benefit the research which will hopefully contribute in recommending a public transport system that will effectively serve the commuters in the Kingdom of Bahrain and ultimately reduce traffic congestion.

This will take few minutes, thank you for sharing your precious time.

يرجى ملء هذا الاستبيان الذي يتم إجراؤه من قبل طالبة هندسة مدنية في جامعة البحرين ، بإشراف د. أونيب قازدير للحصول على رسالة ماجستير. ستساعد نتائج هذا الاستبيان في البحث الذي نأمل أن يساهم في التوصية بنظام نقل عام يخدم بشكل فعال الركاب في مملكة البحرين ويقلل من الازدحام المروري

سيستغرق هذا بضع دقائق ، شكرًا لك على مشاركتنا وقتك الثمين.

marwah.jazi@gmail.com Switch account

Not shared

* Indicates required question

1. Gender / الجنس *

2. Age / العمر *

- Under 18 / أقل من 18
- 18 - 25
- 26 - 35
- 36 - 45
- 45 +

3. Nationality / الجنسية *

- Bahraini / بحريني
- Non - Bahraini / غير بحريني

4. Occupation / المهنة *

- ◯ Student / طالب
- ◯ Employee / موظف
- ◯ Unemployed / عاطل عن العمل
- ◯ Retired / متقاعد
- ◯ Other / أخرى

5. Do you have a driving licence? / هل لديك رخصة قيادة؟ *

- ◯ Yes / نعم
- ◯ No / لا

6. How many cars do you own? / كم عدد السيارات التي تمتلكها؟ *

- ◯ 0
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 3+

7. How much is your salary? / كم هو راتبك؟ *

- ◯ Under 100 BD
- ◯ 100 - 500 BD
- ◯ 500 - 1000 BD
- ◯ 1000 - 1500 BD
- ◯ 1500 - 2000 BD
- ◯ 2000 + BD

8. Where does your trip usually originate from? / من أين تبدأ رحلتك عادة؟ *

- ◯ Northern Governorate / المحافظة الشمالية
- ◯ Capital Governorate / محافظة العاصمة
- ◯ Southern Governorate / المحافظة الجنوبية
- ◯ Muharraq Governorate / محافظة المحرق

9. Purpose of this trip / الغرض من هذه الرحلة *

- ◯ Work / العمل
- ◯ Education / التعليم
- ◯ Shopping/Leisure / التسوق / الترفيه
- ◯ Other / أخرى

10. Where is your destination for this trip located at? / أين تقع وجهتك لهذه الرحلة؟ *

- ◯ Northern Governorate / المحافظة الشمالية
- ◯ Capital Governorate / محافظة العاصمة
- ◯ Southern Governorate / المحافظة الجنوبية
- ◯ Muharraq Governorate / محافظة المحرق

11. Travel time: How long does this trip take in minutes? / كم تستغرق هذه الرحلة بالدقائق؟ *

- ◯ 0 - 10
- ◯ 11 - 20
- ◯ 21 - 30
- ◯ 31 - 40
- ◯ 41 - 50
- ◯ 51 - 60
- ◯ 61 - 120
- ◯ 120 +

12. What is the average cost of this trip in BD? / * ما هو متوسط تكلفة هذه الرحلة بالدينار بحريني؟

Your answer

13. What is your current mode of travel? / ما هي وسيلة النقل الحالية الخاص بك؟ *

- ◯ Car / السياره
- ◯ Sharing Car / مشاركة السياره
- ◯ Public Bus / الحافلة العامة
- ◯ Walking / المشي
- ◯ Cycling / ركوب الدراجات
- ◯ Other:

Future Public Transport Modes

14. For the purpose of trip specified before: Which of the following public transport modes are you most likely to use in the future if they are made available in Bahrain? Tick all possible answers. / لغرض الرحلة المحدد من قبل: أي من وسائل النقل العام التالية من المرجح أن تستخدمها في المستقبل إذا كانت متوفره في مملكة البحرين؟ ضع علامة على جميع الإجابات الممكنة. *

☐ Public Bus / الحافلة العامة

☐ Metro: A rapid city transit system that primarily or traditionally runs below the surface and that operate on an exclusive right-of-way, which cannot be accessed by pedestrians or other vehicles. Average speed is 100 - 160 km/hr. / مترو: نظام عبور سريع داخل المدينة يعمل بشكل أساسي أو تقليدي تحت السطح ويعمل على حق طريق حصري ، والذي لا يمكن للمشاه أو المركبات الأخرى الوصول إليه . متوسط السرعة 100 - 160 كم / ساعة.

☐ Tram: A light rail public transportation mode that runs in tramway tracks along urban streets. It uses electrical power, travels at an average speed of 60-50 km/hr on a specialized lane or a highway's median and is given priority at intersections. / الترام: وسيلة نقل عام بالسكك الحديدية الخفيفة تعمل في مسارات الترام. تستخدم الطاقة الكهربائية ، وتنتقل بمتوسط سرعة 60-50 كم / ساعة على ممر متخصص أو متوسط طريق سريع وتعطى الأولوية عند التقاطعات.

☐ None / لا شيء مما ذكر

15. With reference to your answer to the above question: How would you choose to travel to and from public transport stations? Tick all possible answers. / بالإشارة إلى إجابتك على السؤال أعلاه: كيف تختار السفر من وإلى محطات النقل العام؟ ضع علامة على جميع الإجابات الممكنة. *

☐ Car / السيارة

☐ Bus / الحافلة

☐ Walking / المشي

☐ Cycling / ركوب الدراجات

☐ None / لا شيء مما ذكر

Future Scenarios

15. Based on a previous survey analysis' result, the following scenarios (Public Transportation Systems) are obtained. Each Scenario will include pedestrian-friendly areas with wider sidewalks and better lighting to encourage walking, bike rental stations for those who prefer to cycle for a small fee and an app to plan and book services provided by the system.
Which one of them would you prefer to use?

بناءً على نتائج تحليل استبيان سابق ، تم استنتاج .السيناريوهات التالية
سيشمل كل سيناريو مناطق صديقة للمشاة مع أرصفة أوسع وإضاءة أفضل لتشجيع المشي ومحطات تأجير الدراجات لأولئك الذين يفضلون ركوب الدراجات مقابل رسوم رمزية وتطبيق هاتف لتخطيط وحجز الخدمات التي يوفرها النظام . أي منهم تفضل استخدامه؟

- ◯ Metro: fee is 0.5 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates. Payment done using smart card system. The metro system would operate on a regular schedule with frequent trains arriving and departing every 5 - 10 minutes. Parking lots near metro stations are available for short-term and long-term parking (0.2 BD per hour, first hour is free, monthly subscription available at discounted rates). Metro run between the main areas in the country including airport, shopping malls, educational area and universities.

- ◯ المترو: الرسوم 0.5 دينار بحريني في اليوم. تتوفر أيضًا بطاقات أسبوعية وشهرية وسنوية بأسعار مخفضة. يتم الدفع باستخدام نظام البطاقة الذكية. سيعمل نظام المترو وفقًا لجدول زمني منتظم مع وصول القطارات المتكرره ومغادرتها كل 5-10 دقائق. تتوفر مواقف للسيارات بالقرب من محطات المترو لمواقف السيارات لفترات قصيره وطويلة المدى (0.2 BD للساعة، الساعة الأولى مجانا، يتوفر اشتراك شهري بسعر مخفض). يعمل المترو بين المناطق الرئيسية في المملكة بما في ذلك المطار ومراكز التسوق والمنطقة التعليمية والجامعات.

- ◯ Tram: fee is 0.3 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates. Payment done using smart card system. The trams run on dedicated tracks making them faster than regular road transport. Parking lots near tram stations are available for short-term and long-term parking (0.2 BD per hour, first hour is free). There are multiple tram lines in the country each with several stops along the way. The lines link major landmarks in each city.

- ◯ الترام: تبلغ الرسوم 0.3 دينار بحريني للرحلة. تتوفر أيضًا بطاقات أسبوعية وشهرية وسنوية بأسعار مخفضة. يتم الدفع باستخدام نظام البطاقة الذكية. يعمل الترام على مسارات مخصصة مما يجعلها أسرع من النقل العادي. تتوفر مواقف للسيارات بالقرب من محطات الترام لمواقف السيارات لفترات قصيره وطويلة (0.2 BD للساعة، الساعة الأولى مجانا). هناك العديد من خطوط الترام في الدولة مع توقفات عديده على طول الطريق. تربط الخطوط بين المعالم الرئيسية في كل مدينة.

- ◯ Public Bus: fee is 0.3 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates. Payment done using smart card system. The system is designed to serve all areas of the country.

- ◯ الحافلة العامة: الرسوم 0.3 دينار بحريني للرحلة. تتوفر أيضًا بطاقات أسبوعية وشهرية وسنوية بأسعار مخفضة. يتم الدفع باستخدام نظام البطاقة الذكية. النظام مصمم لخدمة جميع مناطق الدولة.

Metro and Tram: Tram fee is 0.3 BD per trip and Metro fee is 0.5 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates to provide unlimited rides on both trams and metro trains. Payment done using smart card system. The trams run on dedicated tracks making them faster than regular road transport. They act as a connecting link between the residential and commercial areas. The metro trains are the backbone of the system and run between the main districts, including airport, educational area, and other important locations. The metro and tram system would operate on a regular schedule with frequent trains arriving and departing every 5 - 10 minutes.

/المترو والترام: رسوم الترام 0.3 دينار بحريني للرحلة ورسوم المترو 0.5 دينار بحريني للرحلة. تتوفر أيضًا تذاكر أسبوعية وشهرية وسنوية بأسعار مخفضة لتوفير رحلات غير محدوده في كل من الترام وقطارات المترو. يتم الدفع باستخدام نظام البطاقة الذكية. يعمل الترام على مسارات مخصصة مما يجعلها أسرع من النقل العادي ويعمل كحلقة وصل بين المناطق السكنية والتجارية. قطارات المترو هي العمود الفقري للنظام وتعمل بين المناطق الرئيسية ، بما في ذلك المطار والمنطقة التعليمية والمواقع المهمة الأخرى. سيعمل نظام المترو و الترام وفقًا لجدول زمني منتظم مع وصول القطارات المتكرره ومغادرتها كل 5-10 دقائق.

Metro and Public Bus: Bus fee is 0.3 BD and Metro fee is 0.5 BD. Weekly, monthly and yearly passes are also available at discounted rates to provide unlimited rides on both public buses and metro trains. Payment done using smart card system. The bus routes are designed to connect neighbourhoods and feed the metro. The metro trains cover longer distances and offer quick transportation across the country. The metro and bus system would operate on a regular schedule with frequent trains arriving and departing every 5 - 10 minutes.

المترو والحافلات العامة: رسوم الحافلات 0.3 دينار بحريني ورسوم المترو 0.5 دينار بحريني. تتوفر أيضًا تذاكر أسبوعية وشهرية وسنوية بأسعار مخفضة لتوفير رحلات غير محدوده في كل من الحافلات العامة وقطارات المترو. يتم الدفع باستخدام نظام البطاقة الذكية. تم تصميم خطوط الحافلات لربط الأحياء وتغذية المترو. قطارات المترو تغطي مسافات أطول وتوفر النقل السريع عبر البلاد. سيعمل نظام المترو والحافلة وفقًا لجدول زمني منتظم مع وصول القطارات المتكرره ومغادرتها كل 5-10 دقائق.

Metro, Tram, and Public Bus: fee is 0.4 BD per trip per mode or 1 BD daily for unlimited trips for all modes. Weekly, monthly, and yearly passes are also available at discounted rates to provide unlimited rides on all modes. Payment done using smart card system. The trams run on dedicated tracks and connect the downtown areas with surrounding urban neighbourhoods. While buses cover the more remote, residential, and suburban areas. The metro trains are the backbone of the system and run between the main districts, including airport, educational area, and other important locations. The system would operate on a regular schedule with frequent trains and buses arriving and departing every 5 - 10 minutes.

المترو والترام والحافلات العامة: تبلغ الرسوم 0.3 دينار بحريني لكل رحلة لكل وسيلة أو 1 دينار بحريني يوميًا للرحلات غير المحدوده لجميع وسائل النقل العام . تتوفر أيضًا تذاكر أسبوعية وشهرية وسنوية بأسعار مخفضة لتوفير رحلات غير محدوده في جميع الأوضاع. يتم الدفع باستخدام نظام البطاقة الذكية. تعمل عربات الترام على مسارات مخصصة وتربط مناطق وسط المدينة بالأحياء الحضرية المحيطة. بينما تغطي الحافلات المناطق النائية والسكنية والضواحي. قطارات المترو هي العمود الفقري للنظام وتعمل بين المناطق الرئيسية ، بما في ذلك المطار والمنطقة التعليمية والمواقع المهمة الأخرى. سيعمل النظام وفقًا لجدول زمني منتظم مع وصول القطارات والباصات المتكرره ومغادرتها كل 5-10 دقائق. تتوفر مواقف للسيارات بالقرب من محطات المترو والترام لمواقف السيارات لفترات قصيره وطويلة.

None / لا شيء مما ذكر

Received 10.01.2024; accepted in revised form 15.05.2024