TRIP GENERATION MODELLING USING LINEAR REGRESSION AND NEURAL NETWORKS (MLP)

TABLE OF CONTENTS

01

OBJECTIVES

- Develop trip generation models using Linear Regression & MLP.
- Compare model performance.
- Identify key variables affecting daily trip generation.

Literature Review

PAPER-Analysis of Freight Trip Generation Model for Food and Beverage in Belo Horizonte (Brazil)

OBJECTIVE:

To create a statistical model to predict the number of daily freight trips attracted by pubs and restaurants in Belo Horizonte.

Variables Used

- (NT): Number of daily freight trips attracted by an establishment.
- A: The Area of the establishment, measured in square meters (m^2).
- E: The Number of employees at the establishment.

Regression Model Equations

1. **Model 1 (Area):**
   - NT = 0.98 + 0.0016A
   - R^2 = 0.48
2. **Model 2 (Employee):**
   - NT = 1.04 + 0.019E
   - R^2 = 0.55
3. **Model 3 (Multiple Regression):**
   - NT = 1.01 + 0.014E + 0.00049A
   - R^2 = 0.55

Methodology

MODELING STRATEGY

PHASE 1: BASELINE MODEL LINEAR REGRESSION

- Applied Linear Regression to establish an initial benchmark.

PHASE 2: ADVANCEMENT TO MLP NEURAL NETWORK

- Introduced MLP to learn deeper, non-linear patterns.

DATASET

DATA SOURCE: Household travel characteristics survey.

OBSERVATIONS: 1,813 rows of data.

Dependent Variable: Number of Trips

Independent Variables: A mix of 15+ variables, including:

- Ward, HT, Commercial, Residential, Income, EarnMem, Hsize, Area_sqm, Vehicles, etc.

DATA PRE-PROCESSING

Step 1: Label Encoding

- **Input:** Text categories (e.g., Own House, Rented)
- **Action:** Converted to numerical labels
  - Own House → 1
  - Rented → 2
- **Purpose:** Makes categorical attributes usable for ML models.

Step 2: Unit Normalization

- **Input:** Mixed area units (GAJ, Sq. Ft, etc.)
- **Action:** Standardized all units to **Square Meters**
  - GAJ → sqm using: **Area (sqm) = Area × 0.828**
  - Sq. Ft → sqm using: **Area (sqm) = Area × 0.0929**
- **Purpose:** Ensures consistent scale for accurate model training.

Step 3: Feature Engineering

- **Action:** Created derived ratio variables to capture density and socio-economic structure.
- Examples:
  - **EarnRatio = Earning Members / Household Size**
  - **IncomeR = Income / Household Size**
  - **VehiclesR = Vehicles / Household Size**

Purpose: Helps ML models learn deeper patterns beyond raw variables.

Linear Regression 1

Correlation Coefficient: 0.442

R Square (R^2): 0.196

Sample Size (N): 1813

Adjusted R^2: 0.189

Std. Error of Estimate: 1.879

Linear Regression 2

Correlation Coefficient	0.433
R Square (R^2)	0.187
Sample Size (N)	1813
Adjusted R^2	0.184
Std. Error of Estimate	1.885

Linear Regression 2 was developed after removing statistically insignificant variables from MLR Model 1.

Linear Regression: Performance Analysis

The Linear Regression model explained less than 20% of the variation ($R^2 = 0.196$), indicating a very weak predictive ability.

Low R^2 and high error values show that the relationships between variables are non-linear and cannot be captured by a straight-line model.

Because Linear Regression fails to model these complex patterns, an advanced model like MLP/ANN is required to achieve higher accuracy.

Advanced Model

1. **Neural Network Multilayer Perceptron (MLP) Model**

Methodology

- **Dataset:** 1,813 rows of observations.
- **Target Variable:** Trips (Number of daily trips)
- **Independent Variables:** A mix of 15+ variables, including:
    - Ward, HT, Commercial, Residential, Income, EarnMem, Hsize, Area_sqm, Vehicles, etc.
- **Validation:** To prevent overfitting, all models were validated using either a **70/30 Train/Test split.**

Neural Network Multilayer Perceptron (MLP) Model

- An MLP is a powerful machine learning model, inspired by the human brain, that is excellent at finding the **complex, non-linear patterns**.

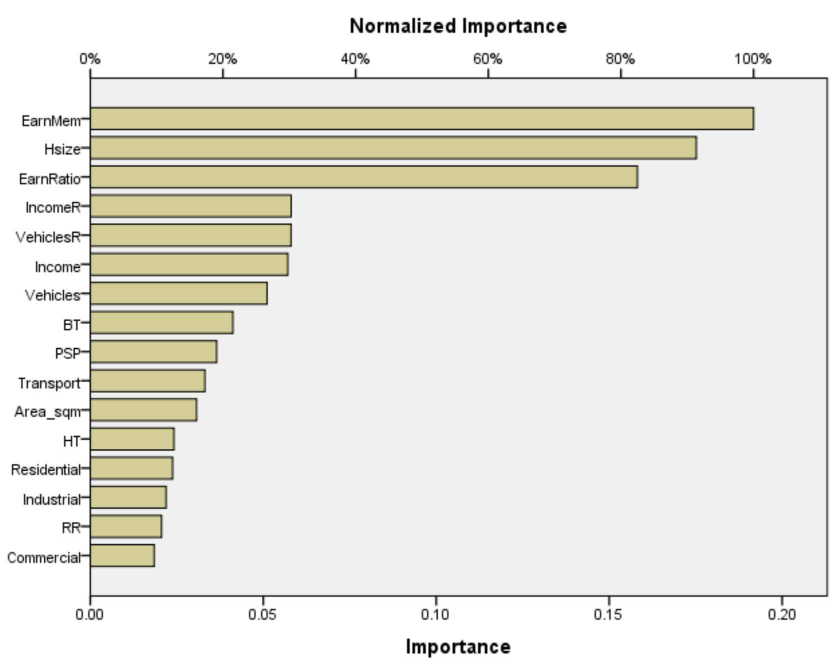
R2=1−(SSE/SST)
SST = Total sum of squares
SST=(N−1)×Variance
Variance of trip= 4.375

Metric	R^2
Training R^2	0.942
Testing R^2	0.973

Model Summary		
Training	Sum of Squares Error	456.541
	Relative Error	.733
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.33
Testing	Sum of Squares Error	213.216
	Relative Error	.781

Dependent Variable: Trips
a. Error computations are based on the testing sample.

Independent Variable Importance



Independent Variable Importance		
	Importance	Normalized Importance
BT	.041	21.5%
HT	.024	12.6%
Hsize	.175	91.4%
EarnMem	.192	100.0%
Vehicles	.051	26.6%
Income	.057	29.8%
Area_sqm	.031	16.0%
EarnRatio	.158	82.5%
VehiclesR	.058	30.3%
IncomeR	.058	30.3%
Commercial	.018	9.6%
Industrial	.022	11.4%
PSP	.037	19.1%
RR	.021	10.7%
Residential	.024	12.4%
Transport	.033	17.3%

COMPARISON

R Square (R^2)

19.6%

Linear Regression

Poor Fit (Data is not linear)

R Square (R^2)

94.2%

Neural Network (MLP)

Improvement over linear model

CONCLUSIONS

Performance of Linear Regression vs. MLP Neural Network

- **Linear Models are Insufficient:** The standard Linear Regression model was a very poor fit **($R^2 < 20\%$)** for this dataset, proving that the relationships between household/land-use data and trip generation are not linear.
- **MLP Models are Superior:** The Neural Network (MLP) model was extremely successful **($R^2 > 94\%$)**, demonstrating its ability to capture the complex, non-linear patterns that the linear model missed.
- **Household Factors are Key Drivers:** The MLP model identified that household characteristics, specifically 'Earning Members' and 'Household Size', are the most important factors in predicting trips.

REFERENCE

1. de Oliveira, L. K., de Albuquerque Nóbrega, R. A., & Ebias, D. G. (2017). Analysis of freight trip generation model for food and beverage in Belo Horizonte (Brazil). *Region*, *4*(1), 17-30.
2. Oludolapo, O. A., Jimoh, A. A., & Kholopane, P. A. (2012). Comparing performance of MLP and RBF neural network models for predicting South Africa's energy consumption. *Journal of Energy in Southern Africa*, *23*(3), 40-46

THANK YOU