# MODELING PUBLIC TRANSPORT USERS' TRIP PRODUCTION IN HAWASSA CITY, ETHIOPIA

Girma Gebre[1*] and Emer T. Quezon[2]

[1]Department of Civil Engineering, Hawassa University, Hawassa, Ethiopia
[2]Deparment of Civil Engineering, School of Architecture & Civil Engineering, Addis Ababa Science and Technology University, Addis Ababa City, Ethiopia

*Abstract* — Today, overcrowded public transport demand, resulting in huge costs in an urban area. Similarly, there are a lot of people who use public transport in Hawassa city. This study aimed to develop public transport users' trip production models at the household level. Some socio-economic characteristics and trip detail of the public transport users were collected randomly from the different households through a questionnaire survey. The data gathered was fed into IBM SPSS package version 20 to develop linear regression models. The developed models are associated with trips for purpose and time intervals of trips made. The developed linear regression models, general trips, work trips, educational trips, and trips made before 8:00 AM and after 4:00 PM had good explanatory power. The value of explanatory power comprised of 0.656, 0.722, 0.549, 0.610 and 0.510. These values indicated the explanation power of the socio-economic characteristics on the trips made. It means the daily trips production was significantly affected by the number of working individuals, the different age brackets, cars and motorcycles, and the monthly income per household. The most frequent public transport users' trips production regarding the trip purpose and time are work trips and occurred after 4:00 PM. This scenario represented a good model developed in this study. Hence, it is suggested that Hawassa city's traffic management office use the developed models to predict the future trips demand to provide a proper scheme to avoid congestion during the peak hour of the day.

## 1.0 INTRODUCTION

Transportation is vital to people, goods, and services to transport from a place to another under a desirable condition. Transportation planning is needed to estimate the travel demand behavior that will face in the future. This is because the travel demand behavior of the people is affected by different factors which are making the trips dynamic through time [1]. The most commonly used transportation planning method from the 1950s until today was the four analytical steps method. This method had four sub-models describing trip generation, trip distribution, modal split, and route assignment [2]. But in this study, only trip generation (the first step) was considered to analyze in the study area of Hawassa City, Ethiopia. Trip generation means predicting and determining the volume of trips produced by and attracted to a geographical district [3]. Those volumes of trips depend on various factors. At the disaggregated level (household level), these factors somehow govern the per-person trips made in a household [4]. Trip generation models may be developed for different trip purposes, different periods, and other modes of transport. Despite this, the basic concept remains the same, whatever the type of trip we are studying [5].

Now a day, in Hawassa City, there was massive people who used public transport. This happened due to many reasons, such as the increases in population growth in urban areas of the city because of both the natural increase and migration from especially rural areas and smaller towns. This was combined with somewhat a limited development of public transport facilities and services within the city. It was observed that there was overcrowding and congestion of the public transport demands every year in the city's urban area. Since public transport demands in the future were expected to continue increasing, therefore the need for public transportation planning process appeared to be essential for the city. Different literature was reviewed about

the development of trip generation models with the consideration of general trip generation and trip generation relative to public transport. But there was no specialized study related to quantifying and modeling public transport demand in the case of Hawassa City. So, this paper considered providing a base for the transportation planning process in the study area of the city by modeling the public transport users' trip production and its various categories. This study also identified and discussed the significant socio-economic factors that affect the public transport users' trip production and the most frequent public transport users' trip production regarding the trip purpose and time at the household level.

The most commonly used methodologies for performing the trip generation model are growth factor (expansion), cross-classification (categorical), and linear regression analysis. However, in the growth factor method, it was difficult to determine precisely the independent variables' growth factor to predict the future trip, which causes erroneous predictions. In cross-classification analysis, there was no facility for testing the statistical significance of various explanatory variables; the cell by cell calculation reduces cell values' reliability. Also, the uncertainty increases when there are cells with small samples and significant variances. Hence, this method required large sample sizes that induce much cost and time [6].

The linear regression analysis method was selected from this trip-generation analyzing method because of its clear and straightforward structure and its easiness of application [7]. Thus, this research study sought a regression model analysis that significantly impacted the relationship between those parameters [8]. A trip generation has two parts; these are trip production and trip attraction. Trip production describes the total number of trips produced in a zone without considering those trips' destinations. Whereas trip attraction means the total number of trips that a zone attracts irrespective of the trip origin. The factors that affect trip production are related to the household characteristics because the household or residential area is the main source of trip production [9]. Further, this research focused on the trip production of public transport users. Public transport "consists of regularly scheduled vehicle trips, open to primary paying passengers, with the capacity to carry multiple passengers whose trips may have different origins, destinations, and purposes" [10].

Nevertheless, there was a chance of variation in routes and schedules. This type of transportation had a vital role in cities [11]. The analysis of trip production could be analyzed on the aggregate level at the zone or area level or the disaggregated level at the household or a personal level. Analysis at the level of households was accepted in this research. This is because analyzing at this level, more detail and accurate information can be obtained [12].

## 2.0 RESEARCH METHODS

2.1. Description of the study area

Hawassa City is located in the Southern Nation Nationalities and Peoples Region (SNNPR) of Ethiopia in the Great Rift Valley. The city has a distance of 273 km south of Addis Ababa, 1,125 km north Nairobi, Kenya, and lies on the Trans-African High Way-4: an international road stretched from Cairo (Egypt) to Cape Town (S. Africa).

Nowadays, Hawassa City is one of the fastest-growing cities in terms of building infrastructure and population size. A large part of this rapid population growth was natural and high immigration levels from other areas, especially rural areas [13].

Thus, there is a fast growth of public transport users to travel from home to another place or from another location to home within the city to achieve their daily activities or needs. So, modeling public transport users' trip production was needed to predict the future number of public transport users based on the public transportation planning process.

2.2. Study subject (Zoning system)

The zoning system defines the geographic boundary area or land use of the study area. It divides it into smaller, simple zones to relate the people's travel demand with the socio-economic characteristics. Scientific literature had established its own rules and guidelines to define the zoning system; one of them stated that zoning systems based on homogeneity should contribute to the improvement of the estimates of trip generation and the reduction of intra-zonal trips for various trip purposes [14]. But, it is challenging to give equal consideration to and implement all the rules in a single process of Traffic Analysis Zone (TAZ) design or zoning system because some rules contradict the

other [15]. Besides, there was no clear rule on carrying out a zoning system in an optimal way. The dominating practice is to do it based on experience, mix a certain degree of within zone homogeneity, and the convenience of using administrative borders zone limits [16].

This research considered the study area's land use compatibility with the city's administrative divisions into simple smaller areas (sub-cities) to consider as Traffic Analysis Zones (TAZs). The study area had eight sub-cities from these seven considered internal zones (urban sub-cities), and one was considered an external zone (rural area).

As shown in Table 1, the Hawella-Tula sub-city was considered the external zone, while the rest zones (sub-cities) were deemed internal zones. This is because Hawela Tula sub-city is found in the city's rural area and far away from its urban area. Internal zones mean the study subject zones, while the external zone means the region outside the study subject.

**Table 1** The area of Traffic Analysis Zones (Sub-cities)

| Sub-Cities (TAZs) | Area (Km$^2$) |
|---|---|
| Addis Ketema | 8.8835 |
| Bahil Adarash | 0.7556 |
| Hawela Tula | 122.0580 |
| Hayek Dar | 2.5113 |
| Mehal Ketema | 0.7753 |
| Menaharia | 5.5608 |
| Misrak | 3.1749 |
| Tabor | 14.6375 |

Source: [Hawassa City Municipality 2021]

2.3. Population of the Study

The study considered a population of urban sub-cities of Hawassa City, which is shown in Table 2 [17].

**Table 2** Urban Population by Sub-city

| Sub-Cities (TAZs) | Population size |
|---|---|
| Addis Ketema | 30,296 |
| Bahil Adarash | 25,237 |
| Hayek Dar | 29,539 |
| Mehal Ketema | 24,885 |
| Menaharia | 41,645 |
| Misrak | 39,431 |
| Tabor | 74,057 |
| Total | 265,090 |

Different strategies can be used according to the necessity of the research work [18]. In this study, two sample size determinations were used to get a representative sample size from the target population. These are using Aczel Method and using Statistical Formula Method.

The formula of Aczel Method

$$n = \frac{N}{1+N(\alpha)^2} \tag{1}$$

Where, N: number of population (number of households in the study subject)
  α: assume 95% confidence level, then alpha (α) = 1- 0.95 = 0.05
  n: number of samples

The formula of Statistical Formula Method is:

$$n = \left[(Z^2\,xpxq) + ME^2\right] \Big/ \left[ME^2 + {Z^2\,xpxq}\big/{N}\right]$$

$(2)$

Where, n:  sample size (sample households)

Z: critical standard score (assume normal distribution, Z =1.96)

N: total population size (number of households in the study subject).

ME: margin of error (assume a margin of error, ME = $\pm$ 5%)

p: population proportion (if not sure about the value of p, assumed the population proportion, p = 0.5)

q: (q = 1 − p)

The sample size calculated using the Aczel method was 396, which was higher than the 381 sample size determined from Statistical Formula.

So, the approach of Aczel was considered to take sample size. A larger sample size is recommended to get a higher degree of accuracy. Therefore, 400 household samples were distributed over the TAZs or study subjects, as shown in Table 3.

**Table 3** Expected Number of Households for each TAZ and their Respective Sample Size Required

| TAZs (Sub-cities) | *Population number | #Expected number of Households | % of Housing Size | Sample size |
|---|---|---|---|---|
| Addis Ketema | 30,296 | 3,874 | 11 | 44 |
| Hayek Dar | 29,539 | 3,777 | 11 | 44 |
| Mehal Ketema | 24,885 | 3,182 | 9 | 36 |
| Bahil Adarash | 25,237 | 3,227 | 10 | 40 |
| Misrak | 39,431 | 5,042 | 15 | 60 |
| Menehariya | 41,645 | 5,325 | 16 | 64 |
| Tabor | 74,057 | 9,470 | 28 | 112 |
| Total | 265,090 | 33,899 | 100 | 400 |

The household size in each TAZ was found by dividing the TAZ population size by the average population size per household in the study area (7.82 people/household, got from the sample field survey)

Based on the distributed sample size, the required information was collected from the respective TAZs.  There are two types of required information that were gathered from the determined sample households. The first type was the socio-economic data and the second one was the trip detail of the household's members associated with public transport. Most of the time, two-part of questionnaires was used by most studies the first questionnaire part recoding general household data and the second part recording travel trip detail [19]. Similarly, a home questionnaire survey and an interview are conducted in the city's selected study area using the pre-designed questionnaire to capture the socio-economic characteristics (explanatory variables) and the public transport trip detail characteristics (dependent variables) of the city residents. These variables are listed in Tables 4 and 5 as independent variables and dependent variables, respectively.

**Table 4** The Possible Independent / Explanatory Variables used in Modeling Process

| Symbol | Description/Explanation |
|---|---|
| $X_1$ | Number of persons in the household |
| $X_2$ | Number of males in the household |
| $X_3$ | Number of females in the household |
| $X_4$ | Number of persons receiving education in the household |
| $X_5$ | Number of employed persons in the household |
| $X_6$ | Number of persons under 16  years old in the household |
| $X_7$ | Number of persons between 17 and 30 years old in the household |
| $X_8$ | Number of persons between 31 and 50 years old in the household |
| $X_9$ | Number of persons between 51 and 64 years old in the household |

| | |
|---|---|
| $X_{10}$ | Number of persons above 65 years old in the household |
| $X_{11}$ | Number of cars owned by a household |
| $X_{12}$ | Number of motorcycles owned by a household |
| $X_{13}$ | Monthly household income (in ETB) |

**Table 5** The Dependent Variables used in the Modeling Process

| Symbol | Explanation |
|---|---|
| Y | Number of daily trips made by household |
| $Y_1$ | Number of daily work trips made by household |
| $Y_2$ | Number of daily educational trips made by household |
| $Y_3$ | Number of daily shopping trips made by household |
| $Y_4$ | Number of daily social trips made by household |
| $Y_5$ | Number of daily recreational trips made by household |
| $Y_6$ | Number of daily trips made by household before 8 AM |
| $Y_7$ | Number of daily trips made by household between 8–9 AM |
| $Y_8$ | Number of daily trips made by household between 9 AM-12 PM |
| $Y_9$ | Number of daily trips made by household between 12 PM - 4 PM |
| $Y_{10}$ | Number of daily trips made by the household after 4 PM |

## 2.4. Method of data processing and analysis

Linear regression analysis was one of the methods that widely used for establishing a relationship between several trips produced or attracted (dependent variables) and the causal factors (independent variables) [20]. This research also adopted this method regarding its advantages when compared with the other methods.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n \qquad (3)$$

The parameters $\beta_1$, $\beta_2$, $\beta_3$,.......,$\beta_n$ are the regression coefficients related with $X_1$, $X_2$, $X_3$,........, $X_n$ respectively, and $\alpha$ is a constant term which reflects there was dependent or study variable even if there was no an independent or explanatory variable.

The forward selection approach method was used to build public transport users' trip generation models. It means the approach begins with a regression model that contains only the constant term. Then, in each step, the explanatory variable that results in the greatest change in the $R^2$ value was added to the model. This process was recurring until there is no more variable result in a significant change in $R^2$ (the overall fitness) [20].

In this study, trip production was produced using public transport services to create a different trip production models. These models were divided into three trip production model groups. Public transport users' general trip production model was the first group. Public transport users' trip production models based on trip purpose were the second group. Public transport users' trip production models based on time were the third and the last group.

This study used the most commonly used statistical tests: T-Test, Correlation Matrix, and VIF- Test, R-Squared ($R^2$) -Test, mean square or mean square error (MSE), root mean square error (RMSE), or estimated standard error (Std. error) and F-Test in the process of model selection, and logical senses also used in model selection.

Generally, according to a rule of thumb, if the model had an equal or greater $R^2$ value of 0.75, 0.50, and 0.25, then the model described as a substantial (strong), moderate (good), and weak (poor) model, correspondingly [21].

Using figures, the effect of significant explanatory variables (socio-economic characteristics) on the number of daily public transport users' trip production at the household level (dependent variable) was analyzed one after the other by feeding the collected data into the IBM SPSS software version 20. In addition to this, identifying the most frequent public transport users' trip production per day at the household level regarding trip purpose and time was also performed.

## 3.0 RESULTS AND DISCUSSIONS

3.1. General trip production model for public transport users

The general trip production model was developed based on several equations using regression analysis. The independent variables that had the highest $R^2$ are added into the model through iteration in model building analysis. After all, from the SPSS package version 20, the essential regression results for public transport users' general trip production model are summarized in Table 6. In comparison, other relevant values are also presented in Table 7.

As shown in Table 6, the regression coefficient of the independent variables (number of employed persons in the household, the size of people aged within 17 and 30 years old, and the numeral of people aged within 31 and 50 years old in the household or $X_5$, $X_7$, and $X_8$ respectively) are indicated a positive value. This indicated increasing one, two, or all of the principal to increase the average number of daily public transport users' trips production in the household (Y). On the other hand, the regression coefficients of the size of persons aged under 16 years old in the household ($X_6$), the number of people above 65 years old in the household ($X_{10}$), the number of cars owned by a household ($X_{11}$), the number of motorcycles owned by a household ($X_{12}$) and the monthly income of a household in 000s ETB ($X_{13}$) are negative values. These indicated that one of the explanatory variables ($X_6$, $X_{10}$, $X_{11}$, $X_{12}$, or $X_{13}$) decreases could increase the number of daily public transport users' trips production at a household (Y) vice versa.

The coefficients of the independent variables ($X_5$, $X_6$, $X_7$, $X_8$, $X_{10}$, $X_{11}$, $X_{12}$, and $X_{13}$) had produced t-values of 7.594, 2.686, 12.584, 8.033, 2.625, 4.278, 7.763, and 2.114 in absolute value correspondingly. These values are statistically different from zero at the 95% level of significance. It showed that the alternative hypothesis explained to those mentioned independent variables separately. It impacted the number of daily public transport users' trips production per household was accepted rather than the null hypothesis stated that the mentioned independent variables individually had no impact on the number of daily public transport users' trips production per household. In other words, based on the t-value, the null hypothesis was rejected by accepting the alternative hypothesis.

From the developed general public transport users' trip generation model, each independent variable had a variance inflation factor (VIF) value of less than 10 (see Table 6). The co-linearity among them was less than 0.8. So, within the developed general public transport users' trip generation model, there was no multicollinearity problem among the explanatory variables.

The general public transport users' trip production model had an $R^2$ value of 0.656, as shown in Table 6. This $R^2$ value indicated that the independent variables (those mentioned above) included in the model explained 65.6% of the variation in the general daily public transport users' trips produces per household. Besides, it had MSE of 15.906 and RMSE of 3.988. It means that the model could predict the actual number of trips with the average squared error of 15.906 and the average error of 3.988 trips from the actual regardless of the direction (see Table 7). So, the model is considered to be a good model.

The F-value of the model was 93.084. This value was statically significant (far away from zero) at a 95 % level of significance. So, the alternative hypothesis stated that the independent variables ($X_5$, $X_6$, $X_7$, $X_8$, $X_{10}$, $X_{11}$, $X_{12}$, and $X_{13}$) all together affect the general public transport users' trips in the household (Y) was accepted by rejecting the null hypothesis stated that the included independent variables mutually had no impact on the general public transport users' trips in the household at 95% level of significance.

**Table 6** General Public Transport users' Trip Production Model Regression Results

| Intercept and variables | Coefficient | Standard error | T – value | Significance | VIF |
|---|---|---|---|---|---|
| Intercept | 6.787 | 0.850 | 7.986 | 0.0000 | |
| $X_5$ | 1.654 | 0.218 | 7.594 | 0.0000 | 2.335 |
| $X_6$ | -0.591 | 0.205 | -2.686 | 0.0040 | 1.285 |
| $X_7$ | 1.741 | 0.138 | 12.584 | 0.0000 | 1.409 |
| $X_8$ | 1.396 | 0.174 | 8.033 | 0.0000 | 1.879 |
| $X_{10}$ | -3.578 | 1.363 | -2.625 | 0.0090 | 1.037 |
| $X_{11}$ | -3.887 | 0.908 | -4.278 | 0.0000 | 1.171 |
| $X_{12}$ | -3.154 | 0.406 | -7.763 | 0.0000 | 1.544 |

| | | | | | |
|---|---|---|---|---|---|
| $X_{13}$ | -0.179 | 0.085 | -2.114 | 0.0350 | 1.119 |

| Sample Size (n) | $R^2$ | F – value |
|---|---|---|
| 400 | 0.656 | 93.084 |

So, the developed general daily public transport users' trip production model was:

$$Y = 6.787 + 1.654X_5 - 0.591X_6 + 1.741X_7 + 1.396X_8 - 3.578X_{10} - 3.887X_{11} - 3.154X_{12} - 0.179X_{13} \tag{4}$$

**Table 7** The General Public Transport Users' Trip Generation Model's ANOVA Table

| Source | Sum of squares | Degree of freedom | Mean square | Std. error of the estimate | F – value | Significance |
|---|---|---|---|---|---|---|
| Regression | 11844.578 | 8 | 1480.572 | | 93.084 | 0.000 |
| Residual | 6219.172 | 391 | 15.906 | 3.988 | | |
| Total | 18063.750 | 399 | | | | |

The value of the field surveyed daily public transport users' trips produced in the household was selected randomly from the sample households. The model was said to be verified when it was able to predict. The model verification process was achieved by comparing the field surveyed value and the predicted value from the developed model. When the differences become within the acceptable agreement, then the model was verified. Table 8 shows the verification of the general public transport users' trip production model.

According to the verification test, an acceptable agreement value existed between the predicted general daily public transport users' trips per household and the actual field surveyed total daily public transport users' trips per household, as seen from Table 8. Hence, the developed general public transport users' trip production model was a good model.

**Table 8** Model verification for general public transport users' trip production model

| The selected sample households' code | The actual Y | The predicted Y | Variation |
|---|---|---|---|
| 1 | 10 | 11 | -1 |
| 3 | 14 | 11.32 | 2.68 |
| 23 | 16 | 14.75 | 1.25 |
| 26 | 27 | 24.74 | 2.26 |
| 34 | 4 | 7.64 | -3.64 |
| 47 | 26 | 24.7 | 1.3 |
| 49 | 18 | 15.84 | 2.16 |
| 83 | 9 | 14.4 | -5.4 |
| 94 | 13 | 15.37 | -2.37 |
| 125 | 17 | 13.5 | 3.5 |
| 131 | 10 | 13.74 | -3.74 |
| 150 | 18 | 14.41 | 3.59 |
| 164 | 28 | 24.01 | 3.99 |
| 166 | 8 | 10.18 | -2.18 |
| 171 | 8 | 11.5 | -3.5 |
| 175 | 14 | 17.98 | -3.98 |
| 198 | 27 | 25.12 | 1.88 |
| 203 | 24 | 20.98 | 3.02 |
| 207 | 13 | 13.49 | -0.49 |
| 216 | 9 | 8.69 | 0.31 |
| 255 | 23 | 21.63 | 1.37 |
| 264 | 16 | 13.95 | 2.05 |
| 274 | 8 | 10.03 | -2.03 |

| | | | |
|---|---|---|---|
| 295 | 24 | 19.75 | 4.25 |
| 313 | 13 | 16.79 | -3.79 |
| 342 | 19 | 16.87 | 2.13 |
| 365 | 11 | 13.09 | -2.09 |
| 368 | 27 | 22.7 | 4.3 |
| 373 | 0 | 4.77 | -4.77 |
| 400 | 4 | 8.33 | -4.33 |
| Total | 458 | 461.27 | -3.27 |

## 3.2. Public transport users' trip production models regarding the trip purpose

In this research, the five developed public transport users' trip production models regarding trip purposes are represented by work, education, shopping, social, and recreational models. In the general public transport users' trip production model, after feeding the collected data of the explained and the explanatory variables into the SPSS package, the following trip purpose models are developed through linear regression analysis. These are:

a. $Y_{Work} = -0.451 + 1.215X_5 + 0.385X_7 + 0.726X_8 - 1.559X_{10} - 1.288X_{11} - 1.312X_{12}$ (5)

Where: $R^2 = 0.722$, MSE = 2.37 and RMSE = 1.54

This $R^2$ value indicated that the independent variables ($X_5$, $X_7$, $X_8$, $X_{10}$, $X_{11}$, and $X_{12}$) included in the model explained 72.2% of the variation in work trips produced per household. The value of the model's MSE and RMSE implicated the model could predict the number of work trips with the squared error of 2.37 and the deviation of 1.54 from the observed value since the model was good.

b. $Y_{education} = 1.447 + 0.096X_4 - 0.463X_5 + 0.648X_7$ (6)

Where: $R^2 = 0.549$, MSE = 1.32 and RMSE = 1.15

The $R^2$ of the model (54.9%) is greater than 50%; this indicated the independent variables ($X_4$, $X_5$, and $X_7$) elaborated more than 50% of the variation in education trips produced per household. And also, the MSE and RMSE values of the model are 1.32 and 1.15, respectively. So, the model was considered a good model.

c. $Y_{shopping} = 1.090 + 0.241X_5 + 0.160X_7 + 0.333X_8 - 1.057X_{10} - 1.097X_{11} - 0.623X_{12}$ (7)

Where: $R^2 = 0.339$, MSE = 1.43 and RMSE = 1.19

As seen, the value of $R^2$ is 0.339, directed that the independent variables included in the model explained 33.9 % of the variation in shopping trips produced per household. The value of MSE and RMSE indicated that the model could predict the number of shopping trips with the average squared error of 1.43 and the average deviation of 1.19 shopping trips from the actual value without considering direction correspondingly.

d. $Y_{social} = 2.490 + 0.531X_5 + 0.342X_7 + 0.221X_8 - 0.997X_{11} - 0.413X_{12} - 0.178X_{13}$ (8)

Where: $R^2 = 0.328$, MSE = 3.64 and RMSE = 1.91

The $R^2$ value of the model was 32.8% implied that the independent variables included in the model explained only 32.8% of the variation in social trips produced per household. So, this model predicted the number of social trips made by the resident of Hawassa city per household with the average squared error (MSE) of 3.64 and the average error (RMSE) of 1.91 social trips per household.

e. $Y_{recreation} = 2.478 + 0.085X_4 + 0.199X_5 - 0.536X_6 - 0.828X_{12} - 0.080X_{13}$ (9)

Where: $R^2 = 0.198$, MSE = 3.03 and RMSE = 1.74

This model had an $R^2$ value of 0.198 showed, the variables ($X_4$, $X_5$, $X_6$, $X_{12}$, and $X_{13}$) elaborated only 19.8% of the variation occurred in the recreation trips produced per household. The value of MSE and RMSE concerned how the developed recreation trip model's prediction closes the actual observation.

### 3.3. Trip production models for public transport users' trips over time

In the same manner, after feeding the collected data of the explained and the explanatory variables into the SPSS package, trip time models are developed for public transport users' trips made before 8:00 AM, between 8:00 AM to 9:00 AM, 9:00 AM to 12:00 PM, 12:00 AM to 4:00 PM, and after 4:00 PM through linear regression analysis. These are:

a. $Y_{before\ 8\ am} = -5.239 + 0.437X_4 + 2.470X_5 + 0.928X_7 + 0.408X_8 - 1.650X_{12}$ (10)

Where: $R^2 = 0.610$, MSE $= 14.51$ and RMSE $= 3.81$

The $R^2$ value indicated that 61.0% of the variation in the number of trips made before 8:00 AM was elaborated by the involved variables ($X_4$, $X_5$, $X_7$, $X_8$, and $X_{12}$). Those variables did not elaborate on the remaining variation percent (39.0%). The value of MSE and RMSE directed that the model could predict the number of trips made before 8:00 AM per household in Hawasss city with the average squared error of 14.51 and average deviation of 3.81 trips from the actual number of trips, respectively, since the model is said to be a good model.

b. $Y_{8\ am\ to\ 9\ am} = 0.671 + 0.040X_4 + 0.235X_5 - 0.173X_6 + 0.087X_7 + 0.122X_8 - 0.868X_{10} - 0.462X_{11} - 0.067X_{13}$ (11)

Where: $R^2 = 0.382$, MSE $= 0.86$ and RMSE $= 0.93$

The model had an $R^2$ value of 0.382. So, from the total variation that happened in the number of trips produced between 8:00 AM and 9:00 AM per household, only 38.2% variation was explained by the independent variables involved in the model. This model estimated the number of trips produced within 8:00 AM and 9:00 AM per household in the city with an average squared error of 0.86 and an average deviation of 0.93 trips from the actual observation.

c. $Y_{9\ am\ to\ 12\ pm} = 0.016 - 0.163X_5 + 0.224X_7 + 0.320X_8 + 0.145X_9$ (12)

Where: $R^2 = 0.220$, MSE $= 0.56$ and RMSE $= 0.75$

The variables ($X_5$, $X_7$, $X_8$, and $X_9$) involved in this model elaborated only 22.0% of the total variation that occurred in trips produced from 9:00 AM to 12:00 PM per household ($Y_{9\ am\ to\ 12}$ PM). With 0.56 average squared errors and 0.75 average error trips, the model could project the number of trips produced between 9:00 AM and 12:00 PM according to the results of MSE and RMSE, respectively.

d. $Y_{12\ pm\ to\ 4\ pm} = -0.259 + 0.241X_1 + 0.156X_5 + 0.104X_7 - 0.179X_9 - 0.524X_{12}$ (13)

Where: $R^2 = 0.279$, MSE $= 1.10$ and RMSE $= 1.05$

The model's explanatory (independent) variables explained that 27.9% of the variation occurred in the number of trips produced between 12:00 PM and 4:00 PM based on its $R^2$ value (0.279). The developed model for trips made between 12:00 PM and 4:00 PM should project the trips produced with an average squared error of 1.10 and an average deviation of 1.05 trips from the actual number of trips without considering the direction.

e. $Y_{after\ 4\ pm} = 2.561 + 0.284X_4 + 0.826X_5 + 0.493X_8 - 0.346X_9 - 2.940X_{11} - 0.950X_{12}$ (14)

Where: $R^2 = 0.510$, MSE $= 4.96$ and RMSE $= 2.23$

According to the value of $R^2$, 51.0% of the variation that happened in the trips made after 4:00 PM per household could be explained by the explanatory variables included in the model. Those MSE and RMSE values indicated that the established model for trips made after 4:00 PM could estimate the number of trips that occurred after 4:00 PM with a squared error of 4.96 and error of 2.23 trips. Therefore, the model is good.

**Note:** In the general public transport users' trip production model, the statistical tests for T-Test, VIF- Test, and F-Test are used in model selection for public transport users' trip production models regarding trip purpose time. The verification of the models was also checked. The results of those statistical tests and the verification of those public transport users' trip production models associated with trip purpose and time became good.

3.4. Explanatory variables that significantly influence public transport users' trip production at the household level

The significant explanatory variables that affect the daily public transport users' trips production (Y) at the household level are elaborated on each one using the Figures below and discussion**.**

As seen in Figure 1, if the number of employed persons in the household indicated one, then the number of public transport users' trip production became 9.2418 trips. Likewise, if depicted two, the number of public transport users' trip production became 11.7804 trips. When the number of employed persons increased by one, then the number of public transport users' trip production increased by 2.5386 trips at the household level. So, regardless of the magnitude of the result, it is still logical, as when the number of employed persons in the household increased, they need more public transport trips to achieve their daily activities. Besides this, the number of employed persons in the household separately explained 27.93% of the variation in the number of public transport users' trips production when ignoring the other explanatory variables.
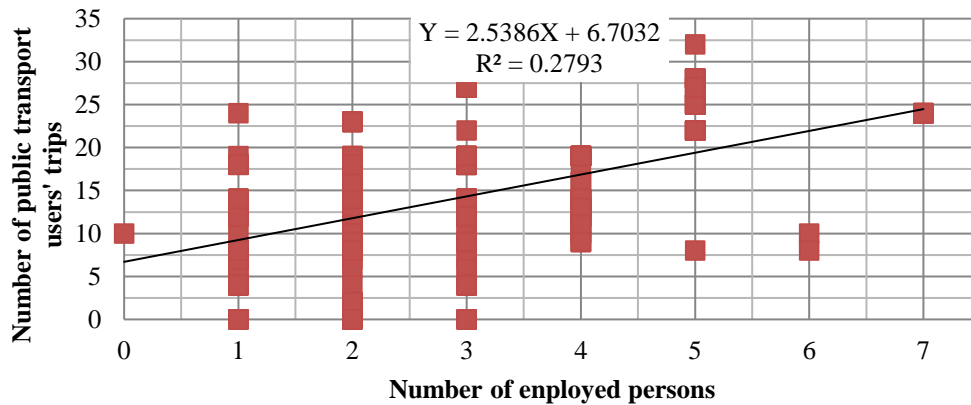


**Figure 1** Daily household public transport users' trips and the number of employed persons in the household

Figure 2 showed the number of public transport users' trip production at the household level. The result indicated 13.4996 trips when the number of people aged under16 years old in the family was one. On the other hand, if the number of people aged less than 16 years old was two in the family, it depicted 11.0262 number of trips would have occurred. Since, with 16.53% explanation power, the number of public transport users' trips production decreased by 2.4734 when the number of people aged less than 16 years increased by one, keeping the other independent variables constant. Ignoring the number, this made sense because the people aged less than 16 years old was expected to be school students in Hawassa city. In most areas of Hawassa city, the places of schools are nearby the students' houses. So, no need to make trips using public transport mode (PTM).
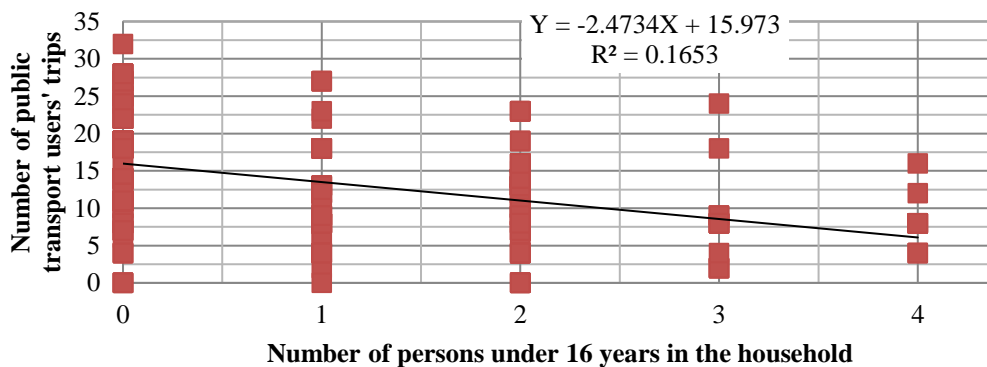


**Figure 2** Daily household public transport users' trips and the number of persons under 16 years in a household

On the other hand, at the time, the number of persons between 17 and 30 years old in the household was one, lead to the number of the daily public transport users' trip production at the household level to be 11.7807 trips. Similarly, the number of trips was 13.8221 when the number of persons between 17 and 30 years old in the family was two (see Figure 3). This relation indicated an $R^2$ value of 0.270, which means the explanatory variable (number of persons within 17 and 30 years old in the household) had 27% of explanatory power over the variation on the dependent variable (number of public transport users' trips production) when keeping the effect of the other explanatory variables zero.
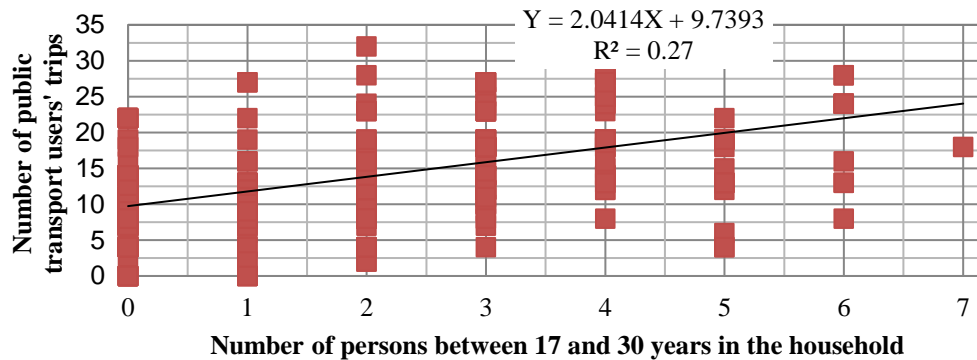


**Figure 3** Daily household public transport users' trips and the number of persons between 17 and 30 years old in the household

Likewise, the elaboration power of 18.08%, the number of public transport users' trip production could be 10.313, 12.1305 and 13.948 trips if the number of people aged between 31 and 50 years in the household is considered to be 0, 1 and 2, respectively (see Figure 4). Since the increasing number of people aged between 31 and 50 years in the household by one principal to increase the number of trips by 1.8175, the effect of the other explanatory variables will become null or void. Typically, in Hawassa city, the people aged under this age group are considered active, productive age groups. They are expected to make more trips to achieve their day-to-day activities next to the people between 17 and 30 years old.
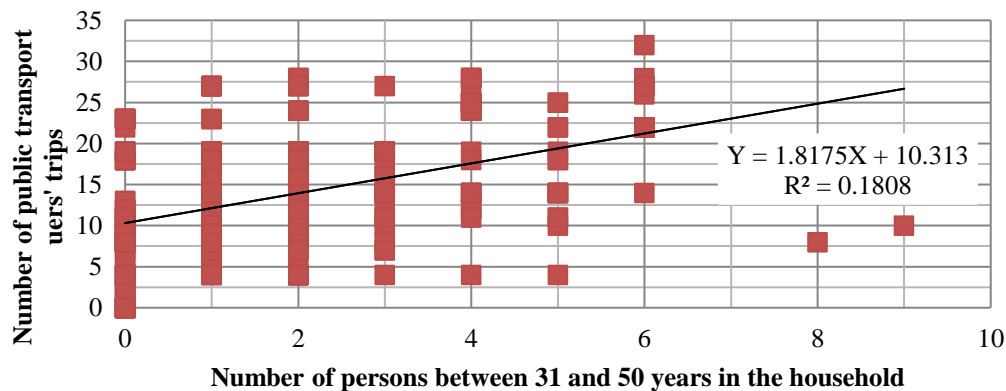


**Figure 4** Daily household public transport users' trips and the number of persons between 31 and 50 years old in the household

As seen in Figure 5, when the number of people aged above 65 years in the household increased from 1 to 2, then the number of public transport users' trips decreased from 9.8978 to 6.2566 trips. This showed an inverse relationship between the size of persons aged above 65 years in the family and the quantity of daily public transport users' trip production in the household. Generally, this is due to the people aged above 65 years old are non-active productive age group or dormant on a fixed place of the city. They expected to make fewer trips to do their daily activities by traveling. Therefore, when the number of persons above 65 years old in the household increased, daily public transport users' trip production at the household level decreased, which is logical and reasonable.
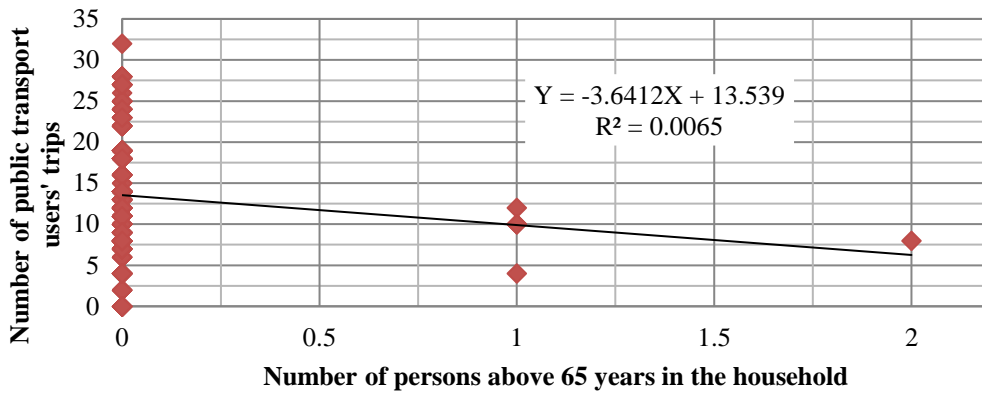
**Figure 5** Daily household public transport users' trips and the number of persons above 65 years old in the household

When a household possessed one car, then 5.3337 public transport users' trips could produce from that household. In the same fashion, 13.995 numbers of trips were expected when the household had no car. So, an increase of a car in the household principal decreased the number of trips using public transport users' by 8.6613 trips with 9.37% explanation power (see Figure 6). As expected, the number of the daily public transport users' trips produced per household and the number of cars owned by that household in the city had a negative relation. Since in general, if the household had their cars, then the household's member made trips more probably using their vehicles rather than using public transport mode (PTM).
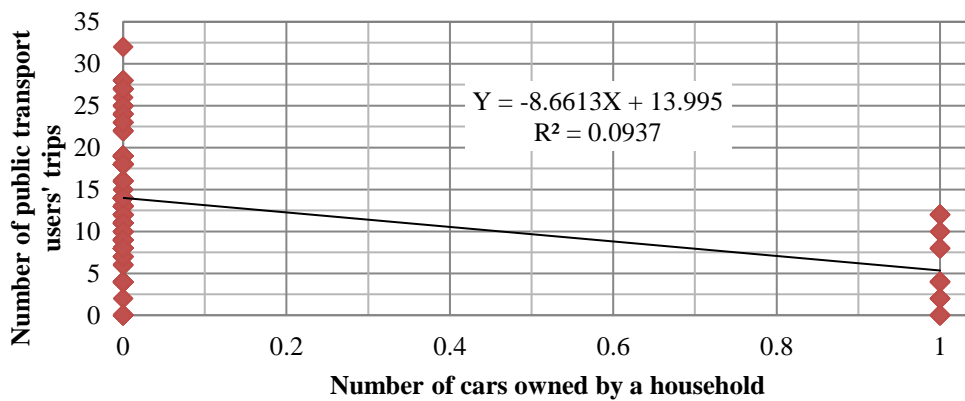


**Figure 6** Number of daily public transport users' trips and the number of cars owned by a household

Figure 7, on the other hand, the number of daily public transport users' trip production in a household (dependent variable, Y) and the number of motorcycles owned by that household had a negative linear relationship. If the household had no motorcycle, then 13.5 public transport users' trips are expected to be made because of the other factors. And when one motorcycle was available in the household, then the number of public transport users' trips production became 13.4304. So, the increasing and decreasing of the availability of motorcycles in the family lead to the decreasing and increasing of the number of public transport users' trips production by 0.0696 amounts correspondingly. This happened because normally, if a household had motorcycles, then the people in the household expected to make more trips using their own motorcycles rather than using public transport service. This is reasonable and significant.
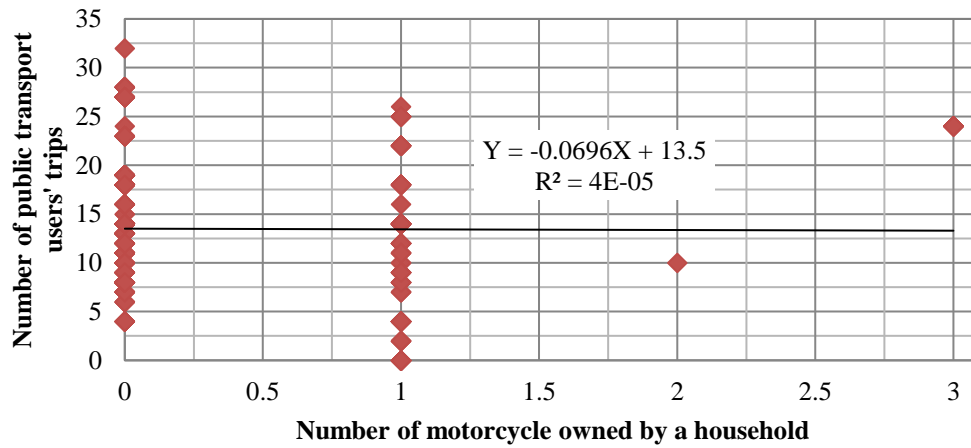
**Figure 7** Number of daily public transport users' trips and the number of motorcycles owned by a household

As shown in Figure 8, the household's monthly income affected the public transport users' trips production with 4.95% explanation power. For instance, the production of the number of public transport users' trips changed from 14.82 to 13.62 when the family's monthly income increased from ETB 5,500 to ETB 7,500. This linear relationship between the household's monthly income and the number of daily public transport users' trip production at the household level indicated a negative. Based on the city's observation, the household's monthly income increased the chance of buying their vehicles for transportation. So, if a household had owned cars, then the household members expected to make more trips using their cars rather than public transport service. However, this is not always the case. There may be a probability of direct (positive) linear relation between the monthly household income and the number of daily transport users' trip production at the household level. This research depicted that they had an inverse relationship when disregarding the effect of the other independent variables.
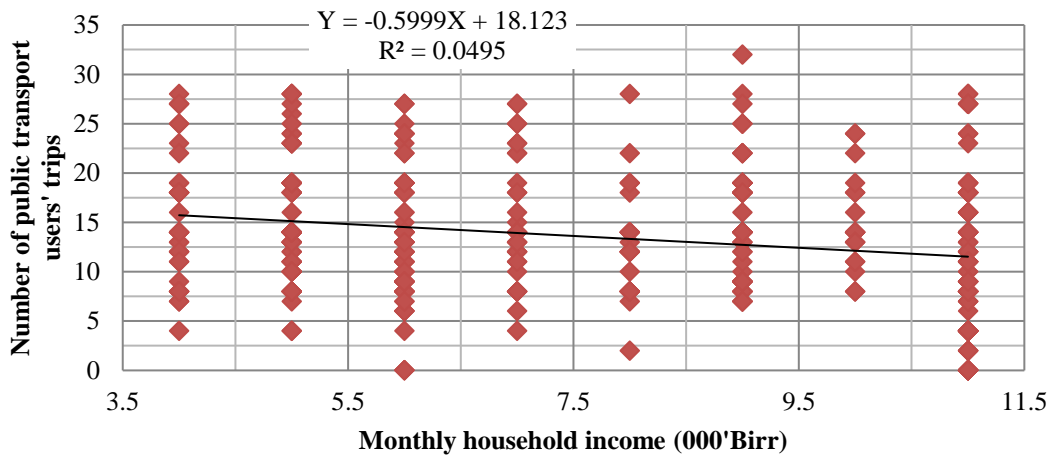


**Figure 8** Number of daily public transport users' trips and monthly income per household

3.5. Frequency of public transport users' trip produced per day regarding the trip purpose and time-based at the household level

Under this sub-heading, it discussed which trip purpose was the most frequent public transport users' trip, and at which time interval the most frequent trips are made?

Classification of trips according to trip purpose is vital for transportation planning. Since for un-alike purposes, people can make trips. This study found out that the highest public transport users' trip frequency according to trip purpose was a work trip, which had 31% of the general public transport users' trips. The subsequent highest public transport users' trip frequency regarding trip purpose was on a social trip. It covered 25% of the general public transport users' trips (see Figure 9).
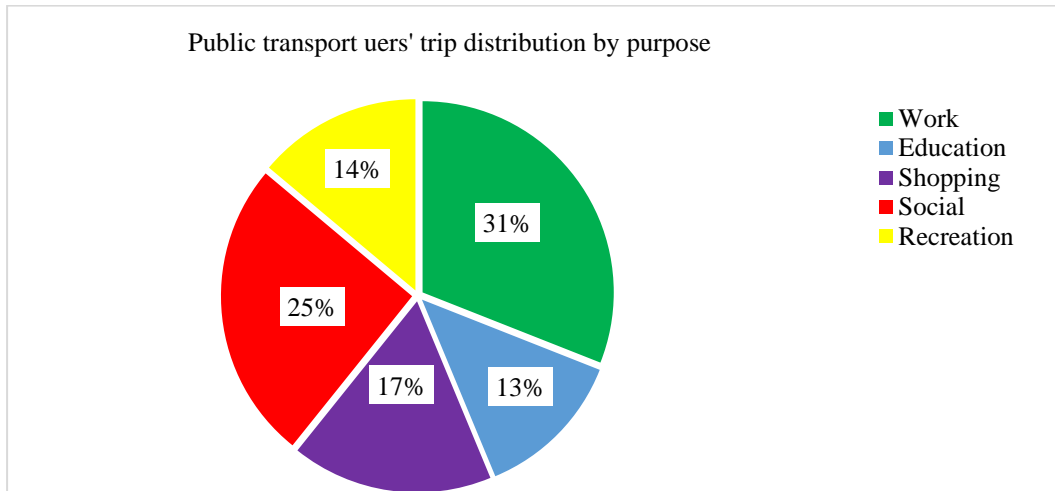
**Figure 9** Daily household public transport users' trip distribution according to trip purpose

Within a day, at a different time interval, a different number of trips occurred. While a typical working day, the peak trip rate period usually occurred during the morning and the late afternoon. As shown in Figure 10, the morning peak period or hours of the public transport users' trip production occurred before 8:00 AM, so, for this peak period, the model was developed ($Y_{before\ 8}$ AM) that linked the number of trips made in this time interval with corresponding socio-economic characteristics of the household (see model or equation 10). The number of trips made in this period comprised 34% of the general public transport users' trip production. Likewise, the afternoon peak hours of the public transport users' trip production occurred after 4:00 PM. It showed 43% of the general public transport users' trip production per day, and it was also the peak period of the day. The model was established for trips in the late afternoon peak period or after 4:00 PM (see model or equation 14). For off-peak period trips, the models were also developed to relate the number of trips made in the off-peak period and the socio-economic characteristics of the residents in Hawassa city at the household level (see from equation or model 11 to 13).
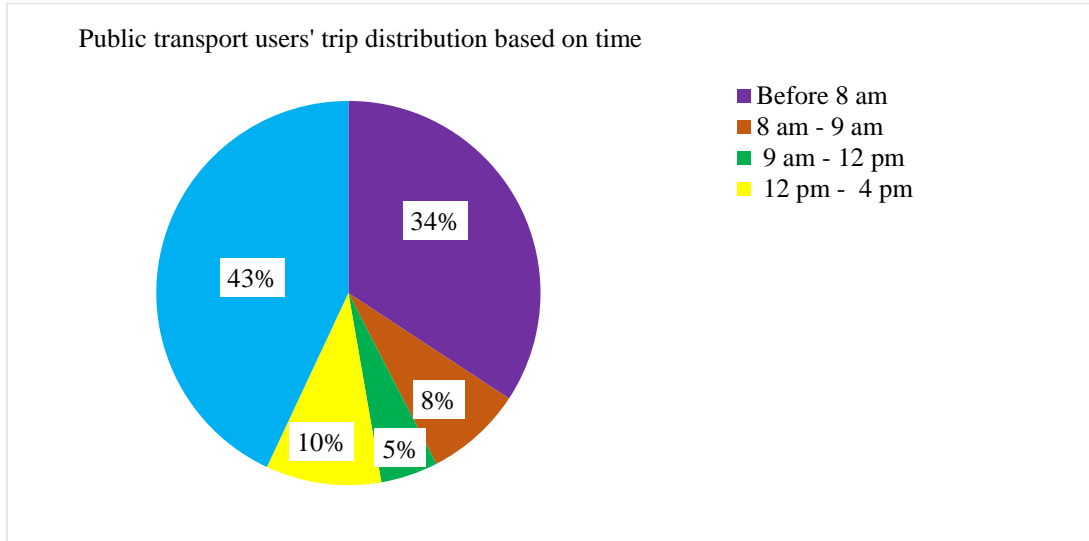


**Figure 10** Daily household public transport users' trip distribution based on time

## 4.0 CONCLUSION

The general public transport users' trip production model indicated an $R^2$ value of 0.656 in addition to MSE of 15.906 and RMSE of 3.988. It specified that the independent variables used in this model had an explanatory power of 65.6% of the total variation on the daily public transport users' trips production at the household level with 3.988 trips deviation from the actual observation. Similarly, based on trip purpose, public transport users' work trip production model and education trip production model had $R^2$, MSE, and RMSE values of 0.722, 2.37, 1.54, and 0.549, 1.32 1.15, respectively. Besides, the trips made based on time, the $R^2$, MSE, and RMSE value of the public transport users' trip production model for trips made before 8:00 AM and for trips made after 4:00 PM indicated

0.610, 14.51 and 3.81, and 0.51, 4.96 and 2.23. The model developed are considered as a good model due to their $R^2$ value.

On the other hand, the significant explanatory variables that affect the number of daily public transport users' trips production per household are represented; such as the number of working people in the household, the number of persons under 16 years old in the household, the number of people within 17 and 30 years old in the family, 31 and 50 years old in the family, above 65 years old in the household, the number of cars owned by a household, motorcycles owned by a household and the monthly household income. Therefore, the trip purpose revealed that the most frequent public transport users' trip production was a work trip. Similarly, the most frequent public transport users' trip production was based on time, a trip made before 8:00 AM and after 4:00 PM. The trips made after 4:00 PM were also observed during the peak period of a typical working day, affecting the travel time of the commuters due to traffic congestion.

## Acknowledgment

## References

[1]     Oyedepo and Makinde (2009). Regression model of household trip generation of Ado Ekiti township in Nigeria. European Journal of Scientific Research, 28 (1), 131-140. https://www.researchgate.net/publication/292567680

[2]     Krasić, D.and Novačko, L. (2015). The impact of public transport network accessibility on trip generation model. Promet – Traffic & Transportation, 27(2), 165-172. https://doi.org/10.7307/ptt.v27i2.1591

[3]     Transport Research Board (TRB), (2011) "Trip generation definition," Washington D.C. Available on: http://www.TRB.org.

[4]     NPTEL (2007). Trip generation," in Introduction to Transportation Engineering, Mumbai, pp. 7.1-7.5. http://nptel.ac.in/course.html

[5]     Eric Ivan Pas (1974). An Analysis of Residential Trip Generation in Cape Town, MSc thesis in Faculty of Engineering, Department of Civil Engineering, at the University of Cape Town, Cape Town. http://hdl.handle.net/11427/17777

[6]     Ortu´zar J. and Willumsen L. (1996). Modeling transport. 4th Ed., Chichester, West Sussex, UK, Wiley.

[7]     Aloc D. S. and Amar J. A. C. (2013). Trip generation modeling of Lipa city. Seminar and research methods in civil engineering research program, University of Philippines Diliman. doi: 10.13140/2.1.2171.7126.

[8]     Assen, S. M. and Quezon, E. T. (2019). Model-based urban road network performance measurement using travel time reliability: A Case Study of Addis Ababa City, Ethiopia. American Journal of Civil Engineering and Architecture, 7(5), 202-207. http://dx.doi.org/10.12691/ajcea-7-5-2.

[9]     Arasan, H.V. (2012) Lectures, Department of Civil Engineering, Indian Institute of Technology Madras.

[10]    Walker, J. (2012). Human Transit: How clearer thinking about public transit can enrich our communities and our lives. Island Press.

[11]    Motta, R.A., Marques da Silva, P. C. & De Sequeira Santos, M. P. (2013). Crisis of public transport by bus in developing countries: A Case Study from Brazil. International Journal on Sustainable Development Plan, 8(3), 348–361. doi:10.2495/SDP-V8-N3-3-348-361.

[12]    Dodeen, A. M. Y. (2014). Developing trip generation models utilizing linear regression analysis: Jericho City as a Case Study. MSc. thesis Department of road and transport engineering, Faculty of Graduate Studies, An-Najah NationalUniversity, Nablus, Palestine.
        https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.semanticscholar.org/paper/Developing-Trip-Generation-Models-Utilizing-Linear

[13]    Shiferaw, Y., Legesse, D., and Denato, M. (2015). Modal shift towards urban public transport Bajaj Taxi in Hawassa. Evaluating the role and factors that impede or facilitate. Proceeding of the 36th annual research review workshop office of vice president for research and technology transfer proceeding, Hawassa University, Hawassa, Ethiopia, pp. 101-112. https://www.academia.edu/ 42013957/

[14]    Khatib, Z., Ou, Y. and Chang, K. (1998) "GIS and Transportation Planning in Medium-Proceedings of the 6th National Conference on Transportation Planning for Small and Sized Communities", Idaho.

[15]    Ding, C., Choi, K. and Kim, T. J. (1993) "GIS-Based Traffic Analysis Zone Design," In Proceedings of the 3rd International Conference on Computers in Urban Planning and Urban Management, Atlanta.

[16]    Miguel Martínez, L., Viegas, J. M. & Silva, E. A. (2007). Zoning decisions in transport planning and their impact on the precision of results. Transportation Research Board, 1994(1), 58–65. https://doi.org/10.3141/1994-08

[17]    Socio-Economic Profile, SEP of Hawassa City (2018). Available on: Web site: www.hawassafinance.com.

[18]    Gauhati University, School of Applied Mathematics and Statistical Sciences (2012) "Determination of appropriate Sample Size" published in the Bulletin of the Gauhati University, Vol. 12.

[19]    Khattak, A.J., Stone, J., Letchworth, W.E., Rasmussen B. K. and Schroeder, B.J. (2004), "Traditional Neighborhood Development Trip Generation Study," Carolina Transportation Program, April. https://www.yumpu.com/s/1dnIdw9ClGoTbn8Y

[20]    Walpole, R. E., Myers, R. H. Myers, S. L. and Ye, K. (2012). Probability & statistics for engineers and scientists, ninth edition, pp. 379 – 394.   http://ndl.ethernet.edu.et/bitstream/123456789/78643/1/Reference.pdf

[21]    Henseler, J., Ringle, C., and Sinkovics, R. (2009). The use of partial least squares path modeling in international marketing. Advances in international marketing (AIM), 20, 277-320. https://doi.org/10.1108/S1474-7979(2009)0000020014