

FORECASTING TRIP ATTRACTION BASED ON COMMERCIAL LAND USE CHARACTERISTICS

Pretina George¹, Gymmy Joseph Kattor²

¹ M Tech Scholar, ² Professor, Civil Department, Rajiv Gandhi Institute of Technology, Kerala, India,
Ipretinageorge@gmail.com, gjkattor@yahoo.co.in

Abstract

Trip Attraction is fundamental in planning of transportation facilities. It is important to the traffic engineer and urban planner in considering the impact of new development such as office complex, shopping center and residential development. This paper focuses its study on the formulation of a trip attraction model using multiple regression technique for the commercial land use in medium sized towns of Kerala. Questionnaire surveys are conducted for the eight selected commercial nodes from the three medium sized towns Tirur, Perinthalmanna, and Ponnani in Kerala. Socioeconomic surveys are conducted for the selected towns for obtaining the origin-destination data. The correlation and regression analysis are performed based on these surveyed data. The study showed that the multiple regression model with the three independent variables namely the number of commercial establishments, percentage of shops in the commercial node and percentage of banks in the commercial node with the R^2 and Adjusted R^2 value of 0.991 and 0.9844 respectively gives the better estimate of trip attraction. After model formulation, the model accuracy was also tested by examining the assumptions of multiple regression. The model is found to be accurate during the assumption examination process. This model would be very useful for estimating the trips attracted to a new or existing commercial center in any medium sized towns in Kerala, and thus aid to assess the traffic impact of the commercial center on the geometric design of roadways in the surrounding area.

Index Terms: correlation, multicollinearity, multiple regressions, trip attraction

-----***-----

1. INTRODUCTION

Transportation planning process relies on travel demand forecasting which is essential for the design of transportation facilities and services, and also for planning, investment, and policy development. In this process, planners develop information about the impacts of implementing alternative courses of action involving transportation services, such as new highways, bus route changes, or parking restrictions. Detailed data on current travel pattern and traffic volumes are needed for developing travel forecasting/prediction models.

One of the most common approaches used by Transportation Engineers to forecast transportation demand is the four-step process. The four-step process is divided into four phases: trip generation, trip distribution, modal split, and traffic assignment. The first step of travel demand forecasting is the trip generation which includes trip production and trip attraction. Trip generation aims at predicting the total number of trips generated and attracted to each zone of the study area. Trip generation serves as an input to modeling transportation requirements and traffic flows influenced by the development. For trip attraction models the multiple regression technique was preferred based on the variables used [1]. Equations based on regression analysis represent 'best fits' through the data points [2].

Several studies in western countries have already been done that can be related to trip generation modeling.

Fillone and Tecson developed models using multiple regressions which showed that floor area was related to trip attraction. Thirty condominiums within Metro Manila were randomly selected and they conducted a questionnaire survey to collect the data related to the condominiums and for modeling purpose. The data gathered include, available residential floor area, number of commercial establishments, parking slots, commercial floor area, occupancy rates, number of floors, years in operation, maintenance fee [3].

Waloejo et al determined the characteristics of trip attraction at the market land use and its influencing factors and he also formulated a model of trip attraction and a model of trip delay analysis at Dinoyo market area. For model formulation correlation and regression analysis are used. Other supporting analyses are ANOVA, parking systems, road capacity and analysis of the level of services [4].

Uddin et al estimated the trip attraction rates of shopping centers at Dhanmondi Area in Dhaka City using trip rate analysis method instead of multiple regression due to the small sample size. They related the trip attraction of the shopping center as a function of the physical features of the shopping

center, e.g. total parking spaces, gross floor area, and the number of shops, employees in the shopping center [5].

Baumeler et al, in their conference paper aim to identify crucial spatial characteristics of neighbourhoods and of shopping destinations, which has an impact on the distance, destination and modal split of shopping trips. They were taken the data for this study from the Swiss national travel survey, which contains individual information on the socioeconomic status and the travel behaviour of approximately 30,000 people. The important independent variables they selected for their study includes the number of different stores like, supermarket, consumer market, small and large stores, sum of stores in the neighbourhood [6].

Yu and Lawrence formulated a trip production model and a trip attraction model. For the formulation of models, households and trips information were retrieved from the Australian Bureau of Statistics Census 2006, the Perth Strategic Transport Evaluation Model (presently calibrated to 2001) and the Perth and Regions Travel Surveys 2001 - 2005, since there were no suitable and available model output and survey data source in Albany [7].

Innes et al conducted a study to determine the major factors affecting shopper's destination choice. They included the factors, store hours of operation, quality of goods offered, availability of parking, price of goods, accessibility to the shopping area, selection of goods offered, and protection from environmental influences for the modeling [8].

Trip attraction is important to the traffic engineer and urban planner in considering the impact of new development such as office complex, shopping center and residential development.

Even though a number of trip generation models are developed using multiple regression analysis to develop the prediction equations for the trip generated by various types of land use in different countries, it is not possible to implement these models in all places as the models are developed based on the local conditions of the study area. Commercial areas in a city can take up about 5% of a city's land. Majority of the trips are attracted to this commercial area.

This study develop a suitable trip attraction model using multiple regression analysis, to forecast the future trips that will be attracted to a commercial node of certain characteristics for a medium sized town in Kerala. So, this model will be beneficial in finding the impact of any development in the commercial nodes on the traffic and thereby, proper planning of transportation facilities and services can be effectively done by the urban planners and traffic engineers.

2. METHODOLOGY

This study used a combination of analytical and descriptive methods. As a first step in conducting a trip generation model analysis, commercial nodes are selected from the medium sized towns in Kerala using the land use data of the corresponding selected medium sized towns. Then the selected towns are divided into appropriate traffic analysis zones in order to simplify the travel pattern. Origin-Destination (O-D) travel data is collected from the socio economic surveys that are conducted for the selected towns so that the O-D matrix can be prepared which shows the trips that are attracted towards each commercial zone from different traffic zones.

A number of relevant independent variables that influence the trip attraction towards the commercial nodes are identified. And the data related to the identified independent variables from each commercial node are collected by conducting questionnaire survey.

After data collection, the characteristics of each of the selected commercial nodes are analyzed. The correlation and regression analysis are done using the identified variables and the best regression model is formulated for predicting the trips attracted towards the commercial nodes in medium sized towns in Kerala. Then a check for the validation of the formulated multiple regression model is done to verify whether the formulated model is valid or not.

3. STUDY AREA DEFINITION

The selected medium sized towns for this study include Tirur, Perinthalmanna, and Ponnani from Malappuram district in Kerala. The study area is selected in such a way as to satisfy the criteria that the commercial nodes should be taken from a minimum of three medium sized towns in Kerala having the latest origin destination data. The map showing the position of towns in Malappuram district is shown in fig 1.



Fig -1: Location of selected medium sized towns

The commercial nodes are selected from the towns by analyzing the concentration of commercial land use and the category of commercial nodes in the town. A total of eight commercial nodes are selected from these towns. The details of the selected commercial nodes are given in the table 1.

Table -1: Summary of selected commercial nodes

Name of commercial node	Town
City Junction	Tirur
Central Junction	Tirur
Thazhepalam Junction	Tirur
Main Junction	Perinthalmanna
Bypass Junction	Perinthalmanna
Hospital Junction	Perinthalmanna
Chamravattam Junction	Ponnani
Chandapadi	Ponnani

4. DATA COLLECTION

4.1 Dependent Variables

The dependent variables selected for the multiple regression modeling is the number of trips attracted towards the commercial nodes from the divided zones of the selected towns, y. The selected towns are divided into zones such that the trips attracted to the selected commercial nodes from these zones can be taken into account. In the case of socio economic survey, 10% sample households are randomly selected from each town and survey is conducted to get the household and travel data.

From the socio-economic survey data, the origin destination details of the members of the household on a working day are collected. Socio economic survey was conducted for the selected towns by the Town and Country planning office, Malappuram in the year 2013. The trips attracted to the selected commercial nodes are given in the chart 1.

4.2 Independent Variables

Independent variables are identified based on the review of several related papers and also based on the influence of these variables on the trip attracted. The selected independent variables are collected from the eight commercial nodes by conducting questionnaire survey using a properly designed questionnaire. The entire commercial node is taken into consideration for the data collection of the independent variable.

The selected independent variables are listed below:

- Total number of commercial establishments
- Type of building.
- Type of commodity sold.
- Commercial floor area in the commercial node, m².
- Percentage of commercial establishments having one, two or more than two number of floors.
- Total number of employees in the commercial node.
- Total number of parking lots in the commercial node.
- Average width of the major corridor, m
- Year of operation.

5. STATISTICAL ANALYSIS OF DATA

Multiple linear regression and correlation techniques involve the seeking of a functional relationship between two or more related variables. Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data.

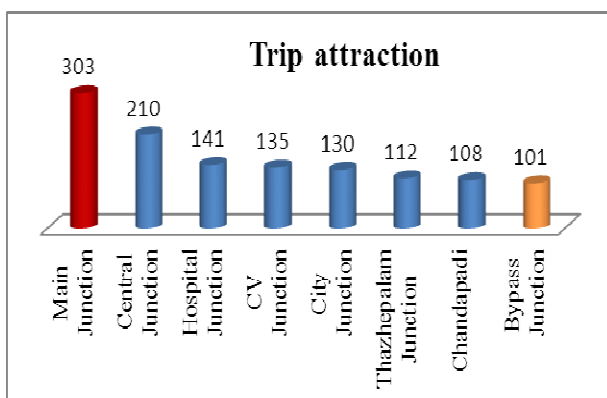


Chart -1: Trips attracted to the selected commercial nodes

Table -2: Correlation Result

SL No.	Independent variables	Correlation Test values
1	Total number of employees (x_1)	0.94777**
2	Total number of commercial establishments in the commercial node (x_2)	0.92354**
3	Commercial area, m^2 (x_3)	0.73403*
4	Average width of the major corridor in meter (x_4)	0.1495
5	Number of parking lots (x_5)	-0.0615
6	Percentage of number of shops in the commercial node (x_6)	0.51418*
7	Percentage of number of offices in the commercial node(x_7)	-0.513*
8	Percentage of number of banks in the commercial node(x_8)	-0.4604
9	Percentage of commercial establishments with only one floor (x_9)	0.51176*
10	Percentage of commercial establishments with two floors. (x_{10})	-0.3548
11	Percentage of commercial establishments with more than two floors (x_{11})	-0.6068*
12	Percentage of commercial establishments with year of operation between 0 to 5 years (x_{12})	0.11372
13	Percentage of commercial establishments with year of operation between 5 to 10 years (x_{13})	-0.1867
14	Percentage of commercial establishments with year of operation between 10 to 20 years (x_{14})	0.09046
15	Percentage of commercial establishments with year of operation above 20 years (x_{15})	-0.0576
16	Percentage of shops where stationary items are sold (x_{16})	-0.3184
17	Percentage of shops where food items are sold (x_{17})	0.71179*
18	Percentage of shops where textile items are sold (x_{18})	-0.2777

5.1 Model formulation

For model formulation, the eight selected commercial nodes is taken. Correlation and regression analysis is performed using Microsoft excel for model formulation.

5.1.1. Correlation analysis

A "correlation coefficient" is a value that indicates whether there is a linear relationship between two variables. The absolute value of the correlation coefficient will be in the range 0 to 1. A value of 0 indicates that there is no relationship whereas a value of 1 indicates that there is a perfect correlation and the two variables vary together. The sign of the correlation coefficient will be negative if there is an inverse relationship between the variables.

As a first step of model formulation, for obtaining the correlation among the selected variables, correlation analysis is performed using Microsoft excel. The Table III shows the correlation of each independent variable on the dependent variable, along with their notations. Sign (**) indicates a strong correlation ($r < 0.9$), whereas the sign (*) indicates a

moderate correlation ($0.5 < r < 0.9$), where r is the correlation coefficient.

The highest correlated independent variable with dependent variable that is the number of trips attracted to commercial node is the number of employees, x_1 (0.947) and the number of commercial establishments, x_2 (0.92354).

Also, the independent variables such as the percentage of shops where food items are sold, x_{17} (0.71179), the percentage of shops, x_6 (0.51418), total commercial area, x_3 (0.73403), and the percentage of commercial establishments with only one floor, x_9 (0.5118) are moderately correlated with the dependent variable. All the above independent variables are positively correlated with the dependent variable, y . This means that as these independent variables increases, the trip attracted to the commercial node increases.

The independent variables such as the percentage of commercial establishments with more than two floors, x_{11} (-0.6068) and the percentage of number of offices in the commercial node, x_7 (-0.513) are negatively and moderately

correlated with the trip attraction. Only the above mentioned eight independent variables which are moderately and strongly correlated with the trip attraction are considered for the further analysis.

5.1.2 Multicollinearity

Multicollinearity which is a statistical phenomenon is a problem in multiple regression that develops when one or more of the independent variables is highly correlated with one or more of the other independent variables. If one independent variable is a perfect linear combination of the other independent variables; that is, if it is regressed on the other independent variables and the resulting R² = 1.0, then the matrix of inter correlations among the independent variables is singular and there exists no unique solution for the regression coefficients.

Correlation matrix is used to detect the possible multicollinearity. It is assumed that, multicollinearity will be diagnosed to be present if the absolute value of the correlation between the independent variables (off-diagonal elements) are larger than 0.4, otherwise there is no multicollinearity or collinearity. The correlation matrix of the selected independent variables is given in the table 4. Most of the independent variables are correlated in this case and the combinations of such independent variables are avoided for the regression analysis.

5.1.3 Regression Analysis

After correlation analysis and considering the multicollinearity, regression analysis is performed with the suitable subsets. The model with high R, R² and Adjusted R² value, minimum standard error of estimate, low significance F value and low p value for the coefficients of independent variables and y intercept is selected as the best model using the regression analysis. The best model obtained is the multiple linear regression model with the independent variables, total number of commercial establishments in the commercial node (x₂), percentage of number of shops in the commercial node (x₆) and percentage of number of bank in the commercial node. The regression statistics of the formulated model is

shown in the table 3. ANOVA test result and regression coefficients are given in the table 5 and 6 respectively.

The predictive equation from the multiple regression model is
 $y = -130.316 + (0.305 \times x_2) + (2.949 \times x_6) + (-15.758 \times x_8)$
 where,
 y = 10% trips attracted to the commercial node
 x₂ = Total number of commercial establishments in the commercial node
 x₆ = Percentage of number of shops in the commercial node
 x₈ = Percentage of number of bank in the commercial node.

The first term in the prediction equation (-130.316) is a constant that represents the predicted criterion value when both predictors equal zero. The values of 0.305, 2.949 and -15.758 represent regression weights or regression coefficients of the selected independent variables, total number of commercial establishments in the commercial node, percentage of number of shops in the commercial node, and percentage of banks in the commercial node respectively.

Table -3: Regression Statistics

Multiple R	0.995551
R Square	0.991122
Adjusted R Square	0.984463
Standard Error	8.574424
Observations	8

Table -4: Multicollinearity – Correlation matrix

	x1	x2	x3	x6	x7	x8	x9	x11	x17
x1	1								
x2	0.95659	1							
x3	0.88395	0.8482	1						
x6	0.23113	0.2783	-0.133	1					
x7	-0.2197	-0.289	0.0997	-0.93	1				
x8	-0.4443	-0.2404	-0.465	-0.06	0.1494	1			
x9	0.61469	0.68	0.5561	0.014	0.1389	0.0634	1		
x11	-0.4374	-0.4943	-0.024	-0.764	0.6245	-0.135	-0.442	1	
x17	0.73123	0.6603	0.6342	0.349	-0.247	-0.241	0.1908	-0.355	1

Table -5: ANOVA results

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	32829.92	10943.31	148.8465	0.000147
Residual	4	294.083	73.52075		
Total	7	33124			

Table -6: Regression coefficients

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-130.316	38.5983	-3.37622	0.027883	-237.482	-23.1503
Number of establishments	0.305245	0.02119	14.4049	0.000135	0.246411	0.364078
Shop (%)	2.949226	0.469839	6.277099	0.003288	1.644744	4.253708
Bank (%)	-15.7589	2.8793	-5.47318	0.005423	-23.7531	-7.7647

5.2 Examination of Regression Assumptions

The examination of regression assumptions is very important. Because, when the assumptions of regression are satisfied, it means that all of the information available from the patterns in the data is used. When an assumption is violated, it usually means that there is a pattern to the data that is not included in our model, and we could actually find a model that fits the data better.

There are two strategies for testing the conformity of a particular relationship to the assumptions:

- Examining the degree to which the variables satisfy the criteria, .e.g. normality and linearity, before the regression is computed by plotting relationships and computing diagnostic statistics
- Studying plots of residuals and computing diagnostic statistics after the regression has been computed.

For this study the second approach is adopted. The details related to the examination of regression assumptions are given below:

5.2.1. Examining Linearity

Let’s consider the model for trip attraction that includes number of commercial establishments, percentage of shop

and percentage of banks. Violations of linearity are extremely serious, if we fit a linear model to data which are nonlinearly related, the predictions are likely to be seriously in error, especially when we extrapolate beyond the range of the sample data.

Nonlinearity is usually most evident in a plot of the observed versus predicted values or a plot of residuals versus predicted values or any of the independent variable, which are a part of standard regression output. The points should be symmetrically distributed around a diagonal line in the former plot or a horizontal line in the latter plot.

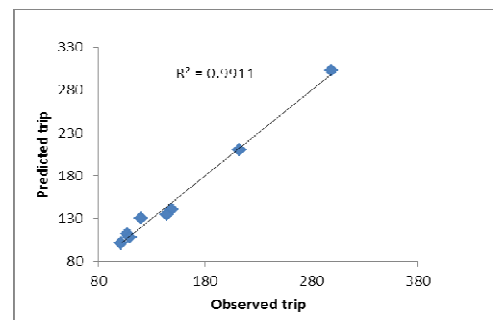


Fig -2: Predicted Trips Corresponding to the Observed Trips

It is clear from the plot of the observed versus predicted values given in figure 2, that all the points are within the diagonal line, without any unsymmetry. So, the linearity assumption is satisfied for this model.

5.2.2. Examining independence

This assumption concerns another type of systematic error in the residuals that is produced by estimated values of the dependent variable that are correlated from one case to the next, i.e. serially correlated. This is done by plotting the residuals versus each of the independent variable. If the residuals are randomly distributed, the independence assumption is satisfied. Residuals should appear independent of each other. Figure 3, 4 and 5 shows the plot of residual versus number of commercial establishments, percentage of shop and that with percentage of banks respectively.

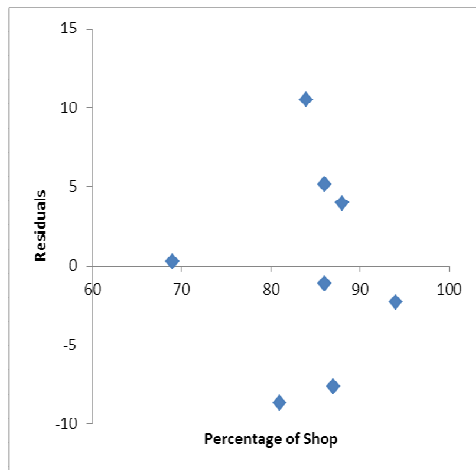


Fig -3: Residuals corresponding to the Independent variable, percentage of shops

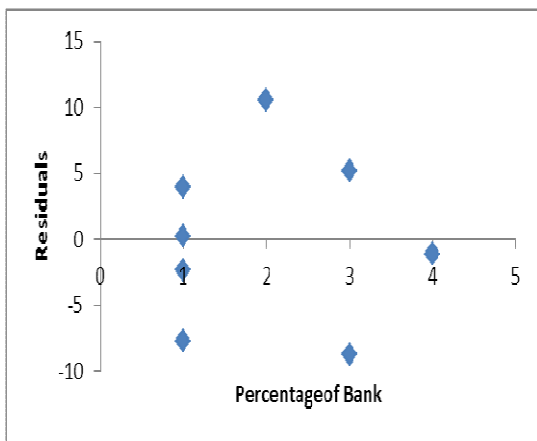


Fig -4: Residuals corresponding to the Independent variable, percentage of banks

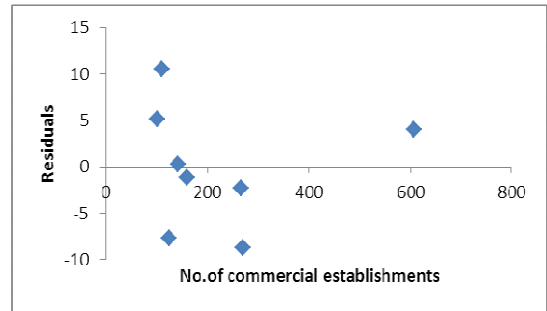


Fig -5: Residuals corresponding to the Independent Variable, number of commercial establishments

The residual are very small and we can say that in these plots, the residuals look almost random and that there is significant correlation with the variable x. So, for this formulated model, independence assumption is said to be satisfied.

5.2.3. Examining equal variance

This assumption has a variety of names: homogeneity of variance, homoscedasticity, uniform variance, etc. All of these refer to the pattern of the errors, or residuals, when plotted against the predicted values. If the assumption is met, the pattern of residuals will have about the same spread on either side of the horizontal line drawn through the average residual.

For detecting equal variance, look at plots of residuals versus predicted value. If the residuals are evenly spread out in a random manner along the horizontal line, then they are of equal variance.

The residual plot of the residuals versus the predicted values of the dependent variables of the developed equations is shown in Figure 6. The residuals are more or less evenly spread out in a random manner along the horizontal line. As the number of observations is less, the variations are not clearly seen. But as the residuals are very small for all the observations, approximately, it is clear that, there is only a small and even variance among the dependent variables. And we can say that the assumption is satisfied here.

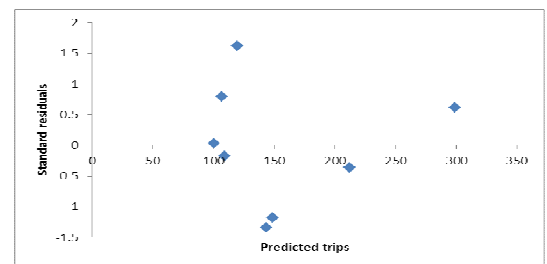


Fig -6: Residuals corresponding to the Predicted Trips

5.2.4. Examining normality

The normal probability plot for the standardized residuals represents the most straightforward test for determining the validity of the assumption that the error term has a normal distribution. The plotted points should cluster closely around the 45-degree line. In general, the more closely the points are clustered about the 45-degree line, the stronger the evidence supporting the normality assumption. Any substantial curvature in the plot is evidence that the residuals have not come from a normal distribution.

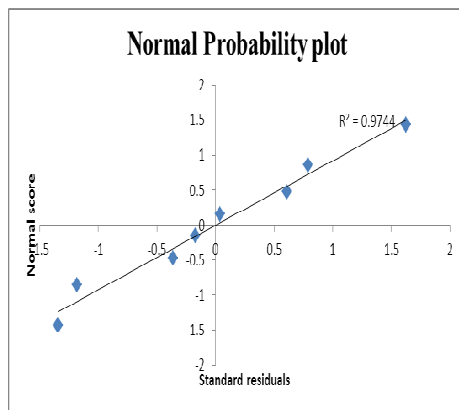


Fig -7: Normal Probability Plot of Standard Residuals

Figure 7 shows the normal probability plot of the standard residuals. Here the relationship between the theoretical percentiles (normal score) and the sample percentiles (residuals) is approximately linear ($R=0.974$) and the plotted points are clustered closely around the 45-degree. Therefore, the normal probability plot of the standard residuals suggests that the error terms are indeed normally distributed. Thus we conclude that the assumption of the error term having a normal distribution is reasonable.

5.2.5. Check for Multicollinearity

Collinearity can be checked either by noting the correlation coefficient between the independent variables in the model. Here the correlation coefficient between the independent variables is less than $|0.4|$, which shows the absence of collinearity.

So, examining all the regression assumptions for the formulated model, it is clear that, all the regression assumptions which are related to linearity, independence, equal variance, normality and multicollinearity are satisfied and hence the formulated model is accurate to predict the future trips.

CONCLUSIONS

Accurate trip generation results will lead to more accurate traffic assignment when forecasting travel demand. For this study, the latest origin - destination data was taken for model formulation, which ensures the accuracy of the model. This study developed a suitable trip attraction model using multiple regression analysis, to forecast future trips attracted to a commercial node of certain characteristics for a medium sized town in Kerala.

Similar to the related studies, this study also found that the number of employees is highly correlated to the trip attraction and total commercial area is moderately correlated to trip attraction. Other factors which found to be correlated to the trip attraction are the number of commercial establishments, percentage of shops where food items are sold, percentage of office, shops and banks in the commercial node and the percentage of commercial establishments with only one floor and more than two floors.

This study gives a vast idea about the factors affecting trip attraction towards commercial land use, thus providing a potential aid to the travel model developers. This model will be beneficial in finding the impact of any development in the commercial nodes on the traffic and thereby, proper planning of transportation facilities and services can be effectively done by the urban planners and traffic engineers.

Since most of the work related to traffic and transportation planning requires an effective framework for the analysis of the present and future travel demand pattern, a model forecasting the trip attracted based on the above mentioned characteristics shall be advantageous for a speedy travel demand forecast.

The main limitations in this study are the restricted number of commercial nodes and the assumption in regression analysis. The accuracy of the regression model can be ensured by including more representative commercial nodes from different medium sized towns in Kerala. Since, there are no comprehensive studies regarding trip attraction of commercial centers in Kerala, this study will help the future researcher. Also more factors that affect trip attraction can be studied for further research purpose.

REFERENCES

- [1]. AASHTO and FHWA. "Quick-Response Urban Travel Estimation Techniques and Transferable Parameters," Users, Guide, National Research Council, Washington, D.C., 1978
- [2]. Institute of Transportation Engineers, "Transportation impact analyses for site development: an ITE proposed recommended practice," Washington, DC, Institute of Transportation Engineers, 2005.
- [3]. A.M. Fillone and M.R. Tecson, "Trip Attraction of Mixed-Use Development in Metropolitan Manila," Proc.,

Eastern Asia Society for Transp. Studies, Vol.4, October, 860-868, 2003.

[4]. B.S. Waloejo, Surjono, and H. Sulistio, (2012), "The Influence of Trip Attraction on the Road's Level of Service (LOS) at Traditional Market Land Use," J. Applied Envir. and Biological Sci., February, pp. 92-96.

[5]. M.M. Uddin, M.R. Hasan, I. Ahmed, P. Das, M.A. Uddin, and T. Hasan, , "A Comprehensive Study on Trip Attraction Rates of Shopping Centers in Dhanmondi Area", International Journal of Civil & Environmental Engineering IJCEE-IJENS Vol:12 No:4

[6]. M. Baumeler, A. Simma, and R. Schlich, "Impact of Spatial Variables on Shopping Trips," STRC fifth Swiss Transport Research Conference, March 9-11, 2005.

[7]. C. Yu and P. Lawrence, "Trip Generation Model Development for Albany," thirteenth conference of Australian Institutes of Transport Research (CAITR), December, 2008.

[8]. J.D. Innes, M.C. Ircha, and D.A. Badoe, "Factors affecting automobile shopping trip destinations," J. Urb. Plan. and Develop., Dec., Vol. 116, No. 3, 1990. [4].

BIOGRAPHIES



Prof. Gymmy Joseph Kattor is working as Professor in the Department of Civil Engineering at Rajiv Gandhi Institute of Technology, Kottayam. He received his M.Tech. Degree in Transportation Engineering from IIT Madras and B.tech. degree in Civil Engineering from Kerala University. He has worked in Military Engineering Service for 3 years and has a teaching experience of 27 years. He published paper in international journal. His area of interest is transportation planning.



Ms. Pretina George has obtained her B.Tech in Civil Engineering from M.G. University, Kottayam. Presently she is pursuing her M.Tech in Transportation Engineering from Rajiv Gandhi Institute of Technology, Kottayam. She published paper in international journal. Her area of interest is in Transportation planning