

University of Bahrain

College of Engineering



**Department of Civil
Engineering**

**Forecasting Public Transport Mode Choice Behavior
of Commuters in Bahrain Using Logit and
Classification Tree Models: A Comparative Study**

**A Thesis Submitted in Partial Fulfilment of the Requirements for the
M.Sc. Degree in Civil Engineering**

Submitted by:

Marwa Jazi Gharaibeh

20152606

Supervised by:

Dr Uneb Gazder

University of Bahrain

September 2023



**University of Bahrain
College of Engineering
Civil Engineering Department**

Forecasting Public Transport Mode Choice Behavior of Commuters in Bahrain Using Logit and Classification Tree Models: a Comparative Study

**A Thesis Submitted in Partial Fulfilment of the Requirements for the
M.Sc.Degree in Civil Engineering**

Submitted by:

Marwa Jazi Gharaibeh

20152606

Supervised by:

Dr Uneb Gazder

University of Bahrain

September 2023



قرار لجنة مناقشة أطروحة

اسم الطالب: _____ الرقم الجامعي: _____
اسم الكلية: _____ القسم: _____
اسم البرنامج: _____
عنوان الأطروحة: _____

اجتمعت لجنة المناقشة يوم _____ الموافق _____ بقاعة رقم _____ وبعد مناقشة الطالب المذكور أعلاه علناً توصي لجنة المناقشة بالتالي:

يمنح الطالب تقدير: _____ في مقرر الأطروحة: _____

- ☐ دون إجراء أية تعديلات على الأطروحة.
☐ على أن يتم إجراء التعديلات المرفقة مع هذا القرار.

تصديق أعضاء لجنة المناقشة:

الاسم	المشرف على الأطروحة	التوقيع
1. _____	الممتحن الداخلي	_____
2. _____	الممتحن الخارجي	_____
3. _____		_____

Abstract

Transportation plays a key role in facilitating trade and social interactions worldwide. However, transportation systems in most countries, including Bahrain, face various serious challenges. This has made it essential to have a thorough comprehension of travel demands and driver's characteristics for the development of engineering solutions that can facilitate the advancement of these transportation systems. This thesis 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, with the model building process conducted in two phases. The first phase concentrated on modeling the current transportation system considering the modes of transportation currently available (Car and Bus), while phase 2 focused on modelling the future transportation system, taking into account potential public transport modes (Car, 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 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 the logit model demonstrated higher accuracy compared to the classification tree when modelling binary responses. However, when it comes to capturing more complex patterns, as in the second phase, the classification tree outperforms the multinomial logit model. Overall, the thesis provides valuable insights into mode choice in Bahrain and highlights the important factors influencing commuting decisions. The results obtained 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.

Dedication

I would love to dedicate this thesis to my parents, specifically, and my siblings. Your consistent support and encouragement have been the greatest reinforcement in my academic journey. Thank you for always being there for me, for pushing me to aim for the best and for the continuous unconditional love.

I would also like to dedicate this thesis to the residents of Bahrain. Thank you for participating in the surveys and ultimately making this research possible. It is my sincere hope that the findings of this thesis can help improve the transportation system in Bahrain and provide its residents with a higher quality of commuting.

This thesis is dedicated to everyone who played a part, small or big, in making this research complete. Thank you for your support and inspiration.

Acknowledgement

All praise and thanks belong to Allah for the completion of this work.

I would like to express a note of gratitude to my supervisor, Dr. Uneb Gazder, for providing me with guidance and unwavering support during the preparation of my thesis. I am deeply grateful for your continuous encouragement, invaluable advice and the countless hours you employed to read drafts, giving constructive comments and support.

My sincere appreciation goes to the staff and faculty of University of Bahrain for providing a great academic environment that fosters intellectual growth and support research endeavors.

I am forever grateful to my family, specifically my parents, for the boundless love and continuous support.

Lastly, I deeply hope that this thesis will support in a way the advancement of knowledge in my field.

Table of Contents

.....	
Abstract	B
Dedication	C
Acknowledgement	D
Table of Contents	E
List of Tables	H
List of Figures	J
1 Chapter One :Research Background	1
1.1 Background	2
1.2 Problem Statement	3
1.3 Research Questions	4
1.4 Objectives of the Study	4
1.5 Significance of the Study	5
1.6 Limitations of the Study	5
1.7 Summary of Chapters	6
2 Chapter Two: Literature Review.....	8
2.1 Introduction	9
2.2 Factors Affecting Travel Demand.....	9
2.2.1 Travel demand and mode choice	9
2.2.2 Travel demand and vehicle ownership.....	10
2.2.3 Travel demand and land use.....	10
2.3 Travel Demand Modelling	11
2.3.1 Trip Generation	12
2.3.2 Trip Distribution	13
2.3.3 Modal Split.....	13
2.3.4 Traffic Assignment	13
2.4 Mode Choice Modelling.....	14
2.4.1 Discrete Choice Models	15
2.4.2 Machine Learning/Artificial Intelligence Models.....	19
2.5 Implementation of Logit Model and Decision Trees in Mode Choice Modeling.....	22
2.5.1 Logit Model for Mode Choice Modelling.....	22
2.5.2 Decision Trees for Mode Choice Modelling.....	24
2.6 Chapter Summary.....	25
3 Chapter Three: Data Collection	27
3.1 Introduction	28

3.2	Data Extracted From Previous Surveys.....	28
3.2.1	Survey 1: Travel Demand Forecasting Using Traditional Four Steps Transportation Modelling Approach: A Case Study of Bahrain	28
3.2.2	Survey 2: Measures to Increase Ridership for Public Transport in Kingdom of Bahrain	29
3.2.3	Survey 3: Evaluation of a Proposed Tram Network Linking Manama’s Main Recreational Centers Through the Implementing of VISSIM Simulation Software	30
3.2.4	Survey 4: Optimizing Routes for Public Transport in Bahrain.....	31
3.2.5	Survey 5: Design Of School Metro.....	31
3.2.6	Survey 6: Determination of Travel Demand for Design of Traffic Facilities for Schools in Bahrain	32
3.2.7	Survey 7: Design Of a Suitable Feeder System For Metro Bahrain	32
3.3	Compiling Responses Extracted from Previous Surveys	33
3.4	Data Collected from a Recent Survey	37
3.4.1	Structure of the Questionnaire	38
3.4.2	Methodology Used to Administrate the Survey.....	38
3.5	Creating the Final Dataset	40
3.5.1	Sample Size.....	41
3.6	Chapter summary	42
4	Chapter Four: Research Methodology	43
4.1	Introduction	44
4.2	Statistical Analysis	44
4.2.1	Commuters’ General Information:.....	45
4.2.2	Trip-related Items:.....	49
4.2.3	Future Mode-related Items.....	54
4.3	Model Building.....	58
4.3.1	Phase 1	58
4.3.2	Phase 2	77
4.4	Validation Survey.....	101
4.4.1	Survey Structure, Characteristics and Collection Method.....	101
4.4.2	Models Evaluation:	105
4.4.3	Future Scenarios Preferences	107
4.5	Chapter Summary.....	107
5	Chapter Five: Results and Discussion.....	109
5.1	Introduction	110
5.2	Logit Models	110
5.2.1	Phase 1: Binary Logit Model	110
5.2.2	Phase 2: Multinomial Logit.....	112

5.3	Classification Tree Models.....	115
5.3.1	Phase 1	115
5.3.2	Phase 2	118
5.4	Logit Model Vs. Classification Tree Model.....	122
5.5	Chapter Summary.....	123
6	Chapter Six: Conclusions and Recommendations	124
6.1	Introduction	125
6.2	Conclusion.....	125
6.3	Recommendations	127
6.3.1	Implications of Research Findings.....	127
6.3.2	Avenues for further Research	128
	References.....	130
7	Appendix A	136
8	Appendix B	145
9	الملخص	150
		151

List of Tables

Table (3.1) Approximate Percentages of Data Available for Each Item	37
Table (3.2) The Updated Approximate Percentages of Data Available for Each Item	40
Table (4.1) Gender Counts and Percentages	45
Table (4.2) Age Counts and Percentages	45
Table (4.3) Nationality Counts and Percentages	46
Table (4.4) Occupation Counts and Percentages	47
Table (4.5) Driving License Counts and Percentages	47
Table (4.6) Car Ownership Counts and Percentages	48
Table (4.7) Salary Counts and Percentages	49
Table (4.8) Origin Counts and Percentages	50
Table (4.9) Destination Counts and Percentages	50
Table (4.10) Purpose of Trip Counts and Percentages	51
Table (4.11) Travel Mode Counts and Percentages	52
Table (4.12) Travel Time Counts and Percentages	52
Table (4.13) Average Cost Metrics	53
Table (4.14) Public Bus Users Counts	54
Table (4.15) Public Bus Users' Feeders Preferences	55
Table (4.16) Metro Users Counts	55
Table (4.17) Metro Users' Feeders Preferences	56
Table (4.18) Tram Users Counts	57
Table (4.19) Tram Users' Feeders Preferences	57
Table (4.20) Variables for Modeling Phase 1	59
Table (4.21) Binary Logit Model Coefficient Analysis	62
Table (4.22) Odd Ratios for Continuous Predictors (Phase 1)	63
Table (4.23) Odd Ratios for Categorical Predictors (Odds ratio for level A relative to level B) (Phase 1)	63
Table (4.24) Binary Logit Model Goodness-of-fit	64
Table (4.25) Node Rules for Car Mode (Phase 1)	69
Table (4.26) Node Rules for Bus Mode	70
Table (4.27) Correlation Coefficient for Binary Logit Model Items	72
Table (4.28): Computing the Adjusted Total Correlation Coefficient for Binary Logit Model	73
Table (4.29) Correlation Coefficient for Classification Tree Model Items (Phase 1)	73
Table (4.30) Computing the Adjusted Total Correlation Coefficient for Classification Tree 1	74
Table (4.31) 10-Fold Cross Validation Results for the Binary Logit Model	75
Table (4.32) Confusion Matrix Rates for Classification Tree 1	76
Table (4.33) Misclassification Rates for Phase 1 Classification Tree	76
Table (4.34) Data used for Logit Model training phase 2	78
Table (4.35) Data used for Logit Model testing phase 2.	78
Table (4.36) Coefficient Analysis for Multinomial Logit Model.	80
Table (4.37) Odds Ratio Multinomial Logit Model	81
Table (4.38) Multinomial Logit Model Data Goodness-of-fit.	82
Table (4.39) Node Rules for Car Mode (Phase 2)	86
Table (4.40) Node Rules for Metro Mode (Phase 2)	86
Table (4.41) Node Rules for Public Bus Mode (Phase 2)	87
Table (4.42) Node Rules for Tram Mode (Phase 2)	88
Table (4.43) Correlation Coefficient for Multinomial Logit Model Items	89
Table (4.44) Computing the Adjusted Total Correlation Coefficient for Multinomial Logit Model	89
Table (4.45) Correlation Coefficient for Classification Tree Model Items	90

Table (4.46) Computing the Adjusted Total Correlation Coefficient for Multinomial Logit Model	91
Table (4.47) A Sample of the Multinomial Logit Model Resultant Probabilities	92
Table (4.48) True and False Positive Rates (Training Data)	93
Table (4.49) True and False Positive Rates (Testing Data)	93
Table (4.50) Misclassification Error for the Classification Tree Model Phase 2	98
Table (4.51) Pilot Survey Results	104
Table (4.52) Phase 1 Logit Model Confusion Matrix (Validation Survey)	105
Table (4.53) Phase 1 Classification Tree Model Confusion Matrix (Validation Survey)	106
Table (4.54) Phase 2 Logit Model Confusion Matrix (Validation Survey)	106
Table (4.55) Phase 2 Classification Tree Model Confusion Matrix	106
Table (5.1) Multinomial Logit Model Findings Summary	114
Table (5.2) Phase 1 Classification Tree Car trips and users' Characteristics	116
Table (5.3) Phase 1 Classification Tree Bus trips and users' Characteristics	117
Table (5.4) Phase 2 Classification Tree Car trips and users' Characteristics	119
Table (5.5) Phase 2 Classification Tree Metro trips and users' Characteristics	120
Table (5.6) Phase 2 Classification Tree Public Bus trips and users' Characteristics	121
Table (5.7) Phase 2 Classification Tree Tram trips and users' Characteristics	121
Table (5.8) Logit Model Vs. Classification Tree Model	122

List of Figures

Figure (1.1) Recent Trend in Vehicle Growth. Source: Statistics from General Directorate of Traffic, Bahrain and Al Bilad News (S. Waleed, 2019).	4
Figure (2.1) The Four-stage Model Stages, (Xintong, M., 2021)	12
Figure (2.2) O-D Matrix, (Modi, K. B., 2011).....	13
Figure (2.3) Probability Density Function for Gumbel and Normal Distributions (same mean and variance) (Koppelman, 2006).	17
Figure (2.4) Artificial neural network framework (Pineda-Jaramillo, 2019).	20
Figure (2.5) Basic Configuration of a Fuzzy Logic System (Ratrout, 2014).....	20
Figure (2.6) Decision Tree Example (Pineda-Jaramillo, 2019).....	21
Figure (3.1) Sample of the Future Public Transport Modes and Feeders for 13 Respondents	36
Figure (3.2) Locations Where Personal Interviews were Conducted.	39
Figure (4.1) Age Data Percentages	45
Figure (4.2) Nationality Data Percentages	46
Figure (4.3) Occupation Data Percentages	46
Figure (4.4): Driving License Data Percentages	47
Figure (4.5) Car Ownership Data Percentages.....	48
Figure (4.6) Salary Data Percentages.....	48
Figure (4.7) Origin Data Percentages Figure (4.8) Destination Data Percentages	50
Figure (4.9) Purpose of Trip Data Percentages	51
Figure (4.10) Purpose of Trip Data Percentages.....	51
Figure (4.11) Travel Time Data Percentages	52
Figure (4.12) Average Cost of Trip Counts	53
Figure (4.13) Public Bus Users Counts.....	54
Figure (4.14) Metro Users Counts	55
Figure (4.15) Tram Users Counts	56
Figure (4.16) Car (1) Utility Functions	60
Figure (4.17) Bus (2) Utility Functions	61
Figure (4.18) Normal Probability Plot (Percent vs. Residual) (Phase 1).....	65
Figure (4.19) Residuals vs. Fitted Values Plot (Devers vs. Fits)	65
Figure (4.20) Misclassification Cost vs. Number of Terminal Nodes Plot (Phase 1).....	66
Figure (4.21) Relative Variable Importance (Phase 1)	67
Figure (4.22) Simplified View of the Classification Tree Model (Phase 1)	68
Figure (4.23) A Sample View of the Detailed Tree (Phase 1).....	68
Figure (4.24) ROC for Phase 1 Classification Tree.....	76
Figure (4.25) Misclassification Cost vs. Number of Terminal Nodes Plot (Phase 2).....	83
Figure (4.26) Relative Variable Importance (Phase 2)	84
Figure (4.27) A Sample View of the Detailed Tree Phase 2.....	85
Figure (4.28) Car ROC (Training Data) (Phase 2).....	94
Figure (4.29) Public Bus ROC (Training Data) (Phase 2)	94
Figure (4.30) Tram ROC (Training Data) (Phase 2).....	95
Figure (4.31) Metro ROC (Training Data) (Phase 2).....	95
Figure(4.32) Car ROC (Testing Data) (Phase 2)	96
Figure (4.33) Public Bus ROC (Testing Data) (Phase 2).....	96
Figure (4.34) Tram ROC (Testing Data) (Phase 2)	97
Figure (4.35) Metro ROC (Testing Data) (Phase 2)	97
Figure (4.36) Car Classification Tree ROC. (Phase 2)	99
Figure (4.37) Tram Classification Tree ROC. (Phase 2)	99
Figure (4.38) Metro Classification Tree ROC	100

Figure (4.39) Public Bus Classification Tree ROC..... 100

Figure (4.40) Validation Survey Current Modes. 104

Figure (4.41) Validation Survey Gender..... 104

Figure (4.42) Validation Survey Age..... 105

Figure (4.43) Hypothetical Scenarios Results Percentages 107

Chapter One :Research Background

- 1.1 Background**
- 1.2 Problem Statement**
- 1.3 Research Questions**
- 1.4 Objectives of the Study**
- 1.5 Significance of the Study**
- 1.6 Limitation of the study**
- 1.7 Summary of Chapters**

1.1 Background

Since the dawn of humanity, transportation has played an essential role in facilitating trade and social interactions by utilizing a considerable share of time and resources. The significance of transportation is derived from the primary need to transfer people and goods from one location to another for the fulfilment of basic survival and economical objectives, including traveling in search for food, work, the exchange of goods and expedition purposes. In fact, the available transportation system's capabilities, namely speed, cost and capacity, have a direct and momentous impact on the economic strength of an area and the ability to maximize the utilization of its natural resources. For example, countries with advanced transportation systems, like the USA, Canada and Europe, take the lead in the industrial and commercial sectors (Nicholas J, 2009).

However, transportation systems in most developing countries currently face serious challenges such as congestion, accidents and air and noise pollution (Juan de Dios, 2011). These challenges arise due to the rapid population growth and the increase in car ownership levels, which consequently led the demand for these systems to surpass the available supply of transportation facilities and services. Therefore, it becomes necessary to establish efficient policies and to have a thorough comprehension of travel demands, travel patterns and driver characteristics for the development of engineering models that can facilitate and promote the advancement of transportation systems (Mwale, 2022).

Travel forecasting models are a crucial aspect of transportation planning and serve as a measure to identify the travel needs of cities (Sowjanya, D., 2014). These models usually employ mathematical equations and algorithms to simulate travel patterns and behaviors (Waghmare, A., 2022). One of the most notable models used for demand forecasting in transportation planning is modal split. This modelling operation, also known as mode choice modelling, focuses on

predicting and managing the transport modes' travel demands in a specific system by determining the key variables affecting the travelers' mode choice decision-making process (Pineda-Jaramillo, 2019).

This research aims to explore the influential factors concerning travel mode choice in Bahrain, study the current travel patterns, 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.

1.2 Problem Statement

Urban development and population growth in the Kingdom of Bahrain eventualized a substantial increment in car ownership levels leading to further congestion on the Kingdom's road network. As illustrated in Figure 1.1 below, the number of cars in Bahrain surged from around 400,000 in 2009 to approximately 700,000 by the end of the year 2019. This increase, accompanied by the scarcity and near absence of a public transportation system made it improbable to navigate through the country without traffic delay reaching several hours during peak flow conditions. Due to this, designing and launching an effective public transport service that would eventually contribute to minimizing congestion, delay and traffic accidents in the country became a necessity. However, the question that remains is: what public transport mode is best fit to the mode choice behavior 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 traveler's decision-making process.

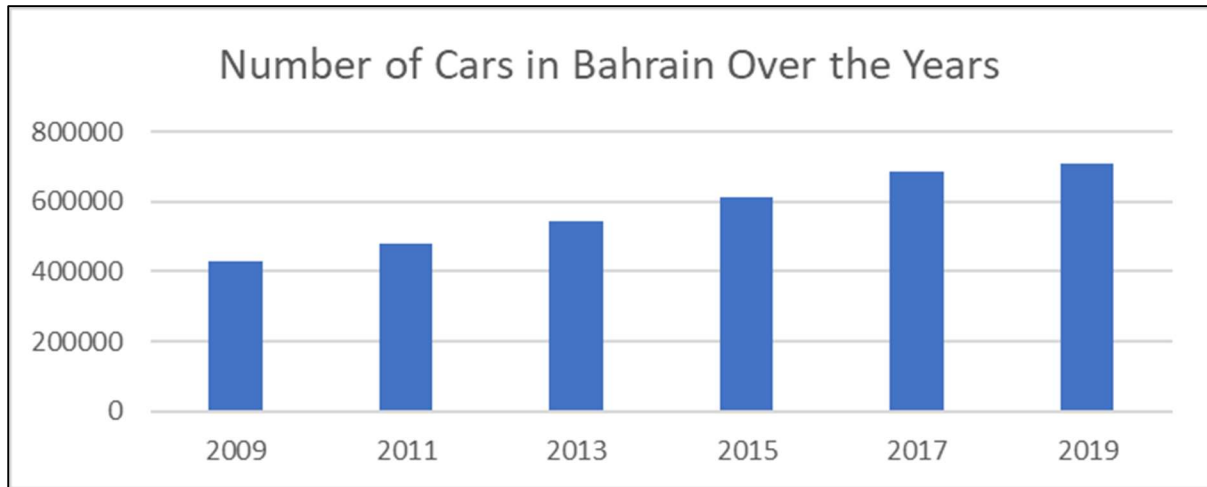


Figure (1.1) Recent Trend in Vehicle Growth. Source: Statistics from General Directorate of Traffic, Bahrain and Al Bilad News (S. Waleed, 2019).

1.3 Research Questions

1. What factors have the greatest influence on travelers' mode choice decisions in Bahrain?
2. What are the current governing characteristics of commuters who choose to travel by currently available modes (car, bus, etc.)?
3. Is it possible to build a logit model and a classification tree model to forecast probable utilities and probabilities of various future public transport modes, including bus, metro and tram?
4. What is the public transport service that best fit trip-makers in Bahrain?

1.4 Objectives of the Study

1. Explore the influential factors and the decision-making process of the trip maker concerning travel mode choice in Bahrain.
2. Study the current transportation system and identify the key characteristics of car users and bus users in Bahrain.

3. Utilize Logit model and Classification Tree model to forecast the utilities and probabilities of various potential future public transport modes including bus, metro and tram.

1.5 Significance of the Study

The importance of this study is directly related to the “Problem Statement” discussed above. The obtained results from the research will hopefully contribute to recommending a public transport system that will effectively serve commuters in the Kingdom of Bahrain and ultimately reduce traffic congestion and create a positive qualitative leap in the economy, social life and environment of the country.

Adding to that, apart from Bahrain, this analysis can have broader implications for urban transportation planning and policy-making worldwide. By identifying the factors that govern mode choice behavior in Bahrain, valuable insights and recommendations can be offered for the development of effective public transportation systems in similar urban areas around the world. Furthermore, the findings of this study can also provide beneficial knowledge that would aid the advancement of the transportation field in general.

1.6 Limitations of the Study

Despite the significant findings from this research, there are several limitations that need to be acknowledged and addressed. These limitations can be summarized as follows:

- The data used for modeling in this study is a combination of data extracted from previous surveys and data obtained from a questionnaire conducted as part of the current research. Therefore, merging data from different

questionnaires bounded the list of factors that can be utilized in model building.

- Since the research depends on data mainly collected through online questionnaires, respondents may have misunderstood certain questions or answered without full comprehension of all choices.
- Interviews were done at the public bus stations to amplify their involvement in the survey; however, it was difficult to communicate with many of the commuters due to language barriers.
- Minitab, the statistical software used for model building, excludes the responses that have empty values for the Logit models reducing the number of records utilized.
- When building the models, the "non-motorized transportation" users were excluded since they were very few compared to the Car and Bus users and the models generated with them had a poor fit or would not converge.

1.7 Summary of Chapters

This report is organized into six chapters as follows:

1. Chapter one: The first chapter discusses the statement of problem, research justification and objectives.
2. Chapter two: the second chapter presents a literature review that highlights the theoretical framework in which the research is based on, and briefly reviews similar studies.
3. Chapter three: This chapter discusses the details of data collection and handling.
4. Chapter four: This Chapter demonstrates the methodology of the research, the statistical analysis of data and the details of Logit and Classification Tree models built for both current and future scenarios.

5. Chapter five: This chapter reviews the results obtained from the analysis done in chapter four.
6. Chapter six: This chapter presents the recommended public transport system that will effectively serve commuters in the Kingdom of Bahrain based on the conclusions derived from the results as well as recommendations for future research and development.

Chapter Two: Literature Review

2.1 Introduction

2.2 Factors Affecting Travel Demand

2.3 Travel Demand Modelling

2.4 Mode Choice Modelling

2.5 Implementation of Logit Model and Decision Trees in Mode Choice Modelling

2.6 Chapter Summary

2.1 Introduction

Transportation is one of the most prominent contributors to the development of urban areas due to its focal function of providing users with efficient accessibility to services and mobility. In fact, implementation of proper transportation planning prompts and improves the economic growth of cities, which consequently boosts income and population growth. This subsequently generates an increase in both passenger and freight demands making it necessary to update the current transportation system to satisfy the new travel demand while maintaining equilibrium (Modi, K. B., 2011, Waghmare, A., 2022).

Travel demand is defined as the number of persons or vehicles per unit time that can be predicted to use a certain segment of a transportation system under specific conditions such as availability, quality and cost (Nicholas J., 2009). Forecasting travel demand is a vital step in the transportation planning process and since it is undergoing a continuous increase, more emphasis is given to understanding its relationship with factors affecting it such as: mode choice, vehicle ownership, trip and route frequencies and land use (L Ester A, 2011).

2.2 Factors Affecting Travel Demand

2.2.1 Travel demand and mode choice

Mode choice is considered one of the most prominent aspects of travel demand and therefore it is important to understand factors affecting it. In developing cities most transport modes available are private cars, public buses, trains, taxis, walking and cycling. According to Mwale (2022), affordability, availability, age and income are the main factors influencing mode choice. Private cars, for instance, are more likely to be chosen by households with high income while those with lower income are generally more attracted to public transport modes that are cheap and with great availability (Lekshmi et al, 2016).

2.2.2 Travel demand and vehicle ownership

The scarce supply and decreasing quality of public transport led to an increase in private cars' ownership in many developing countries (Etminani-Ghasrodashti, 2016). This shift is also highly influenced by household size, income and distances travelled. The chances in which a household would own a car increase with the presence of school-age children, as well as income and education level increase. In addition to that, lifestyle patterns, travel attitudes and socioeconomic status are considered the key influential factors on trip and route frequency (Mwale, 2022). As reported by Dissanayake and Morikawa, 2002, the travel attitude of husbands, within a household, is more likely to control the choice of owning a car while the wives are more involved in decisions concerning trip and route frequency.

2.2.3 Travel demand and land use

Land use has a significant impact on travel demand since it plays a major role in guiding road users through determining their preferred mode choice and travel preferences. Locations that are classified as highly populated areas demonstrate a positive correlation with walking and use of public transport while on the contrary, they have a negative association with private cars (Mwale , 2022). Considering that travel demand emerges from the need to perform daily trips and activities such as traveling to work, school or hospital, the distribution of these activities in space is recognized as a key factor affecting travel patterns and demand. Therefore, geographic data of activity locations is crucial for the development of accurate travel demand estimates (Rocha, 2021).

2.3 Travel Demand Modelling

Principally, models are defined as being a simplified representation of real-life aspects or phenomena. In transportation planning, travel forecasting models play a core role in the process and are considered as a measure to detect travel needs of cities. They can be used to predict the effect of changes happening in the aspects of the transportation system and the system's users on travel demand (Sowjanya, D., 2014). These models usually employ mathematical equations and algorithms to simulate travel patterns and behaviors (Waghmare, A., 2022). In practice, it is only after World War II, in the 1960s, that travel-modeling application commenced and the classic four-stage model was gradually constructed. (Mladenovic, 2014).

The four-stage model breaks down the study area into homogenous traffic analysis zones in which the data required for building the model is collected (Xintong, M., 2021). The modeling process can be divided into two main phases. The first phase focuses on collecting, evaluating, calibrating and validating data to develop a travel demand, while the second phase loads this demand onto the network to formulate equilibration of route choice. With this in mind, these two phases can be further divided into four distinct stages known as trip generation, trip distribution, modal split and traffic assignment (Figure 2.1) (McNally, M. G., 2017).

Since the current research focuses on mode choice modelling, section four will provide detailed review of its process, nonetheless, the four stages of the classic travel model are briefly explained below.

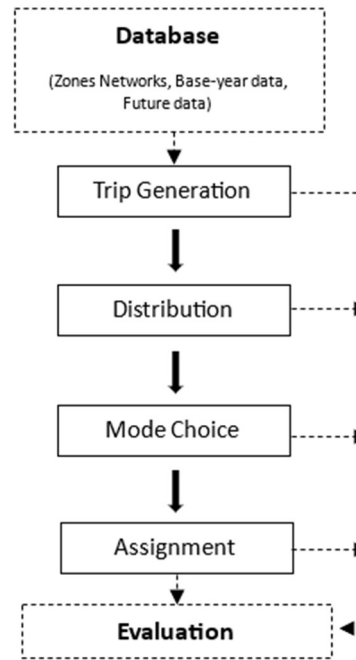


Figure (2.1) The Four-stage Model Stages, (Xintong, M., 2021)

2.3.1 Trip Generation

Trip Generation is the first stage of the classical four-stage demand model. This stage concentrates on forecasting the total number of daily trips generated from and attracted to each zone in the model system (Sowjanya, D., 2014). Trips originating from a zone are referred to as Production and the term Attraction is used for the trips attracted to the zone (Modi, K. B., 2011). Usually, Production relies upon the population aspects of a zone while Attraction depends on its employment characteristics and land use (Transport and Infrastructure Council, 2016). In general, Trip generation utilizes econometric analysis, such as linear or log-linear regression models, and time series analysis to approximate the relationship between travel demand and the socioeconomic factors affecting it (Tskeris, T., 2011).

2.3.2 Trip Distribution

Trip distribution is the stage in which the trips collected from the trip generation process are allocated to origin and destination points based on different factors such as the distances between the zones and their land use. The outcome of this stage can be demonstrated through an origin-destination (O-D) matrix as shown in Figure 2.2 below (Tskeris, T., 2011).

Zones	1	2	3		j		n	O _i
1	T ₁₁	T ₁₂	T ₁₃		T _{1j}		T _{1n}	O ₁
2	T ₂₁	T ₂₂	T ₂₃		T _{2j}		T _{2n}	O ₂
.	.	.	.		.		.	.
.	.	.	.		.		.	.
i	T _{i1}	T _{i2}	T _{i3}		T _{ij}		T _{in}	O _i
.	.	.	.		.		.	.
.	.	.	.		.		.	.
n	T _{n1}	T _{n2}	T _{n3}		T _{nj}		T _{nn}	O _n
D _j	D ₁	D ₂	D ₃		D _j		D _n	T

Where, $D_j = \sum_i T_{ij}$ $O_i = \sum_j T_{ij}$ $T = \sum_{ij} T_{ij}$

Figure 2.2) O-D Matrix, (Modi, K. B., 2011)

2.3.3 Modal Split

Mode choice or modal split is the third stage of the four-stage model. This stage focuses on the traveler's behavior in relation to the selection of travel mode (Sowjanya, D., 2014). Traditionally, 'Logit' model is considered the most frequently used form of modal split modelling. This model assumes that each travel mode, such as car, bus and train, has a specific level of utility based on certain trip, commuters and mode aspects which subsequently affects the traveler's decision and preferences (Transport and Infrastructure Council, 2016).

2.3.4 Traffic Assignment

The final step of the modelling process is called traffic assignment. This stage inputs the forecasted origin-destination demand per mode into the network according to the supply conditions available (Tskeris, T., 2011). The traveler's

choice of route is influenced by various factors including journey time, distance, cost, delay, type of maneuvers required and the traveler's habits (J. de Dios Ortúzar, 2011). Traffic assignment is mainly accomplished through three basic approaches: diversion curves, minimum time path assignment and minimum time path with capacity restraint (Garber, N. J., 2018).

2.4 Mode Choice Modelling

Mode choice modeling is one of the most prominent components of the transportation planning process, specifically the four-stage model. This modelling stage focuses on forecasting and managing travel demand of transport modes in a specific system through determining the key variables that affect the travelers' mode choice decision-making process (Pineda-Jaramillo, 2019). In terms of transportation, the travelers tend to favor transport modes that fit their preferences and traveling habits best and thus the factors that control their decision can be arranged into three categories (Chen, 2013; Ratrout, 2014).

These categories are as follows (Ratrout, 2014):

- Characteristics of trip makers: These are the socio-economic and cultural aspects of a traveler, such as automobile ownership, income, employment, gender, age and personal desires.
- Characteristics of the trip: Mainly concerning the purpose of the trip, the time of the day at which it is made and the land-use of the area in which it is made.
- Characteristics of the transport mode or service: Including travel time, travel cost, comfort, safety and availability.

With this in mind, automobile ownership, public transport availability and land-use are considered to employ a larger influence on the traveler's choice than the

other mentioned factors. For instance, a traveler who owns a car in an area with limited access to public transportation would normally become “auto-dependent” and inversely, a traveler who does not own a car but have good access to other modes would naturally depend on public transportation, cycling or walking (Ratrou, 2014). Considering that accurate predication of travelers’ mode choice behavior depends on several variables, it is very crucial to select the right and most suitable approach of modeling (Chen, 2013).

Mode choice modelling can be achieved through various approaches that can be divided into two main groups: traditional discrete choice models and machine learning/artificial intelligence models.

2.4.1 Discrete Choice Models

Since the 1960s, transportation planning generally relied on discrete choice models for modelling travelers’ mode choice behavior. These models employ the theory of utility maximization and assume that a traveler is most likely to choose the mode that presents the most benefits to its users (Ratrou, 2014; Hillel, 2019). In addition to that, according to the nature of the data used in model building, discrete models can be further classified as either Logit Model, Probit Model or General Extreme Value Model (Ratrou, 2014).

2.4.1.1 Logit Model:

Through the utilization of simple mathematical techniques, Logit model manages to simulate complicated travel patterns of any population and thus it is considered the most used approach for mode choice modelling (Sekhar, 2014). According to this model, the utility function of a mode is demonstrated by a linear equation that links the factors influencing mode choice as independent variables and the utility of a particular mode as the dependent variable.

The utility function of a mode, and the probability that a traveler chooses it, is expressed by the following equations (1 and 2) respectively (Puan, 2019 ; Ratrout, 2014):

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

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

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

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

Generally, based on the number of alternatives included in the model, Logit model can be either of two forms: Binary Logit model or Multinomial Logit model.

- **Binary Logit Model:** This model is utilized when commuters have two options of travel modes to choose from; for example, private car and public bus. The probability of choosing mode “A” in a binary logit model can be obtained through Equation 3, which is a simpler form of Equation 2, while Equation 4 represents the probability of choosing the other alternative (B) (Puan, 2019):

$$P_A = \frac{e^{U_A}}{1 + e^{U_A}} \quad (3)$$

$$P_B = 1 - P_A \quad (4)$$

Where:

$P_{A,B}$ = Probability of trip maker choosing mode A, B

U_A = utility of alternative A

- **Multinomial Logit Model:** When commuters have more than two modes to choose from, multinomial logit model is used to build the utility functions and estimate probabilities as per Equations 1 and 2 (Puan, 2019). Multinomial logit model is based on the theoretical assumptions that the error elements follow Gumbel distribution rather than normal distribution (figure 2.3), error elements for each alternative are comparable and independently distributed and that error elements are similar and independently distributed for each observation (Hussain, 2017).

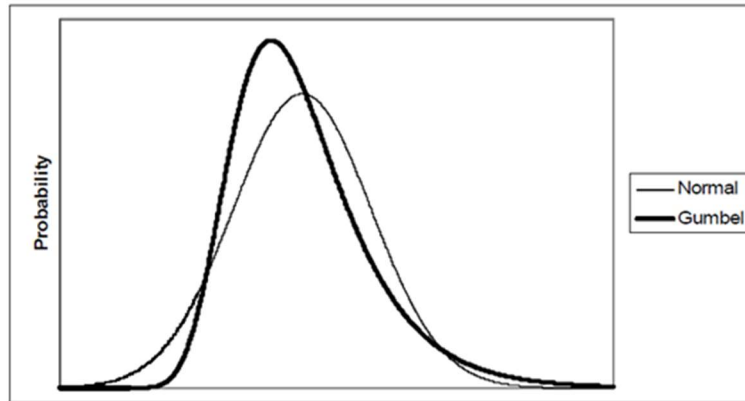


Figure 2.3) Probability Density Function for Gumbel and Normal Distributions (same mean and variance) (Koppelman, 2006).

2.4.1.2 Probit Model

Probit models are also common in transportation planning for mode choice modeling (Bunch, 1991). In this model, the utility of an alternative is given by the following equation (5):

$$U_i = V(X_i, S) + \varepsilon_i \quad (5)$$

Where U_i is the utility of mode i , V is the systematic component derived from the weighted sum of the variables, X is the vector of explanatory variables, S is the vector of coefficients, and ε is the error element of the utility function (Ratrou, 2014).

Probit models are established on the assumption that error elements follow a normal distribution. This assumption, however, subsequently leads to complex and notable losses in the model flexibility and interpretation (Sekhar, 2014; Hussain, 2017). Therefore, logit model, which adopt Gumbel distribution, is viewed by transportation planners as superior to probit models and generally have better goodness-of-fit measures and calibration (Ghareib, 1996).

2.4.1.3 General Extreme Value Models

These models are considered a significant simplification of multinomial logit models. They treat error elements according to type one extreme value distribution and regard them as independent but non-identical (Sekhar, 2014). A study done by Bhat 1995 suggests that extreme value models are found to be more useful in modeling intercity mode choice than multinomial logit model (Bhat, 1995).

2.4.2 Machine Learning/Artificial Intelligence Models

Although transportation planning exclusively relied on discrete models in the past, due to the introduction of new transportation technologies, which allowed a wider-range of data collection, and the remarkable development in machine learning research that demonstrated numerous successes of its application, transportation planners began to explore more advanced alternative approaches (Hillel, 2019). Machine learning is a term used to refer to a group of algorithms that support computers in mechanizing data and utilizing it to build models that aim to overcome the challenges imposed by the growing travel demands on the planning process through detecting statistical patterns (Bhavsar, 2017).

In order to provide insight of Machine Learning methods, the following sections discuss Artificial Neural Networks, Fuzzy Logic and Decision Trees, which are three of the main algorithms used to explore mode choice modeling.

2.4.2.1 Artificial Neural Networks

Artificial Neural Network (ANN) is a term used to indicate to a group of classifiers that imitate the function and structure of the brain. ANN can have numerous forms depending on the type of input data it is handling; however, the Feed-Forward Neural Network (FFNN) is generally used for mode choice operations (Hillel, 2019).

The elementary unit of ANN is called a Neuron and it is responsible to compute an output for an input parameter and its weight through the implementation of an activation function. When creating a multi-layer ANN, the following should be considered (Ratrout, 2014):

- The number of hidden layers
- The number of neurons in each layer
- Activation and error functions for the neurons

- Parameters and their weights
- Initialization technique and conditions
- Input characteristics

Figure 2.4 below illustrates an example of an ANN having a hidden layer that has a specific number of neurons accountable for output calculations (Pineda-Jaramillo, 2019).

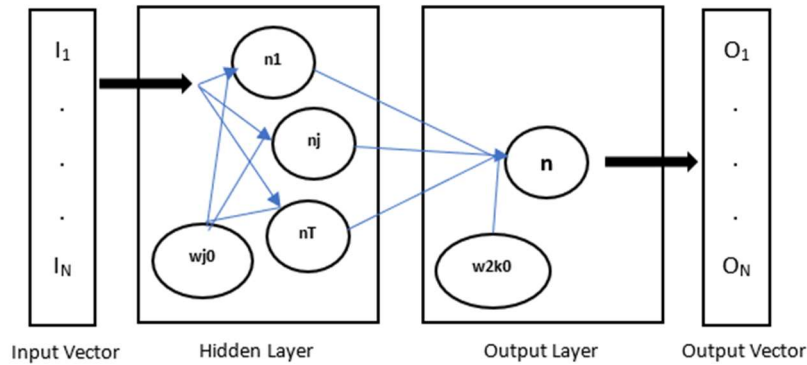


Figure (2.4) Artificial neural network framework (Pineda-Jaramillo, 2019).

2.4.2.2 Fuzzy Logic

Fuzzy logic can prepare a standardized and productive framework to integrate linguistic fuzzy data from a human expert. The fundamental structure of a fuzzy logic system includes a fuzzifier, fuzzy IF-THEN functions, a fuzzy inference engine and a defuzzifier. Using the IF-THEN functions, the fuzzy inference engine executes a mapping through a process called fuzzification from an input vector and then by the means of defuzzification prepare a mapping of the fuzzy decision to an output vector as shown in Figure 2.5 below (Ratrout, 2014).

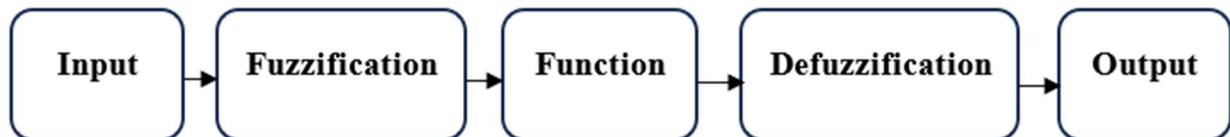


Figure 2.5) Basic Configuration of a Fuzzy Logic System (Ratrout, 2014).

It is noteworthy to mention that the formulation of fuzzy logic rules and functions demand a vast amount of data and knowledge and thus its application may be challenging to researchers (Ratrout, 2014). Nonetheless, Fuzzy artificial neural

networks have been successfully implemented for mode choice modeling and their comparison with a Logit model shows that the former can provide better accuracies (Vythoukas, 2003).

2.4.2.3 Decision Trees

Decision trees, such as Classification trees and Regression trees, are non-parametric methods that classify data into groups and model the relationships between features and possible outputs through following a structure identical to a flowchart or to a tree (Pineda-Jaramillo, 2019). The most frequently used form of Decision trees are the trees that use binary splits. To compute each split, the data is classified depending on the features specified and then for each feature the possible binary split points are tested (Hillel, 2019).

Figure 2.6 below illustrates a decision tree used to evaluate if making a trip by bike from a certain origin to a specific destination is a good idea. The decision undergoes three levels of binary splits assessments considering the features of the trip (rain, travel time and topography). The feasible outcomes of these assessments are represented as Yes or No (Pineda-Jaramillo, 2019).

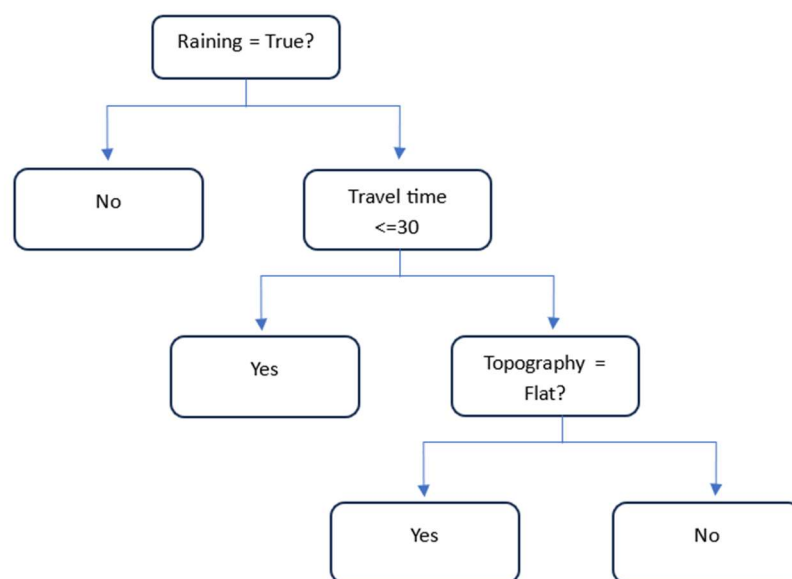


Figure (2.6) Decision Tree Example (Pineda-Jaramillo, 2019).

2.5 Implementation of Logit Model and Decision Trees in Mode Choice Modeling

This purpose of this thesis is to address the research gap in the literature concerning public transport mode choice behavior of commuters in Bahrain. While existing studies have investigated a variety of factors influencing mode choice, there is still a lack of thorough understanding of the complex decision-making processes that individuals consider when choosing a mode of transportation in this specific context. To fill this gap, the use of the logit model and decision trees is justified. The logit model, which has been widely used in transportation planning due to its simplicity, is well-suited for capturing the probability of choosing a specific mode of transport based on individual and trip characteristics, while decision trees offer a non-parametric approach to mode choice modeling by providing a flexible and interpretable framework for analyzing complex decision-making processes. By employing these models, this research seeks to provide deeper insights into the factors influencing public transport mode choice and contribute to the development of effective transportation policies and strategies in Bahrain.

This section will review several studies that have applied these models in mode choice modeling and discuss their strengths and limitations.

2.5.1 Logit Model for Mode Choice Modelling

2.5.1.1 Mode Choice of Work Trips in Developing Cities (Gaza)

Almasri and Alraee (Essam, 2013) conducted a study in 2013 in Gaza to determine the key factors that influence employees' choice of transport modes. The aim of the study was achieved by developing a Multinomial Logit model using two thirds of a total sample of 552 questionnaires, while the remaining third was used for validation purposes. The model involved four transport modes,

namely car, taxi, motorcycle and bicycle, and the results obtained indicated that travel time, trip cost per income, ownership, distance travelled, age and monthly income had the most significant effect on employees' mode choice.

2.5.1.2 Mode Choice Behavior for Travel Between Saudi Arabia and Bahrain

In 2014, a study by N. Ratrout and U. Gazder (Ratrout, 2014) investigated the present travel trends and the changes in border mode choice behavior 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 implemented it in the development of Logit models. It was recognized that the majority of travelers favored 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 travelers, it was observed that the train service was considered more attractive than the ferry service, while the latter being more appealing to single travelers compared to car. Moreover, it was concluded that travel cost had the greatest impact on both modes' utilities, whereas rigidity of schedule affected only the utility of train service. The researchers performed a sensitivity analysis as well, determining that the optimum fare values at which a higher probability for the train and ferry to be chosen by travelers over car are SAR 52 and SAR 78 respectively.

2.5.1.3 Public Transport Versus Solo Travel During COVID-19

On account of the outbreak of the COVID-19 pandemic, public transport experienced a severe drop in the number of its users worldwide (Abdullah, 2021). This issue inspired Abdullah and other researchers to investigate and analyze travelers' behavior 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.

2.5.2 Decision Trees for Mode Choice Modelling

2.5.2.1 Freight Mode Choice Decision Tree

Samimi, Razi-Ardakani and Mohammadian (Samimi, 2012) studied freight mode choice trends in the US using five separate decision tree models i.e., CART, CHAID, E.CHAID, QUEST and C5.0, and a binary Logit model. The models were based on data collected through a survey conducted in 2009 that comprised of 881 domestic shipments and considered two mode choices: truck only and rail. In regard to the G-Mean value, the C5.0, CART and logit demonstrated the best forecasting power. According to the C5.0 and CART trees, shipments that are heavy and traveling for a long distance have a greater chance for rail transport. In addition, it was concluded that shipment weight is the key factor determining freight mode choice.

2.5.2.2 Decision Trees for Estimating Mode Choice

L. O. Oral and V. Tecimb (Oral, 2013) conducted a study to forecast mode choice of trips in district Buca in Izmir, Turkey using C4.5 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, driver license, 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, origin and destination of trips had no major influence on mode choice.

2.5.2.3 Estimating Motorized Travel Mode Choice Using Classifiers

A study done in São Paulo Metropolitan Area, Brazil by A. Lindner, C. Pitombo and A. Cunha compared the prediction efficiency of ANN and Classification Tree models with that of a binary Logit model (Lindner, 2017). 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%).

2.6 Chapter Summary

Based on the literature reviewed, the following hypotheses can be formulated regarding commuter's mode choice, which were tested during this study for Bahrain:

1. H0: Gender, age, origin, destination, income, trip purpose, travel cost, travel time, occupation, driving license and car ownership will have a significant effect on mode choice.
H1: These factors have a minimal or no effect on mode choice.
2. H0: Logit models can accurately and efficiently identify the characteristics of travelers who currently choose to travel by car/bus, as well as predict commuters' future public transport mode choices.
H1: Logit models are not effective in identifying the characteristics of travelers who currently choose to travel by car/bus and predicting commuters' future public transport mode choices.

3. H0: Classification Trees model is a powerful tool for accurately identifying and predicting current and future mode choice in Bahrain.
H1: Classification Tree model is not a powerful tool for accurately identifying and predicting current and future mode choice in Bahrain.
4. H0: Classification Tree model will have better accuracy than Logit model.
H1: Classification Tree model will not have better accuracy than Logit model.

Chapter Three: Data Collection

3.1 Introduction

3.2 Data Extracted From Previous Surveys

3.3 Compiling Responses Extracted from Previous Surveys

3.4 Data Collected From a Recent Survey

3.5 Creating the Final Dataset

3.6 Chapter Summary

3.1 Introduction

This chapter provides a comprehensive overview of the data sources and collection methods used to obtain the necessary information for modeling mode choice and evaluating the feasibility of implementing various public transport systems in Bahrain, including tram, public bus and metro. 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 research.

3.2 Data Extracted From Previous Surveys

To gain a better understanding of the travel behavior of a specific population, previous surveys can be an invaluable resource. This section reviews data collection methods of surveys conducted by civil engineering students at the University of Bahrain from 2017 to 2021, with a focus on using this information to attain the objectives of this research.

3.2.1 Survey 1: Travel Demand Forecasting Using Traditional Four Steps Transportation Modelling Approach: A Case Study of Bahrain

Aiming to gain knowledge of the urban transport modelling process and to recognize travel behavior between governorates in Bahrain, a group of students conducted a questionnaire in 2017 to collect the required data. The questionnaire was distributed among 544 residents of Bahrain and consisted of 17 simple questions. The data collected included: nationality, gender, occupation salary, car ownership, travel mode/s, Origin governorate, number of trips per day, purpose of trip, Origin-Destination, travel time and cost, distance traveled and route major highways.

3.2.2 Survey 2: Measures to Increase Ridership for Public Transport in Kingdom of Bahrain

In 2018, a research was done to analytically study possible measures that can be taken to increase public transportation ridership in Bahrain. To accomplish this, the study was conducted in two stages each including two questionnaires: private transport questionnaire and public transport questionnaire.

3.2.2.1 Stage 1

- Private transportation:

The questionnaire was distributed online on social media and achieved 320 total responses. It collected data concerning gender, age, education, monthly income, occupation, Origin-Destination, travel time, travel cost, as well as the respondent's opinion on how to improve public transportation ridership. Besides that, the survey included an evaluation question where respondents were asked to rate reasons for why they choose not to travel by public transportation, such as, I enjoy driving my own vehicle, I do not have information about bus schedule, I may get lost, public transportation is not punctual, etc.

- Public transportation:

Data related to public transportation users was obtained through questionnaire-based interviews done at public bus stations. The questionnaire consisted of 12 questions focusing on gender, age, education, monthly income, occupation, Origin-Destination, travel time, travel cost, walking time, waiting time and car ownership. A total of 114 responses were collected.

3.2.2.2 Stage 2

- Private transportation:

For the second stage, 202 responses were collected and besides the original questions, a question concerning rating potential public transportation modes was added. The options for this question comprised of the existing bus service with improvements, metro with parking facilities, metro with bus feeder and metro with cycling service.

- Public transportation:

Similarly, besides the original questions in the first stage, an evaluation question was added asking respondents to rate reasons for why they choose not to travel by car. Such as, I cannot afford a car, bus station is near my house, bus provide a direct route to my destination, I can relax during my journey, etc. A total of 106 responses were collected.

3.2.3 Survey 3: Evaluation of a Proposed Tram Network Linking Manama's Main Recreational Centers Through the Implementing of VISSIM Simulation Software

To study the feasibility of a proposed tram network connecting four major recreational centers in the Kingdom of Bahrain (City Centre Bahrain Mall, Bab Al Bahrain, The Avenues and Bahrain National Museum), 857 questionnaire responses were collected in 2019. The responses were achieved by distributing the questionnaire through social media and by conducting personal interviews at the targeted destinations.

The questionnaire included 16 questions that can be divided into the following categories:

- General questions: gender, age, residency, driving license, preferred type of transportation, number of owned vehicles and the regularly visited places.
- Evaluation questions: respondents were asked to rate the traffic congestion levels and parking conditions in different circumstances by providing their trip timings, trip days and visits frequency.
- Tram-related questions: respondents' willingness to use the proposed tram network and ticket pricing suggestions.

3.2.4 Survey 4: Optimizing Routes for Public Transport in Bahrain

Aiming to determine the LOS of the current public bus service, a survey was conducted in 2020. The survey collected data from public bus users traveling through the main stations: Manama main station, Isa Town main station, Muharraq main station and Alsalmaniya bus stop. Consequently, 248 responses were achieved providing information relating to gender, Origin-Destination, waiting time, walking time, travel time and whether the trip was direct or transfer.

3.2.5 Survey 5: Design Of School Metro

An online survey done during the COVID19 pandemic (2020) aiming to provide data that could aid the design of a school metro service, collected 265 responses. It inquired information concerning gender, age, school and public buses use and respondents' willingness to use the proposed metro in the future.

3.2.6 Survey 6: Determination of Travel Demand for Design of Traffic Facilities for Schools in Bahrain

To determine the travel demand attracted to pooled and scattered schools in Bahrain, a survey was conducted in 2020. The survey included private and public schools and was distributed electronically targeting students, parents and employees in both languages, Arabic and English.

The questionnaire consisted of 25 questions that can be divided into the following categories:

- General questions: occupation, number of family members, number of students in the family, income, parents' level of education and Student's grade.
- School-related questions: school's name and location, reasons for choosing the school, the availability of parking lots and school bus services.
- Travel-related questions: travel time, distance travelled, travel mode, the respondents' willingness to use public transportation, congestion levels and possible solutions for it.

A total number of 271 responses was collected.

3.2.7 Survey 7: Design Of a Suitable Feeder System For Metro Bahrain

A project that intended to study travel patterns in Bahrain and design a feeder system for metro Bahrain, organized an online survey in 2021. The questionnaire consisted of 18 questions in both English and Arabic languages and achieved 814 total responses.

The questions can be categorized as follows:

- General questions: age, gender, nationality, occupation, driving license, and number of owned vehicles.
- Trip questions: Number of trips per day, mode of transportation, origin, destination, and purpose of the trip.
- Metro and feeder related questions: Respondents willingness to use the metro, the reasons they would or wouldn't consider using the metro, the purpose of using metro, feeder choice and reasons for choosing it.

3.3 Compiling Responses Extracted from Previous Surveys

In order to establish effective and sustainable transportation systems, it is critical to apprehend how people make decisions regarding their choice of travel mode. The previous surveys, reviewed above, provide valuable insights on this topic, but compiling and analyzing this data can be challenging. This section delves into the responses extracted from these surveys with the aim to create a dataset that can be utilized to better understand current conditions for mode choice in Bahrain and make informed forecasts about probable public transport mode choice in the future.

By looking at the responses collectively, it can be concluded that information concerning gender, age, nationality, occupation, salary, driving license and car ownership appear 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 discussed below.

- Gender:

The letter M is used to refer to males, whereas the letter F is used for females.

- Age:

A group of the surveys collected age information through a multiple-choice question providing the respondents with various ranges, while others asked them to type in their ages. 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:

Similar to age, some surveys requested the respondents to type in the exact locations while others asked them to choose the governorate corresponding to their origin and destination points. To standardize the data, corresponding governorates are used (Capital, Muharraq, Northern and Southern) mentioning the exact locations, if available, between brackets. It is worth to note that since one of the surveys was conducted at a time when Bahrain was divided into five governorates instead of four, “Central” governorate option is also available as the exact locations were not mentioned and hence it was not possible to distribute them as per the new system. In addition, Saudi Arabia and Kuwait are common origins for non-resident respondents.

- 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 BD to above 2000 BD.

- Driving License:

“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:

Although a few of the surveys had no data collected within its structure for trip purposes, some of the missing data are derived from the nature and objective of the study. For example, survey 3 was performed to aid in designing a tram network that connects shopping and recreational centers, hence the purpose of trip is “Shopping/Leisure”. This being said, 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 public transport modes and feeders:

Bus	Metro	Tram	Feeder : Car	Feeder: Bus	Feeder: Walking	Feeder: Cycling
				x		
	x		x	x		
	x				x	
x	x				x	
	x				x	
	x				x	
	x		x	x		
x	x	x				
x						

Figure (3.1) Sample of the Future Public Transport Modes and Feeders for 13 Respondents

Figure 3.1 above shows a sample of the future public transport modes and feeders each respondent is inclined to use in the future with “x” referring to a positive response. 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, as illustrated in Table 3.1, the number of data available for items range from 3722 to 1343 responses with “Salary” having the lowest value. Insufficient data undermines the accuracy and reliability of any analysis; therefore, it was necessary to conduct an additional survey in order to increase the overall data and ultimately lead to more informed and effective decision-making.

Table (3.1) Approximate Percentages of Data Available for Each Item

Item	Data Available	Percentage
Gender	3466	93%
Age	3220	86%
Nationality	1666	45%
Origin	2334	62%
Destination	3335	89%
Current mode	3722	99%
Purpose of trip	2873	77%
Travel time (min)	2542	68%
Total cost of trip	2071	55%
Occupation	2370	63%
Average Salary	1343	36%
Driving License	2168	58%
Car Ownership	1885	50%
Future Modes	2246	60%
Feeder Mode	1585	42%

3.4 Data Collected from a Recent Survey

To improve the accuracy and reliability of available data and the understanding of mode choice behavior in Bahrain, a recent survey was conducted to collect data on commuters’ travel patterns and preferences. This section concentrates on detailing the structure of the questionnaire and the methodology used to administrate the survey.

3.4.1 Structure of the Questionnaire

To gather comprehensive information, the questionnaire was organized to ensure that all relevant questions are 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 license 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 would they choose to travel to and from public transport stations.

Furthermore, it is noteworthy to mention that respondents were provided with 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 exception of trip cost, to facilitate easy collection of data. For the full structure of the questionnaire, please refer to appendix (A).

3.4.2 Methodology Used to Administrate the Survey

The methodology utilized to conduct a survey is a key aspect to consider, as it directly influences the quality and integrity of the data collected. In order to obtain sufficient and diverse responses, the questionnaire was distributed through various online channels and was used as basis for interviews done at several locations across the country. As a result, 409 complete responses were successfully collected.

The particulars regarding the distribution avenues of the questionnaire and the designated locations of the personal interviews are enumerated below:

- Online distribution channels:
 - WhatsApp
 - Instagram
 - Students UOB online forms
 - Online communities
- Personal interviews' locations (figure 3.2):
 - University of Bahrain – Sakhir campus.
 - Isa Town Bus Terminal
 - Manama Terminal 1
 - Water Garden City
 - The Avenues
 - The City Center Mall
 - Seef Mall – Isa Town

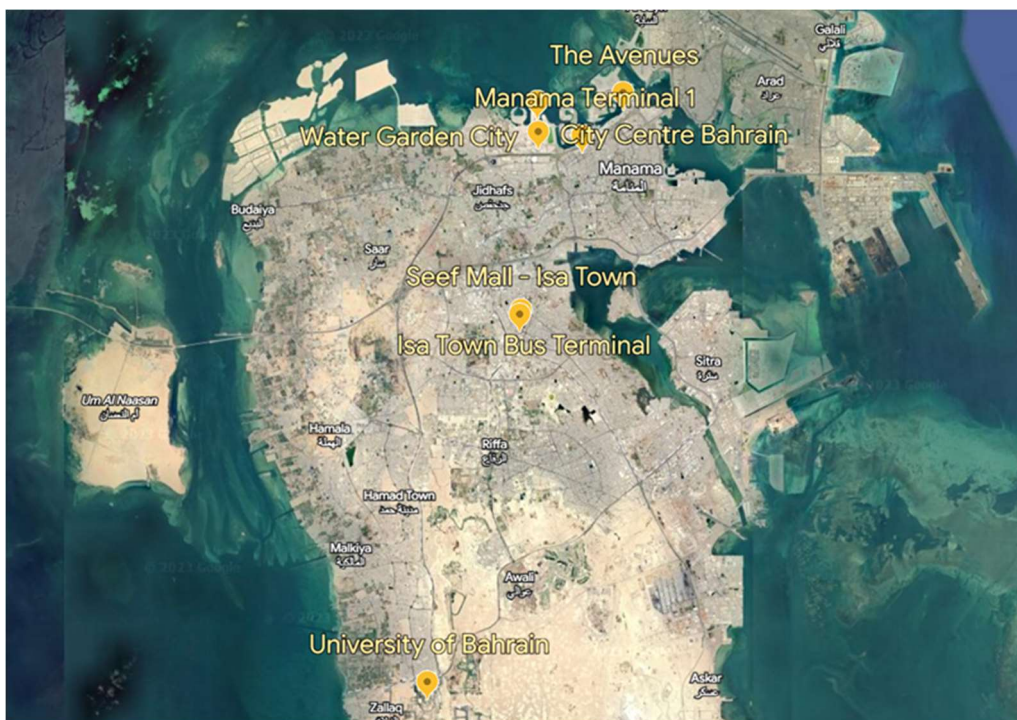


Figure (3.2) Locations Where Personal Interviews were Conducted.

3.5 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, in order 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. Table 3.2 below illustrates the updated approximate percentages of data available for each data item.

Table 3.2) The Updated Approximate Percentages of Data Available for Each Item

Item	Data Available	Percentage
Gender	3594	93%
Age	3346	87%
Nationality	2016	52%
Origin	3330	86%
Destination	3680	95%
Current mode	3864	100%
Purpose of trip	3051	79%
Travel time (min)	2817	73%
Total cost of trip	2482	64%
Occupation	2716	70%
Average Salary	1744	45%
Driving License	2479	64%
Car Ownership	2198	57%
Future Modes	2481	64%
Feeder Mode	1903	49%

3.5.1 Sample Size

Based on the studies discussed in the literature review, it is noticed that there is considerable variation in the sample sizes used in mode choice modelling. For example, (Essam, 2013) conducted in Gaza used 522 responses (0.025% of total population), (Abdullah, 2021) in Lahore had 1516 responses (0.014% of total population) and (Samimi, 2012) investigated 881 US domestic shipments. Given that Bahrain's population and transportation system may bear no resemblance to these locations, it is essential to carefully study several factors when justifying the sample size used in this research.

These factors include:

- **Population:** Bahrain has an approximate population of more than 1.6 million people with 73.2% of which are above the age of 20 (Ministry of Information) . As the sample size of a study should be representative of the population being investigated, a sample size of 3864 responses (0.24% of total population, 0.33% of total adult population) for Bahrain is fairly reasonable compared to the total population and is anticipated to decrease the likelihood of a biased sample.
- **Significance:** The research has a sample size of 3864 responses, which is greatly larger than that of the other studies reviewed in chapter 2. This suggests the possibility of achieving a higher level of statistical significance in the results of the analysis.
- **Heterogeneity:** It is crucial to consider the heterogeneity within the total population being investigated. Therefore, since Bahrain's population is diverse in terms of socio-economic characteristics and travel patterns, a large sample size becomes essential to adequately account for this diversity. By including a wide range of respondents in this study (3864), it is expected that the potential to capture this diversity is competent as the

final dataset includes adequate data records for all genders, age groups, nationalities and occupations (refer to Chapter 4 for details).

3.6 Chapter summary

This chapter focused on demonstrating the data collection methods performed for this research which intends to model current mode choice conditions in Bahrain, as well as forecast future public transport mode choice. Data was utilized from a number of sources, including previous surveys conducted by civil engineering students at the University of Bahrain, as well as a more recent survey done as part of the current research. In total, the final dataset comprised 3864 responses, which will be used as the basis for building Logit and Classification Tree mode choice models. Besides that, the chapter emphasized the importance of careful data collection and compiling in ensuring the accuracy and relevance of the resulting predictions and models.

Chapter Four: Research Methodology

4.1 Introduction

4.2 Statistical Analysis

4.3 Model Building

4.4 Validation Survey

4.5 Chapter Summary

4.1 Introduction

As previously mentioned in the earlier chapters, mode choice modeling is a pivotal tool for forecasting and comprehending commuter's mode choice behavior. This chapter delves into the methodology followed by this research to achieve conclusive and definitive mode choice models, which can be confidently used to recommend a public transportation system that aligns with Bahraini commuters' preferences.

The procedure involves a statistical analysis of the obtained dataset and model building conducted in two phases. The first phase focuses on modeling the current mode choice behavior of the sample population, while the second phase utilizes the data to forecast future public transport mode choice. In particular, each phase comprises the development of two models, namely the Logit model and the classification tree model, and outlines the measures taken to ensure the accuracy and reliability of each. Finally, a validation survey was conducted to test the competency of the models generated.

4.2 Statistical Analysis

Statistical analysis is crucial for interpreting complex data sets, as it provides an organized framework from which meaningful insights can be extracted. This section highlights key statistical metrics such as counts, percentages, mean and mode that would contribute in providing a better understanding of the characteristics of the data items and their different options, as well as serve as a valuable information about the central tendency of the data.

4.2.1 Commuters' General Information:

The findings of the statistical analysis conducted on gender, age and socio-economic characteristics of commuters are presented in the following figures 4.1 – 4.6 and tables 4.1 – 4.7.

- Gender:

Table (4.1) Gender Counts and Percentages

Gender	Count	Percentage
M	1820	51%
F	1774	49%
Sum	3594	100%

- Age:

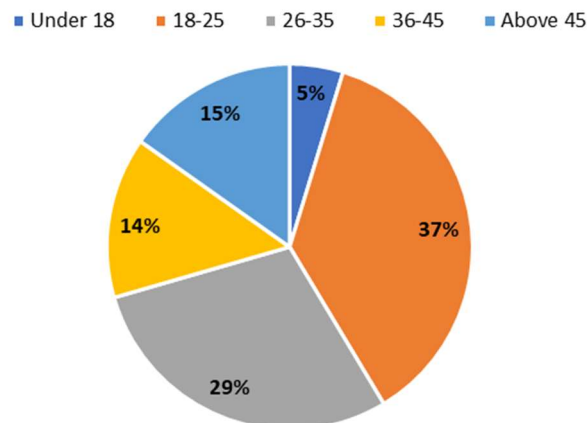


Figure (4.1) Age Data Percentages

Table 4.2) Age Counts and Percentages

Age	Count	Percentage
Under 18	158	5%
18-25	1227	37%
26-35	975	29%
36-45	477	14%
Above 45	509	15%
Sum	3346	100%

- Nationality:

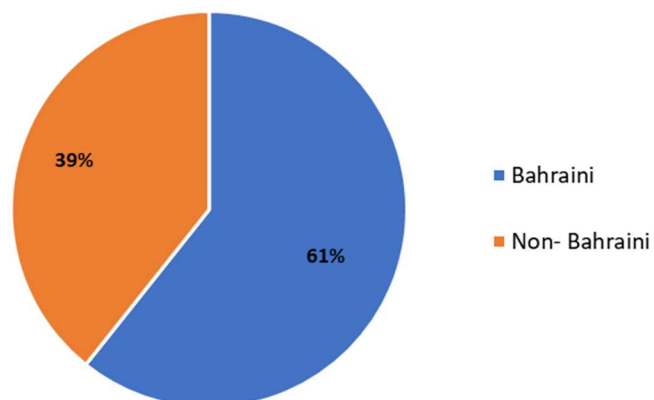


Figure 4.2) Nationality Data Percentages

Table (4.3) Nationality Counts and Percentages

Nationality	Count	Percentage
Bahraini	1224	61%
Non- Bahraini	792	39%
Sum	2016	100%

- Occupation:

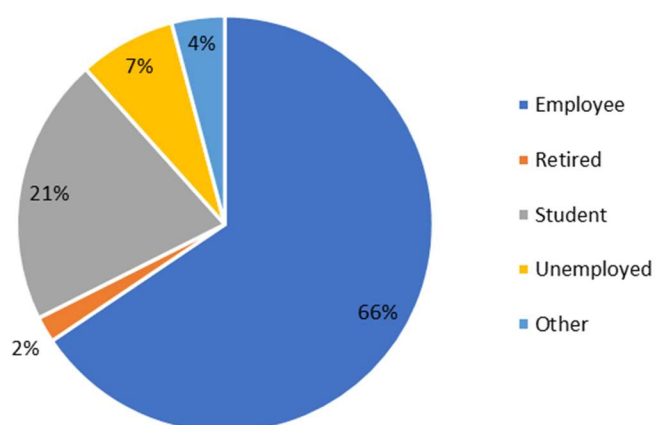


Figure (4.3) Occupation Data Percentages

Table 4.4) Occupation Counts and Percentages

Occupation	Count	Percentage
Employee	1781	66%
Retired	55	2%
Student	564	21%
Unemployed	204	8%
Other	112	4%
Sum	2716	100%

- Driving License:

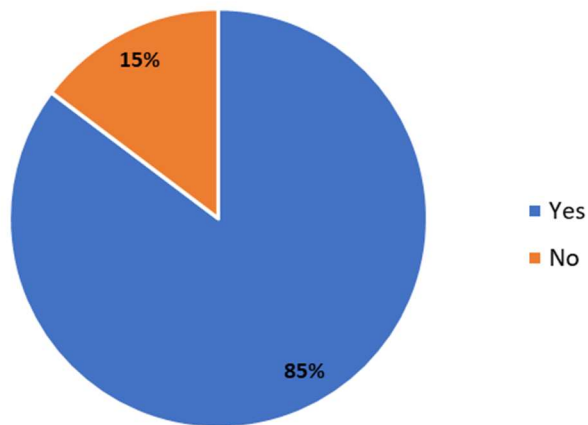


Figure (4.4): Driving License Data Percentages

Table (4.5) Driving License Counts and Percentages

Driving Licence	Count	Percentage
Yes	2114	85%
No	365	15%
Sum	2479	100%

- Number of Cars Owned:

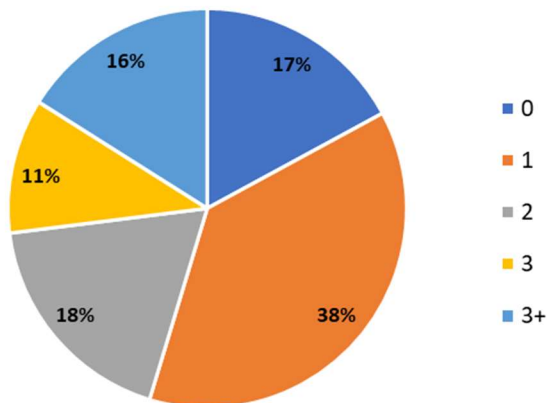


Figure 4.5) Car Ownership Data Percentages

Table 4.6) Car Ownership Counts and Percentages

Car Ownership	Count	Percentage
0	376	17%
1	826	38%
2	403	18%
3	240	11%
3+	353	16%
Sum	2198	100%

- Salary:

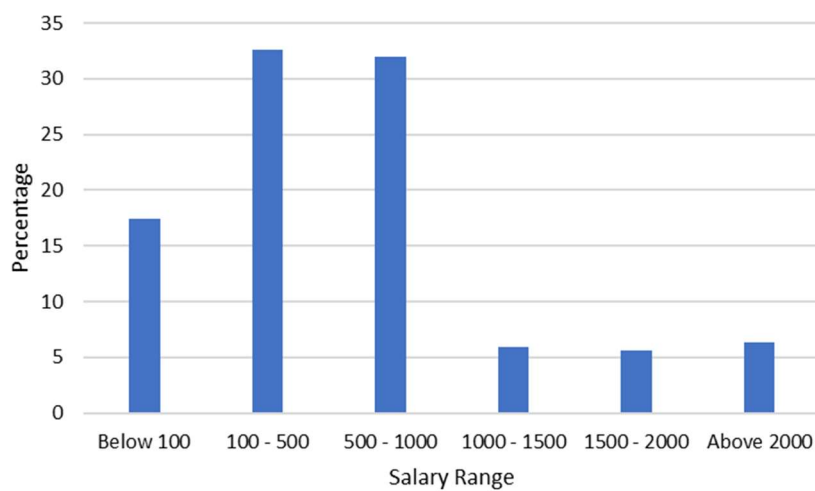


Figure (4.6) Salary Data Percentages

Table (4.7) Salary Counts and Percentages

Range	Count	Percentage
Below 100	304	17%
100 – 500	569	33%
500 – 1000	558	32%
1000 – 1500	104	6%
1500 – 2000	98	6%
Above 2000	111	6%
Sum	1744	100%

Inspecting figures 4.1 – 4.6 and tables 4.1– 4.7, the following can be deduced:

- The data represents both genders almost equally with no significant difference between them.
- The majority of respondents belongs to the age group of 18 – 25, constituting 37% of the data available for age.
- A notable proportion, 39% of 2016 responses, comprises non-Bahraini nationals.
- The greater part of the sample population are employees, accounting for 66% of the data available for occupation, and around 21% reported being students.
- 85% of 2479 respondents reported holding a driving license.
- The majority of the sample own 1 car only.
- The most common ranges of salary are 100 – 500 BD and 500 – 1000 BD, accounting for 33% and 32% of 1744 responses, respectively.

4.2.2 Trip-related Items:

The results of the statistical analysis of trip-related items including Origin-Destination, purpose of trip, travel mode, travel time and total cost of trip are displayed in the subsequent figures 4.7 – 4.12 and tables 4.8 – 4.13.

- Origin-Destination:

Table (4.8) Origin Counts and Percentages

Origin	Count	Percentage
Northern	795	24%
Southern	747	22%
Capital	932	28%
Muharraq	712	21%
Central	118	4%
Kuwait	2	0.1%
Saudi Arabia	23	1%
Sum	3329	100%

Table 4.9) Destination Counts and Percentages

Destination	Count	Percentage
Northern	526	14%
Southern	974	26%
Capital	1662	45%
Muharraq	358	10%
Central	160	4%
Sum	3680	100%

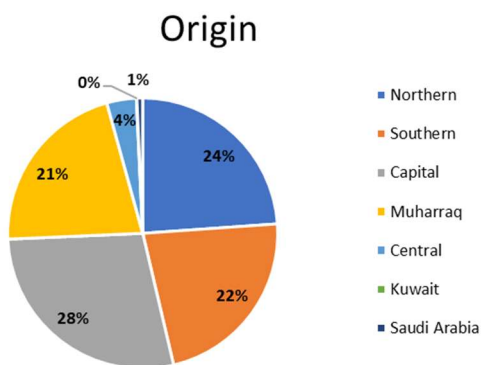


Figure (4.7) Origin Data Percentages

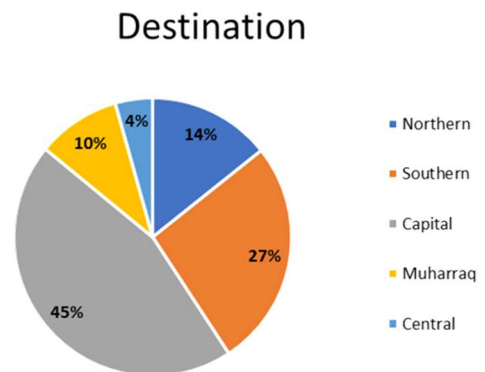


Figure 4.8) Destination Data Percentages

- Purpose of Trip:

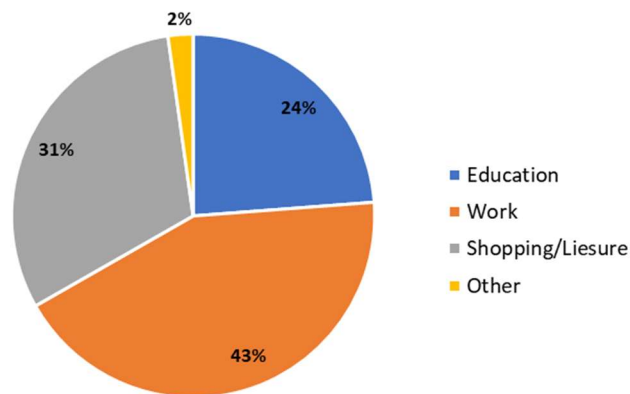


Figure 4.9) Purpose of Trip Data Percentages

Table (4.10) Purpose of Trip Counts and Percentages

Purpose	Count	Percentage
Education	726	24%
Work	1311	43%
Shopping/Leisure	946	31%
Other	68	2%
Sum	3051	100%

- Travel Mode:

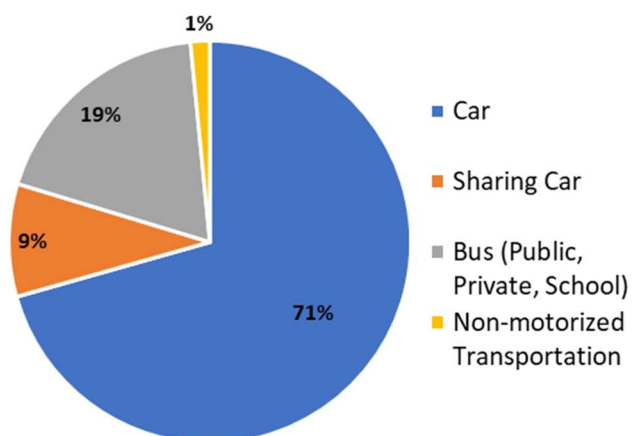


Figure (4.10) Purpose of Trip Data Percentages

Table 4.11) Travel Mode Counts and Percentages

Mode	Count	Percentages
Car	2729	71%
Sharing Car	350	9%
Bus (Public, Private, School)	725	19%
Non-motorized Transportation	60	2%
Sum	3864	100%

- Travel Time

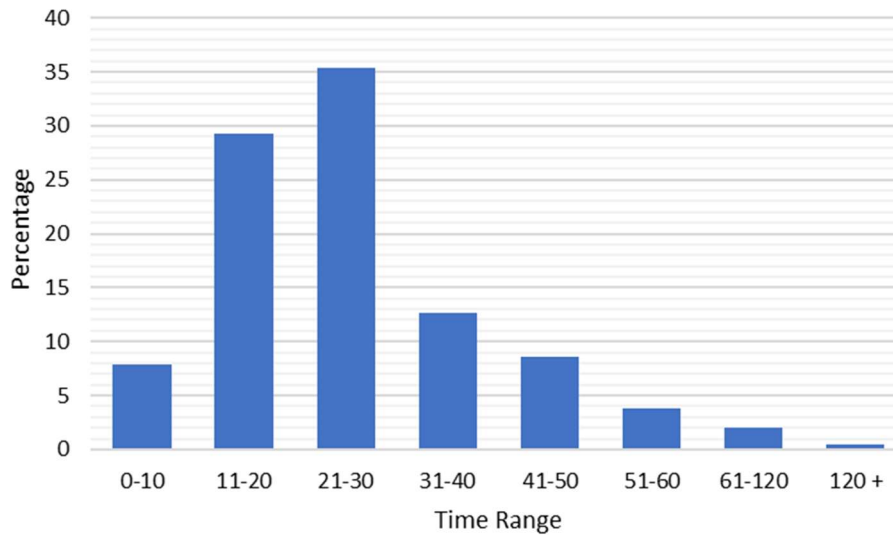


Figure (4.11) Travel Time Data Percentages

Table 4.12) Travel Time Counts and Percentages

Travel Time	Count	Percentage
0-10	233	8%
11-20	867	29%
21-30	1047	35%
31-40	375	13%
41-50	254	9%
51-60	112	4%
61-120	59	2%
120 +	13	0.44%
Sum	2960	100%

- Average Cost of Trip:

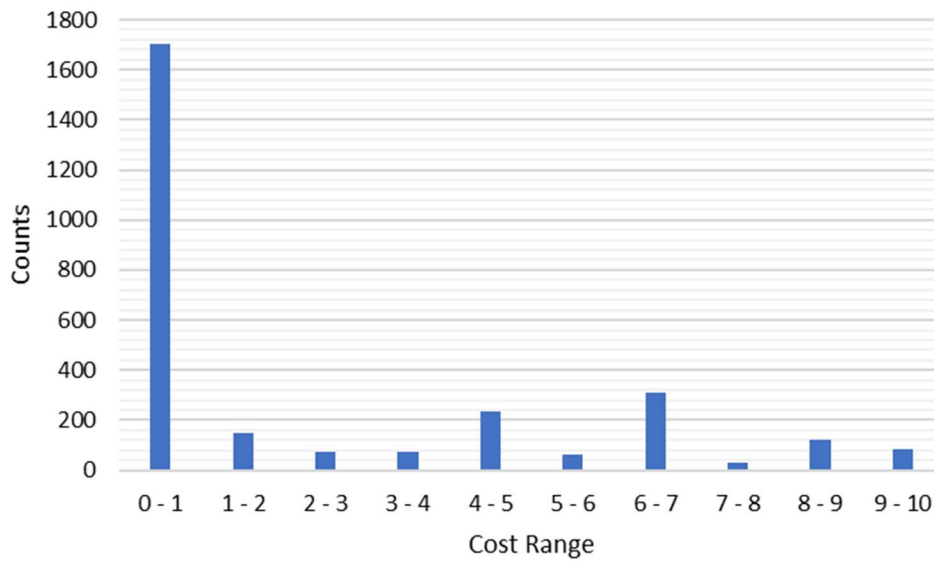


Figure 4.12) Average Cost of Trip Counts

Table (4.13) Average Cost Metrics

Count	2845
Average	2.53
Mode	0.3
Median	0.55

Referring to figures 4.7 – 4.12 and tables 4.8 – 4.13, the following can be concluded:

- The data indicates that trips produced in the country are almost equally distributed among the four governorates.
- The Capital governorate is the most common destination, accounting for 45% of trips recorded.
- Work is the primary purpose of the majority of trips (43%), followed by shopping/Leisure (31%) and education (24%).
- 71% of respondents choose to travel using their private cars while only 19% use bus (public, private, school) services.

- The most common ranges of travel time are 11 – 20 mins and 21 – 30 mins, with roughly one third of the data available in each range.
- The average cost of a trip is 2.53 BD, with 0.3 BD being the most common cost.

4.2.3 Future Mode-related Items

The figures 4.13 – 4.15 and tables 4.14 – 4.19 below demonstrate a thorough summary of counts and percentages pertaining to the commuters' future public transport modes and feeders' preferences.

- Public Bus:

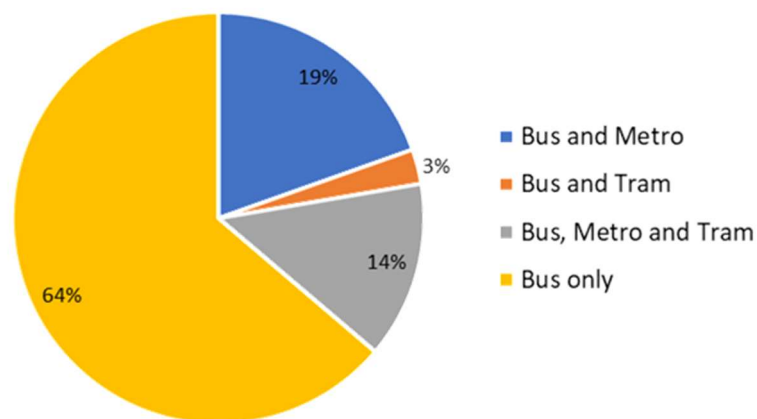


Figure 4.13) Public Bus Users Counts

Table 4.14) Public Bus Users Counts

Mode	Count	Percentage
Bus and Metro	187	20%
Bus and Tram	26	3%
Bus, Metro and Tram	133	14%
Bus only	609	64%
Sum	955	100%

Table (4.15) Public Bus Users' Feeders Preferences

Feeder	Count	Percentage
Car	45	20%
Bus	77	33%
Walking	38	17%
Cycling	23	10%
Car, Bus	7	3%
Car, Walking	13	6%
Car, Bus, Walking	6	3%
Car, Bus, Walking, Cycling	2	1%
Bus, Walking	9	4%
Bus, Cycling	3	1%
Walking, Cycling	7	3%
Sum	230	100%

- Metro:

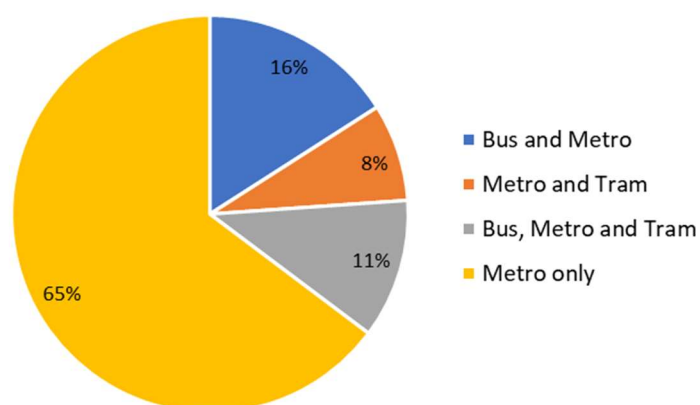


Figure 4.14) Metro Users Counts

Table (4.16) Metro Users Counts

Mode	Count	Percentage
Bus and Metro	187	16%
Metro and Tram	93	8%
Bus, Metro and Tram	133	11%
Metro only	759	65%
Sum	1172	100%

Table 4.17) Metro Users' Feeders Preferences

Feeder	Count	Percentage
Car	282	28%
Bus	144	15%
Walking	210	21%
Cycling	42	4%
Car, Bus	30	3%
Car, Walking	127	13%
Car, Cycling	3	0%
Car, Bus, Walking	22	2%
Car, Bus, Cycling	3	0%
Car, Walking, Cycling	25	3%
Car, Bus, Walking, Cycling	7	1%
Bus, Walking	30	3%
Bus, Cycling	31	3%
Bus, Walking, Cycling	4	0.40%
Walking, Cycling	31	3%
Sum	991	100%

- Tram:

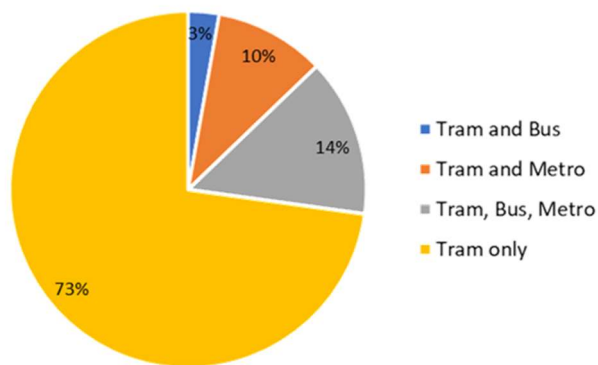


Figure 4.15) Tram Users Counts

Table (4.18) Tram Users Counts

Mode	Count	Percentage
Tram and Bus	26	3%
Tram and Metro	93	10%
Tram, Bus, Metro	133	14%
Tram only	675	73%
Sum	927	100%

Table (4.19) Tram Users' Feeders Preferences

Feeder	Count	Percentage
Car	704	86%
Bus	7	1%
Walking	46	6%
Cycling	5	1%
Car, Bus	7	1%
Car, Walking	22	3%
Car, Cycling	2	0.25%
Car, Bus, Walking	4	0.49%
Car, Bus, Cycling	1	0.12%
Car, Walking, Cycling	6	1%
Bus, Walking	5	1%
Walking, Cycling	7	1%
Sum	816	100%

Inspecting figures 4.13 – 4.15 and tables 4.14 – 4.19, the following can be deduced:

- 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.

4.3 Model Building

The process of developing mode choice models requires a rigorous procedure of data preparation, model building, and validation. This section will demonstrate the methodology employed in generating mode choice models for Bahrain, which is carried out in two phases and includes the development of both logit and classification tree models using ‘Minitab Statistics’. Additionally, the section details the measures taken to establish the reliability of the dataset and the accuracy of the models developed.

4.3.1 Phase 1

The main focus of phase 1 of this research is to build a mode choice model that portrays the present transportation conditions in Bahrain. This involves constructing a binary logit model and a binary classification tree model concerning the modes Car (including Car, Sharing Car) and Bus (including Public, Private and School).

4.3.1.1 Logit Model

4.3.1.1.1 Data Preparation

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 binary logit model, the data variables were classified as either continuous predictors or categorical predictors. Table 4.20 presented below displays a summary of the scales utilized for each category. It is noteworthy to mention that non-motorized modes were excluded from the model due to their diminutive percentage.

Table 4.20) Variables for Modeling Phase 1

Category	Variable	Scale
Continuous	Age	Continuous Number
	Travel Time	Continuous Number
	Trip Cost	Continuous Number
	Salary	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
	Current Mode	1 for Car , 2 for Bus
	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 License	1 for Yes, 0 for No
	Car Ownership	Discrete Numbers (1 - 4)

4.3.1.1.2 The Model

To develop a binary logit model, several models were built utilizing various variables until one that had an adequate goodness of fit and accuracy was achieved. The optimal model was selected through careful evaluation and testing of different variables until the best combination was identified. The details of the model equations, coefficients analysis, odd ratios and goodness-of-fit tests' results are described below.

Regression Equations:

The selected model makes use of 1396 responses, neglecting 2408 responses due to missing values. It employs five variables, three of which are of a continuous nature, namely travel time, trip cost and salary, and two categorical variables: trip purpose and occupation. Adding to that, around 80% and 20% of the data, used for this model, are associated to car and bus users respectively, hence maintaining the percentages obtained from the statistical analysis. Figures 4.16 – 4.17 present the resulting utility functions for both Car and Bus. Note that C12 and C19 refer

to the trip purpose and occupation, respectively. More details about these variables can be observed in Table 4.20.

$$P(1) = \exp(Y') / (1 + \exp(Y'))$$

C12 C19

1	1	$Y' = -0.4280 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
1	2	$Y' = -2.111 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
1	3	$Y' = -1.389 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
2	1	$Y' = 1.379 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
2	2	$Y' = -0.3040 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
2	3	$Y' = 0.4178 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
3	1	$Y' = 2.448 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
3	2	$Y' = 0.7652 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
3	3	$Y' = 1.487 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
4	1	$Y' = 3.513 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
4	2	$Y' = 1.830 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$
4	3	$Y' = 2.552 - 0.01267 \text{ Average Travel Time} + 0.9388 \text{ Total cost of trip} + 0.001054 \text{ Average Salary}$

Figure 4.16) Car (1) Utility Functions

$$P(2) = \exp(Y') / (1 + \exp(Y'))$$

C12 C19

1	1	$Y' = 0.4280 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
1	2	$Y' = 2.111 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
1	3	$Y' = 1.389 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
2	1	$Y' = -1.379 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
2	2	$Y' = 0.3040 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
2	3	$Y' = -0.4178 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
3	1	$Y' = -2.448 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
3	2	$Y' = -0.7652 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
3	3	$Y' = -1.487 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
4	1	$Y' = -3.513 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
4	2	$Y' = -1.830 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$
4	3	$Y' = -2.552 + 0.01267 \text{ Average Travel Time} - 0.9388 \text{ Total cost of trip} - 0.001054 \text{ Average Salary}$

Figure 4.17) Bus (2) Utility Functions

Coefficients Analysis:

By examining the coefficients' analysis thoughtfully, the importance of each variable in the model can be assessed. This is accomplished using Wald approximation tests to calculate the Z-value and P-value of each variable. The Z-value computes the ratio between the coefficient and its standard error indicating whether its estimate is large and precise (far from 0) or small and too imprecise (close to 0). Inspecting table 4.21 below, it can be concluded that all coefficients have precise estimates to be statistically not equal to 0 since the Z-values calculated are sufficiently far from 0. Successively, Minitab utilizes the Z-value to calculate the P-value which is a probability that demonstrates evidence against the null hypothesis which suggests that there is no association between the independent variable and the dependent variable. As commonly used in analysis, a 95% confidence level implies that if the P-value is less than or equal to 0.05, the variable is considered statistically significant. Accordingly, based on the P-values shown in table 4.21, it can be deduced that there is no evidence for the null hypothesis and therefore all variables utilized in the model are significant (Mendenall, 2013).

Table (4.21) Binary Logit Model Coefficient Analysis

Term	Coefficient	Standard Error	Z-Value	P-Value
Constant	-0.428000	0.204	-2.100	0.035
Travel Time	-0.012670	0.004	-3.030	0.002
Total cost of trip	0.938800	0.082	11.420	0.000
Salary	0.001054	0.000	4.330	0.000
C12				
2	1.807000	0.674	2.680	0.007
3	2.876000	0.714	4.030	0.000
4	3.940000	1.120	3.520	0.000
C19				
2	-1.683000	0.680	-2.480	0.013
3	-0.961000	0.411	-2.340	0.019

Odds Ratios:

Odds ratios are prominent in statistical analysis as they can be used to measure the odds of an outcome of interest, occurring with association to a specific variable. In general, the odds ratios for continuous variables demonstrate the change in the odds of the dependent variable with a unit increase in the independent variable, whereas odds ratios of categorical predictors compare the odds of the outcome of non-reference categories to that of the reference category. This subsequently helps in identifying the extent of each variable's effect on the outcome and assists informed decision-making. Tables 4.22 and 4.23 below display the odd ratios for both continuous and categorical predictors utilized in the logit model (Kleinbaum, 1988).

Table 4.22) Odd Ratios for Continuous Predictors (Phase 1)

Odds Ratios for Continuous Predictors		
	Odds Ratio	95% CI
Average Travel Time	0.98	(0.97, 0.99)
Total cost of trip	2.55	(2.17, 3.00)
Average Salary	1.00	(1.00, 1.00)

Table 4.23) Odd Ratios for Categorical Predictors (Odds ratio for level A relative to level B) (Phase 1)

Odds Ratios for Categorical Predictors			
Level A	Level B	Odds Ratio	95% CI
C12			
2	1	6.09	(1.62, 22.84)
3	1	17.74	(4.37, 71.94)
4	1	51.46	(5.75, 460.47)
3	2	2.91	(0.57, 14.68)
4	2	8.45	(0.95, 74.80)
4	3	2.90	(0.26, 31.36)
C19			
2	1	0.18	(0.04, 0.70)
3	1	0.38	(0.17, 0.85)
3	2	2.05	(0.61, 6.84)

Goodness-of-fit:

Predominantly, the goodness-of-fit tests' results shown in table 4.24 suggest that the model fits the data well, as both deviance and Pearson chi-square tests have a high P-value of 1. This implies that the model's estimates are not significantly different from the observed data. However, since the Hosmer-Lemeshow test has a low P-value of 0, this suggests that there is some evidence of lack of fit possibly due to the large sample size used in the analysis (Kutner, 2004 ; Hadi, 2015).

Table (4.24) Binary Logit Model Goodness-of-fit

Test	Degree of Freedom	Chi - Square	P – Value
Deviance	1387	842	1.00
Pearson	1387	1162	1.00
Hosmer – Lemeshow	8	50	0.00

Therefore, to better evaluate the model's goodness-of-fit, the normal probability plot and standard errors vs. fits were generated (figure 4.18 and 4.19). Based on statistical principles and interpretation of diagnostic plots commonly used in regression analysis, the approximate linear form of both plots suggests that the model has an acceptable fit to the data and that the linearity assumption between the dependent and independent variables within the utility functions has been met (Kutner, 2004; Hadi, 2015). Subsequently, despite the lack of fit provided by Hosmer-Lemeshow test, the other tests and diagnostic plots support not rejecting the model generated in this study.

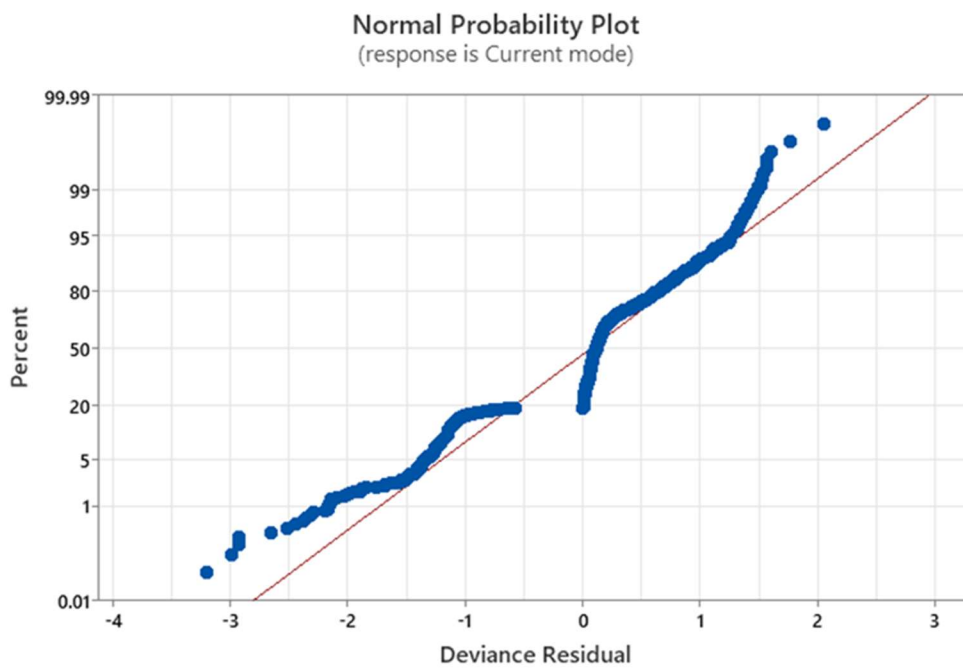


Figure (4.18) Normal Probability Plot (Percent vs. Residual) (Phase 1)

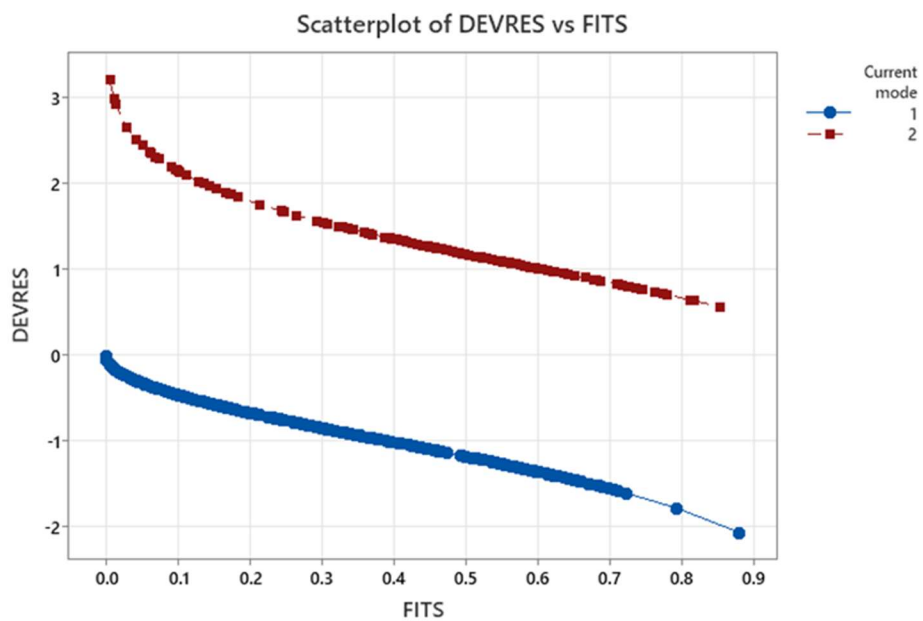


Figure 4.19) Residuals vs. Fitted Values Plot (Devers vs. Fits)

4.3.1.2 Classification Tree Model

4.3.1.2.1 Data Preparation

Since the classification tree model is an AI algorithmic model, no major alteration or adjustment is required to prepare the raw data for analysis. However, to facilitate the model building process and improve its performance, non-motorized transport users, which account for a small percentage, were excluded and the car ownership groups (3) and (3+) were merged together.

4.3.1.2.2 The Model

The software generates various trees within its algorithm, while calculating the misclassification cost (error) of each tree to determine the number of terminal nodes necessary to achieve the optimal model. Figure 4.20 presented below illustrates the plot used to identify the optimal model (35 terminal nodes) accounting for a misclassification cost of 0.2472, which is within 1 standard error of the minimum misclassification cost recorded.

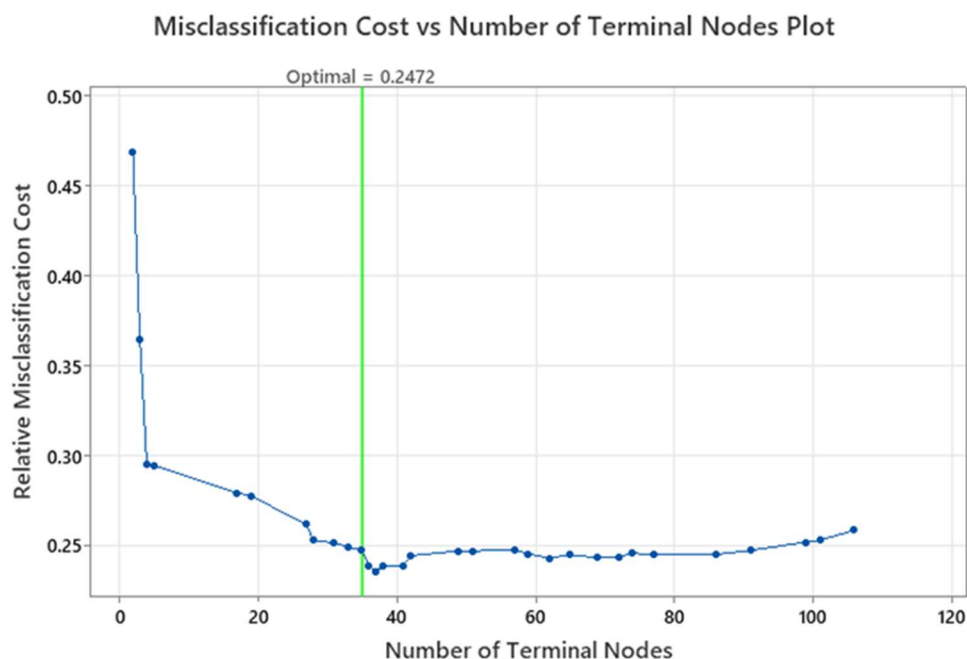


Figure (4.20) Misclassification Cost vs. Number of Terminal Nodes Plot (Phase 1)

The details of the model summary and tree diagram are described below.

Model Summary:

The model utilizes 3804 responses, 3079 (80.94%) and 725 (19.06%) of which belong to Car and Bus users respectively. It consists of 35 decisive nodes and makes use of 11 important predictors, namely trip cost, car ownership, Origin-Destination, age, gender, purpose of trip, travel time, driving license, occupation and Salary. Figure 4.21 demonstrates the relative variable importance, which is the percentage of improvement achieved when splits are made on a predictor, with respect to the top predictor (trip cost).

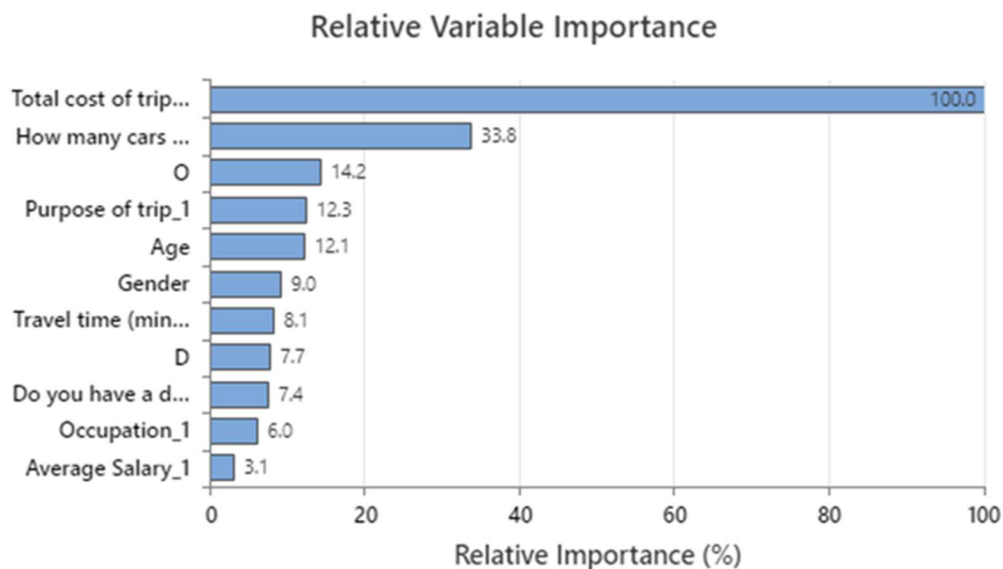


Figure (4.21) Relative Variable Importance (Phase 1)

Tree Diagram:

As mentioned above, the generated classification tree model consists of 35 terminal nodes, out of which 20 represent Car mode and 15 represent Bus mode. A sample view of the detailed tree alongside the simplified view is illustrated in Figures 4.22 and 4.23. Besides that, for easy interpretation of results, the node rules are presented in Tables 4.25 and 4.26.

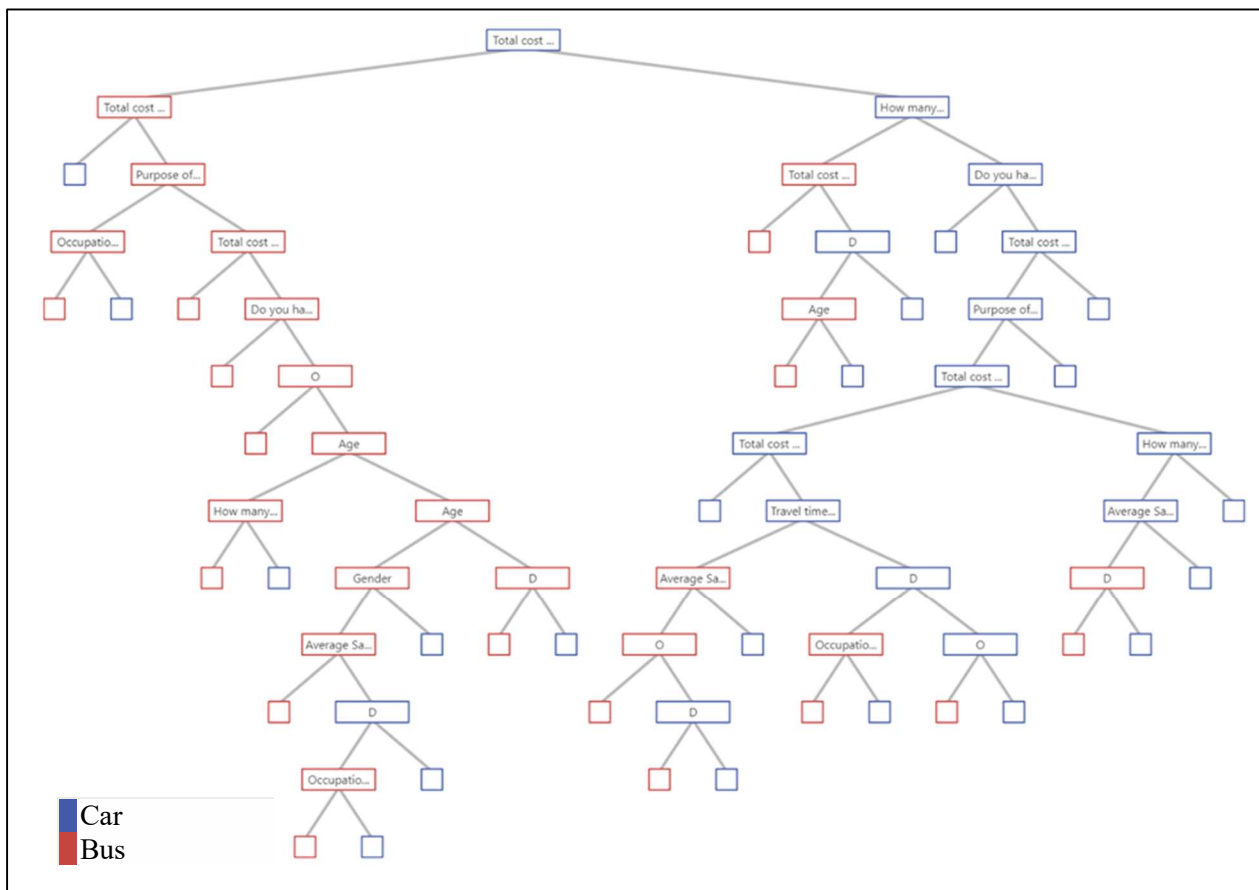


Figure 4.22) Simplified View of the Classification Tree Model (Phase 1)

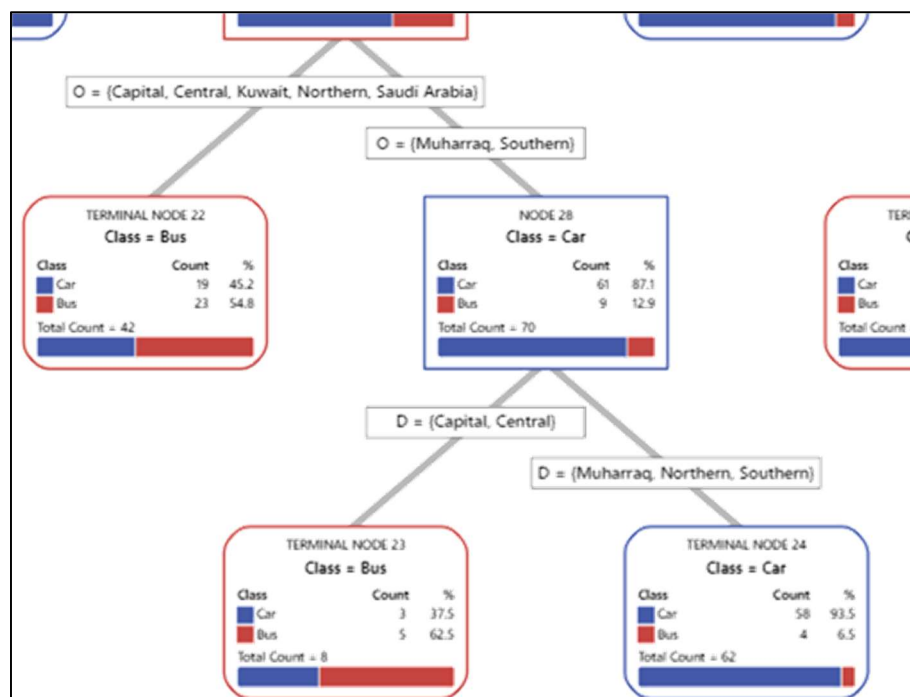


Figure (4.23) A Sample View of the Detailed Tree (Phase 1)

Table (4.25) Node Rules for Car Mode (Phase 1)

No	Car
1	Total cost of trip_1 ≤ 0.175
2	0.175 < Total cost of trip_1 ≤ 0.325, Purpose of trip_1 = {Other, Shopping/Leisure, Work}, Occupation_1 = {Unemployed}
3	0.2875 < Total cost of trip_1 ≤ 0.325, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, Age = {18 - 25, 26 - 35, Below 18}, How many cars do you own_1 = {1}, O = {Capital}
4	0.2875 < Total cost of trip_1 ≤ 0.325, Gender = {M}, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, Age = {36 - 45}, O = {Capital}
5	0.2875 < Total cost of trip_1 ≤ 0.325, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, Age = {Above 45}, D = {Muharraaq, Southern}, O = {Capital}
6	0.2875 < Total cost of trip_1 ≤ 0.325, Gender = {F}, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, Age = {36 - 45}, D = {Southern}, Average Salary_1 = {1000 - 1500 BD, 1500 - 2000, Above 2000}, O = {Capital}
7	0.2875 < Total cost of trip_1 ≤ 0.325, Gender = {F}, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, Occupation_1 = {Employee}, Age = {36 - 45}, D = {Capital, Central, Muharraaq, Northern}, Average Salary_1 = {1000 - 1500 BD, 1500 - 2000, Above 2000}, O = {Capital}
8	Total cost of trip_1 > 0.91665, D = {Northern, Southern}, How many cars do you own_1 = {0}
9	Total cost of trip_1 > 0.91665, Age = {26 - 35, 36 - 45, Above 45}, D = {Capital, Central, Muharraaq}, How many cars do you own_1 = {0}
10	Total cost of trip_1 > 0.325, Do you have a driving licence_1 = {No}, How many cars do you own_1 = {1, 2, 3}
11	Total cost of trip_1 > 3.875, Do you have a driving licence_1 = {Yes}, How many cars do you own_1 = {1, 2, 3}
12	0.325 < Total cost of trip_1 ≤ 3.875, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Other, Shopping/Leisure}, How many cars do you own_1 = {1, 2, 3}
13	0.325 < Total cost of trip_1 ≤ 0.48335, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, How many cars do you own_1 = {1, 2, 3}
14	0.48335 < Total cost of trip_1 ≤ 0.8833, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, D = {Muharraaq, Northern, Southern}, How many cars do you own_1 = {1, 2, 3}, Average Salary_1 = {100 - 500 BD, 1500 - 2000, 500 - 1000 BD, Below 100 BD}, O = {Muharraaq, Southern}, Travel time (min)_1 = {0 - 10, 11 - 20, 120 +, 31 - 40, 41 - 50, 51 - 60, 61 - 120}
15	0.48335 < Total cost of trip_1 ≤ 0.8833, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, How many cars do you own_1 = {1, 2, 3}, Average Salary_1 = {1000 - 1500 BD, Above 2000}, Travel time (min)_1 = {0 - 10, 11 - 20, 120 +, 31 - 40, 41 - 50, 51 - 60, 61 - 120}

16	0.48335 < Total cost of trip_1 <= 0.8833, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, Occupation_1 = {Student, Unemployed}, D = {Capital, Central, Muharraq}, How many cars do you own_1 = { 1, 2, 3}, Travel time (min)_1 = {21 - 30}
17	0.48335 < Total cost of trip_1 <= 0.8833, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, D = {Northern, Southern}, How many cars do you own_1 = { 1, 2, 3}, O = {Muharraq, Northern, Southern}, Travel time (min)_1 = {21 - 30}
18	0.8833 < Total cost of trip_1 <= 3.875, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, D = {Muharraq, Northern}, Average Salary_1 = {100 - 500 BD}
19	0.8833 < Total cost of trip_1 <= 3.875, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, Average Salary_1 = {1000 - 1500 BD, 1500 - 2000, 500 - 1000 BD, Above 2000, Below 100 BD}
20	0.8833 < Total cost of trip_1 <= 3.875, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, How many cars do you own_1 = {1, 2, 3}

Note: O = Origin. D = Destination

Table 4.26) Node Rules for Bus Mode

No	Bus
1	0.175 < Total cost of trip_1 <= 0.325, Purpose of trip_1 = {Other, Shopping/Leisure, Work}, Occupation_1 = {Employee, Other, Student}
2	0.175 < Total cost of trip_1 <= 0.2875, Purpose of trip_1 = {Education}
3	0.2875 < Total cost of trip_1 <= 0.325, Do you have a driving licence_1 = {No}, Purpose of trip_1 = {Education}
4	0.2875 < Total cost of trip_1 <= 0.325, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, O = {Central, Kuwait, Muharraq, Northern, Saudi Arabia, Southern}
5	0.2875 < Total cost of trip_1 <= 0.325, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, Age = {18 - 25, 26 - 35, Below 18}, How many cars do you own_1 = {0, 2, 3}, O = {Capital}
6	0.2875 < Total cost of trip_1 <= 0.325, Gender = {F}, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, Age = {36 - 45}, Average Salary_1 = {100 - 500 BD, 500 - 1000 BD, Below 100 BD}, O = {Capital}
7	0.2875 < Total cost of trip_1 <= 0.325, Gender = {F}, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, Occupation_1 = {Other, Student, Unemployed}, Age = {36 - 45}, D = {Capital, Central, Muharraq, Northern}, Average Salary_1 = {1000 - 1500 BD, 1500 - 2000, Above 2000}, O = {Capital}
8	0.2875 < Total cost of trip_1 <= 0.325, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education}, Age = {Above 45}, D = {Capital, Central, Northern}, O = {Capital}
9	0.325 < Total cost of trip_1 <= 0.91665, How many cars do you own_1 = {0}
10	Total cost of trip_1 > 0.91665, Age = {18 - 25, Below 18}, D = {Capital, Central, Muharraq}, How many cars do you own_1 = {0}
11	0.48335 < Total cost of trip_1 <= 0.8833, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, How many cars do you own_1 = {1, 2, 3}, Average Salary_1 = {100 - 500 BD, 1500 - 2000, 500 - 1000 BD, Below 100 BD}, O = {Capital, Central, Kuwait, Northern, Saudi Arabia}, Travel time (min)_1 = {0 - 10, 11 - 20, 120 +, 31 - 40, 41 - 50, 51 - 60, 61 - 120}

12	0.48335 < Total cost of trip_1 <= 0.8833, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, D = {Capital, Central}, How many cars do you own_1 = {1, 2, 3}, Average Salary_1 = {100 - 500 BD, 1500 - 2000, 500 - 1000 BD, Below 100 BD}, O = {Muharraq, Southern}, Travel time (min)_1 = {0 - 10, 11 - 20, 120 +, 31 - 40, 41 - 50, 51 - 60, 61 - 120}
13	0.48335 < Total cost of trip_1 <= 0.8833, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, Occupation_1 = {Employee, Other}, D = {Capital, Central, Muharraq}, How many cars do you own_1 = {1, 2, 3}, Travel time (min)_1 = {21 - 30}
14	0.48335 < Total cost of trip_1 <= 0.8833, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, D = {Northern, Southern}, How many cars do you own_1 = {1, 2, 3}, O = {Capital, Central, Kuwait, Saudi Arabia}, Travel time (min)_1 = {21 - 30}
15	0.8833 < Total cost of trip_1 <= 3.875, Do you have a driving licence_1 = {Yes}, Purpose of trip_1 = {Education, Work}, D = {Capital, Central, Southern}, Average Salary_1 = {100 - 500 BD}

Note: O = Origin. D = Destination

4.3.1.3 Split-half Reliability

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 (Nunnally, 1994). This section will outline the procedure followed to perform the split-half reliability test on the data used for developing the Logit model and Classification tree model, as well as present the results obtained.

4.3.1.3.1 Procedure

To conduct split-half reliability test using Excel the following steps were taken (Nunnally, 1994):

1. The sample was divided into two equivalent parts by alternating between odd and even responses.
2. Categorical data was assigned scores (such as 1, 2, 3..), while continuous items were not altered.

3. Using the CORREL function, the correlation coefficient was calculated between each column in the first half and the corresponding column in the second half.
4. The square of each correlation coefficient was computed and their sum was determined.
5. To find the total correlation coefficient, the square root of the sum divided by the number of items was obtained.
6. Using the “Spearman-Brown Formula” the adjusted total correlation coefficient was calculated (equation 6). A correlation coefficient of 0 indicates no correlation, while a value of 1 represents a perfect correlation.

$$\text{Adjusted Correlation Coefficient} = \frac{2 * R}{1 + R} \quad (6)$$

where R = total correlation coefficient

4.3.1.3.2 Logit Model Results

In order to determine the internal consistency of the data used to build the Logit model, a split-half reliability analysis was conducted by dividing the data into two samples each containing 698 responses. The correlation coefficients for each item, namely current mode (CM), travel time (AT), trip cost (TC) and salary (AS), were computed. Ultimately, the adjusted correlation coefficient was calculated to be 0.59, indicating a moderate level of consistency. The summary of results is displayed below in tables 4.27 and 4.28.

Table 4.27) Correlation Coefficient for Binary Logit Model Items

Item	Correlation Coefficient
CM	0.65
AT	0.05
TC	0.50
AS	0.02

Table (4.28): Computing the Adjusted Total Correlation Coefficient for Binary Logit Model

Term	Value
r_1^2 (CM)	0.42
r_2^2 (TC)	0.25
r_3^2 (AT)	0.00
r_4^2 (AS)	0.00
Sum	0.69
Total Correlation Coefficient	0.41
Adjusted Correlation Coefficient	0.59

4.3.1.3.3 Classification Tree Results

Similarly, the data used to construct the classification tree model was divided into two halves each consisting of 50% (1902) of the responses. The items included in the analysis were current mode (CM) travel time (TT), gender, origin (O), destination (D), occupation (OC), car ownership (CW), trip cost (TC), driving license (DL) and salary (AS) (table 4.29). As shown in table 4.30, the adjusted correlation coefficient was found to be 0.67 suggesting a moderate to high level of consistency and reliability within the data.

Table 4.29) Correlation Coefficient for Classification Tree Model Items (Phase 1)

Item	Correlation Coefficient
D	0.5028
P	0.7151
CM	0.7632
TC	0.7836
OC	0.1887
DL	0.2543
TT	0.3440
Age	0.3846
CW	0.5776
AS	0.3422
Gender	0.2902
O	0.3931

Table (4.30) Computing the Adjusted Total Correlation Coefficient for Classification Tree 1

Term	Value
r_1^2 (CM)	0.583
r_2^2 (TC)	0.614
r_3^2 (TT)	0.118
r_4^2 (Age)	0.148
r_5^2 (Gender)	0.084
r_6^2 (O)	0.154
r_7^2 (D)	0.253
r_8^2 (P)	0.511
r_9^2 (OC)	0.036
r_{10}^2 (DL)	0.065
r_{11}^2 (CW)	0.333
r_{12}^2 (AS)	0.117
Sum	3.017
Total Correlation Coefficient	0.501
Adjusted Correlation Coefficient	0.670

4.3.1.4 Validation

To evaluate the predictive performance of a model, it is crucial to employ validation methods. Therefore, in this research, to ensure that both the Logit and the Classification Tree models are generated with an adequate accuracy, 10-fold cross validation technique was utilized. This method involves dividing the dataset into ten equal samples (folds), in which the model is trained on nine of them and tested on the remaining one. The process is then repeated ten times, so that each fold serves as the test sample exactly once. As a result, ten varying estimates of the model's performance are generated and by finding its average for the 10 folds, a more accurate estimate can be computed (Hastie, 2009). This section details the results of the 10-fold cross-validation analysis conducted while constructing the Logit and Classification Tree models.

4.3.1.4.1 Logit Model

The 10-fold cross validation analysis conducted on the Logit model resulted in a 10-fold deviance R-sq of 34.24% and a 10-fold area under the receiver operating characteristics (ROC) curve of 0.8837 (table 4.31).

The Area Under Curve (AUC) ROC is a method commonly used to evaluate the performance of binary classifier models. It measures the model's ability to distinguish between classes by plotting the true positive rate against the false positive rate at each threshold value. The area under this plot ranges from 0 to 1, where a value of 0.5 indicates a model with no predictive power and a value of 1 is considered perfect (Fawcett, 2006). Therefore, despite the relatively low deviance R-sq, the AUC ROC results in table 4.31 suggest that the binary model has a good predictive power and can accurately distinguish between the two modes (car, bus). In fact, the low deviance R-sq should not necessarily be interpreted as poor performance of the model as it is a measure that's highly affected by the sample size and is generally used to compare different models alongside other measures rather than evaluating the performance solely on its basis (Fawcett, 2006).

Table 4.31) 10-Fold Cross Validation Results for the Binary Logit Model

Deviance R-Sq	Deviance R-Sq (adj)	AIC	AICc	BIC	Area Under ROC Curve	10-fold Deviance R-Sq	10-fold Area Under ROC Curve
37.53%	36.94%	860	860	907	0.89	34.24%	0.88

4.3.1.4.2 Classification Tree

Referring to the results shown in tables 4.32 – 3.33, the 10-fold cross-validation analysis done for the classification tree model demonstrates that the model has a great ability to accurately distinguish between positive and negative classes for both the training and testing datasets. The AUC ROC (figure 4.24) is calculated as 0.9505 for the training data and as 0.9236 for the testing data. Additionally, the percentage of misclassified cases, which is the overall error rate shown in table 4.33, is largely insignificant being as low as 11% for training and 13.4% for testing.

Table 4.32) Confusion Matrix Rates for Classification Tree 1

Statistics	Training (%)	Test (%)
True positive rate (sensitivity or power)	88.0	86.0
False positive rate (type I error)	6.8	10.8
False negative rate (type II error)	12.0	14.0
True negative rate (specificity)	93.2	89.2

Table (4.33) Misclassification Rates for Phase 1 Classification Tree

		Training			Test		
Actual Class	Count	Misclassified	% Error	Cost	Misclassified	% Error	Cost
Car (Event)	3079	369	12.0	0.11	430	14.0	0.13
Bus	725	49	6.8	0.06	78	10.8	0.10
All	3804	418	11.0	0.09	508	13.4	0.12

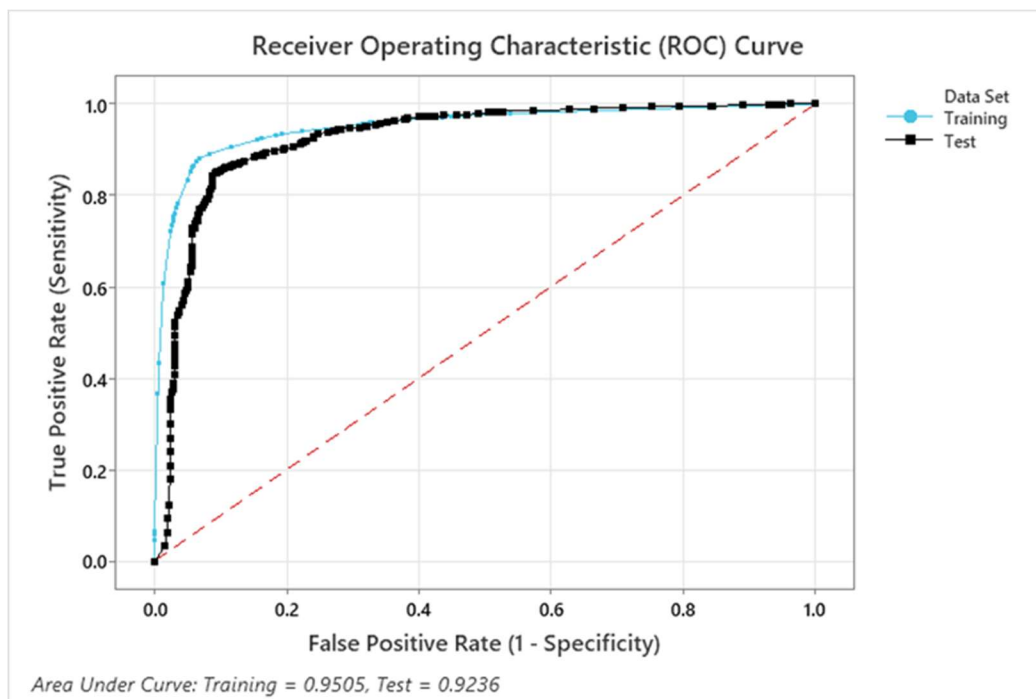


Figure 4.24) ROC for Phase 1 Classification Tree

4.3.2 Phase 2

This section discusses the second phase of model building which focuses on creating a model that can effectively forecast future public transport mode choice in Bahrain. Specifically, this phase concentrates on four modes of transportation - public bus, metro, tram, or none (car) - and utilizes the data collected to develop two models, namely multinomial logistic regression model and multinomial classification tree.

4.3.2.1 Logit Model

4.3.2.1.1 Data Preparation

Similar to phase 1, data was restructured to the scales mentioned in Table 4.20, substituting “current mode” with “future mode”. Following that, it 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 tram and metro, their response was classified under both categories.

4.3.2.1.2 The Model

Since the software does not perform cross-validation for multinomial logit models, a slightly different procedure than the one followed in phase 1 was adopted in this phase. Firstly, 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.

Model Summary:

The selected model makes use of a total responses 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 license and car ownership. Adding to that, the logit model was developed using car as the reference mode. Tables 4.34 – 4.35 present the counts and percentages of each mode for each dataset.

Table (4.34) Data used for Logit Model training phase 2.

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

Table 4.35) Data used for Logit Model testing phase 2.

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

Coefficients Analysis:

Based on the coefficients' analysis results demonstrated below in table 4.36, it can be said that the majority of the predictors utilized in the model are statistically significant. However, it is interesting to note that specific predictors were deduced to be significant for some modes but insignificant for others. For example:

- Gender "Male" was not significant for tram and public bus but was significant for metro.
- Owning a driving license is significant for public bus only.
- Car ownership was significant for public bus and metro but not tram.
- Age was significant for tram only.

Table (4.36) Coefficient Analysis for Multinomial Logit Model.

Predictor	Coef	SE Coef	Z	P	Significance
Logit 1: (Tram/Car)					
Gender					
1	-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					
1	0.051	0.248	0.210	0.837	No
Age	0.008	0.013	0.610	0.545	No
TT	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					
1	0.615	0.182	3.380	0.001	YES
Age	-0.015	0.010	-1.540	0.123	No
TT	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

Odds Ratios:

Table 4.37 displayed below highlights the odds ratios for both continuous and categorical predictors utilized in the multinomial logit model.

Table (4.37) Odds Ratio Multinomial Logit Model.

Predictor	Odds Ratio	95% CI	
		Lower	Upper
Logit 1: (Tram/Car)			
Gender			
1	0.82	0.58	1.16
Age	0.96	0.94	0.97
TT	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			
1	1.05	0.65	1.71
Age	1.01	0.98	1.03
TT	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			
1	1.85	1.29	2.64
Age	0.99	0.97	1.00
TT	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

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 table 4.38 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. This being said, 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.

Table (4.38) 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

4.3.2.2 Classification Tree Model

4.3.2.2.1 Data Preparation

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 together.

4.3.2.2.2 The Model

Using Minitab, a multinomial classification tree was developed having 40 terminal nodes and a misclassification cost of 0.4487. Figure 4.25 presented below illustrates the plot used to identify the optimal model.

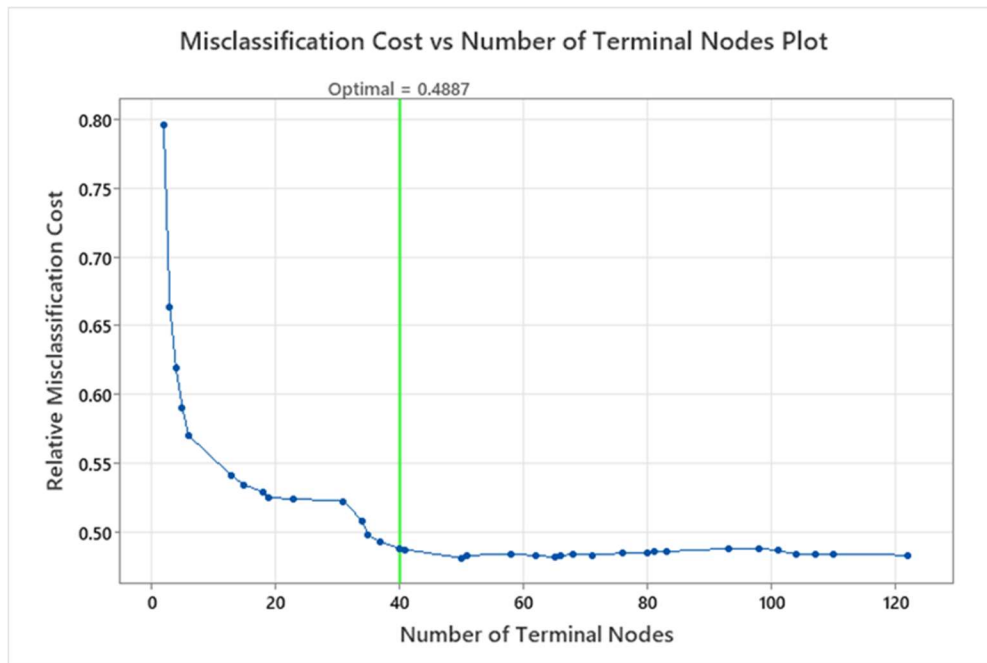


Figure 4.25) Misclassification Cost vs. Number of Terminal Nodes Plot (Phase 2)

The details of the model summary and tree diagram are described below.

Model Summary:

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 license, occupation and Salary. Figure 4.26 demonstrates the relative variable importance for each predictor with respect to the top predictor, which is trip cost.

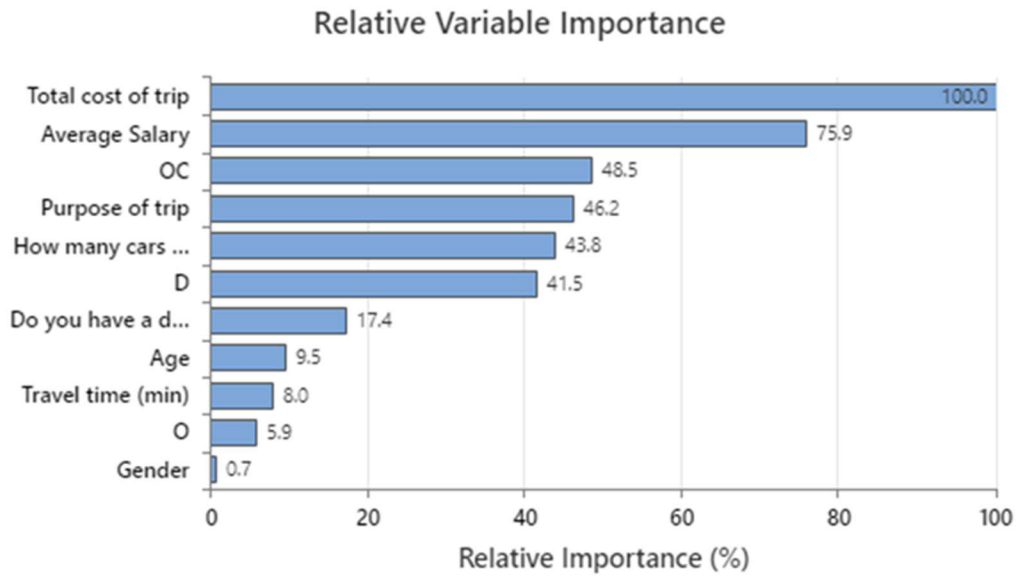


Figure 4.26) Relative Variable Importance (Phase 2)

Tree Diagram:

As previously mentioned, 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 alongside the node rules are presented in Figure 4.27 and Tables 4.39 – 4.42 .

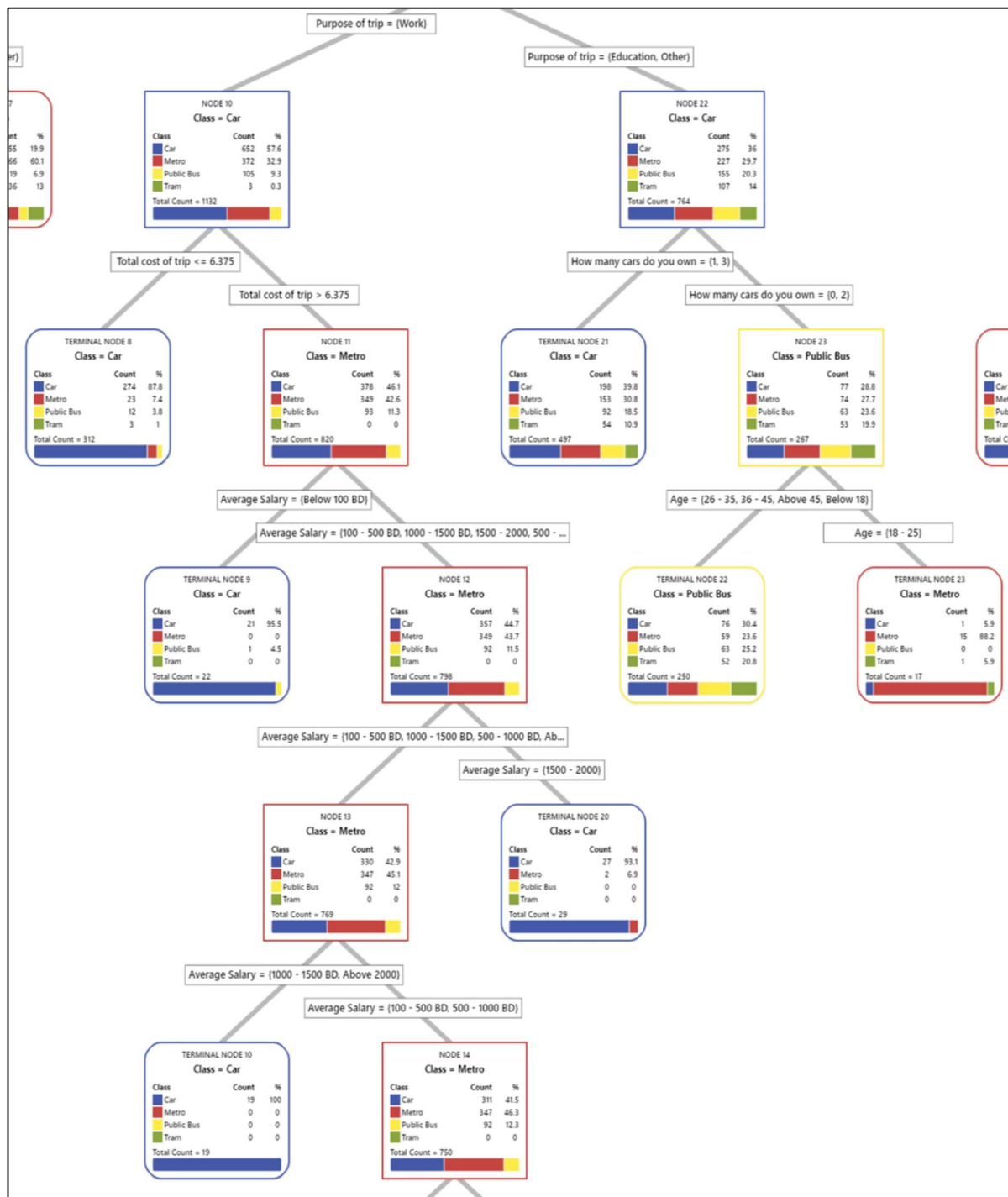


Figure (4.27) A Sample View of the Detailed Tree Phase 2

Table (4.39) Node Rules for Car Mode (Phase 2)

No	Car
1	Total cost of trip ≤ 0.05 , Purpose of trip = {Education, Other, Work}
2	$0.633333 < \text{Total cost of trip} \leq 3.15$, Purpose of trip = {Work}, Average Salary = {100 - 500 BD}, O = {Capital, Kuwait, Muharraq, Northern, Saudi Arabia, Southern}
3	$0.633333 < \text{Total cost of trip} \leq 3.15$, Purpose of trip = {Work}, O = {Central}
4	$3.15 < \text{Total cost of trip} \leq 6.375$, Purpose of trip = {Work}
5	Total cost of trip > 6.375 , Purpose of trip = {Work}, Average Salary = {Below 100 BD}
6	Total cost of trip > 6.375 , Purpose of trip = {Work}, Average Salary = {1000 - 1500 BD, Above 2000}
7	$6.75 < \text{Total cost of trip} \leq 8.5$, Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Capital, Central, Muharraq, Northern}, Average Salary = {500 - 1000 BD}
8	Total cost of trip > 8.5 , Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Capital, Central, Muharraq, Northern}, Average Salary = {500 - 1000 BD}, Travel time (min) = {0 - 10, 11 - 20, 21 - 30, 51 - 60}
9	Total cost of trip > 6.375 , Do you have a driving licence = {Yes}, Purpose of trip = {Work}, Average Salary = {100 - 500 BD}
10	Total cost of trip > 6.375 , Purpose of trip = {Work}, Average Salary = {1500 - 2000}
11	Total cost of trip > 3.15 , Purpose of trip = {Education, Other}, How many cars do you own = {1, 3}

Table (4.40) Node Rules for Metro Mode (Phase 2)

No	Metro
1	$0.633333 < \text{Total cost of trip} \leq 3.15$, Purpose of trip = {Work}, Average Salary = {1000 - 1500 BD, 1500 - 2000, 500 - 1000 BD, Above 2000, Below 100 BD}, O = {Capital, Kuwait, Muharraq, Northern, Saudi Arabia, Southern}
2	$0.633333 < \text{Total cost of trip} \leq 3.15$, Purpose of trip = {Education, Other}
3	Total cost of trip > 6.375 , Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {2, 3}, Average Salary = {500 - 1000 BD}
4	Total cost of trip > 6.375 , Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Southern}, Average Salary = {500 - 1000 BD}
5	$6.375 < \text{Total cost of trip} \leq 6.75$, Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Capital, Central, Muharraq, Northern}, Average Salary = {500 - 1000 BD}
6	Total cost of trip > 6.375 , Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {0}, Average Salary = {500 - 1000 BD}
7	Total cost of trip > 6.375 , Do you have a driving licence = {No}, Purpose of trip = {Work}, Average Salary = {100 - 500 BD, 500 - 1000 BD}
8	Total cost of trip > 3.15 , Purpose of trip = {Education, Other}, How many cars do you own = {0, 2}, Age = {18 - 25}
9	Total cost of trip > 0.4 , Purpose of trip = {Shopping/Leisure}, D = {Central, Muharraq, Northern, Southern}

10	Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student, Unemployed}, D = {Capital}, Average Salary = {1500 - 2000, Above 2000, Below 100 BD}
11	Purpose of trip = {Shopping/Leisure}, OC = {Employee}, D = {Capital}, Average Salary = {500 - 1000 BD}
12	0.325 < Total cost of trip <= 0.525, Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2}, Age = {36 - 45, Above 45, Below 18}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}
13	Total cost of trip <= 0.625, Purpose of trip = {Shopping/Leisure}, OC = {Employee, Unemployed}, How many cars do you own = {0, 1}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}
14	Total cost of trip <= 0.625, Purpose of trip = {Shopping/Leisure}, OC = {Other, Student}, How many cars do you own = {0}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}
15	Total cost of trip > 0.625, Purpose of trip = {Shopping/Leisure}, How many cars do you own = {0, 1}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}

Table 4.41) Node Rules for Public Bus Mode (Phase 2)

No	Public Bus
1	0.05 < Total cost of trip <= 0.633333, Purpose of trip = {Education, Other, Work}, Average Salary = {100 - 500 BD, 1500 - 2000, 500 - 1000 BD, Below 100 BD}
2	Total cost of trip > 8.5, Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Capital, Central, Muharraq, Northern}, Average Salary = {500 - 1000 BD}, Travel time (min) = {120 +, 31 - 40, 41 - 50, 61 - 120}
3	Total cost of trip > 3.15, Purpose of trip = {Education, Other}, How many cars do you own = {0, 2}, Age = {26 - 35, 36 - 45, Above 45, Below 18}
4	Total cost of trip <= 0.4, Purpose of trip = {Shopping/Leisure}, D = {Central, Muharraq, Northern, Southern}
5	0.225 < Total cost of trip <= 0.325, Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2, 3}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}

Table (4.42) Node Rules for Tram Mode (Phase 2)

No	Tram
1	0.05 < Total cost of trip ≤ 0.633333, Purpose of trip = {Education, Other, Work}, Average Salary = {1000 - 1500 BD, Above 2000}
2	Purpose of trip = {Shopping/Leisure}, OC = {Other}, D = {Capital}, Average Salary = {1500 - 2000, Above 2000, Below 100 BD}
3	Purpose of trip = {Shopping/Leisure}, OC = {Other, Student, Unemployed}, D = {Capital}, Average Salary = {500 - 1000 BD}
4	Purpose of trip = {Shopping/Leisure}, OC = {Other, Unemployed}, How many cars do you own = {2, 3}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}
5	Total cost of trip ≤ 0.225, Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2, 3}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}
6	0.325 < Total cost of trip ≤ 0.525, Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {3}, Age = {36 - 45, Above 45, Below 18}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}
7	0.325 < Total cost of trip ≤ 0.525, Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2, 3}, Age = {18 - 25, 26 - 35}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}
8	Total cost of trip > 0.525, Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2, 3}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}
9	Total cost of trip ≤ 0.625, Purpose of trip = {Shopping/Leisure}, OC = {Other, Student}, How many cars do you own = {1}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}

Note: O = Origin, D = Destination, OC =

4.3.2.3 Split-half Reliability

In accordance with the procedure detailed for split-half reliability test in phase 1, the reliability coefficients of the models constructed in phase 2 were computed. The results of the analysis are outlined below.

4.3.2.3.1 Logit Model Results

The dataset used for this model was divided into two equal samples each containing 898 responses. The analysis covered six items, namely Future mode (FM), travel time (AT), gender, age, driving license (DL) and car ownership (CW). 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. The summary of results is displayed below in tables 4.43 and 4.44.

Table 4.43) Correlation Coefficient for Multinomial Logit Model Items

Item	Correlation Coefficient
Gender	0.16
FM	0.99
Age	0.50
AT	0.32
DL	0.21
CW	0.59

Table (4.44) Computing the Adjusted Total Correlation Coefficient for Multinomial Logit Model

Term	Value
r_1^2 (Gender)	0.02
r_2^2 (FM)	0.99
r_3^2 (Age)	0.25
r_4^2 (AT)	0.10
r_5^2 (DL)	0.04
r_6^2 (CW)	0.35
Sum	1.78
Total Correlation Coefficient	0.54
Adjusted Correlation Coefficient	0.71

4.3.2.3.2 Classification Tree Results

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 (FM), travel time (TT), gender, origin (O), destination (D), occupation (OC), car ownership (CW), trip cost (TC), driving license (DL) and salary (AS) (table 4.45). As shown in table 4.46, the adjusted correlation coefficient was calculated to be 0.68 suggesting a moderate to high level of consistency and reliability within the data.

Table (4.45) Correlation Coefficient for Classification Tree Model Items

Item	Correlation Coefficient
TT	0.30
Age	0.36
FM	0.99
TC	0.76
Gender	0.25
O	0.34
OC	0.15
DL	0.27
D	0.49
P	0.65
CW	0.57
AS	0.36

Table (4.46) Computing the Adjusted Total Correlation Coefficient for Multinomial Logit Model

Term	Value
r_1^2 (FM)	0.99
r_2^2 (TC)	0.59
r_3^2 (TT)	0.09
r_4^2 (Age)	0.13
r_5^2 (Gender)	0.06
r_6^2 (O)	0.12
r_7^2 (D)	0.24
r_8^2 (P)	0.43
r_9^2 (OC)	0.02
r_{10}^2 (DL)	0.07
r_{11}^2 (CW)	0.33
r_{12}^2 (AS)	0.13
Sum	3.24
Total Correlation Coefficient	0.52
Adjusted Correlation Coefficient	0.68

4.3.2.4 Validation

In phase 2, 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 in the manner described in phase 1. In this section, the results obtained from the analysis of both models are presented.

4.3.2.4.1 Logit Model

The probabilities for each mode in both the training and testing datasets were computed and then utilized to generate ROC curves (table 4.47). The ROC curves were developed by calculating the true positive rates (equation 7) and false positive rates (equation 8) at probability thresholds ranging from 0 – 0.95 as shown in tables 4.48 – 4.49 (Fawcett, 2006). By plotting this information and

fitting a mathematical equation to the curves, the AUC values were obtained (figures 4.28 – 4.35).

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 car were relatively low. That being said, focusing on public transport modes, it can be concluded that the multinomial logit model developed within this research has an acceptable accuracy and a good predictive power.

$$\text{True Positive Rate} = \frac{\text{True positive}}{\text{Total Actual Positive}} \quad (7)$$

$$\text{False Positive Rate} = \frac{\text{False positive}}{\text{Total Actual Negative}} \quad (8)$$

Table (4.47) A Sample of the Multinomial Logit Model Resultant Probabilities

No.	Car	Tram	PB	Metro	Max	Predicted Mode
1	3%	24%	6%	67%	67%	Metro
2	14%	21%	5%	61%	61%	Metro
3	10%	50%	6%	34%	50%	Tram
4	7%	15%	27%	51%	51%	Metro
5	10%	14%	44%	32%	44%	Public Bus
6	6%	9%	10%	75%	75%	Metro
7	21%	21%	11%	47%	47%	Metro
8	10%	17%	40%	34%	40%	Public Bus
9	7%	11%	33%	49%	49%	Metro
10	5%	11%	8%	75%	75%	Metro
11	7%	11%	33%	49%	49%	Metro
12	9%	23%	33%	35%	35%	Metro
13	7%	15%	27%	51%	51%	Metro
14	9%	23%	33%	35%	35%	Metro
15	7%	11%	33%	49%	49%	Metro

Table (4.48) True and False Positive Rates (Training Data)

Car			Public Bus			Tram			Metro		
Cutoff	FPR	TPR	Cutoff	FPR	TPR	Cutoff	FPR	TPR	Cutoff	FPR	TPR
0	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
0.05	0.94	0.99	0.05	0.67	0.91	0.05	1.00	1.00	0.05	0.78	1.00
0.10	0.70	0.89	0.10	0.19	0.49	0.10	0.97	0.99	0.10	0.61	0.99
0.15	0.39	0.58	0.15	0.09	0.34	0.15	0.82	0.96	0.15	0.56	0.97
0.2	0.15	0.27	0.20	0.07	0.32	0.20	0.66	0.90	0.20	0.48	0.94
0.25	0.06	0.14	0.25	0.05	0.28	0.25	0.47	0.84	0.25	0.44	0.93
0.30	0.03	0.07	0.30	0.03	0.21	0.30	0.38	0.81	0.30	0.41	0.90
0.35	0.01	0.02	0.35	0.01	0.08	0.35	0.28	0.76	0.35	0.34	0.80
0.40	0.00	0.00	0.40	0.00	0.03	0.40	0.22	0.73	0.40	0.31	0.76
0.45	0.00	0.00	0.45	0.00	0.00	0.45	0.22	0.72	0.45	0.26	0.66
0.50	0.00	0.00	0.50	0.00	0.00	0.50	0.19	0.69	0.50	0.16	0.49
0.55	0.00	0.00	0.55	0.00	0.00	0.55	0.15	0.65	0.55	0.14	0.44
0.60	0.00	0.00	0.60	0.00	0.00	0.60	0.12	0.59	0.60	0.13	0.41
0.65	0.00	0.00	0.65	0.00	0.00	0.65	0.10	0.56	0.65	0.04	0.12
0.70	0.00	0.00	0.70	0.00	0.00	0.70	0.08	0.48	0.70	0.02	0.07
0.75	0.00	0.00	0.75	0.00	0.00	0.75	0.07	0.38	0.75	0.01	0.05
0.80	0.00	0.00	0.80	0.00	0.00	0.80	0.05	0.31	0.80	0.00	0.00
0.85	0.00	0.00	0.85	0.00	0.00	0.85	0.01	0.06	0.85	0.00	0.00
0.90	0.00	0.00	0.90	0.00	0.00	0.90	0.00	0.01	0.90	0.00	0.00
0.95	0.00	0.00	0.95	0.00	0.00	0.95	0.00	0.00	0.95	0.00	0.00

Table (4.49) True and False Positive Rates (Testing Data)

Car			Public Bus			Tram			Metro		
Cutoff	FPR	TPR	Cutoff	FPR	TPR	Cutoff	FPR	TPR	Cutoff	FPR	TPR
0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
0.02	1.00	1.00	0.05	0.52	1.00	0.05	1.00	1.00	0.05	0.69	0.96
0.04	0.99	1.00	0.10	0.26	0.93	0.10	0.97	1.00	0.10	0.46	0.95
0.06	0.93	0.98	0.15	0.13	0.67	0.15	0.82	0.97	0.15	0.35	0.95
0.08	0.83	0.94	0.20	0.12	0.67	0.20	0.60	0.96	0.20	0.26	0.91
0.10	0.62	0.86	0.25	0.10	0.63	0.25	0.37	0.94	0.25	0.23	0.90
0.12	0.52	0.75	0.30	0.09	0.47	0.30	0.36	0.93	0.30	0.21	0.88
0.14	0.46	0.61	0.35	0.04	0.03	0.35	0.33	0.93	0.35	0.14	0.67
0.16	0.34	0.25	0.40	0.01	0.00	0.40	0.31	0.91	0.40	0.14	0.64
0.18	0.24	0.18	0.45	0.00	0.00	0.45	0.30	0.90	0.45	0.13	0.58
0.20	0.15	0.12	0.50	0.00	0.00	0.50	0.29	0.88	0.50	0.11	0.36
0.22	0.10	0.10	0.55	0.00	0.00	0.55	0.27	0.87	0.55	0.08	0.26
0.24	0.07	0.06	0.60	0.00	0.00	0.60	0.23	0.82	0.60	0.08	0.24
0.26	0.05	0.06	0.65	0.00	0.00	0.65	0.21	0.78	0.65	0.01	0.08
0.28	0.03	0.06	0.70	0.00	0.00	0.70	0.17	0.72	0.70	0.00	0.03
0.30	0.03	0.06	0.75	0.00	0.00	0.75	0.13	0.57	0.75	0.00	0.02
0.32	0.02	0.02	0.80	0.00	0.00	0.80	0.09	0.36	0.80	0.00	0.00
0.34	0.02	0.02	0.85	0.00	0.00	0.85	0.02	0.12	0.85	0.00	0.00
0.36	0.01	0.00	0.90	0.00	0.00	0.90	0.00	0.04	0.90	0.00	0.00
0.38	0.00	0.00	0.95	0.00	0.00	0.95	0.00	0.00	0.95	0.00	0.00

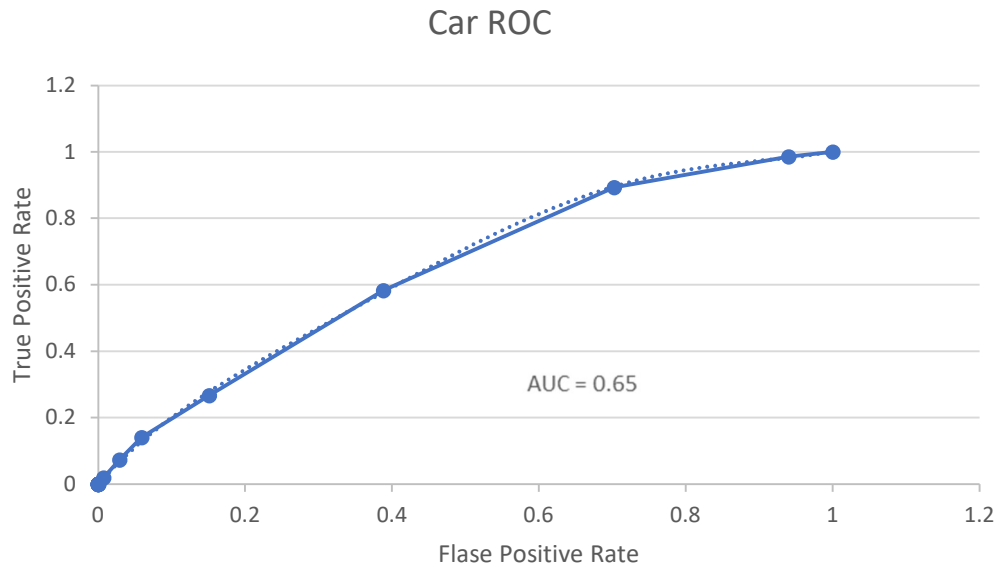


Figure 4.28) Car ROC (Training Data) (Phase 2)

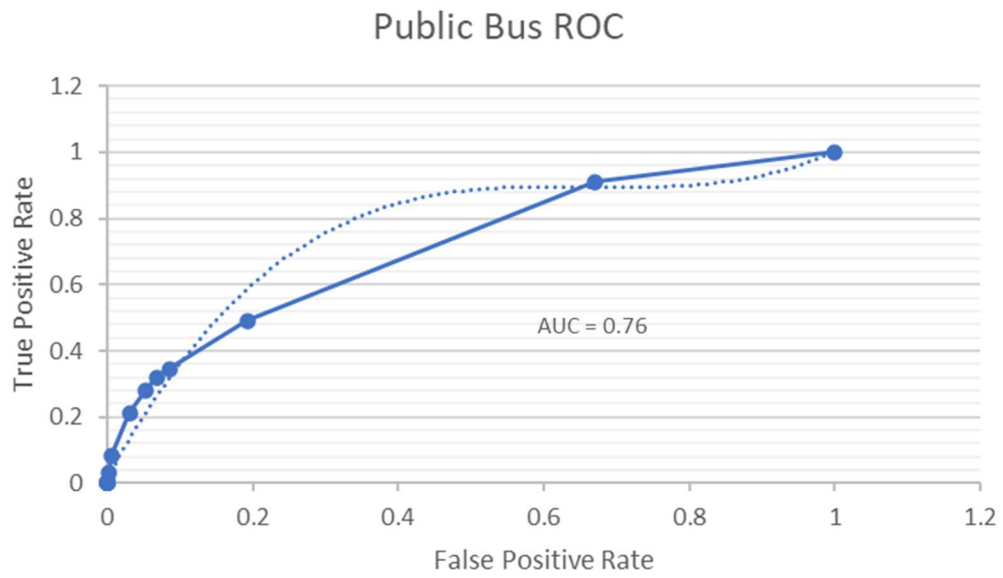


Figure 4.29) Public Bus ROC (Training Data) (Phase 2)

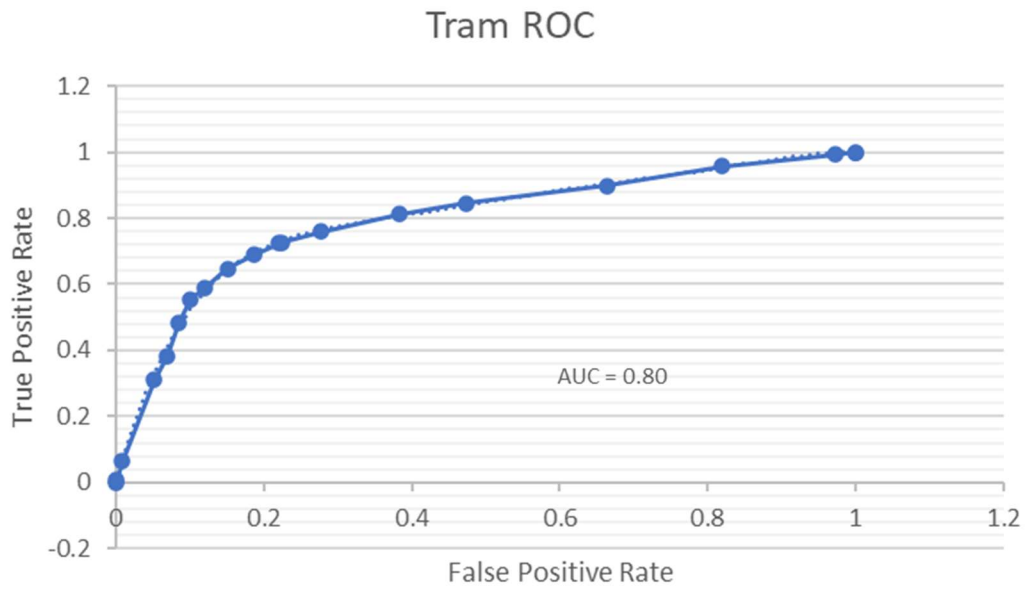


Figure 4.30) Tram ROC (Training Data) (Phase 2)

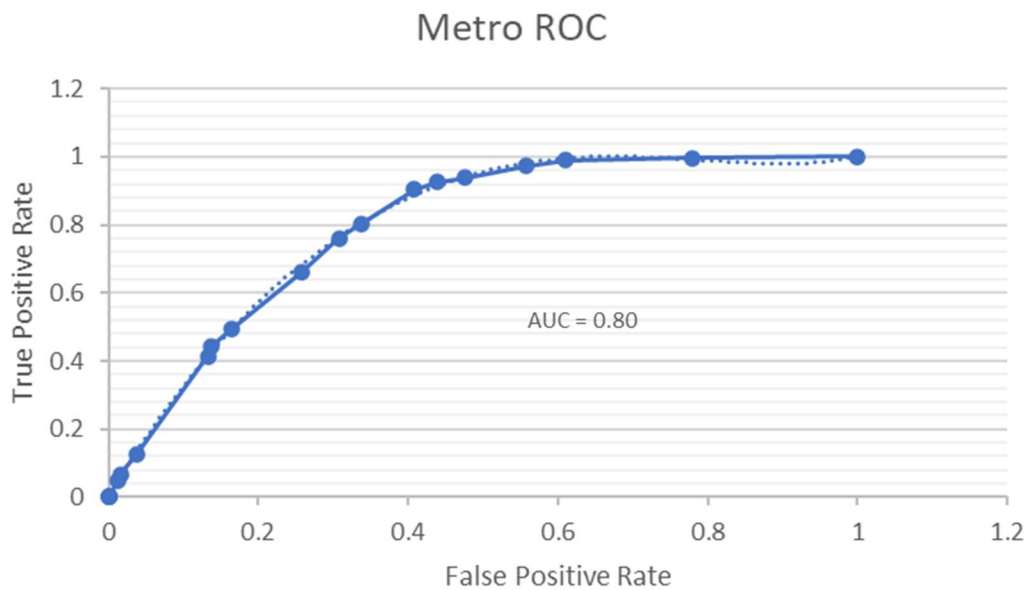
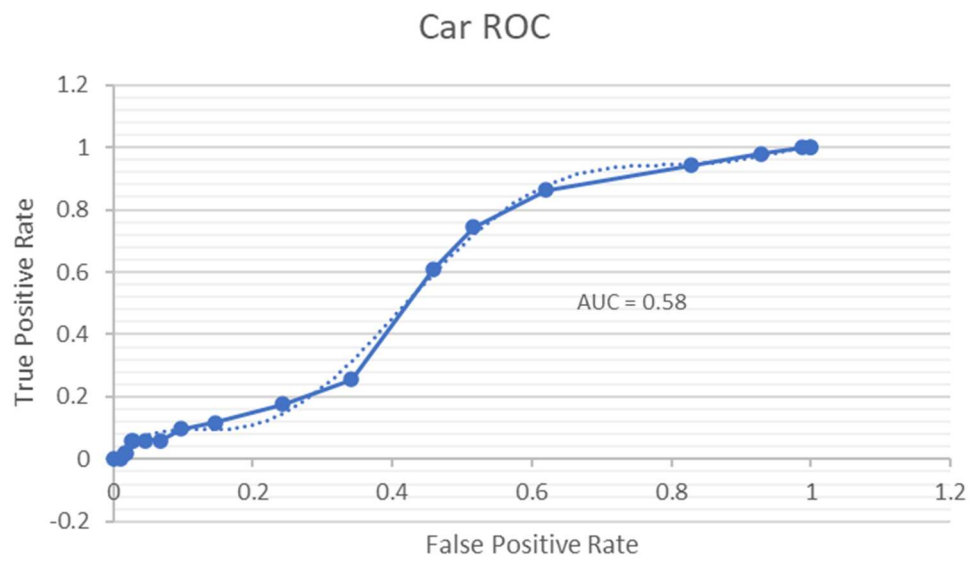


Figure (4.31) Metro ROC (Training Data) (Phase 2)



Figure(4.32) Car ROC (Testing Data) (Phase 2)

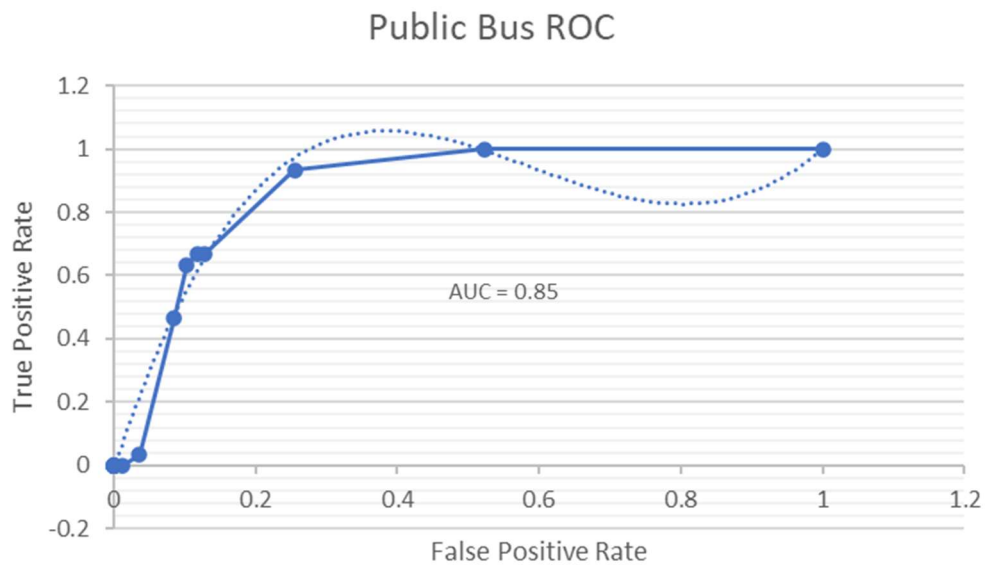


Figure (4.33) Public Bus ROC (Testing Data) (Phase 2)

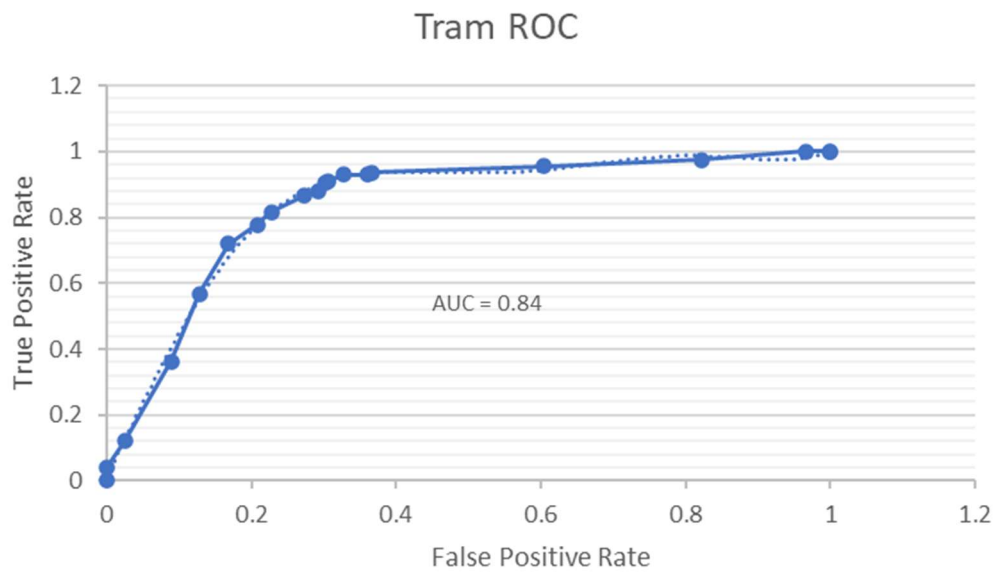


Figure (4.34) Tram ROC (Testing Data) (Phase 2)

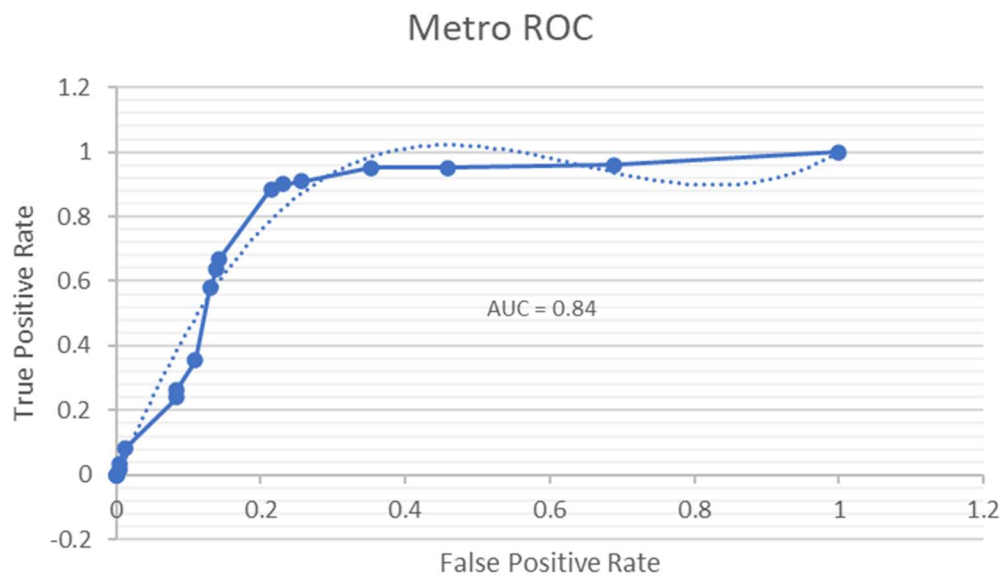


Figure (4.35) Metro ROC (Testing Data) (Phase 2)

4.3.2.4.2 Classification Tree

Upon inspection of the results shown in table 4.50 and table 4.33, it can be deduced that the classification tree generated in phase 2 did not perform as good as the one generated in phase 1. This change in performance is possibly due to the larger sample size used for phase 2 and the inclusion of more transport modes.

Table (4.50) Misclassification Error for the Classification Tree Model Phase 2

		Training			Test		
Actual Class	Count	Misclassified	% Error	Cost	Misclassified	% Error	Cost
Car	1357	549	40.5	0.40	638	47	0.47
Metro	1147	475	41.4	0.41	481	41	0.41
Public Bus	947	288	30.4	0.30	275	29	0.29
Tram	905	252	27.8	0.27	259	28	0.28
All	4356	1564	35.9	0.35	1653	37	0.36

The overall misclassification error for the training and testing data were noted to be higher than expected, being 35.9% and 37.9% respectively. It is interesting to mention that, by studying table 4.50, 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 (figures 4.36 – 4.39).

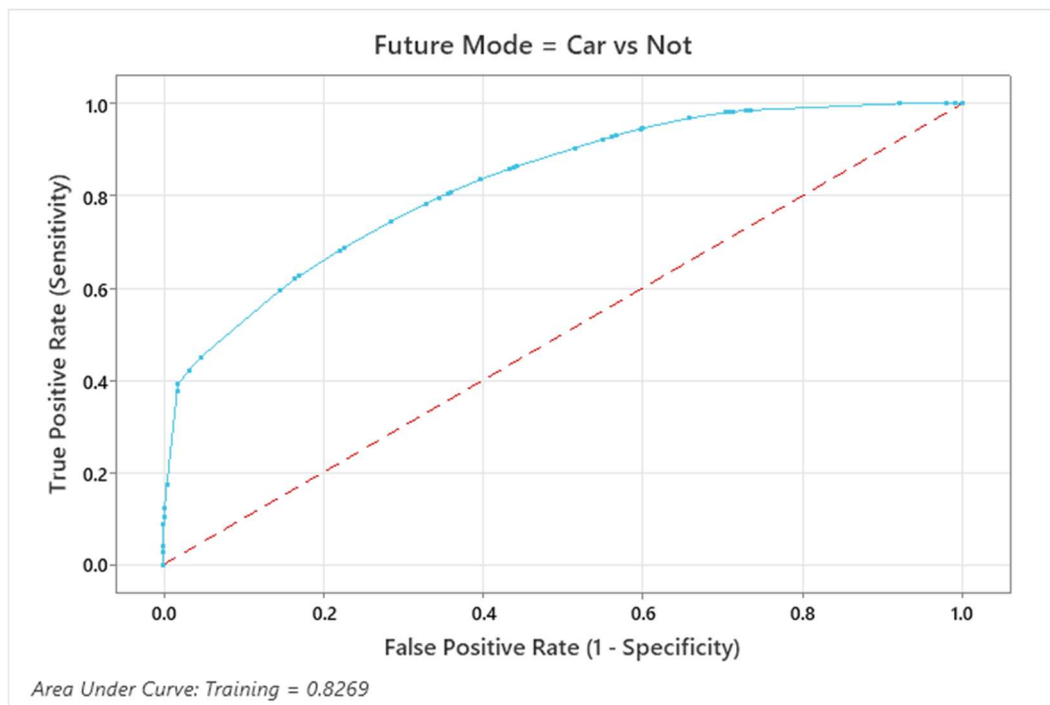


Figure 4.36) Car Classification Tree ROC. (Phase 2)

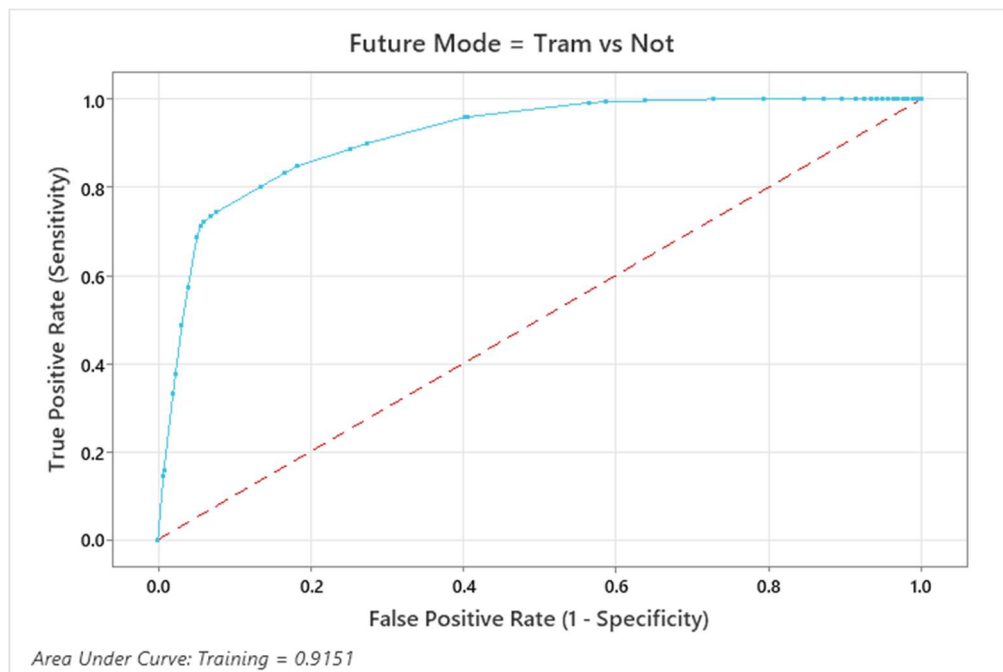


Figure (4.37) Tram Classification Tree ROC. (Phase 2)

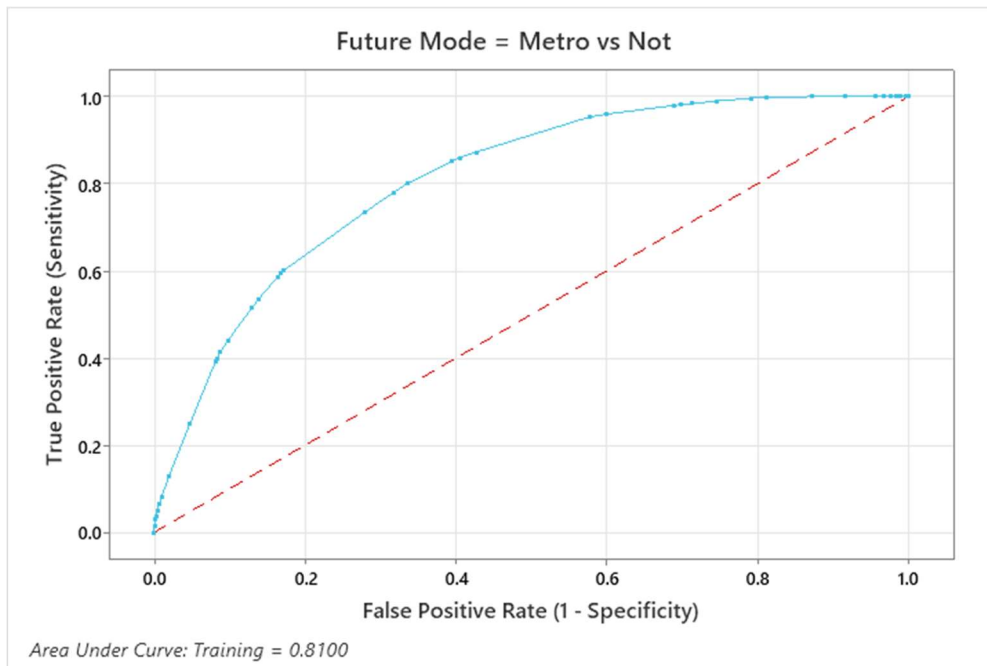


Figure (4.38) Metro Classification Tree ROC

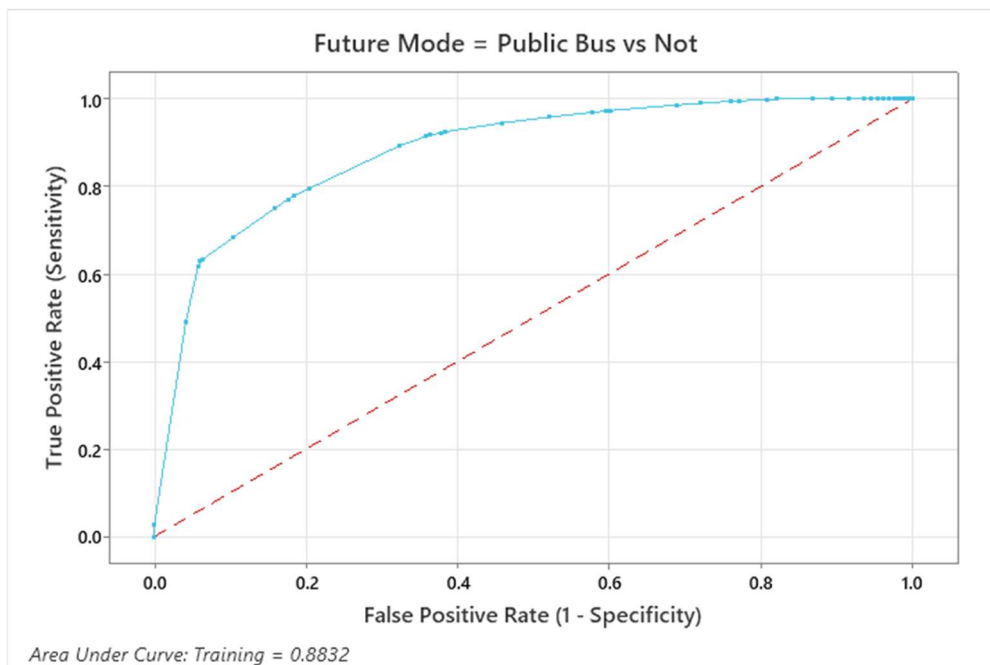


Figure (4.39) Public Bus Classification Tree ROC.

4.4 Validation Survey

In order to evaluate the competency of the models developed and their accuracy in forecasting current and future public transport mode choice in Bahrain, a validation survey was conducted. The survey was administered 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. The accuracy of the models in predicting mode choice was computed using this data. This section will outline the survey structure and characteristics, data collection method and major statistical metrics used to evaluate the accuracy of the models.

4.4.1 Survey Structure, Characteristics and Collection Method

The validation survey (appendix (A)) adapted a similar structure of the questionnaire described in chapter 3. 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 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. 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 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.

- Tram: fee is 0.3 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates. Payment is 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.
- Public Bus: fee is 0.3 BD per trip. Weekly, monthly, and yearly passes are also available at discounted rates. Payment is done using 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 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.
- 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 smart card system. The bus routes are designed to connect neighborhoods 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 smart card system. The trams run on dedicated tracks and connect the downtown areas with surrounding urban neighborhoods. 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.

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 as a result 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 (Table 4.51). Consequently, this led to the decision to include the scenarios as an independent section. This will allow 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.

Table 4.51) 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 (figure 4.40). As shown in Figure 4.41, male respondents accounted for 61% of the sample while females made up the remaining 39%. Besides, the majority of 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 (figure 4.42).

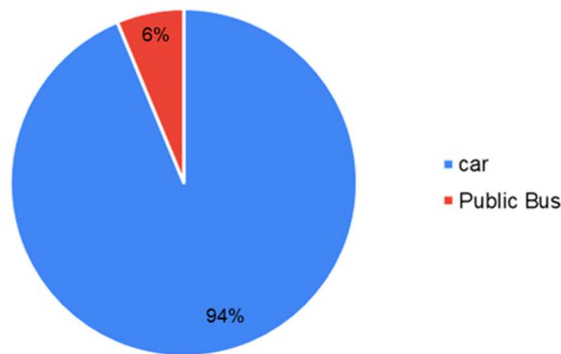


Figure 4.40) Validation Survey Current Modes.

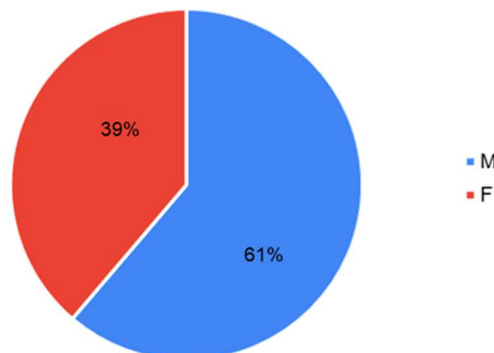


Figure 4.41) Validation Survey Gender.

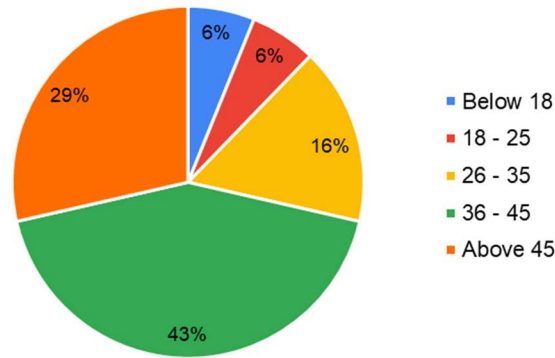


Figure (4.42) Validation Survey Age.

4.4.2 Models Evaluation:

Inputting the data collected from the validation survey into the models, demonstrated that the models performed differently for the two phases. For phase 1 models, while both models performed well, the logit model surpassed the classification tree with an accuracy of 92% compared to an accuracy of 81% for the latter. In the contrast, when considering phase 2 models, the classification tree outperformed the multinomial logit model with accuracy rates of 67% and 55% respectively. It is worth noting that for phase 2, the models were tested to accurately predict at least one of the modes the respondents are willing to use in the future. Besides, the fall noticed in accuracy when comparing phase 2 to phase 1 can possibly be a result of involving more modes and more complicated patterns and interrelations within its models. Tables 4.52 – 4.55 summarizes the findings of the analysis. See appendix (B) for full analysis tables.

Table (4.52) Phase 1 Logit Model Confusion Matrix (Validation Survey)

Actual/Predict	Car	Public Bus
Car	41	4
Public Bus	0	3
Total Positive	44	
Accuracy	92%	

Table (4.53) Phase 1 Classification Tree Model Confusion Matrix (Validation Survey)

Actual/Predict	Car	Public Bus
Car	37	8
Public Bus	1	2
Total Positive	39	
Accuracy	81%	

Table 4.54) Phase 2 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%			

Table 4.55) Phase 2 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%			

4.4.3 Future Scenarios Preferences

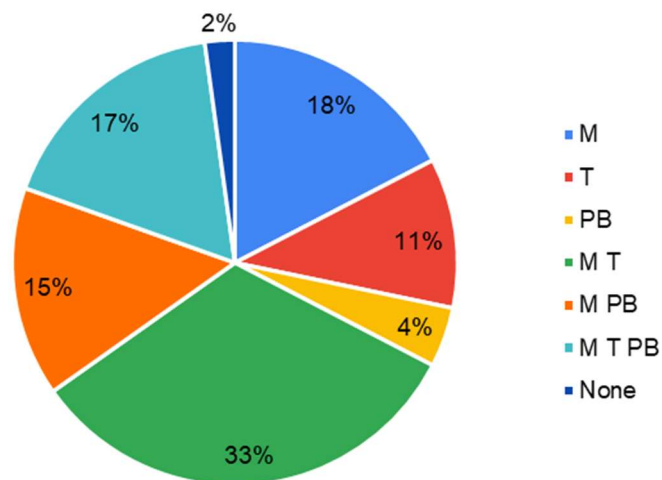


Figure (4.43) Hypothetical Scenarios Results Percentages

Upon Inspecting the percentages shown in Figure 4.43, 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% favored a combination of metro and public bus. However, only 11% and 4% of participant chose 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.

4.5 Chapter Summary

This chapter outlined the methodology utilized to model mode choice in Bahrain. It provided a detailed account of the major statistical characteristics of the dataset's items and the process followed to build the models, which was divided into two phases. Phase 1 focused on modeling the current mode choice while phase 2 aimed to forecast the future public transport mode choice. Both phases included the development of logit and classification tree models and utilized split-

half reliability to test the internal consistency of the data used to build them. In addition, cross-validation, AUC ROC method and a validation survey were used to validate both the process of model building and the performance of the models generated. Based on the results obtained, the logit model was found to be more accurate than the classification tree in phase 1 but less accurate in phase 2. Overall, the chapter mapped in depth the methods used to achieve the research's objectives.

Chapter Five: Results and Discussion

5.1 Introduction

5.2 Logit Model Odds Ratios Discussion

5.3 Classification Tree Results Discussion

5.4 Logit Model Vs. Classification Tree Model

5.5 Chapter Summary

5.1 Introduction

To comprehend mode choice behavior in Bahrain, Logit and Classification Tree models were developed, as detailed in the preceding chapter. These models aimed to predict the current mode choice as well as forecast the future public transport mode choice. This chapter intends to provide an in-depth discussion of the models' odds ratios emphasizing the variables that have the greatest influence, identify the characteristics of commuters that govern their choice behavior and compare the performance of the two types of models.

Fundamentally, this chapter is believed to provide valuable insights for mode choice in Bahrain and serve as a robust foundation for future research.

5.2 Logit Models

5.2.1 Phase 1: Binary Logit Model

Upon reviewing the odds ratios demonstrated in tables 4.22 and 4.23, the following summarizes the effects of both continuous and categorical variables on the present transportation decisions :

- Continuous Variables:
 - For each one-unit increase in travel time, the odds for choosing car mode over bus mode decreases by 1.26%.
 - A one-unit increase in cost leads to 155.69% increase in the odds of choosing car over bus.
 - For each one-unit increase in Salary, the odds for choosing car mode over bus mode increases by a percentage of 0.11%.

- Categorical Factors:
 - Trips related to education, shopping and Other-purposes lead to 6.0924, 17.7474 and 51.4671 times higher odds of choosing car over bus compared to work-related trips, respectively.
 - Shopping and Other-purposes trips result in 2.9131 and 8.4478 times higher odds of choosing car over bus than education- related trips.
 - The odds of choosing car over bus for Other-purposes trips are 2.9 times higher than shopping-related trips.
 - Employees are more likely to choose a car over bus compared to students and Others (unemployed, retired and other).
 - Students are more likely to choose bus over car than Others.

According to these observations, it can be concluded that trip cost is the most significant factor influencing mode choice behavior of commuters in Bahrain. In fact, for each one-unit increase in cost, the chances of choosing to travel by car over bus services increases by approximately 156%. Similar findings have been noted by Kumar et al. (Kumar, 2004) in a study that focused on intercity bus services in India where travel expenses were concluded to have a negative impact on bus travelers' utility. Additionally, students have a higher likelihood to travel by bus compared to other occupations and interestingly, trips that are produced for purposes other than work, education and shopping are 51 times more likely to be done by car than work trips.

5.2.2 Phase 2: Multinomial Logit

Based on the odds ratios shown in Table 4.37, the following statements can be concluded regarding the variables affecting forecasting public transport mode choice (refer to table 5.1 for findings summary):

- Tram / Car:
 - The odds of a male choosing Tram over Car is 18% lower than female.
 - Each one-unit increase in age leads to a decrease in the odds of choosing Tram over Car by 4%.
 - The odds of choosing Tram over Car increases by 2% with every unit increase in Travel Time.
 - The chances of a person with a driving license to choose Tram over Car is 34% lower than someone not having a driving license.
 - The likelihood of a person owning 1 car, 2 cars, 3 cars and 3+ cars choosing Tram over Car is -18%, 107%, 103% and 109% times the likelihood of someone who doesn't own a car, respectively.

Accordingly, it can be concluded that 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 license and car ownership, table 4.36 states that these variables are considered statistically insignificant for tram.

- Public Bus / Car:
 - The odds of a male choosing Public Bus over Car is 5% higher than female.
 - For every unit increase in age results in an increase of the odds of choosing Public Bus over Car by 1%.
 - The odds of choosing Public Bus over Car increases by 2% with every unit increase in Travel Time.
 - A person with a driving license is 58% less likely choosing Public Bus over Car than someone not having a driving license.
 - The likelihood of a person owning 1 car, 2 cars, 3 cars and 3+ cars choosing Public Bus over Car is 71%, 80%, 88% and 97% times less the likelihood of someone who does not own a car, respectively.

In light of these observations, it can be concluded that car ownership and possession of a driving license are the main factors that affect choosing 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 driving license are more likely to use it, particularly for longer trips. This trend is also captured by (Chee, 2013) 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 public bus over car, the p-values for these variables shown in table 4.36 suggest that they are statistically insignificant.

- Metro / Car:
 - The odds of a male choosing Metro over Car is 85% higher than a female.
 - The odds of choosing Metro over Car decreases by 1% with every unit increase in age.

- With every unit increase in travel time, the odds of choosing Metro over Car increase by 2%.
- The odds of a person with a driving license choosing Metro over Car are 68% higher than someone not having a driving license.
- The odds of a person owning 1 car, 2 cars, 3 cars and 3+ cars choosing Metro over Car is 62%, 84%, 99% and 95% times less the odds of someone who doesn't own a car.

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

Table 5.1) Multinomial Logit Model Findings Summary

Mode	Gender (relative to F)	Age	Travel Time	Driving License (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

5.3 Classification Tree Models

5.3.1 Phase 1

The node rules extracted from the binary classification tree have provided beneficial insights into the attributes of commuters in Bahrain and their trips that can be used to improve the effectiveness of the current transportation policies and systems. These rules were classified for each mode based on trip cost as shown in Table 5.2 and 5.3.

Upon reviewing the rules, the following was noticed:

- Education and work are the primary purposes for car trips which are mostly undertaken by commuters who own 1, 2 or 3 cars.
- The majority of car trips originate from the Capital governorate.
- Bus is mainly used for educational purposes.
- Surprisingly, the majority of bus users have driving license. This can be attributed to several reasons:
 - Having a driving license provides individuals with flexibility to choose between different modes of transportation. Some may use bus to avoid parking troubles or high fuel costs but possess a driving license as an alternative option.
 - Young individuals who recently obtained their driving license may still prefer to travel by bus, especially if they have limited financial resources.
 - They may possess a driving license as a job requirement but cannot financially afford a car for their personal use.

Table 5.2) Phase 1 Classification Tree Car trips and users' Characteristics

Trip Cost	Details
≤ 0.175	-
$0.175 < TC \leq 0.325$	Purpose of trip = {Other, Shopping/Leisure} and Occupation = {Unemployed}
	Do you have a driving licence= {Yes}, Purpose of trip= {Education}, Age = {Below 18 - 35}, How many cars do you own = {1}, O = {Capital}
	Gender = {M}, Do you have a driving licence = {Yes}, Purpose of trip = {Education}, Age = {36 - 45}, O = {Capital}
	Do you have a driving licence = {Yes}, Purpose of trip = {Education}, Age = {Above 45}, D = {Muharraq, Southern}, O = {Capital}
	Gender = {F}, Do you have a driving licence = {Yes}, Purpose of trip = {Education}, Age = {36 - 45}, D = {Southern}, Average Salary = {1000 - Above 2000}, O = {Capital}
	Gender = {F}, Do you have a driving licence= {Yes}, Purpose of trip= {Education}, Occupation= {Employee}, Age = {36 - 45}, D = {Capital, Central, Muharraq, Northern}, Average Salary= {1000 - Above 2000}, O = {Capital}
> 0.325	Do you have a driving licence = {No}, How many cars do you own= { 1, 2, 3 }
$0.325 < TC \leq 0.48335$	Do you have a driving licence = {Yes}, Purpose of trip= {Education, Work}, How many cars do you own_1 = {1, 2, 3}
$0.48335 < TC \leq 0.8833$	Do you have a driving licence= {Yes}, Purpose of trip= {Education, Work}, D = {Muharraq, Northern, Southern}, How many cars do you own= {1, 2, 3}, Average Salary= {Below 100 - 1000, 1500 - 2000}, O = {Muharraq, Southern}, Travel time= {0 - 20, 120 +, 31 - 120}
	Do you have a driving licence={Yes}, Purpose of trip = {Education, Work}, How many cars do you own = {1, 2, 3}, Average Salary = {1000 - 1500 BD, Above 2000}, Travel time = {0 - 20, 120 +, 31 - 120}
	Do you have a driving licence = {Yes}, Purpose of trip= {Education, Work}, Occupation = {Student, Unemployed}, D = {Capital, Central, Muharraq}, How many cars do you own = {1, 2, 3}, Travel time = {21 - 30}
	Do you have a driving licence={Yes}, Purpose of trip = {Education, Work}, D = {Northern, Southern}, How many cars do you own= {1, 2, 3}, O = {Muharraq, Northern, Southern}, Travel time = {21 - 30}
> 0.91665	D = {Northern, Southern}, How many cars do you own = {0}
	Age = {26 - Above 45}, Destination = {Capital, Central, Muharraq}, How many cars do you own= {0}

0.325 < TC ≤ 3.875	Do you have a driving licence= {Yes}, Purpose of trip= {Other, Shopping/Leisure}, How many cars do you own= {1, 2, 3}
0.8833 < TC ≤ 3.875	Do you have a driving licence= {Yes}, Purpose of trip = {Education, Work}, D = {Muharraq, Northern}, Average Salary= {100 - 500 BD}
	Do you have a driving licence= {Yes}, Purpose of trip= {Education, Work}, Average Salary = {Below 100, 500 - Above 2000}
	Do you have a driving licence= {Yes}, Purpose of trip= {Education, Work}, How many cars do you own = {1, 2, 3}
> 3.875	Do you have a driving licence= {Yes}, How many cars do you own_1 = {1, 2, 3}

Table 5.3) Phase 1 Classification Tree Bus trips and users' Characteristics

Trip Cost	Details
0.175 < TC ≤ 0.2875	Purpose of trip = {Education}
0.175 < TC ≤ 0.325	Purpose of trip = {Other, Shopping/Leisure, Work}, Occupation= {Employee, Other, Student}
0.2875 < TC ≤ 0.325	Do you have a driving licence= {No}, Purpose of trip= {Education}
	Do you have a driving licence = {Yes}, Purpose of trip = {Education}, O = {Central, Kuwait, Muharraq, Northern, Saudi Arabia, Southern}
	Do you have a driving licence= {Yes}, Purpose of trip= {Education}, Age = {Below 18 -35}, How many cars do you own= {0, 2, 3}, O = {Capital}
	Gender = {F}, Do you have a driving licence = {Yes}, Purpose of trip= {Education}, Age = {36 - 45}, Average Salary = {Below 100 - 1000}, O = {Capital}
	Gender = {F}, Do you have a driving licence= {Yes}, Purpose of trip= {Education}, Occupation = {Other, Student, Unemployed}, Age = {36 - 45}, D = {Capital, Central, Muharraq, Northern}, Average Salary= {1000 - Above 2000}, O = {Capital}
	Do you have a driving licence= {Yes}, Purpose of trip= {Education}, Age = {Above 45}, D = {Capital, Central, Northern}, O = {Capital}
0.48335 < TC ≤ 0.8833	Do you have a driving licence = {Yes}, Purpose of trip = {Education, Work}, How many cars do you own = {1, 2, 3}, Average Salary = {Below 100 - 1000, 1500 - 2000}, O = {Capital, Central, Kuwait, Northern, Saudi Arabia}, Travel time (min) = {0 - 20, 120 +, 31 - 120}

	Do you have a driving licence= {Yes}, Purpose of trip= {Education, Work}, D = {Capital, Central}, How many cars do you own = {1, 2, 3}, Average Salary = {Below 100 - 1000 , 1500 - 2000}, O = {Muharraq, Southern}, Travel time (min)= {0 - 20, 120 +, 31 - 120}
	Do you have a driving licence= {Yes}, Purpose of trip= {Education, Work}, Occupation = {Employee, Other}, D = {Capital, Central, Muharraq}, How many cars do you own = {1, 2, 3}, Travel time (min)= {21 - 30}
	Do you have a driving licence= {Yes}, Purpose of trip = {Education, Work}, D = {Northern, Southern}, How many cars do you own= {1, 2, 3}, O = {Capital, Central, Kuwait, Saudi Arabia}, Travel time = {21 - 30}
$0.325 < TC \leq 0.91665$	How many cars do you own = {0}
$0.8833 < TC \leq 3.875$	Do you have a driving licence= {Yes}, Purpose of trip= {Education, Work}, D = {Capital, Central, Southern}, Average Salary= {100 - 500}
> 0.91665	Age = {Below 18 - 25}, D = {Capital, Central, Muharraq}, How many cars do you own= {0}

5.3.2 Phase 2

Similarly, 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. Tables 5.4 – 5.7 demonstrate this information for each mode.

Studying the findings, it was observed that:

- Commuters, who declined the use of public transportation in the future, mainly own 1 car and travel for work purposes.
- 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 (Anwar, 2023), aimed at exploring the factors influencing car users to switch to 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.

- Public bus is more probable to be used nearly equally for all purposes.
- 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. 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.

Table 5.4) Phase 2 Classification Tree Car trips and users' Characteristics

Trip Cost	Details
≤ 0.05	Purpose of trip = {Education, Other, Work}
$0.633333 < TC \leq 3.15$	Purpose of trip = {Work}, Average Salary = {100 - 500}, O = {Capital, Kuwait, Muharraq, Northern, Saudi Arabia, Southern}
	Purpose of trip = {Work}, O = {Central}
> 3.15	Purpose of trip = {Education, Other}, How many cars do you own = {1, 3}
$3.15 < TC \leq 6.375$	Purpose of trip = {Work}
> 6.375	Purpose of trip = {Work}, Average Salary = {Below 100}
	Purpose of trip = {Work}, Average Salary = {1000 - 1500, Above 2000}
	Do you have a driving licence = {Yes}, Purpose of trip = {Work}, Average Salary = {100 - 500}
	Purpose of trip = {Work}, Average Salary = {1500 - 2000}
$6.75 < TC \leq 8.5$	Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Capital, Central, Muharraq, Northern}, Average Salary = {500 - 1000}
> 8.5	Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Capital, Central, Muharraq, Northern}, Average Salary = {500 - 1000}, Travel time (min) = {0 - 30, 51 - 60}

Table (5.5) Phase 2 Classification Tree Metro trips and users' Characteristics

Trip Cost	Details
$0.325 < TC \leq 0.525$	Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2}, Age = {36 - Above 45, Below 18}, D = {Capital}, Average Salary = {100 - 500, 1000 - 1500}
> 0.4	Purpose of trip = {Shopping/Leisure}, D = {Central, Muharraq, Northern, Southern}
≤ 0.625	Purpose of trip = {Shopping/Leisure}, OC = {Employee, Unemployed}, How many cars do you own = {0, 1}, D = {Capital}, Average Salary = {100 - 500, 1000 - 1500}
	Purpose of trip = {Shopping/Leisure}, OC = {Other, Student}, How many cars do you own = {0}, D = {Capital}, Average Salary = {100 - 500, 1000 - 1500}
> 0.625	Purpose of trip = {Shopping/Leisure}, How many cars do you own = {0, 1}, D = {Capital}, Average Salary = {100 - 500, 1000 - 1500}
$0.633333 < TC \leq 3.15$	Purpose of trip = {Work}, Average Salary = {500 - Above 2000, Below 100 BD}, O = {Capital, Kuwait, Muharraq, Northern, Saudi Arabia, Southern}
	Purpose of trip = {Education, Other}
> 3.15	Purpose of trip = {Education, Other}, How many cars do you own = {0, 2}, Age = {18 - 25}
> 6.375	Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {2, 3}, Average Salary = {500 - 1000}
	Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Southern}, Average Salary = {500 - 1000}
	Do you have a driving licence = {No}, Purpose of trip = {Work}, Average Salary = {100 - 1000}
	Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {0}, Average Salary = {500 - 1000}
$6.375 < TC \leq 6.75$	Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Capital, Central, Muharraq, Northern}, Average Salary = {500 - 1000}
-	Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student, Unemployed}, D = {Capital}, Average Salary = {1500 - 2000, Above 2000, Below 100 BD}
-	Purpose of trip = {Shopping/Leisure}, OC = {Employee}, D = {Capital}, Average Salary = {500 - 1000 BD}

Table (5.6) Phase 2 Classification Tree Public Bus trips and users' Characteristics

Trip Cost	Details
0.05 < TC ≤ 0.633333	Purpose of trip = {Education, Other, Work}, Average Salary = {Below 100 - 1000, 1500 - 2000}
0.225 < TC ≤ 0.325	Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2, 3}, D = {Capital}, Average Salary = {100 - 500, 1000 - 1500}
≤ 0.4	Purpose of trip = {Shopping/Leisure}, D = {Central, Muharraq, Northern, Southern}
> 3.15	Purpose of trip = {Education, Other}, How many cars do you own = {0, 2}, Age = {26 - Above 45, Below 18}
> 8.5	Do you have a driving licence = {Yes}, Purpose of trip = {Work}, How many cars do you own = {1}, D = {Capital, Central, Muharraq, Northern}, Average Salary = {500 - 1000}, Travel time (min) = {31 - 50, 61 - 120+}

Table (5.7) Phase 2 Classification Tree Tram trips and users' Characteristics

Trip Cost	Details
0.05 < Total cost of trip ≤ 0.633333	Purpose of trip = {Education, Other, Work}, Average Salary = {1000 - 1500, Above 2000}
≤ 0.225	Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2, 3}, D = {Capital}, Average Salary = {100 - 500, 1000 - 1500}
0.325 < Total cost of trip ≤ 0.525	Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {3}, Age = {36 - Above 45, Below 18}, D = {Capital}, Average Salary = {100 - 500, 1000 - 1500}
	Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2, 3}, Age = {18 - 35}, D = {Capital}, Average Salary = {100 - 500, 1000 - 1500}
> 0.525	Purpose of trip = {Shopping/Leisure}, OC = {Employee, Student}, How many cars do you own = {2, 3}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}
≤ 0.625	Purpose of trip = {Shopping/Leisure}, OC = {Other, Student}, How many cars do you own = {1}, D = {Capital}, Average Salary = {100 - 500, 1000 - 1500}
-	Purpose of trip = {Shopping/Leisure}, OC = {Other}, D = {Capital}, Average Salary = {1500 - 2000, Above 2000, Below 100}
-	Purpose of trip = {Shopping/Leisure}, OC = {Other, Student, Unemployed}, D = {Capital}, Average Salary = {500 - 1000}
-	Purpose of trip = {Shopping/Leisure}, OC = {Other, Unemployed}, How many cars do you own = {2, 3}, D = {Capital}, Average Salary = {100 - 500 BD, 1000 - 1500 BD}

5.4 Logit Model Vs. Classification Tree Model

The performance of logit model and classification tree model in modeling and forecasting mode choice in Bahrain was measured in two phases. In the first phase, both models suggested that trip cost is the top variable affecting mode choice and that bus services are mainly used for education-related trips. The AUC computed for the logit model was 0.8837, while the classification tree achieved a greater AUC value of 0.9236. Nevertheless, using the data obtained from the validation survey, the logit model outperformed the classification tree model. In comparison, the logit model successfully identified bus users with a small error for car users and attained a remarkable 92% accuracy, whereas the classification tree had slightly more error in identifying both modes and achieved an accuracy of 81% (Table 5.8).

Table 5.8) Logit Model Vs. Classification Tree Model

Phase	Model	AUC	Initial validation	Validation-survey Accuracy
Phase 1	Logit Model	0.8927	10-fold AUC = 0.8837	92%
	Classification Tree	0.9236	Accuracy = 86%	81%
Phase 2	Logit Model	Car = 0.6500 Public Bus = 0.7600 Tram = 0.8000 Metro = 0.8000	AUC Values: Car = 0.5800 Public Bus = 0.8500 Tram = 0.8400 Metro = 0.8400	55%
	Classification Tree	Car = 0.8269 Public Bus = 0.8832 Tram = 0.9151 Metro = 0.8100	Accuracy = 62%	67%

As for the second phase, the logit model indicated that age, gender, car ownership, possession of a driving license and travel time were the key variables shaping commuters' future mode choice preferences. Conversely, since the classification tree model utilized considerably more data and more predictors, it implied that trip cost and trip purpose were the major variables governing the splits with a

noticeable contribution of car ownership. As summarized in Table 5.8, the classification tree model was observed to have greater AUC values and a higher validation-survey accuracy of 67% compared to the logit model that achieved a relatively low 55% accuracy.

It is noteworthy to mention that phase 2 models performed less effectively than the models generated in phase 1 due to the more complex patterns being modeled.

5.5 Chapter Summary

This chapter analyzed the results obtained from modeling current and future mode choice in Bahrain using logit and classification tree models. The odds ratios of the logit models and node rules of the classification trees were thoroughly explained to highlight the personal and trip-related characteristics governing commuters' choice behavior. Lastly, the performance of the models was evaluated through comparing their AUC values and validation accuracy rates.

Chapter Six: Conclusions and Recommendations

6.1 Introduction

6.2 Conclusion

6.3 Recommendations

6.1 Introduction

To conclude, this thesis has concentrated on putting forward the demanding need for thorough modeling of the current mode choice in Bahrain and 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' behavior 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. This chapter 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.

6.2 Conclusion

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:

- Trip production in the country is distributed in almost equal proportions among the four governorates.
- The Capital governorate attracts the majority of trips of various purposes.
- The majority of trips are primarily undertaken for work purposes, followed by shopping and education.

- Trip cost is identified to be the key predictor governing current mode choice and future mode choice in Bahrain.
- Commuters with high socio-economic status are more likely to travel by Car than use Bus services.
- Travel time and speed of travel play a considerable role in shaping mode choice decisions; trips taken by Bus account for longer travel time compared to Car trips.
- Bus services are recognized to be mainly used by students for education-related trips.
- 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.
- Tram is established as an immensely renowned mode of transportation for trips undertaken for shopping and leisure purposes.
- 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.
- The research indicates that direct and quick travel, accessibility and conveniences play a significant role in influencing the choice of transport mode.
- The accuracy of logit model in modeling binary responses has been proven to outperform that of classification tree, however, when it comes to capturing more complex relationships and multinomial responses, the classification tree model surpasses the multinomial logit model.

6.3 Recommendations

6.3.1 Implications of Research Findings

In reference to the conclusions obtained from the study, the following are valuable recommendations for transportation authorities in Bahrain:

- **Prioritize investment in the development of a Metro-Tram system:**
Considering that the Metro-Tram combination is identified as the top choice for future public transportation, the establishment of an efficient metro and tram system must be a top priority for transportation authorities. Besides, taking into account that tram is a popular choice for shopping and leisure trips, tram networks should be introduced in areas with heavy commercial and recreational traffic.
- **Improve the accessibility and affordability of the current bus services:**
Since buses are mainly used by students, investment should be considered in developing more frequent and convenient services to fulfill their needs. In addition, providing more bus stops, reducing trip costs and establishing inclusive premium services would help increase ridership from other groups.
- **Improve the transportation system infrastructure:**
Since walking and cycling were popular among respondents as feeder modes, provide better lighting, wider sidewalks and specialized cycling lanes to encourage the use of sustainable transportation modes, such as walking and cycling.
- **Introduce congestion charges:**
As trip cost is recognized as the top variable affecting mode choice in Bahrain, making private transportation less affordable and public transportation more affordable would help increase the ridership of the latter. This can be done by investigating the introduction of congestion

charges in highly congested highways and areas and having special measures during peak hours.

In general, the efficiency, affordability and speed of public transportation options play a prominent role in influencing mode choice decisions. Subsequently, giving more consideration and efforts to improve these aspects would lead to a more sustainable and efficient transport system in Bahrain.

6.3.2 Avenues for further Research

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

- Extend the analysis to incorporate more variables:
Researchers can investigate the influence of other factors, especially qualitative, on mode choice such as comfort, safety, and network characteristics.
- Study the effect of transport measures on mode choice:
Researchers can explore the impact of transport measures and policies, such as public transport subscriptions, road pricing, parking fees and congestion charges on mode choice.
- Examine the connection between sustainable transportation and other sustainable objectives:
Researchers can investigate the interactions between transportation and health, climate change and energy.

- Study the role of urban design in shaping mode choice:

Researchers can study how urban design, such as lighting, pedestrian and cycling friendly streets could affect the mode choice of commuters in Bahrain.

References

Nicholas J. Garber Lester A. Hoel. (2009) Traffic and Highway Engineering (fourth edition), University of Virginia.

Juan de Dios Ortúzar, Luis G. Willumsen. (2011) Modelling Transport (fourth edition), John Wiley & Sons, Ltd.

Mwale, M., Luke, R., & Pisa, N. (2022). Factors that affect travel behaviour in developing cities: A methodological review. *Transportation Research Interdisciplinary Perspectives*, 16, 100683.

Sowjanya, D., Tahlyan, D., & Sekhar, C. R. (2014). Travel Demand Modelling for a Metropolitan City. In *International Conference on Recent Trends and Challenges in Civil Engineering* (pp. 19-40).

Waghmare, A., Yadav, G., & Tiwari, K. (2022). Four Step Travel Demand Modeling for Urban Transportation Planning. *Sci. Eng. Technol.*, 5, 1254.

Pineda-Jaramillo, J. D. (2019). A review of Machine Learning (ML) algorithms used for modeling travel mode choice. *Dyna*, 86(211), 32-41.

S. Waleed, 2019, “The Ministry of Works is implementing a package of projects to reduce traffic congestion”. *Al Bilad News*, Kingdom of Bahrain.

Retrieved from

<http://www.albiladpress.com/news/2019/4016/bahrain/602413.html>. Accessed on 19th Oct 2019.

Modi, K. B., Zala, L. B., Umrigar, F. S., & Desai, T. A. (2011, May).

Transportation planning models: a review. In *National Conference on Recent Trends in Engineering and Technology*, Gujarat India.

L Ester A. Hoel, Nicholas J. Garber, Adel W. S Adek. (2011) Transportation Infrastructure Engineering A Multimodal Integration SI Edition.

Lekshmi, G. A., Landge, V. S., & Kumar, V. S. (2016). Activity based travel demand modeling of Thiruvananthapuram urban area. *Transportation Research Procedia*, 17, 498-505.

Etminani-Ghasrodashti, R., & Ardeshiri, M. (2016). The impacts of built environment on home-based work and non-work trips: An empirical study from Iran. *Transportation Research Part A: Policy and Practice*, 85, 196-207.

Dissanayake, D., & Morikawa, T. (2002). Household travel behavior in developing countries: Nested logit model of vehicle ownership, mode choice, and trip chaining. *Transportation Research Record*, 1805(1), 45-52.

Rocha, S. S., Pitombo, C. S., Costa, L. H. M., & de França Marques, S. (2021). Applying optimization algorithms for spatial estimation of travel demand variables. *Transportation Research Interdisciplinary Perspectives*, 10, 100369.

Mladenovic, M., & Trifunovic, A. (2014). The shortcomings of the conventional four step travel demand forecasting process. *Journal of Road and Traffic Engineering*, 60(1), 5-12.

Xintong, M. (2021). A critical overview of four-stage model under the background of the rise of ride-sharing. In *E3S Web of Conferences* (Vol. 253, p. 03020). EDP Sciences.

McNally, M. G. (2007). The four-step model. In *Handbook of transport modelling*. Emerald Group Publishing Limited.

Transport and Infrastructure Council (2016). *Travel Demand Modelling*. Australia

Tskeris, T., & Tsekeris, C. (2011). Demand forecasting in transport: Overview and modeling advances. *Economic research-Ekonomska istraživanja*, 24(1), 82-94.

Garber, N. J., & Hoel, L. A. (2018). *Traffic and highway engineering*. Cengage Learning.

Chen, X., Liu, X., & Li, F. (2013). Comparative study on mode split discrete choice models. *Journal of Modern Transportation*, 21(4), 266-272.

Ratrout, N. T., Gazder, U., & Al-Madani, H. M. (2014). A review of mode choice modelling techniques for intra-city and border transport. *World Review of Intermodal Transportation Research*, 5(1), 39-58.

Hillel, T., Bierlaire, M., & Jin, Y. (2019). A systematic review of machine learning methodologies for modelling passenger mode choice.

Sekhar, C. (2014). Mode Choice Analysis: The Data, the Models and Future Ahead. *International Journal for Traffic & Transport Engineering*, 4(3).

Puan, O. C., Hassan, Y. A. H., Mashros, N., Idham, M. K., Hassan, N. A., Warid, M. N. M., & Hainin, M. R. (2019, May). Transportation mode choice binary logit model: A case study for Johor Bahru city. In *IOP conference series: Materials science and engineering* (Vol. 527, No. 1, p. 012066). IOP Publishing.

Hussain, H. D., Mohammed, A. M., Salman, A. D., Rahmat, R. A. B. O. K., & Borhan, M. N. (2017). Analysis of transportation mode choice using a comparison of artificial neural network and multinomial logit models. *ARPN Journal of Engineering and Applied Sciences*, 12(5), 1483-1493.

Koppelman, F. S., & Bhat, C. (2006). *A self instructing course in mode choice modeling: multinomial and nested logit models*.

- Bunch, D. S. (1991). Estimability in the multinomial probit model. *Transportation Research Part B: Methodological*, 25(1), 1-12.
- Ghareib, A. H. (1996). Evaluation of logit and probit models in mode-choice situation. *Journal of transportation engineering*, 122(4), 282-290.
- Bhat, C. R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological*, 29(6), 471-483.
- Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). Machine learning in transportation data analytics. In *Data analytics for intelligent transportation systems* (pp. 283-307). Elsevier.
- Vythoulkas, P. C., & Koutsopoulos, H. N. (2003). Modeling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks. *Transportation Research Part C: Emerging Technologies*, 11(1), 51-73.
- Essam, A., & Sadi, A. (2013). Factors affecting mode choice of work trips in developing cities—Gaza as a case study. *Journal of Transportation Technologies*, 2013.
- Ratrout, N. T., & Gazder, U. (2014). Effects of potential modes on mode choice behaviour for travel between Saudi Arabia and Bahrain. *World Review of Intermodal Transportation Research*, 5(2), 125-148.
- Abdullah, M., Ali, N., Javid, M. A., Dias, C., & Campisi, T. (2021). Public transport versus solo travel mode choices during the COVID-19 pandemic: Self-reported evidence from a developing country. *Transportation engineering*, 5, 100078.

Samimi, A., Razi-Ardakani, H., & Mohammadian, K. A. (2012). A Decision Tree Approach to Analyze Freight Mode Choice Decisions. In Proceedings of the 1st European Symposium on Quantitative Methods in Transportation Systems.

Oral, L. O., & Tecim, V. A. H. A. P. (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.

Lindner, A., Pitombo, C. S., & Cunha, A. L. (2017). Estimating motorized travel mode choice using classifiers: An application for high-dimensional multicollinear data. *Travel Behaviour and Society*, 6, 100-109.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2013). Introduction to probability and statistics. Cengage Learning.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). Odds ratio. In *Applied regression analysis and other multivariable methods* (pp. 104-118). PWS-Kent.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear statistical models* (5th ed.). McGraw-Hill/Irwin.

Hadi, A. S., & Chatterjee, S. (2015). *Regression analysis by example*. John Wiley & Sons.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.
- Fawcett, T. (2006). Introduction to receiver operator curves. *Pattern Recognit. Lett*, 27, 861-874.
- Kumar, C. P., Basu, D., & Maitra, B. (2004). Modeling generalized cost of travel for rural bus users: a case study. *Journal of Public Transportation*, 7(2), 59-72.
- Chee, W. L., & Fernandez, J. L. (2013). Factors that influence the choice of mode of transport in Penang: A preliminary analysis. *Procedia-Social and Behavioral Sciences*, 91, 120-127.
- Anwar, A. M., Oakil, A. T., Muhsen, A., & Arora, A. (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.

Appendix A

Questionnaire Format

Forecasting Public Transport Mode Choice in Bahrain: Part 2

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 غازدير 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 *

- ☐ Male / ذكر
- ☐ Female / أنثى

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 / السيارة
- ☐ Sharing Car / مشاركة السيارة
- ☐ Public Bus / الحافلة العامة
- ☐ Walking / المشي
- ☐ Cycling / ركوب الدراجات
- ☐ Other: _____

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 متوفرة 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 للمشاة 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 50-60 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.

☐

السيارة / Car

☐

الحافلة / Bus

☐

المشي / Walking

☐

ركوب الدراجات / Cycling

☐

لا شيء مما ذكر / None

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 المتكررة ومغادرتها كل 5-10 دقائق. تتوفر مواقف للسيارات بالقرب من محطات المترو لمواقف السيارات لفترات قصيرة وطويلة المدى (0.2 BD للساعة، الساعة الأولى مجانية، 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 قصيرة وطويلة (0.2 BD للساعة، الساعة الأولى مجانية). هناك العديد من خطوط الترام في الدولة مع توقفات عديدة 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 أيضاً تذاكر أسبوعية وشهرية وسنوية بأسعار مخفضة لتوفير رحلات غير محدودة 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 وفقاً لجدول زمني منتظم مع وصول القطارات المتكررة 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 أيضاً تذاكر أسبوعية وشهرية وسنوية بأسعار مخفضة لتوفير رحلات غير محدودة 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 وفقاً لجدول زمني منتظم مع وصول القطارات المتكررة 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 المحدودة لجميع وسائل النقل العام. تتوفر أيضاً تذاكر أسبوعية وشهرية وسنوية بأسعار مخفضة لتوفير رحلات غير محدودة في جميع الأوضاع. يتم الدفع باستخدام نظام البطاقة الذكية. تعمل عربات الترام على مسارات مخصصة وترتبط 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 وفقاً لجدول زمني منتظم مع وصول القطارات والباصات المتكررة ومغادرتها كل 5-10 دقائق. تتوفر مواقف للسيارات بالقرب من محطات المترو والترام لمواقف السيارات لفترات قصيرة وطويلة.

- ☐ لا شيء مما ذكر / None

Appendix B

Validation Survey Analysis Tables

Logit Model Phase 1 Prediction and Probabilities

Current Mod ▾	Utility	Car %	Bus %	Mode Predicted
Car	1.01	73	27	Car
Car	4.86	99	1	Car
Car	0.63	65	35	Car
Car	0.38	59	41	Car
Car	5.71	100	0	Car
Car	1.82	86	14	Car
Car	3.83	98	2	Car
Car	5.98	100	0	Car
Car	1.79	86	14	Car
Car	1.35	79	21	Car
Car	-0.03	49	51	Bus
Car	2.50	92	8	Car
Car	4.47	99	1	Car
Car	6.17	100	0	Car
Car	1.64	84	16	Car
Car	2.29	91	9	Car
Car	4.86	99	1	Car
Car	2.77	94	6	Car
Car	1.79	86	14	Car
Car	3.83	98	2	Car
Car	1.10	75	25	Car
Car	3.11	96	4	Car
Car	5.42	100	0	Car
Car	0.79	69	31	Car
Car	5.98	100	0	Car
Car	1.10	75	25	Car
Car	1.54	82	18	Car
Car	1.77	85	15	Car
Car	-1.87	13	87	Bus
Car	1.51	82	18	Car
Car	-0.14	46	54	Bus
Public Bus	-0.21	45	55	Bus
Public Bus	-0.21	45	55	Bus
Public Bus	-0.97	27	73	Bus
Car	11.58	100	0	Car
Car	1.10	75	25	Car
Car	0.69	67	33	Car
Car	1.77	85	15	Car
Car	4.39	99	1	Car
Car	-0.03	49	51	Bus
Car	4.47	99	1	Car
Car	0.67	66	34	Car
Car	2.50	92	8	Car
Car	1.64	84	16	Car
Car	2.29	91	9	Car
Car	1.64	84	16	Car
Car	1.64	84	16	Car
Car	2.11	89	11	Car

Classification Tree Model Phase 1 Prediction and Probabilities

Current Mode	Predicted Mode	Car %	Bus %
Car	Bus	44%	56%
Car	Car	100%	0%
Car	Car	100%	0%
Car	Car	100%	0%
Car	Car	99%	1%
Car	Car	100%	0%
Car	Car	99%	1%
Car	Car	99%	1%
Car	Car	100%	0%
Car	Car	94%	6%
Car	Bus	16%	84%
Car	Car	100%	0%
Car	Car	99%	1%
Car	Car	100%	0%
Car	Car	100%	0%
Car	Car	99%	1%
Car	Car	94%	6%
Car	Car	100%	0%
Car	Car	99%	1%
Car	Car	100%	0%
Car	Car	99%	1%
Car	Car	100%	0%
Car	Car	100%	0%
Car	Car	99%	1%
Car	Car	94%	6%
Car	Car	99%	1%
Car	Car	100%	0%
Car	Car	99%	1%
Car	Bus	44%	56%
Car	Car	93%	7%
Car	Car	100%	0%
Public Bus	Car	100%	0%
Public Bus	Bus	44%	56%
Public Bus	Bus	44%	56%
Car	Bus	44%	56%
Car	Car	100%	0%
Car	Car	100%	0%
Car	Car	98%	2%
Car	Bus	44%	56%
Car	Car	100%	0%
Car	Bus	16%	84%
Car	Car	99%	1%
Car	Bus	63%	38%
Car	Car	100%	0%
Car	Car	100%	0%
Car	Car	99%	1%
Car	Car	100%	0%
Car	Bus	16%	84%

Logit Model Phase 2 Prediction and Probabilities

Future Mode	Predicted Mode	Car %	Tram %	PB %	Metro %
Metro	Metro	11%	16%	14%	60%
Metro	Metro	18%	12%	8%	62%
Tram	Metro	18%	12%	8%	62%
Tram	Tram	16%	40%	8%	36%
Public Bus	Metro	13%	11%	10%	66%
Metro	Metro	16%	11%	9%	64%
Metro	Metro	19%	19%	13%	49%
Metro	Tram	27%	55%	2%	16%
Metro	Metro	14%	29%	9%	48%
Metro	Metro	7%	34%	24%	35%
Public Bus	Metro	16%	16%	6%	62%
Metro	Tram	19%	39%	7%	35%
Metro	Metro	29%	19%	9%	43%
Public Bus	Metro	18%	36%	9%	37%
Metro	Metro	15%	13%	8%	64%
Metro	Metro	19%	19%	13%	49%
Public Bus	Tram	14%	48%	14%	24%
Metro	Metro	18%	12%	8%	62%
Car	Metro	15%	20%	14%	51%
Metro	Metro	12%	13%	9%	66%
Metro	Metro	19%	19%	13%	49%
Car	Metro	25%	20%	10%	45%
Metro	Tram	16%	40%	8%	36%
Metro	Metro	22%	34%	8%	36%
Metro	Metro	7%	7%	13%	74%
Metro	Tram	27%	55%	2%	16%
Metro	Metro	21%	26%	8%	45%
Metro	Tram	14%	55%	8%	22%
Metro	Metro	8%	12%	19%	61%
Metro	Metro	9%	22%	5%	64%
Metro	Tram	19%	39%	7%	35%
Tram	Metro	16%	16%	6%	62%
Metro	Metro	3%	12%	9%	76%
Metro	Metro	4%	23%	21%	52%
Metro	Tram	2%	36%	28%	35%
Metro	Metro	4%	12%	12%	71%
Metro	Metro	21%	26%	8%	45%
Metro	Tram	23%	37%	7%	33%
Metro	Metro	8%	12%	19%	61%
Metro	Tram	19%	46%	5%	31%
Public Bus	Metro	16%	16%	6%	62%
Metro	Metro	29%	19%	9%	43%
Metro	Metro	15%	13%	8%	64%
Metro	Tram	19%	39%	7%	35%
Public Bus	Metro	15%	13%	8%	64%
Metro	Metro	19%	19%	13%	49%
Metro	Metro	15%	13%	8%	64%
Tram	Metro	19%	19%	13%	49%
Tram	Metro	13%	11%	10%	66%

Classification Tree Model Phase 2 Prediction and Probabilities

Future Mode	Predicted Mode	Car %	Merto %	PB %	Tram %
Metro	Public Bus	3%	16%	76%	5%
Metro	Car	88%	7%	4%	1%
Tram	Metro	23%	42%	15%	20%
Tram	Metro	23%	42%	15%	20%
Public Bus	Metro	20%	60%	7%	13%
Metro	Metro	20%	60%	7%	13%
Metro	Metro	20%	60%	7%	13%
Metro	Metro	31%	69%	0%	0%
Metro	Metro	23%	42%	15%	20%
Metro	Metro	20%	60%	7%	13%
Public Bus	Public Bus	3%	16%	76%	5%
Metro	Metro	23%	42%	15%	20%
Metro	Metro	20%	60%	7%	13%
Public Bus	Car	30%	24%	25%	21%
Metro	Metro	23%	42%	15%	20%
Metro	Metro	27%	44%	2%	27%
Public Bus	Public Bus	3%	16%	76%	5%
Metro	Car	88%	7%	4%	1%
Car	Car	40%	31%	19%	11%
Metro	Metro	23%	42%	15%	20%
Metro	Metro	20%	60%	7%	13%
Car	Metro	23%	42%	15%	20%
Metro	Metro	23%	42%	15%	20%
Metro	Metro	20%	60%	7%	13%
Metro	Metro	23%	42%	15%	20%
Metro	Metro	31%	69%	0%	0%
Metro	Metro	23%	42%	15%	20%
Metro	Metro	0%	53%	22%	25%
Metro	Metro	0%	53%	22%	25%
Metro	Public Bus	3%	16%	76%	5%
Metro	Metro	23%	42%	15%	20%
Tram	Metro	23%	42%	15%	20%
Metro	Public Bus	3%	16%	76%	5%
Metro	Public Bus	3%	16%	76%	5%
Metro	Public Bus	3%	16%	76%	5%
Metro	Car	40%	31%	19%	11%
Metro	Metro	23%	42%	15%	20%
Metro	Metro	36%	64%	0%	0%
Metro	Metro	0%	53%	22%	25%
Metro	Car	88%	7%	4%	1%
Public Bus	Public Bus	3%	16%	76%	5%
Metro	Metro	20%	60%	7%	13%
Metro	Metro	23%	42%	15%	20%
Metro	Metro	23%	42%	15%	20%
Public Bus	Metro	23%	42%	15%	20%
Metro	Metro	27%	44%	2%	27%
Metro	Metro	23%	42%	15%	20%
Tram	Tram	19%	21%	2%	57%
Tram	Tram	6%	6%	10%	78%

الملخص

يلعب النقل دورًا رئيسيًا في تسهيل التجارة والتفاعلات الاجتماعية في جميع أنحاء العالم. ومع ذلك ، فإن أنظمة النقل في معظم البلدان ، بما في ذلك البحرين ، تواجه تحديات خطيرة مختلفة. وقد جعل ذلك من الضروري أن يكون لديك فهم شامل لحاجة التنقل (Travel Demand) وخصائص السائق لتطوير الحلول الهندسية التي يمكن أن تسهل تطوير أنظمة النقل هذه. تهدف هذه الأطروحة إلى استكشاف العوامل المؤثرة فيما يتعلق باختيار وضع السفر في البحرين واستخدام نماذج اختيار الوضع للتنبؤ بالكفاءة المحتملة لمختلف وسائل النقل العام في المستقبل.

تستخدم الدراسة 3864 سجلاً متنوع البيانات استخرجت من استبيانات سابقة بالإضافة إلى استبيان آخر تم إجراؤه ضمن هذا البحث. بعد ذلك ، باستخدام برنامج Minitab ، تم بناء نوعين من النماذج الهندسية، وهما نموذج اللوغاريتم ونموذج شجرة التصنيف ، مع إجراء عملية البناء على مرحلتين. ركزت المرحلة الأولى على نمذجة نظام النقل الحالي مع الأخذ في الاعتبار وسائل النقل المتاحة حالياً (السيارة والحافلة) ، بينما ركزت المرحلة الثانية على نمذجة نظام النقل المستقبلي ، مع الأخذ في الاعتبار وسائل النقل العام المحتملة (السيارة ، والحافلات العامة ، والمترو ، والترام). حدد تحليل البيانات أن "تكلفة الرحلة" هو المتنبئ الأول وأن السفر المباشر والسريع وإمكانية الوصول والراحة هي العوامل الأكثر تأثيراً على اختيار وسيلة النقل في البحرين. بالإضافة إلى ذلك ، تشير النتائج إلى أن المترو هو الخيار الأفضل لوسائل النقل العام المستقبلية ، مع ملاحظة تفضيل قوي لنظام النقل الذي يشمل كلا من المترو والترام. يشير البحث أيضاً ، من حيث أداء النموذج ، إلى أن نموذج اللوغاريتم أظهر دقة أعلى مقارنة بشجرة التصنيف عند نمذجة الاستجابات الثنائية. ومع ذلك ، عندما يتعلق الأمر بالنقاط أنماط أكثر تعقيداً ، كما في المرحلة الثانية ، فإن شجرة التصنيف تتفوق على نموذج اللوغاريتم متعدد الحدود. بشكل عام ، توفر الأطروحة رؤى قيمة حول اختيار وسيلة النقل في البحرين وتسلط الضوء على العوامل المهمة التي تؤثر على قرارات التنقل. يمكن أن تساعد النتائج التي تم الحصول عليها في تطوير نظام نقل عام فعال يلبي احتياجات وتفضيلات الركاب في البحرين وإنشاء بنية تحتية مستدامة.



جامعة البحرين
كلية الهندسة
قسم الهندسة المدنية

التنبؤ بسلوك اختيار وسيلة النقل العام للركاب في البحرين باستخدام نموذج اللوغاريتم و شجرة التصنيف: دراسة مقارنة

أطروحة مقدمة كجزء من متطلبات الحصول على درجة الماجستير في الهندسة المدنية

اعداد:

مروة جازي غرايبة

20152606

اشراف:

دكتور أونيب قازدير

جامعة البحرين

سبتمبر 2023