Bangladesh University of Engineering and Technology

CE 457
Term Project

A Report on
Developing Regression Models for Weekday Trip Generation

Submitted To:

Dr. Sk. Md. Mashrur
Assistant Professor, Department of Civil Engineering

Submitted by:

Group 27	
Akash Saha	1904082
Afif Al Masduk	1904117
S.M Abdullah Al Jobair Raihan	1904145

Developing Regression Models for Weekday Trip Generation

1. Introduction

This report focuses on developing a regression model to predict the total number of daily weekday trips generated by students using selected socio-economic variables. The dependent variable is the *number of daily weekday trips*, while independent variables include factors such as current employment status, household's living status, dwelling type of household, number of household's possessed driver's license and Household's average monthly income. The aim is to evaluate the model's performance and its suitability for forecasting.

2. Variable Selection Rationale

The data analysis was conducted using Python, where one-hot encoding was applied to handle categorical variables, and the resulting dummy variables were consolidated. Certain columns, such as ID, home location, and latitude of home, were excluded as explanatory variables, along with columns containing uniform values. To identify optimal combinations of variables, five groups of explanatory variables were systematically combined and ranked based on their adjusted R-squared values. However, some of the top-ranked combinations did not align with real-world context or logical expectations. Therefore, the most appropriate combination was selected based on its relevance and consistency with practical considerations.

The selected variables were chosen based on their theoretical relevance and availability:

- **Current employment status**: Influences trip generation as employed individuals typically commute for work.
- **Household's living status**: Reflects household structure, which can impact trip frequency (e.g., single vs. family households).
- **Dwelling type of household**: Indicates residential characteristics that may affect travel patterns and accessibility.
- **Households with driver's licenses**: Represents vehicle access, which directly influences the number of trips made.
- **Household's average monthly income**: Affects affordability of travel modes and trip-making capacity.

While not all variables were statistically significant, they were retained to evaluate their practical impact.

3. Model Summary and Results

The regression model results are summarized as follows:

- **R-Squared:** 0.3360 (33.6% variability explained)
- **Adjusted R-Squared:** 0.2032
- **F-Statistic:** 2.5302 (p-value = 0.0091)

- **AIC:** 268.62 | **BIC:** 298.40

The model identifies key predictors of trip generation and their statistical significance.

4. Regression Analysis

n_trips_wk = 0.2098 + 0.0022 (curr_emp_unemp) -2.6596 (curr_emp_part) - 0.5411 (living_stat_1nf) +1.5532 (living_stat_fc0) + 1.5748 (living_stat_fcm) + 2.3195 (living_stat_fjoin) + 0.6655 (dwell_type_sfr) - 0.5639 (dwell_type_apb) - 0.8616 (dwell_type_ap) + 2.4377 (dwell_type_temp) - 0.274 (lics_per) + 8.59E-06 (avg_hh_inc)

	Variable	**Coefficient**	**Interpretation**
Current employment status	curr_emp_unemp	0.0022	Unemployment generates more trips as they don't have any rigid work schedule so they have more flexibility.
	curr_emp_part	-2.6596	Due to time constraints and income limitation, their ability to make trips reduces.
Household's living status	living_stat_1nf	-0.5411	Buet students have to share household responsibilities, leading to fewer trips.
	living_stat_fc0	1.5532	They have more leisure time to dine out and shopping, so more trips are likely to be generated.
	living_stat_fcm	1.5748	With multiple children, parents may need to make multiple or frequent trips for school, healthcare and shopping.
	living_stat_fjoin	2.3195	Wider variety of trip purposes resulting in more trips.
Dwelling type of household	dwell_type_sfr	0.6655	More personal space leading to higher trip generation.
	dwell_type_apb	-0.5639	Many facilities such as gym, swimming and grocery stores are integrated in a complex, so fewer trips are produced.
	dwell_type_ap	-0.8616	
	dwell_type_temp	2.4377	Students in a hall make more trips for tuition, grocery and essentials.
Households with driver's licenses	lics_per	-0.274	Shared responsibility among household members results in trip consolidation and strategic trip making reducing the cost.
Household's average monthly income	avg_hh_inc	8.59E-06	More household income means more ability to makes trips, so high trip generation.

3. Goodness-of-Fit and Model Evaluation

The model has a moderate goodness-of-fit:

- **Adjusted R-Squared (0.2032):** Indicates 20.3% of the variance in trip generation is explained by the predictors.
- **F-Statistic (2.5302, p = 0.0091):** Confirms overall model significance at a 1% level.

While the model explains some variability, a lower Adjusted R-Squared suggests room for improvement with additional predictors or alternative functional forms.

5. Forecasting Potential

The model has limited forecasting potential due to its relatively low explanatory power (Adjusted R-Squared = 0.2032). However, significant variables like **employment status** and **household income** can offer insights into trip behavior trends.

5.1 Justification

- **Strengths:**, statistically valid F-statistic.
- **Limitations:** Low R-Squared, insignificance of some variables, and potential multicollinearity.

The model can be used for short-term forecasting with caution, provided similar socio-economic conditions persist.

6. Conclusion

The regression model identifies employment status and average household income as significant predictors of weekday trip generation. Despite moderate fit, the results provide valuable insights into factors influencing student travel behavior. Further model refinements are recommended to improve accuracy.

Appendix

Statistics for the 12th Best Model (Based on Adjusted R-squared)

- **R-squared:** 0.3360
- **Adjusted R-squared:** 0.2032
- **AIC:** 268.6202
- **BIC:** 298.3961
- **F-statistic:** 2.5302
- **F-statistic p-value:** 9.0853e-03

Model Coefficients and Statistics

Variable	Coefficient	Std Error	t-Statistic	P-Value	CI Lower 0.025	CI Upper 0.975
const	0.2098	1.565	0.134	0.894	-2.921	3.34
curr_emp_unemp	0.0022	0.481	0.005	0.996	-0.96	0.964
curr_emp_part	-2.6596	0.905	-2.939	0.005	-4.47	-0.85
living_stat_1nf	-0.5411	0.556	-0.973	0.335	-1.654	0.571
living_stat_fc0	1.5532	1.161	1.338	0.186	-0.769	3.875
living_stat_fcm	1.5748	1.098	1.434	0.157	-0.621	3.771
living_stat_fjoin	2.3195	1.32	1.758	0.084	-0.32	4.959
dwell_type_sfr	0.6655	1.253	0.531	0.597	-1.842	3.173
dwell_type_apb	-0.5639	1.26	-0.448	0.656	-3.084	1.956
dwell_type_ap	-0.8616	1.42	-0.607	0.546	-3.702	1.979
dwell_type_temp	2.4377	1.493	1.633	0.108	-0.548	5.423
lics_per	-0.274	0.178	-1.542	0.128	-0.629	0.081
avg_hh_inc	8.59E-06	3.71E-06	2.316	0.024	1.17E-06	1.60E-05

Correlation Matrix

	curr_emp_unemp	curr_emp_part	living_stat_1nf	living_stat_fc0	living_stat_fcm	living_stat_fjoin	dwell_type_sfr	dwell_type_apb	dwell_type_ap	dwell_type_temp	lics_per	avg_hh_inc
curr_emp_unemp	1.000000	-0.087199	0.006572	-0.095975	0.057159	-0.087199	-0.041483	0.018792	0.037380	0.013743	0.080625	-0.073844
curr_emp_part	-0.087199	1.000000	-0.009870	0.264379	-0.135968	-0.042857	0.132263	-0.003984	-0.056136	-0.043818	0.060475	0.152441
living_stat_1nf	0.006572	-0.009870	1.000000	-0.346205	-0.488499	-0.153975	-0.104890	-0.476485	-0.201683	0.668930	-0.194696	-0.294079
living_stat_fc0	-0.095975	0.264379	-0.346205	1.000000	-0.305719	-0.096362	-0.065644	0.418617	0.015535	-0.411225	0.026896	0.136643
living_stat_fcm	0.057159	-0.135968	-0.488499	-0.305719	1.000000	-0.135968	0.207735	0.280951	0.294670	-0.580242	0.249737	0.263032
living_stat_fjoin	-0.087199	-0.042857	-0.153975	-0.096362	-0.135968	1.000000	0.132263	0.141435	-0.056136	-0.182892	0.003142	0.050003
dwell_type_sfr	-0.041483	0.132263	-0.104890	-0.065644	0.207735	0.132263	1.000000	-0.270633	-0.101686	-0.331295	0.248017	-0.061994
dwell_type_apb	0.018792	-0.003984	-0.476485	0.418617	0.280951	0.141435	-0.270633	1.000000	-0.195695	-0.637577	-0.068995	0.336680
dwell_type_ap	0.037380	-0.056136	-0.201683	0.015535	0.294670	-0.056136	-0.101686	-0.195695	1.000000	-0.239560	-0.070982	0.168366
dwell_type_temp	0.013743	-0.043818	0.668930	-0.411225	-0.580242	-0.182892	-0.331295	-0.637577	-0.239560	1.000000	-0.231261	-0.329867
lics_per	0.080625	0.060475	-0.194696	0.026896	0.249737	0.003142	0.248017	-0.068995	-0.070982	-0.231261	1.000000	-0.060483
avg_hh_inc	-0.073844	0.152441	-0.294079	0.136643	0.263032	0.050003	-0.061994	0.336680	0.168366	-0.329867	-0.060483	1.000000

Covariance Matrix

	curr_emp_unemp	curr_emp_part	living_stat_1nf	living_stat_fc0	living_stat_fcm	living_stat_fjoin	dwell_type_sfr	dwell_type_apb	dwell_type_ap	dwell_type_temp	lics_per	avg_hh_inc
curr_emp_unemp	0.129756	-0.006279	0.001142	-0.013318	0.009513	-0.006279	-0.004947	0.003234	0.003425	0.002473	0.035198	-1.353215e+03
curr_emp_part	-0.006279	0.039954	-0.000951	0.020358	-0.012557	-0.001712	0.008752	-0.000381	-0.002854	-0.004376	0.014650	1.550133e+03
living_stat_1nf	0.001142	-0.000951	0.232496	-0.064307	-0.108828	-0.014840	-0.016743	-0.109779	-0.024734	0.161149	-0.113775	-7.213661e+03
living_stat_fc0	-0.013318	0.020358	-0.064307	0.148402	-0.054414	-0.007420	-0.008371	0.077055	0.001522	-0.079148	0.012557	2.677892e+03
living_stat_fcm	0.009513	-0.012557	-0.108828	-0.054414	0.213470	-0.012557	0.031773	0.062024	0.034627	-0.133942	0.139840	6.182458e+03
living_stat_fjoin	-0.006279	-0.001712	-0.014840	-0.007420	-0.012557	0.039954	0.008752	0.013508	-0.002854	-0.018265	0.000761	5.084665e+02
dwell_type_sfr	-0.004947	0.008752	-0.016743	-0.008371	0.031773	0.008752	0.109589	-0.042808	-0.008562	-0.054795	0.099505	-1.044045e+03
dwell_type_apb	0.003234	-0.000381	-0.109779	0.077055	0.062024	0.013508	-0.042808	0.228311	-0.023782	-0.152207	-0.039954	8.183980e+03
dwell_type_ap	0.003425	-0.002854	-0.024734	0.001522	0.034627	-0.002854	-0.008562	-0.023782	0.064688	-0.030441	-0.021880	2.178463e+03
dwell_type_temp	0.002473	-0.004376	0.161149	-0.079148	-0.133942	-0.018265	-0.054795	-0.152207	-0.030441	0.249619	-0.140030	-8.384228e+03
lics_per	0.035198	0.014650	-0.113775	0.012557	0.139840	0.000761	0.099505	-0.039954	-0.021880	-0.140030	1.468798	-3.729072e+03
avg_hh_inc	-1353.215373	1550.133181	-7213.660578	2677.891933	6182.458143	508.466514	-1044.044901	8183.980213	2178.462709	-8384.227549	-3729.071537	2.588028e+09

Group Contributions

Member Name	Member ID	Role
Akash Saha	1904082	Data preprocessing and variable selection
S.M. Abdullah Al Jobair Raihan	1904145	Model development and statistical analysis
Afif Al Masduk	1904117	Report writing and formatting