Journal of Civil Engineering, Science and Technology

Volume 13, Issue 1, April 2022, 6 – 22

A SOFT COMPUTING APPROACH TO TRIP GENERATION ESTIMATION IN LAGOS METROPOLIS, NIGERIA

Olanrewaju Oluwafemi Akinfala[1*], Folorunso Oladimeji Ogunwolu[2] and Chidi Onyedikam[2]
[1]Department of Geography and Planning, University of Lagos, 101245, Lagos, Nigeria
[2]Department of Systems Engineering, University of Lagos, 101245, Lagos, Nigeria

Date received: 19/08/2021 Date accepted: 11/12/2021
*Corresponding author's email: lanrewaju.akinfala@gmail.com
DOI: 10.33736/jcest.3821.2022

Abstract — Trip generation is an indispensable component of the four-stage transportation planning process because the subsequent three stages are predicated on its results. Linear regression has been widely adopted to predict trips due to its simplicity and its outperformance of more sophisticated count models and in some cases, soft computing models. The efficacy of regression for estimating trip generation alongside Artificial Neural Networks (ANN) and Fuzzy Expert System (FES) was examined. The performance of each model was evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Coefficient of Determination (R^2) and the capability of predicting average trips. The R^2 for Regression, ANN and FES were all 0.71. The MAE for Regression, FES and ANN were 0.56, 0.55 and 0.49 respectively. The MSE for Regression, ANN and FES were 1.15, 1.16 and 1.15 respectively. Finally, FES and ANN resulted in average trips of 4.5 in comparison to actual average trips of 4.51 per household, while regression produced average trips of 4.51. ANN and FES are not superior alternatives to the linear regression model for trip generation modelling. The performance increments gained from adopting these models are marginal and the extra development and computational effort required to apply such sophisticated approaches may not be justified.

Keywords: Linear regression, model performance, trip rates, fuzzy expert system, artificial neural network

1.0 INTRODUCTION

Travel demand modelling is an indispensable part of transportation planning. Trip generation is the first step of the four-stage travel demand modelling process. It provides the paradigm to estimate the volume of trips made by individuals [1] or between one Traffic Analysis Zone (TAZ) and another for various purposes [2]. Mathematical models are typically adopted for trip generation to express the relationship between certain variables that are known to affect trip generation [2]. The development of accurate and reliable mathematical models for trip generation is crucial because it is the first step in the process and the basis upon which the subsequent steps of travel demand modelling are developed. Therefore, errors and inaccuracies produced at this stage will propagate and perhaps become more compounded in subsequent stages [1], [3]. Additionally, the development of reliable mathematical models permit future planning since all that is needed is the future data on certain variables for future trips to be predicted. At least this holds true assuming the relationship between the explanatory variables and the number of trips generated remains unchanged between the base and predicted year [4].

The number of trips generated is widely acknowledged to be dependent on socio-economic and household characteristics (income levels, household size, car and driver's license ownership, education, gender, age, employment status, and the number of children) [1], [2], [5]–[7]. Other factors include land-use/spatial factors (urban, rural, population density, number of establishments); temporal factors (time of day, year, season) [7] and finally, climatic conditions [1], [7].

In any modelling effort, one or more of these aforementioned factors is/are the independent/input/explanatory variable(s), and the number of trips produced is the response/output/dependent variable. The conventional approach to trip generation modelling for some decades now has been to adopt Regression Analysis. Regression is the most widely used method due to its simplicity [8] and generally acceptable performance [2]. Although some studies have provided descriptions of the methodological deficiencies of regression [2], [9], [10], there is a general consensus that it performs quite well. Nevertheless, while multiple linear regression is able to establish good statistical relationships between explanatory variables and the number of trips generated, transport planning and invariably trip generation naturally involves complex human decision-making processes, many of which are ladened with uncertainty and ambiguity [11], [12]. Uncertainty and complexity are embedded into the trip making process due to perceptions regarding travel time and destination utilities and the inexactness of the relationship between trip generation and explanatory variables [13]. After all, the decision to make a trip is generally not as clear and straightforward as, "I have a car and there are two other household members in my house, therefore I must make two trips today". In many cases, individuals decide to make trips among several other competing factors which are often immeasurable. These uncertainties in the trip making process have been difficult for modellers to capture in conventional models [12] and this offers a unique selling point for soft computing methods. Furthermore, the notion that the relationship between explanatory variables and trips are constant over the prediction year infuses a significant amount of uncertainty to the trip generation modelling process [4]. Despite the clear and unique opportunity to apply soft computing methods for trip generation, this has not necessarily been the case. However, this is not particularly surprising, as soft computing methods including fuzzy logic, neural network, Bayesian networks, adaptive neuro-fuzzy inference systems (ANFIS), genetic algorithms and other intelligent computational methods are scarcely applied in the transportation planning domain in general. Only about 8% of soft computing papers in transportation were attributed to transport planning (the domain that trip generation falls under) in contrast to traffic control and management that accounted for approximately 37% of the application of soft computing [12]. Although the literature is more than fifteen years old, the literature survey in this study has shown a continued shortage of research work on the application of these methods for trip generation. Furthermore, there was very limited literature on their comparative performance. The existing studies using these methods have typically done so by applying one of the soft computing approaches in their work. This makes conclusions on the general performance of soft computing approaches difficult to establish. Therefore, in this work, fuzzy logic and artificial neural network models are applied to estimate trip generation and their outputs are compared with those of a traditional linear regression estimate.

Soft computing methods have been applied to the entire four-stage modelling process. However, the literature of this study focused entirely on its application for the first stage, trip generation. Seyedabrishami and Shafahi [13] developed an expert-guided ANFIS model that incorporated expert judgment and experience into the fuzzy modelling process. It was used specifically to ameliorate data sparseness issues which typically affect predictive performance. One notable finding in their work was that linear regression (R^2 of 0.9241) outperformed ANFIS (R^2 of 0.8861). However, the expert-guided ANFIS (R^2 of 0.9289) outperformed linear regression marginally. Ahmadpour et al., [11] estimated worker trip production using a neuro-fuzzy system and subtractive clustering was adopted to determine the most influential factors necessary for developing the rule antecedents. Abu-Eisheh and Irshaid [14] applied ANFIS to trip generation in Palestine. ANFIS was more accurate than linear regression in terms of R^2 and all measures of predictive errors. Shafahi and Abrishami [3] adopted ANFIS for school trip attraction modelling in Shiraz, Iran and compared the results with a regression model. The training and test data of the optimal ANFIS model had $R^2 > 0.75$. Thus, outperforming R^2's regression of 0.69.

Rassafi et al., [4] developed a fuzzy expert system to estimate work trips in Mashad, Iran and compared the performance of their model to linear regression. The R^2 of linear regression and fuzzy expert system were 0.81 and 0.79 respectively. This showed a slight over-performance by regression. However, the mean square error of the fuzzy system was much lower than the regression model.

Artificial neural networks (ANNs) have also been applied to trip generation. Perhaps one of the earliest works where ANN was applied to trip generation was in [15]. Arliansyah and Hartono [16] used a Radial Basis Function Neural Network (RBFNN) to model trip attraction and compared the results with regression. RBFNN significantly outperformed regression, as it generated $R^2 > 0.7$ for both training and test data, while regression generated R^2 of 0.459.

Similarly, Etu and Oyedepo [17] applied RBFNN to trip generation for a high-density residential area in Akure, Nigeria. RBFNN was significantly superior to regression with R^2 of 0.913 compared to R^2 of 0.552 generated by regression. In contrast, Tillema et al., [18] assessed the performance of neural networks in comparison to regression under two circumstances: data complexity and scarcity. Contrary to their initial hypothesis that neural networks should outperform regression, this was not the case, as the results were not significantly different. They concluded that no additional benefit was to be obtained by adopting neural networks over regression.

The literature review on the performance of soft computing methods in comparison to linear regression has generated mixed results. The evidence so far, therefore, limits the ability to make a general statement regarding the clear superiority of soft computing methods and regression. However, one thing appears to be clear, the superiority in performance of soft computing approaches is more pronounced when it involves trip attraction. This is perhaps a result of trip attraction inherently exhibits more uncertainty than trip production [3]. Therefore, linear regression may tend to underperform more frequently. Additionally, even though linear regression in some instances outperforms a soft computing approach, in the absence of robust data and perhaps extensive data collection resources, soft computing methods may offer significant benefits over the traditional regression [4].

Clearly, a reasonable number of trip generation modelling work has been performed in recent times using soft computing methods. However, a comprehensive comparative study, where one or more soft computing approaches were executed on the same data and the results presented side by side is lacking. Similar comparative works have been carried out for econometric models [2], [7], [10]. This study is the equivalent of these works but from a soft computing perspective. The central question of this work is whether Fuzzy Expert System or Artificial Neural Networks are superior in performance to regression for household trip generation modelling using data from Lagos, Nigeria . Such a study may help provide some context for model performance evaluation and provide better justification for the continued reliance on traditional linear regression or for a shift to one or more soft computing methods.

The rest of this paper is organized as followed. Section 2 discusses the materials and methods adopted. More specifically, section 2 highlights the following: methods and procedures for data collection, data types, regression modelling, development of the Fuzzy Expert System, development and parameter selection for the Artificial Neural Network model, comparative measures used and data imputation methods adopted. Section 3 focuses on the Results and Discussions. In that section, the results produced are explained or rationalized and situated in the context of existing literature. In section 4, conclusions are drawn.

2.0 MATERIALS AND METHODS

2.1 Data: Source, Collection and Collation

The data used for the study was obtained with permission from the Lagos Area Metropolitan Transport Agency (LAMATA) archives. It consists a result of 6,000 household interviews conducted between February and November 2009 as part of a comprehensive Lagos Transport Study that was later extended to the entirety of the Lagos Mega City extending to some parts of adjoining Ogun State, Nigeria. The data consisted of two tranches of household interviews: (1) an initial 2,000 household interviews focused on the corridors of the New Badagry Expressway which was in its teething stage then. This tranche of data collection extends from Obalende/Lagos Island through Iganmu to the traffic corridor to Mile 2 and backward to Apapa-Oshodi expressway corridor and backward to Anthony Interchange. The primary purpose was to gain an understanding of travel behaviour around the then proposed New Lagos/Badagry Expressway corridor (2) The second tranche of 4,000 household interviews were spread to the whole of Lagos state systematically designed to cover all Local Government and Local Council Development Areas (LCDAs). It was conducted on consecutive weekends (Saturday and Sunday) spanning three months where interviewers and the supervisory team were trained for effective data collection.

In total, 26,261 trips from 5,816 households were captured. The captured trips indicated notable trip features such as trip ends (origin and destinations), trip purposes, modes of travel, time of trip and cost of the trip in addition to the general household characteristics obtained from household heads such as household income, number of cars (if any), and occupations of household members. The characteristics of the specific data adopted for modelling are presented in Table 1. Individual members of each household were interviewed to capture the characteristics and features of the trips they have made within the last three days of the interview. Lagos was divided into 734 contiguous zones for travel demand prediction. The interview location (home) and the trip ends (origin and destination locations) for each trip were coded in terms of the zones which they fall.

Table 1 Data characteristics

Variable	Range	Type
Trips	1-21	Discrete-valued
Household size	1-9	Discrete-valued
Car ownership	0-7	Discrete-valued
Income	1-9	Categorical

2.2 Regression Analysis

Regression analysis is a type of predictive modelling technique that investigates the relationship between a dependent (target) and a set of independent variable(s) (predictors). Ordinary Least Square (OLS) regression modelling is adopted from Ortuzar and Willumsen [19] to model the relationship between trips and household characteristics. This model uses a best fit straight line known as regression line that is fitted to the data containing the observed dependent and independent variables using the least square method [19]. The technique establishes a relationship between a dependent variable Y, and one or more independent variables, vector X (Equation 1). The regression model served as a basis of comparison for the other models adopted in this study. The analysis was performed using Statsmodels. Statsmodels is a python package for the implementation of various statistical analyses and tests [20]. The significance level was 0.05 for all the regression analyses performed.

$$Y_i = \beta_0 + \beta^i X_n + \varepsilon_n, \qquad n = 1,2,3,4 \dots N \tag{1}$$

Where Y_i is the estimated trips made by household i, β_0 the constant, β^i a vector of parameters to be estimated from the data, X_n a vector of explanatory variables and ε_n is the error term.

2.3 Artificial Neural Network (ANN)

Artificial Neural Networks, also known as connectionist systems due to their large networks of highly interconnected processing units known as neurons are one of the key methods in soft computing. ANNs are abstracted systems that attempt to mimic the architecture and the operating characteristics of the biological neuron and human brain. The likeness of ANN to the human brain is a result of its parallel local processing and highly distributed architecture [21], as well as its processing units and mechanisms. Although one of its major criticisms is that it is a black-box process, where its processing mechanism does not render itself to traditional analytical or empirical testing and derivations. Nonetheless, it is capable of mapping highly complex and non-linear input to their outputs accurately [18]. More importantly, ANN analysis is performed without assumptions about the underlying distribution of the data given a sufficient number of input/output pairs. In addition to their ability to handle complex non-linear data, ANNs are particularly attractive for modelling due to their tolerance for faulty, noisy and incomplete data. They are self-correcting and they generalize functions quite well [22]. This ability to map complex systems accurately is largely due to the constant adjustment of one of the key elements of ANN, the connection weights, or in biological terms, synapses. Connection weights are constantly adjusted based on the difference between the estimated output and the target output (when the system is a supervised one) [23]. Connection weights also determine the strength and character of the connection between input and output [24]. ANNs typically have three or more layers: input, hidden and output layers. Each layer has one or more nodes. The input and output layers typically have one layer each. However, hidden layers may be more than one. The number of nodes in each layer, the training algorithm and schemes as well as pre-processing steps are described below. The analysis was performed using Keras. Keras is an open-source software that enables neural network implementation in Python. It is part of the larger Scikit-learn package [25].

The nodes in the input layer are the independent variables in the model. In this study, four different networks were trained. In the network with one input, only the household size was adopted. In the two inputs network, two networks (1) with household size and car ownership (2) with household size and income were adopted while for the three inputs network, a network with household size, car ownership and income were adopted. In all four networks, the output layer that predicts the dependent variable has a single node, which was the predicted number of trips as a function of household size, car ownership and or income as the case may be. In contrast to input and output layer structures which are dependent on the variables of interest, there is no strict rule of thumb for the best configuration of hidden layers. Therefore, several hidden layer architectures were experimented with and their performances were monitored and assessed. The complexity of the architectures included those with as many as 50 neurons in each of the two hidden layers and the simplest was 3 neurons in one hidden layer. The results of model training and performance are discussed in section 3.2. The objective was to minimize Mean Square Error (MSE). The training algorithm adopted was the Stochastic Gradient Descent (SGD). The activation function was the Rectified Linear Activation Function (Relu) in the hidden layer and Linear in the output layer.

The data was split into three sets. 70% (4071 household samples) for training, 15% (873 household samples) for testing and (872 household samples) for validation.

2.4 Fuzzy Expert System (FES)

Fuzzy logic is a soft computing technique first proposed by Zadeh [26] to represent the uncertain and imprecise nature of human knowledge. Fuzzy logic is highly beneficial for modelling non-linear and complex systems because it handles imprecision and uncertainty quite robustly and it incorporates expert knowledge naturally [4]. There are four key components of any FES: Fuzzification, rule-base, inference system and defuzzification.

In many modelling processes, one of the early steps is to define the inputs into the model. In FES, the first step is to define the input variables for the model. The crisp input variable(s) are fuzzified using fuzzy membership functions. Fuzzification transforms crisp inputs into fuzzy inputs at various degrees of membership within [0,1] partitioned among the sets based on the linguistic variables defined. For example, for a variable such as income, fuzzy sets can include low, medium and high. The fuzzification process is achieved through membership functions. A wide variety of membership functions exist, including Gaussian, trapezoidal, triangular, generalized bell curve, and sigmoid curve [27], [28]. The assignment of membership functions to fuzzy variables is dependent on expert knowledge and trial and error. However, some automated methods such as a hybrid system which may adopt a neural network or genetic algorithm for membership function assignment may also be used [4].

The rule base is developed using a set of IF-THEN rules to model the relationship between the inputs and the outputs. Rule base development is heavily dependent on expert knowledge in the specific domain. The fuzzy inference system or engine combines the fuzzified inputs and rule-base to produce fuzzy outputs. The final stage of the Fuzzy analysis is defuzzification. The output needs to be transformed from a fuzzy output to a crisp output. A number of methods exist for this including centroid, weighted average maxima, average maximum, height method [27], [28]. The centroid method is the most commonly adopted.

In this study, SciKit-Fuzzy, a fuzzy logic toolbox for Scipy (Scientific Python), a python package for the development and implementation of fuzzy logic algorithms [29] was adopted to implement a fuzzy logic-based trip generation model. One input variable, household size was selected to predict one output variable (number of trips). The universe of discourse for household size and trips/household were (1-9) and (1-21) respectively. Triangular membership functions were selected for fuzzification of both household size and trips produced. Nine fuzzy sets were developed for both household size and trips. This ensured sufficient overlap among the fuzzy sets. The rule base was based on the authors' knowledge about the subject, evaluation of the data set and trial and error. Nine simple rules in the form of (very low household = very low trips, low household size = low trips, medium household = medium trips, high household = high trips, very high household = very high trips), were developed for the modelling process. Finally, the centroid approach was adopted for defuzzification.

2.5 Comparative Analysis

Performance measures such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2) were performed using the metrics module of Scikit-learn library in python. In addition, the average number of trips produced from each of the soft computing methods adopted were compared to the Regression model result and the real-data set by quantifying their deviations. Lower MSE and MAE reflect smaller prediction errors between observed and modelled trip rates. Higher R^2 values reflect higher model explanatory power. Therefore, a model with lower MAE and MSE and a higher R^2 is preferred to one with a lower R^2 and higher MAE and MSE.

2.6 Imputation for Missing Data

Income is a sensitive and difficult measure to capture during surveys and respondents do not typically want to disclose income information [30]. Essig and Winter [31] noted factors such as the socio-psychological dynamics between the interviewer and the respondent, privacy, and cognitive burden may lead to higher non-repsonse rates. Questions such as income tend to produce higher than usual non-response rates in surveys and this may lead to information loss that may be as problematic as losing the entire response unit [31], [32]. A review of the literature showed that income non-response rates varies quite widely. It was between 19-21% in a study carried out by [31]. Non-response rates of 10-15% have been reported by (Moore, Stinson, and Welniak 2000; Moore and Loomis 2001 in [30]). Similarly, non-response rates in Regina and Oliver, [32] were in the range of 5-35% depending on what type of income or wealth variable was enquired. In this study, missing data for income was approximately 38%. Out of the 5,816 households surveyed, 1,633 (28.07 % of the entire responses) refused to provide any information while 627 (10.7 % of the entire responses) responded: "Don't know". The high number of missing values in income may limit conclusions that can be drawn from the analysis based on income.

Initially, a simple method of deleting all the records with missing income data was adopted but this led to noticeable underperformance. Therefore, imputation was adopted to ameliorate some of the missing data deficiencies .

Imputation is a means of inferring missing data from existing data. Imputation in this study was performed using Logistic Regression since income in this study was categorical in nature. This was performed in sklearn python environment. The records with complete data (3,556 records) were extracted as training data for which the Logistic Regression model was fitted. The missing income (2,260 records) data was then predicted using the trained Logistic regression model. All subsequent modelling for Linear regression and Artificial Neural Networks was performed using the imputed dataset.

3.0 RESULTS AND DISCUSSION

3.1 Regression Analysis Model Development and Performance

Four regression models were developed to estimate trip rates based on a step-wise approach where one variable was added to the model at each additional step. The first model adopted household size only as the dependent variable. The second model included auto-ownership as well as household size. In the third, income was added to household size. The final model included all three independent variables (household size labelled as household, auto-ownership labelled as Car, and Income labelled as Income).

Table 2 shows that household size has a positive and statistically significant influence on trip rates with a coefficient of 2.07. Car ownership has a positive impact on trip rates with a coefficient of 0.0033. However, the magnitude of the effect is practically negligible. Moreover, it was not statistically significant. Kitamura et al., [33] also found no relationship between vehicle ownership and person trips in their study in the San Francisco Bay Area. Although intuitively and even from existing literature, a higher number of cars per household would mean an increased number of trips [4], [6], the Lagos case may, however, be peculiar. It must be taken into consideration that car ownership is generally low amongst Lagosians. Based on the data adopted in this study, the average number of trips per household is 4.51 and the car ownership per household is 0.33.

Table 2 Regression coefficients with model comprising household size, car ownership and income

	Coef	std err	T	P>|t|	[0.025	0.975]
Constant	0.0856	0.046	1.874	0.061	-0.004	0.175
Household	2.0741	0.017	118.664	0.000	2.040	2.108
Car	0.0033	0.022	0.153	0.878	-0.039	0.046
Income	0.0084	0.011	0.795	0.427	-0.012	0.029

In another study in Lagos, cars per household was 0.86, with 52% of households owning no vehicle and 26% owning one vehicle [34]. The majority of trips (70%-80%) in Lagos are completed with public transportation [34], [35]. In this study, the proportion of the 26,261 trips completed with private vehicles was approximately 11.7%, while public transportation including buses, tricycles, motorbikes, and taxis accounted for nearly 78% of the total trips. The mode share distribution of the Lagos population is clearly dominated by public transportation and more specifically, buses. Therefore, the presence or absence of cars may not influence trip making noticeably. Since the vast majority of trips will be completed irrespective of whether there is a car available or not. Additionally, the low level of car usage for the large majority of trips in Lagos may have inhibited the detection of any statistical difference particularly if the trip generation rates of those households with cars and those without cars are not markedly different.

Income also showed a positive effect on trip rates based on the observed coefficient. But again, the magnitude of the effect was negligible (0.0084) and it was not statistically significant. Another study in other African cities reported a somewhat similar result. Trip rates did not vary significantly across various income levels in Nairobi, Kenya and Dar es Salaam, Tanzania [36]. Similarly, Xu et al., [37] reported that levels of wealth did not affect mobility patterns noticeably in Boston and Singapore. While no conclusive explanation can be provided for why income is not statistically significant in this study, some possible explanations include the endogeneity of income and car ownership and as explained earlier, the dominant mode share of buses in Lagos. Higher-income households tend to have more cars [19], [38], [39]. These households typically move around more due to increased mobility. However, with the ubiquity of public transportation options in Lagos (Mini buses, tricycle, motorbikes, large buses, taxis) and the large mode share of public transport, the expected accessibility gap is bridged and the differences in trip making behaviour across the various income groups may be minimized. Therefore, because accessibility is not encumbered to the point where there is a discernible difference in trip making behaviour between households that own cars and those that do not, the income variable may have been insignificant. Barbosa et al., [40]; Xu et al., [37] offered similar explanations for the lack of significant variability in trip making characteristics across various income groups in their respective studies.

Table 3 and 4 show the outputs of the regression models with two input variables. The two tables show that household size is the only statistically significant variable. Both car ownership and income were positive but the coefficients were negligible and statistically insignificant.

Table 3 Regression coefficients with model comprising household size and car ownership

	Coef	std err	T	P>|t|	[0.025	0.975]
Constant	0.1030	0.040	2.567	0.010	0.024	0.182
Household	2.0749	0.017	118.917	0.000	2.041	2.109
Car	0.0126	0.018	0.680	0.497	-0.024	0.049

Table 4 Regression coefficients with model comprising household size and Income

	Coef	std err	T	P>|t|	[0.025	0.975]
Constant	0.0847	0.045	1.871	0.061	-0.004	0.173
Household	2.0740	0.017	118.729	0.000	2.040	2.108
Income	0.0093	0.009	1.035	0.301	-0.008	0.027

The model with only one variable was sufficient to provide high explanatory power for trip generation estimation with an R^2 of 0.71. This is considering that only household size was adopted as the independent variable. The parameters of the final regression model with only household size are presented in Table 5. The coefficient observed suggests that one household member increases the conditional mean of trips per day by approximately 2.04 – 2.11. This value can essentially be considered the average daily person-trip rates per day in Lagos. Reasonably similar values were observed in a study in three other African cities. The person-trip rates in Dar es Salaam, Nairobi and Cape Town were 2.52, 2.3 and 1.70 respectively [36]. Finally, the results of the single variable (household size) regression model were compared with the models that included auto-ownership and income. Not surprisingly, all four regression models showed nearly equal performance on all the four performance criteria. This was as a result of the negligible magnitude of the coefficients of car ownership and income. In all the models, only household size had any sizable effect on the predicted output and the coefficient of household size remained fairly similar (2.04 - 2.11) across all models**.** The highly significant effect observed for household size is not particularly new because numerous other studies have also shown that demographic size at any scale (household, zone, city, state) tends to have a significant effect on trip characteristics [2], [17], [33]. Trips are a derived demand because the activities that people typically need to perform are inaccessible at their origins [41]. As a result, there is a need to travel or make trips to destinations to engage in these activities (work, shopping, school). Consequently, the increased presence of people means there is an increased level of activity and by extension, trips.

Table 5 Regression model parameters (household size)

	Coef	std err	T	P>|t|	[0.025	0.975]
Constant	0.1073	0.040	2.707	0.007	0.030	0.185
Household	2.0749	0.017	118.922	0.000	2.041	2.109

Because one independent variable was used for modelling, the regression equation results in (Equation 2).

$$y = 0.1073 + 2.0749\ hh \tag{2}$$

Where y is the number of expected trips and hh is the household size.

Since the addition of income and car ownership to the regression model did not result in any marked improvement, the regression model with only household size as the independent variable is considered the best since it requires a single variable to produce equivalent performance to those with more variables included. The performance of the regression model with one variable which was the best regression model is presented in Table 6. This regression model was used for all subsequent comparisons because it was simple and provided a relatively high performance.

Table 6 Performance of best regression model (household size)

Performance Measure
R^2: 0.71
MAE: 0.57
MSE: 1.15
Average trips: 4.51

3.2 Artificial Neural Network

Four different configurations of inputs were used for prediction. Table 7 presents the performance measures obtained for the best input configuration. However, the performance was quite similar across the various input configurations. The full data of the other ANN models with additional input variables are presented in Table 8. Obviously, Neural networks performed well for modelling trip rates, with each of the four networks resulting in average trips in the range of 4.5 – 4.52 in comparison to the actual average trip rate of 4.51. As discussed in Section 2.3, various hidden and input layer configurations were tested. Table 7 shows the best performing network and configuration. This was the network with a single input (household size) and one hidden layer with three neurons. In this network, one hidden layer was sufficient to achieve equivalent or superior results in comparison to the more complex models. The more complex networks often yielded subpar performance in terms of accuracy and convergence time. It was initially assumed that neural networks may be capable of extracting some complex or hidden information from the two variables which produced statistically insignificant coefficients when modelled with a regression model. However, the results showed that including car ownership and/or income did not result in marked improvement in performance. A case may even be made for diminished performance due to the addition of the other two variables and it is almost certain that the reason for diminished performance is the same as those discussed in section 3.1 regarding car ownership and income. With the exception of the network for household size and income that produced average trips of 4.51, the neural network model with a single input variable produced better performance outcomes in comparison to the other three. Additionally, it converged faster at approximately 20-30 epochs using three neurons in the hidden layer. Using a configurationof a single hidden layer with three neurons, the other three models presented in Table 8 typically converged around 80-90 epochs. Convergence did not improve for these models until the number of hidden layer neurons was increased from three to five. Only after increasing the number of hidden layer neurons, did the convergence time reduce to approximately 20 epochs.

Table 7 Neural Network Performance with one input (household size)

Performance Measure	Training data	Test data	Validation data	Full data
R^2	0.68	0.74	0.81	0.71
MAE	0.58	0.53	0.49	0.49
MSE	1.27	1.06	0.68	1.15
Average trips	4.5	4.49	4.52	4.5

Table 8 Comparison of the Performance of Neural Networks with additional inputs

Performance Measure	Household & car ownership	Household & Income	Household, Income & car ownership
R^2	0.71	0.71	0.71
MAE	0.55	0.56	0.53
MSE	1.15	1.15	1.15
Average trips	4.52	4.51	4.5

In summary, the results in Tables 7 and 8 show that all the models produced somewhat similar performance outcomes. The results clearly show the suitability of neural networks for trip generation modelling and its equal or superior performance to regression, at least by using this data set. For subsequent comparison with other models, the ANN result of the full data and a single input model was adopted.

3.3 Fuzzy Expert System

The fuzzy model was developed using one input (household size) and one output (trip rate). This was because regression and ANN have already shown that including the other two variables (income and car ownership) in the model are not significantly beneficial to model performance. The Fuzzy Expert System trip generation model showed varied performance levels depending on the type of performance measure assessed and the nature of the fuzzification of the model output (i.e. trip rates). Two separate fuzzifications were performed for the trip rates using the triangular membership function. The first produced a convex and normal fuzzy set as shown in Figure 1. While the second produced a non-convex and non-normal fuzzy set as shown in Figure 2.

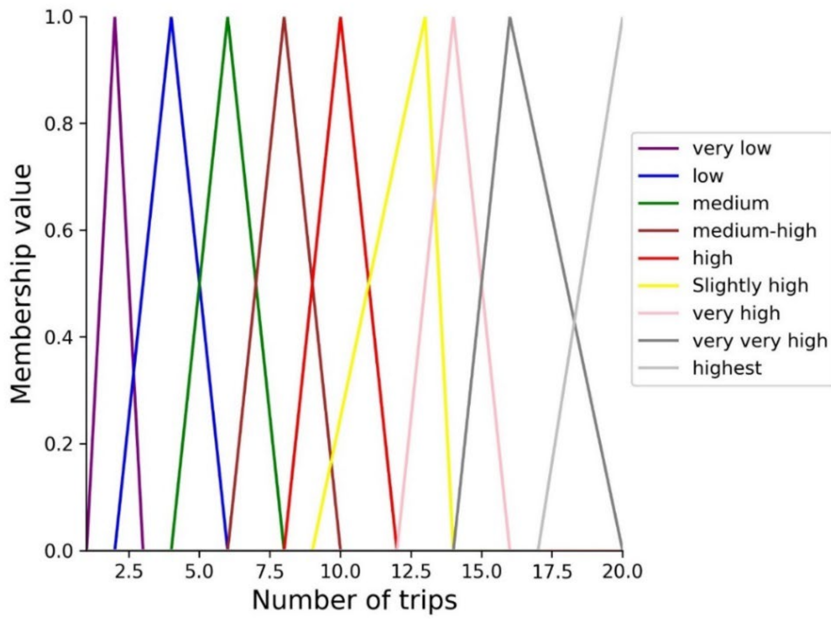

Figure 1 Fuzzification for trips (Convex, Normal)

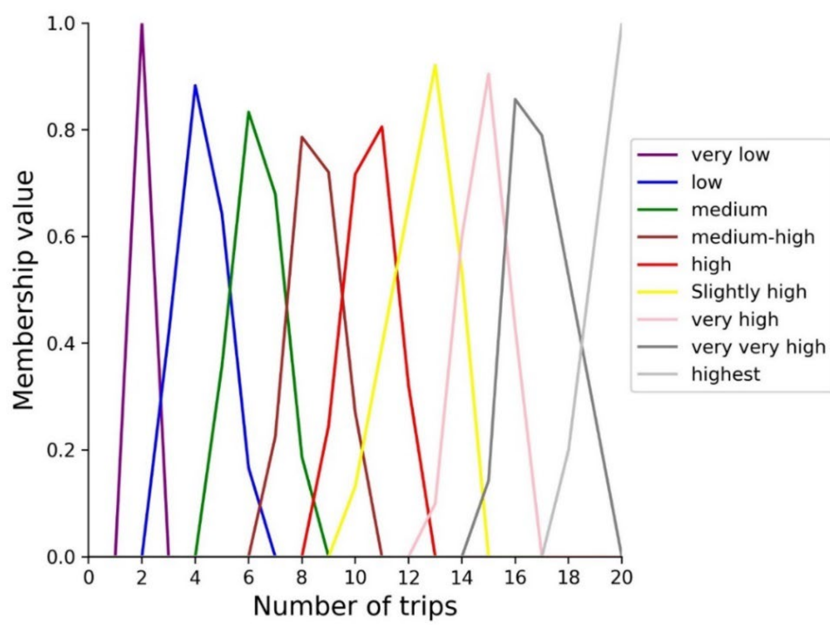

Figure 2 Fuzzification for trips (Non-convex, Non-normal)

The non-convex set was generated using the coefficient of the household size variable in regression modelling. Since regression is considered one of the most superior models for trip generation modelling, our initial hypothesis was that the coefficients would be quite suitable for understanding how household size influences trip rates. The input was the household size and it was fuzzified as a normal convex set in Figure 3.

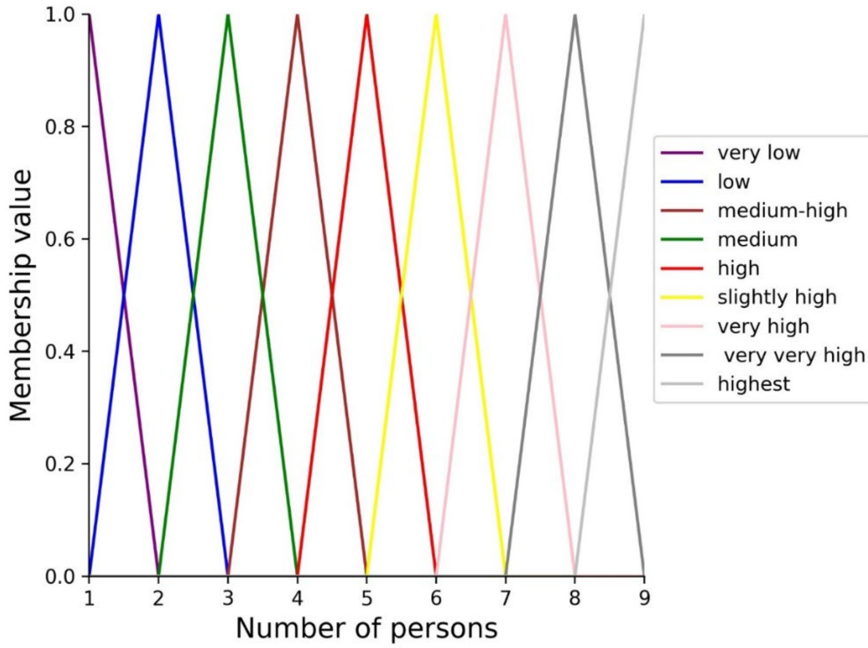

Figure 3 Fuzzification of household size

The results presented in Table 9 show that when the coefficient from regression was the central factor in fuzzification process, the MSE, average trips and R^2 values of the non-convex and non-normal fuzzy sets were of superior performance to those of the normal and convex fuzzy sets.

Table 9 Comparison of Convex normal to non-convex non-normal fuzzy set

Performance measure	Convex normal fuzzy set	Non-convex and non-normal fuzzy set
R^2	0.69	0.71
MAE	0.38	0.55
MSE	1.23	1.16
Average trips	4.25	4.5

The higher MAE produced by the non-convex fuzzy set suggests that it over-predicts and/or under-predicts with larger magnitudes than the convex fuzzy set. However, the deviations are often within similar ranges therefore the errors are cancelled out when MSE is computed. This suggests that non-convex fuzzy sets can offer similar or even better model performance. Despite this, non-convex fuzzy sets are not as widely used for modelling in comparison to convex fuzzy sets [42]. This is perhaps due to the long-standing precedence of restricting fuzzification to standard well-known shapes and forms [43]. The results in this study, together with those of [42], [43] show that non-convex fuzzy sets may just be as good or better than convex fuzzy sets.

3.4 Comparison of Models

The results presented in Table 10 show quite similar performances across the various models. Regression was able to exactly produce the same number of average trips per household and it was a slightly better estimate in comparison to the estimates of ANN and FES. Nevertheless, ANN may have the potential to produce results that are significantly more accurate and precise than FES and Regression.

Table 10 Model performance

Performance Measure	Actual trips	Regression	Fuzzy Expert System (Non-convex)	ANN
Average trips/household	4.51	4.51	4.5	4.5
R^2	-	0.71	0.71	0.71
MAE	-	0.57	0.55	0.49
MSE	-	1.15	1.16	1.15

The validation set in ANN highlights this, as it produced performance outcomes such as R^2 and MSE that were remarkably better in comparison to FES and regression. FES and regression both produced an R^2 of 0.71. In contrast, the validation set of the best neural network produced an R^2 of 0.81, although, the overall data showed very similar performance across the board as shown in Table 7. This marked difference in the validation is a confirmation that indeed, ANN, may offer superior performance in comparison to Regression and FES. FES may also hold some potential as well, considering that the convex fuzzy set model produced a lower MAE of 0.38 as shown in Table 9, in comparison to all the other models that had a MAE of at least 0.49. The findings of this work are therefore similar to the findings of [13], [18], where, although soft computing methods outperformed regression, the performance improvements were marginal.

Based on the experience gained during this study, a statement can be made that Regression, Fuzzy Expert Systems and Artificial Neural Networks all have their peculiar advantages.

FES may be quite useful where data is very limited but the modeller has a very good knowledge of the system's behaviour. Owing to FESs amenability and flexibility, it can be developed to perform prediction at a high level of accuracy by the modeller even in the absence of sufficient data. While this is advantageous, fine-tuning the model may be difficult and laborious as it relies substantially on human experts' knowledge of fuzzy logic and the particular system being modelled. There has been limited application of Non-convex fuzzy sets. The results of this study have shown that they have the potential to offer similar or superior performance in comparison to convex fuzzy sets.

ANN on the other hand requires little input from the modeller but it requires a substantial amount of data for training. It may however hold a lot of potential for modelling more complex trip making scenarios.

Regression is quite simple and the data requirements are quite manageable. Jenkins & Quintana-Ascencio [44] recently showed that as few as 25 observations may be sufficient for reliable model estimates. In addition, regression provides useful parameters such as coefficients, confidence intervals, precision, significance levels which may provide further insight for decision-making. For example, in this work, it was established that the conditional mean of trips increases by a rate of 2.04 - 2.11 per household member. This can be interpreted as the average daily trips per person. FES and ANN lack such added statistics owing to their black-box nature. Unfortunately, when data becomes more complex and non-linear, many of the benefits of regression are jettisoned and the performance of regression may begin to deteriorate.

4.0 CONCLUSION

Trip generation modelling is arguably the most important task in transportation planning. This has resulted in a significant effort being geared towards accurate modelling over the past few decades. A key question that still faces engineers and planners is which modelling approach is better? In this study, the performance of linear regression, Fuzzy Expert System and Artificial Neural Network for trip generation prediction were compared using data from 5,816 households in Lagos, Nigeria.

For reasons which remain unexplainable to date, regression, despite its simplicity and methodological shortcomings, has been the gold standard for modelling trip generation because it continues to outperform more sophisticated count-based models and in some instances, soft computing models. Even in most instances where soft computing outperforms regression, the difference is only marginal and this brings into question the need to adopt more complex soft computing models.

The results of this study have shown that Fuzzy Expert Systems and Artificial Neural Networks both perform quite well and even better than Regression in some instances. However, while these soft computing methods may outperform Regression models for trip generation, the extra modelling and computation time may not be justified. Nevertheless, it is possible that trip making will become increasingly more complex due to advances in technology or extreme behavioural modifications due to situations such as disasters and lockdowns. Simple relationships with explanatory variables such as household size, income, auto-ownership and employment characteristics, may then become insufficient and soft computing models may begin to offer more substantial benefits in comparison to regression.

Despite the results of this study and those from existing literature, caution must be exercised regarding the generalization of comparative model performance. Literature review shows that results from comparative analysis among trip generation models (soft computing methods and linear regression) have been mixed and this is probably due to the varying datasets adopted. The complexity of the parameter specification of each dataset also varies. Some models may be better at specifying the parameters of a particular dataset more than others. Therefore, this does not mean one model is generally superior to another for modelling a particular problem domain. It may just be that the model is better at specifying the parameters of the dataset it was applied to.

The observation that household size alone can explain such a large proportion of trip generation in Lagos, Nigeria may be highly beneficial to transportation planners and engineers since this reduces the amount of data required to produce reasonable trip estimates. In a developing country where resources may be scarce to collect vast amounts of data, having access to household data may therefore be the difference between a successful modelling exercise or no modelling at all for planning purposes. In conclusion, for practical purposes, it will be beneficial for traffic engineers, transportation planners and researchers to first apply regression to trip generation. The performance of the regression model should then inform the decision of whether more complex or sophisticated models such as FES or ANN should be adopted.

Conflict of Interests

The authors report no potential conflict of interest.

Acknowledgement

The authors acknowledge the Lagos Metropolitan Area Transport Agency (LAMATA) for providing the household data used in this work through Advanced Engineering Consultants, Ilupeju, Lagos, Nigeria.

References

[1] Cubucku, K. M. (2001). Factors affecting shopping trip generation rates in metropolitan areas. Studies in Regional & Urban Planning, 9, 51–67.

[2] Chang, J. S., Jung, D., Kim, J., & Kang, T. (2014). Comparative analysis of trip generation models: results using home-based work trips in the Seoul metropolitan area. Transportation Letter: The International Journal of Transportation Research, 6(2), 78–88. https://doi.org/10.1179/1942787514Y.0000000011.

[3] Shafahi,Y. & Abrishami, E. S., (2005). School trip attraction modeling using neural & fuzzy-neural approaches. Vienna, Austria. https://doi.org/10.1109/ITSC.2005.1520199.

[4] Rassafi, A. A., Rezaei, R., & Hajizamani, M. (2012). Predicting urban trip generation using a fuzzy expert system. Iranian Journal of Fuzzy Systems. Iranian Journal of Fuzzy Systems, 9(3), 127–146. https://dx.doi.org/10.22111/ijfs.2012.151

[5] Berki, Z. & Monigl, J. H. (2017). Trip generation and distribution modelling in Budapest. Transportation Research Procedia, 27, 172–179. https://doi.org/10.1016/j.trpro.2017.12.023.

[6] Bwambale, A., Choudhury, C. F., & Hess, S. (2019). Modelling trip generation using mobile phone data: A latent demographics approach. Journal of Transport Geography, 76, 276-286. https://doi.org/10.1016/j.jtrangeo.2017.08.020.

[7] Lim, K.K. & Srinivasan, S. (2011). Comparative analysis of alternate econometric structures for trip generation models. Transportation Research Record: Journal of the Transportation Research Board, 2254(1), 68–78. https://doi.org/10.3141/2254-08.

[8] McCarthy, G. M. (1969). Multiple-Regression Analysis of Household Trip Generation-A Critique. Highway Research Record, 297, 31-43.

[9] Badoe, D. A. (2007). Forecasting travel demand with alternatively structured models of trip frequency. Transportation Planning and Technology, 30(5), 455–475. https://doi.org/10.1080/03081060701599938.

[10] Jang, T. Y. (2005). Count data models for trip generation. Journal of Transportation Engineering, 131(6), 444–450. https://doi.org/10.1061/(asce)0733-947x(2005)131:6(444).

[11] Ahmadpour, M., Yue, W. L., & Mohammadzaheri, M. (2009). Neuro-Fuzzy Modelling of Workers Trip Production. In 32nd Australasian Transport Research Forum, ATRF 2009.

[12] Avineri, E. (2005). Soft Computing Applications in Traffic and Transport Systems: A Review, Soft Computing: Methodologies and Applications, 32(1), 17–25. https://doi.org/10.1007/3-540-32400-3_2.

[13] Seyedabrishami, S. & Shafahi, Y. (2011). Expert knowledge-guided travel demand estimation: Neuro-fuzzy approach. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 15(1), 13–27. http://dx.doi.org/10.1080/15472450.2011.544576.

[14] Abu-Eisheh, S. & Irshaid, M. (2020). Modelling Trip Generation using Adaptive Neuro-Fuzzy Inference System in Comparison with Traditional Multiple Linear Regression Approach. International Journal of Simulation--Systems, Science & Technology, 21(2), 1–6. https://doi.org/10.5013/IJSSST.a.21.02.17.

[15] Faghri, A. & Aneja, S. (1996). Artificial Neural Network–Based Approach to Modeling Trip Production. Transportation Research Record: Journal of the Transportation Research Board, 1556(1), 131–136. https://doi.org/10.1177/0361198196155600115.

[16] Arliansyah, J. & Hartono, Y. (2015). Trip attraction model using radial basis function neural networks. Procedia Engineering, 125, 445–451. https://doi.org/10.1016/j.proeng.2015.11.117.

[17] Etu , J. E. & Oyedepo, J. O. (2018). Forecasting Trip Generation for High Density Residential Zones of Akure, Nigeria: Comparability of Artificial Neural Network and Regression Models. Journal of Civil Engineering, Science and Technology, 9(2), 76–86. https://doi.org/10.33736/jcest.988.2018.

[18] Tillema, F., van Zuilekom, K. M., & van Maarseveen, M. F. A. M. (2004). Trip generation: comparison of neural networks and regression models. Urban Transport X: Urban transport and the environment in the 21st century, 89, Southampton, UK: WIT Press, 121–130.

[19] Ortuzar, J. de D. & Willumsen, L. G. (2011). Modelling Transport, 4th ed. John Wiley & Sons, Inc. https://doi.org/10.1002/9781119993308

[20] Seabold, S. & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference, Austin, TX, 57, 61. https://doi.org/10.25080/Majora-92bf1922-011

[21] Zilouchan, A. & Jamshidi, A. (2000). Intelligent Control Systems Using Soft Computing Methodologies. Florida: CRC Press. https://doi.org/10.1201/9781420058147

[22] Huang, Y., Lan, Y., Thomson, S., Fang, J. A., Hoffman, W. C., & Lacey, R. E. (2010). Development of soft computing and applications in agricultural and biological engineering. Computers and Electronics in Agriculture, 71, 107–127. https://doi.org/10.1016/j.compag.2010.01.001.

[23] Ibrahim, D. (2016). An overview of soft computing. Procedia Computer Science, 102, 34–38. https://doi.org/10.1016/j.procs.2016.09.366.

[24] Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. Artificial Intelligence in Engineering, 9(3), 143–151 https://doi.org/10.1016/0954-1810(94)00011-S

[25] Chollet, F. (2015). Keras. https://keras.io (accessed Oct. 03, 2021).

[26] Zadeh, L. A. (1965). Fuzzy set. Information and Control, 8(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X.

[27] Singh, H. & Lone, Y. A. (2020). Deep Neuro-Fuzzy Systems with Python: With Case Studies and Applications from the Industry, APRESS. Accessed: Mar. 29, 2021. [Online]. https://doi.org/10.1007/978-1-4842-5361-8

[28] Sivanandam, S. N., Deepa, S. N. & Sumathi, S. (2007). Introduction to Fuzzy Logic using MATLAB, 1. Berlin: Springer. https://doi.org/10.1007/978-3-540-35781-0

[29] Scikit-Fuzzy, (2016). https://pythonhosted.org/scikit-fuzzy/overview.html (accessed Mar. 29, 2021).

[30] Davern, M., Holly, R., Beebe, T. J., & Call, K. T. (2005). The Effect of Income Question Design in Health Surveys on Family Income, Poverty and Eligibility Estimates. HSR Health Service Research, 40(5), 1534–1552. https://doi.org/10.1111/j.1475-6773.2005.00416.x.

[31] Essig, L. & Winter, J. K., (2009). Item Non-Response to Financial Questions in Household Surveys: An Experimental Study of Interviewer and Mode Effects. Fiscal Studies, 30(3-4), 367-390. https://doi.org/10.1111/j.1475-5890.2009.00100.x

[32] Riphahn, R. T. & Serfling, O. (2005). Item non-response on income and wealth questions. Empirical Economics, 30(2), 521-538. https://doi.org/10.1007/s00181-005-0247-7

[33] Kitamura, R., Mokhtarian, P. L., & Laidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. Transportation, 24, 125–158. https://doi.org/10.1023/A:1017959825565

[34] Salau, T. (2015). Public transportation in metropolitan Lagos, Nigeria: analysis of public transport users' socioeconomic characteristics. Urban, Planning and Transport Research, 3(1), 132–139. https://doi.org/10.1080/21650020.2015.1124247.

[35] LAMATA, https://lamata.lagosstate.gov.ng/ (accessed Feb. 18, 2021).

[36] Masaoe, E. N., Del Mistro, R. F., & Makajuma, G. (2011). Travel behaviour in Cape Town, Dar Es Salaam and Nairobi cities. 30th Annual Southern African Transport Conference, 11-14 July 2011, CSIR International Convention Centre, Pretoria, South Africa.

[37] Xu, Y., Belyi,A., Bojic, I., & Ratti, C. (2018). Human mobility and socioeconomic status: Analysis of Singapore and Boston. Computers, Environment and Urban Systems, 72, 51–67. https://doi.org/10.1016/j.compenvurbsys.2018.04.001.

[38] Kockelman, K. M. (1997). Travel Behavior as Function of Accessibility, Land Use Mixing, and Land Use Balance Evidence from San Francisco Bay Area. Transportation Research Record: Journal of the Transportation Research Board, 1607(1), 116–125. https://doi.org/10.3141%2F1607-16.

[39] Stead, D. (2001). Relationships between land use, socioeconomic factors, and travel patterns in Britain. Environment and Planning B: Urban Analytics and City Science, 28, 499–528. https://doi.org/10.1068/b2677.

[40] Barbosa, H., Hazarie, S., Dickinson, B. et al. (2021). Uncovering the socioeconomic facets of human mobility, Science Report, 11, Article 8616. https://doi.org/10.1038/s41598-021-87407-4.

[41] Dargay, J. M., & Hanly, M. (2003, October). The impact of land use patterns on travel behaviour. In European Transport Conference, Strasbourg, France.

[42] Hu, D., Jiang, T., & Yu, X. (2020). The construction of non-convex fuzzy sets and its application. Neurocomputing, 393, 175–186. https://doi.org/10.1016/j.neucom.2018.10.111.

[43] Garibaldi, J., Musikasuwan, S., Ozen, T., & John, R. (2004). A Case Study to Illustrate the Use of Non-Convex Membership Functions for Linguistic Terms. IEEE International Conference on Fuzzy Systems, 3, 1403–1408. https://doi.org/10.1109/FUZZY.2004.1375377.

[44] Jenkins, D. & Quintana-Ascencio, P. (2020). A solution to minimum sample size for regressions. PLoS ONE, 15(2), e0229345. https://doi.org/10.1371/journal.pone.0229345.

APPENDIX

The appendix presents the results generated by adopting missing data without imputation and compares it to the results of imputed income data as discussed in section 2.6. In addition, the performance of the best regression model obtained by deleting records with missing income data is compared to those of the imputed data. The performances of the Neural networks on raw missing data and imputed data are also compared. An excerpt of the data used for the analysis is also presented.

1. Regression Analysis with Imputation

Although the results for analysis with missing income data is not presented, imputation did not offer any improvement over raw data because the coefficients produced by both models are very similar. However, while Logistic regression for imputation did not offer any marked improvement, the overall average trips/household produced from the imputed income data produced exactly the same value as the real data set. This suggests that imputation was somewhat beneficial to the regression model. As shown in Table A1, deleting the records which is probably the simplest means of dealing with missing values led to underperformance on all the performance measures when compared to the model with Logistic regression for imputation of missing income data.

Table A1 Performance of regression model based on method of handling missing values

	Logistic Regression	**Deleting no-response income records**
R^2	0.71	0.68
MAE	0.57	0.58
MSE	1.15	1.19
Average trips	4.51	4.48

2. Neural Network with Imputation

Tables A2 and A3 present the performance measures of the imputed data compared to raw (non-imputed) data for training, test and validation as well as the full data. In the tables, negative values in brackets on R^2 performance means the model with imputed values outperformed the model with raw values and vice versa. In the case of MAE and MSE, negative values mean the raw data outperformed the imputed data by the percentage value in the bracket. The results vary from measure to measure and it was concluded that this variation is simply due to the stochastic nature of neural networks and perhaps the negligible impact of income on trip rates in Lagos. No consistent and significant difference in performance was observed between imputed and raw data with missing responses on income. With regards to the prediction of trips, the difference between actual trips and predicted trips were relatively small. The imputed data for household and income presented in Table 8 produced an average trip of 4.51 which is exactly the same as the raw data. Table A2 presents the results for raw data with missing income values. The average was 4.52 with a difference of 0.01.

Table A2 Neural network model (household size and income)

Performance Measure	**Training data**	**Test data**	**Validation data**	**Full data**
R^2	0.7(-1)	0.76(4)	0.72 (1)	0.71(0)
MAE	0.58 (2)	0.49(-7)	0.58(2)	0.58(2)
MSE	1.22 (5)	0.8 (-33)	1.18 (12)	1.15 (0)
Average trips	4.5	4.52	4.58	4.52

Table A3 Neural Network Performance with three inputs (household size, auto-ownership and income)

Performance Measure	Training data	Test data	Validation data	Full data
R^2	0.69 (-1)	0.76(0)	0.73(0)	0.71(0)
MAE	0.54 (-3)	0.54 (0)	0.53 (0)	0.53 (0)
MSE	1.2 (-4)	0.98 (0)	1.07 (0)	1.15 (0)
Average trips	4.51	4.49	4.45	4.5

In Table A3, ANN for raw data with missing income response values and the two other variables of household size and auto-ownership (called imputed data here), resulted in average trips of 4.5.

3. Data adopted for analysis

An excerpt of the data adopted for analysis is presented in Table A4. It shows the household and trip making characteristics of 21 households out of the 5,816 households adopted.

Table A4 Data used for analysis (household size, auto-ownership and income, trips)

Serial number	Household Size	Number of Cars	Income Level	Number of trips
1	3	2	4	6
2	4	1	6	10
3	5	1	8	14
4	1	0	9	2
5	3	0	1	8
6	3	0	9	6
7	4	0	8	8
8	4	0	9	8
9	3	2	8	6
10	3	1	7	6
11	3	0	9	6
12	4	1	8	8
13	3	0	9	6
14	2	1	9	4
15	2	1	4	4
16	2	2	6	4
17	1	0	9	2
18	3	1	6	6
19	3	0	9	6
20	3	1	8	6
21	3	0	9	6