Trip Attraction Rate Estimation Using Open Data Sources for Smart Transportation Planning

Ahmad Baikuni Perdana [a,1*], Rakan Pramoe Izdihar [b,2], Jonathan Badawi Wempi Suek [c,3]

[a,b,c] School of Architecture, Planning and Policy Development, Institut Teknologi Bandung, Bandung, Indonesia
[1*]baikuni.perdana@gmail.com, [2] rakanpramoe@gmail.com, [3] jonathan.suek@gmail.com

[*]corresponding e-mail

ABSTRACT

Trip Attraction Rate (TAR) serves as a load factor utilized in aggregate transportation modeling. Conventional estimation consumes a lot of time and manpower for field surveys. Meanwhile, open data sources, such as Google Maps, offer an alternative approach to conducting a survey. The popular times feature from Google Maps can estimate coffee shop TAR in the city of Bandung. This study focuses on coffee shops in Bandung City. Data was collected using web scraping techniques on Google Maps and produced a sample size of 377 data points after the validity and reliability processes. Subsequently, multiple linear regression was employed to estimate TAR. The findings reveal that variables like building area and distance to public transportation hubs influence TAR simultaneously. Moreover, our approach could also distinguish TAR between weekdays (0.13 people/m2/hour) and weekends (0.14 people/m2/hour) in Bandung City. This result challenges the usage of Institute of Transportation Engineers (ITE) standards, which is often due to limitations of time and workforce in conducting surveys for TAR estimation. The difference in values indicates the need to estimate specific TAR for each city in Indonesia rather than relying on ITE values, and the proposed approach to using open data sources will shorten the estimation process.

Keywords: trip attraction, open data, GIS, smart transportation, coffee shop

A. Introduction

Trip attractions in urban areas are complex phenomena to capture from a transportation planning perspective. The complexity is due to an infinite combination of spatial organizations, one of which is the diverse distribution of activities (Cats & Ferranti, 2022). One of the activities that tends to create trip attractions is coffee shops (Setiati et al., 2015). Coffee shops in the urban landscape serve not only as places to get caffeine but have also become the 'third place' between home and workplace, serving as facilities for building social relationships (Ferreira et al., 2021; Hariandini & Maharani, 2021). Urban communities that tend to use private vehicles for transportation can choose coffee shops according to their individual preferences without always considering distance variables. As a result, cities with current polycentric attractions are less efficient due to the high mobility of the community, which affects their lifestyles (Ismiyati & Hermawan, 2018).

The urban locale of Bandung is replete with a profusion of coffee shops, constituting a reservoir of meaningful data for the purpose of estimating Trip Attraction Rate (TAR) through the utilization of Google Maps' Popular Times feature. TAR is often used to build a model for trip attractions using the person-category approach. The rate is generated from the travel survey results of visitors to a place based on the activity. Some studies in travel behavior do travel surveys to get the trip rate. The survey was conducted separately for each dominant activity in an area or a building (Rahayu et al., 2019). Another survey conducted to estimate trip rates was used by Javed et al. (2020) for several places with similar activities. It needs to be done several times to get a satisfactory result. Travel surveys are always based on some type of sampling. Even if it were possible to survey all travelers on a specific service on a given day, this would only be a sample of travelers making trips in a given week, month, or year. The challenge in sampling design is to identify sampling strategies and sizes that allow reasonable conclusions and reliable and unbiased models without spending excessive resources on data collection (Ortúzar & Willumsen, 2011).

The challenge became a limitation in research efforts to estimate TAR by conducting field measurements. One of the limitations in the study of TAR is related to sample size and an inadequate number of surveys (D. E., 2013). Survey limitations rely on calculations using a standard, one of which is published by the Institute of Transportation Engineers (ITE). ITE is widely used to calibrate TAR calculations with only a few samples (Greene & Kannan, 2011; Maulana & Alvinsyah, 2018). This challenge can be addressed by using GIS Open Data because the data is well-documented, thus reducing the workload from a technical point of view (Klinkhardt et al., 2021).

The efficiency of using open data in calculating TAR opens up opportunities to contribute to the development of smart transportation management. Open data obtained from various devices, such as smartphones, serves as one of the inputs used in a more dynamic transportation management system (Karami & Kashef, 2020). Geolocation data can address the dynamics within the system (Iliashenko et al., 2021), and Google Maps is one of the providers that collects such data. In this article, we explore this potential using the TAR estimation method we employ.

The popular times feature from Google Maps can estimate coffee shop TAR in the city of Bandung as a solution for capturing the complexity of trip attractions in urban areas. Referring to the limitations of previous studies, we would like to highlight the presence of open data in estimating trip rates. City conditions and phenomena that are too dynamic require data that is also constantly changing, so open data is seen as a solution (Milne & Watling, 2019). The purpose of this study is to estimate trip rates based on certain activities as an alternative in one of the transportation planning processes. This research begins by discussing the process of collecting data from open sources, namely Google Maps search results based on keywords. Then the data obtained is cleaned, and a series of tests are carried out to meet the assumptions of classical regression.

B. Method

1. Location of the Research Study

Bandung is one of the many big cities in Indonesia, which is a destination for local and foreign tourists to shop and spend their weekends or vacation time (Yulian & Soegoto, 2020). Data from the Central Statistics Agency (BPS) indicated that the hotels and restaurants sector was the primary contributor to Bandung City's economy, with a sectoral contribution to the Gross Regional Domestic Product (GRDP) of 4.38% in 2021 and keeping growing to 4.72% in 2022 (BPS Kota Bandung, 2023). Furthermore, the growth rate of the hotel and restaurant sector, based on expenditure, is projected to increase by 8.40% in 2022. This suggests that the informal sector, particularly restaurants, will sustain positive growth. Notably, coffee consumption has become ingrained in the lifestyle, as evidenced by the proliferation of coffee shops across Bandung.

One of the habits of the modern lifestyle is the habit of a group of people relaxing and spending their time in the coffee shop. This causes an increase in the number of visitors to a coffee shop.

2. Data Collection with Web Scraping

Google Maps offers various Application Programming Interfaces (APIs) designed to facilitate data retrieval in a customized manner. One such API, the Places API, serves to acquire information concerning specific geographical locations (point-based). However, solely relying on this API to access Popular Times data proves impractical, as it does not furnish such information. An alternative approach involves accessing open data through the client-side web page, which is triggered when a user conducts a place search. This data gathering method necessitates the employment of web scraping techniques. Considering the substantial volume of data involved, this undertaking can be categorized as big data, rendering manual collection, by human perception, a time-consuming endeavor. Furthermore, maintaining a continuous archive of such data tends to stimulate human error.

This process utilizes a Python-based script with Selenium to deploy a bot for automated data retrieval from a web page. The required data is client-side and stored within a designated element, with the page structure primarily composed of HTML and containing unique class identifiers for similar element tags. These classes act as keys to access the data via the bot. Although the bot can automate the process, human supervision is still essential throughout the procedure. This supervision encompasses website analysis, website crawling, and data organization (Krotov et al., 2020). Website analysis involves investigating the HTML structure following the search operation to pinpoint the specific element that stores the Popular Times bar chart. Subsequently, a script is developed to automate the website crawling process. Finally, the resulting data is organized into a readable format, specifically the GeoJSON data format.

Open data from Google search results contains coordinate information in each item's listing. The information provided from Google search results can be formatted into GIS format to be readable by spatial data processing applications, such as QGIS. GeoJSON is a JSON format that binds attributes to coordinates. The notation of

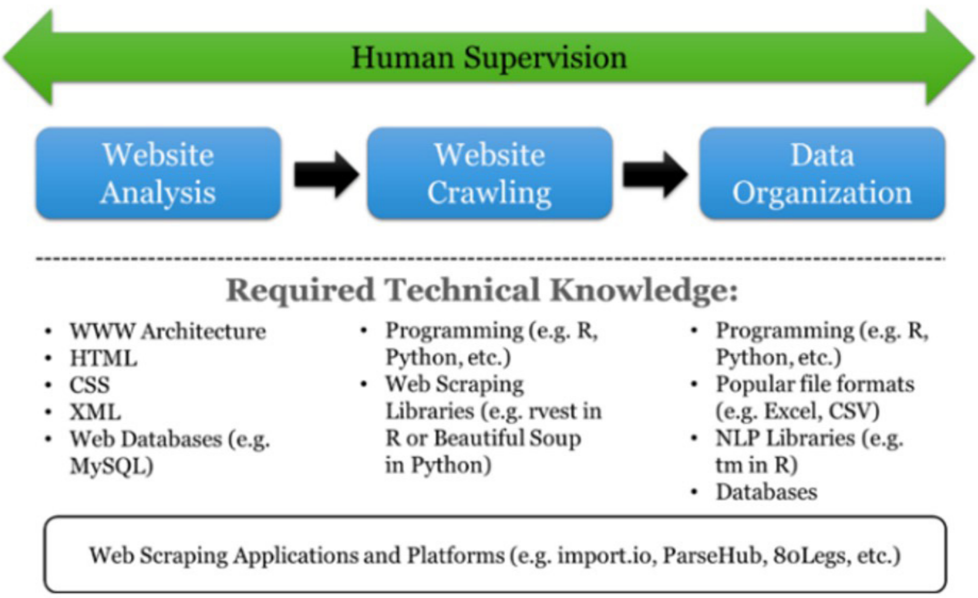


Figure 1. Webscraping Framework (Krotov et al., 2020)

E ISSN 2442-3149 | P ISSN 2355-472X
https://journal.itltrisakti.ac.id/index.php/jmtranslog

GeoJSON must, at the very least, include the keys "type" and "features." The key "features" represents a class that holds spatial objects with the keys "properties" and "geometry." The key "geometry" spatially links attributes stored in "properties," thus forming the GeoJSON format.

3. Data Cleaning

Another necessary feature for estimating the trip rate is the building area, which cannot be obtained through web scraping. Therefore, we computed the building area for each object through satellite image interpretation. The objects that resulted from the cleaning process were then overlaid onto the satellite image map using QGIS. Subsequently, manual digitization was performed to create polygons identified as buildings containing point objects. Next, each polygon's area was calculated in square meters. The building area became one of the variables considered in TRE as part of our analytical approach.

The popular times feature solely presents the cumulative visitation levels in pixel units (representing the height of bars in its bar chart). Several assumptions were made to convert the estimated visitation levels into individual units. First, the popular times indicate the coffee shop's occupancy rate. This assumption represents the fluctuations in popular times for each coffee shop relative to its own location. Secondly, not all areas within the coffee shop premise accommodate visitors; only specific portions are designated for them, with the distribution percentages as outlined in Figure 2. This proportional allocation takes non-service spaces into account, such as the kitchen, restrooms, and others.

The assumption used for the calculation of the number of visitors based on Google Popular Times data is to calibrate between the Popular Times data bar and the area of the café building. The Google Popular Times data has the highest pixel count, at 85 pixels. Then, using the café space requirement approach and the building's effective area, an estimate of the number of visitors is obtained. In this study, using the approach of moving people in

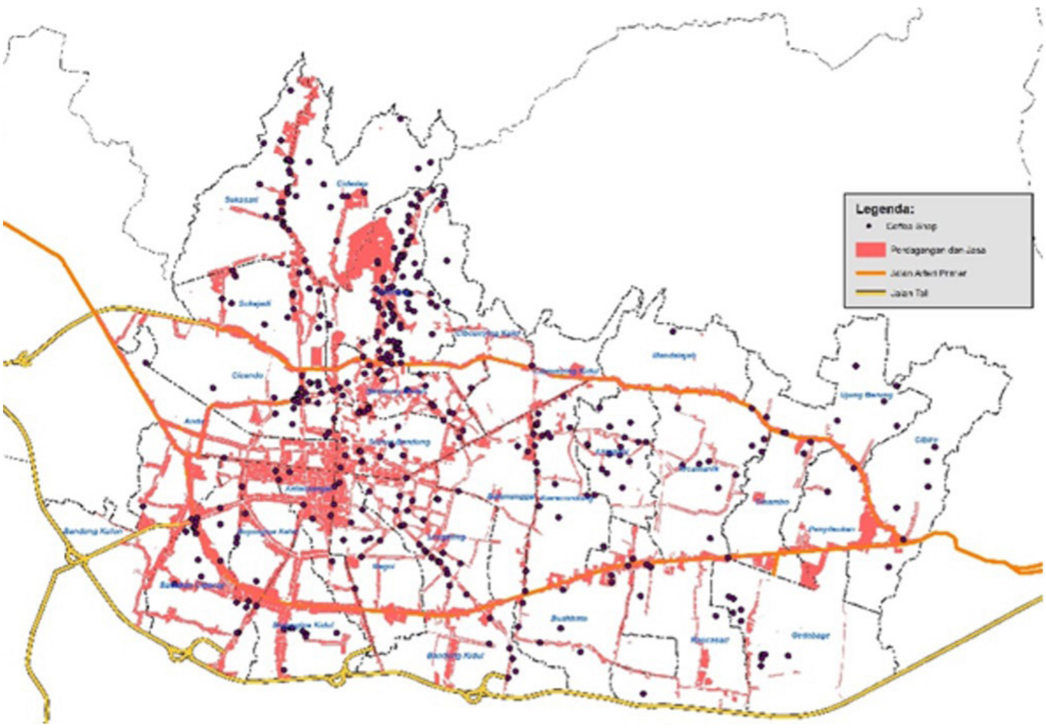


Figure 2. Overlay Coffee Shop on Retail Area in Bandung

a coffee shop of 1.6 m^2/person, the effective area used to calculate the number of visitors is 42% of the total area.

4. Data validation and reliability testing

There are several processes in place to ensure the reliability of the data obtained from web scraping. Data reliability explains how appropriate data is obtained during the web scraping process. The first step involves identifying keywords from the name attribute that are not relevant to coffee shops. The second step involves using QGIS to remove data that is not under the jurisdiction of the Bandung City administration. Meanwhile, to test the validity of the data, the Statistical Normality Test is used with z-scores of skewness and kurtosis.

5. Multiple Linear Regression

The multiple linear regression (MLR) technique is the most widely used method for the analysis of trip generation. This is a statistical technique in which the influence of different independent factors, in association with other independent factors, on the dependent variable is estimated (Ponnuswamy & Victor, 2012). One of the multiple linear regressions for trip generation is used by Basuki et al. (2020) in Semarang. The multiple linear regression analysis technique for evolving trip generation models is based on the following assumptions: (i) Each of the independent variables is linearly related to the dependent variable; (ii) All the variables are normally distributed and continuous; (iii) The independent variables are not highly correlated between themselves; (iv) The influence of each independent variable is additive, i.e., the inclusion of each independent variable would explain part of the variation of the dependent variable.

Trip attraction and mobility are closely intertwined. Therefore, to gauge the influx of customers visiting a cafe, it's essential to consider the cafe's location, the accessibility of public transportation, and the availability of road networks. Additionally, variables such as cafe characteristics, including ratings, pricing, and category, are also examined to discern their impact on trip attraction patterns, particularly in the context of coffee shops in the city of Bandung. Multiple linear regression was chosen to estimate the number of cafe visitor movements based on independent variables in Table 1, which include the area of the café, café ratings, price, café category, type of café, accessibility network, and public transport accessibility. Web scraping results can capture placeholder variables such as ratings, price, category, and type. In addition to the size of the café, accessibility is also an important variable because of the people's tendency to use various modes of travel in Bandung. One of them is the tendency of public transportation to be a substitute for its younger users there (Rizki et al., 2021).

The number of coffee shop visitors as the dependent variable is obtained from popular times, which were originally in the form of bars, and then transposed to get an estimate of the number of visitors. The estimation process involves adjusting the bar level average with the building area, as previously discussed. Trip attraction is dominated by employment availability variables and floor space area for land use characteristics that could attract movement, such as trade, services, and retail.

Office functions emerge as the primary catalysts for trip generation, exerting a significant pull on individuals' mobility. Retail and service functions closely follow in their capacity to attract movement. Within this study's context, coffee shops are regarded as a minor yet influential element shaping urban mobility dynamics in Bandung. This notion is underscored by the spatial configuration delineated by the 2023 Bandung City Regional Spatial Plan (RDTR), revealing the relatively modest prominence of the retail function in relation to other attractions. Nevertheless, it's important to acknowledge that Bandung, as the capital city of West Java, holds a significant share in the Gross Regional Domestic Product (GRDP), primarily attributed to the wholesale and retail trade sector, along with car and

E ISSN 2442-3149 | P ISSN 2355-472X
https://journal.itltrisakti.ac.id/index.php/jmtranslog

Table 1. Predictor Variable and Criterion Variables

Variables	Symbols
number of weekday visitors (person/peak hour)	y1
number of weekend visitors (person/peak hour)	y2
building area (m2)	x1
distance to bus stop (meter)	x2
type of retail	x3
price cafe	x4
category cafe	x5
ratings cafe	x6
road classification	x7

motorcycle repair services, which collectively amounted to IDR 54 trillion in 2022. Furthermore, when examining the employment landscape, the informal sector exerts a more pronounced dominion over labor absorption in the city of Bandung, eclipsing sectors such as public service employees, educators, healthcare professionals, and others.

C. Results and Discussion

1. TAR Estimation Model

A web scraping process was conducted via Google Maps search at 20 distinct locations scattered throughout the city of Bandung. This effort resulted in the acquisition of a total of 3,508 objects. The search was performed using two specific keywords: "café," "coffee shops," and "coffee," representing the name of the activity, along with the term "nearest," which guided the search towards the closest activity from each designated search point. Following the search, the retrieved objects underwent a merging process, and subsequent filtering was applied using a unique ID to ensure the elimination of any duplicates. This step led to the removal of objects located outside the administrative boundaries of the city of Bandung, ultimately reducing the number of objects to 2,858. Further refinement was performed by excluding data that lacked the popular times attribute and some objects that are not coffee shops (we choose them manually by their name), resulting in the retention of only 416 café-related objects for subsequent analysis.

The use of the linear regression method necessitates the fulfillment of classical assumptions to improve the credibility of the analysis results, such as normality, linearity, multicollinearity, and the homoscedasticity test. Errors in fulfilling assumptions can lead to biased estimations in the relationships between variables (Williams et al., 2013). We conducted a normality test on nine variables with a sample size of 416 data points. The normality test was performed by examining the z-scores for Skewness and Kurtosis. Based on the initial normality test, the data were not normally distributed, so we transformed the data into natural logarithm form and removed 39 outlier data points. These efforts ensured that several variables, including the number of visitors both weekday and weekend, the building area, and the distance of bus stops, passed the normality test with a confidence level of 95% for the Skewness and Kurtosis tests (the result is in Table 2). Hereinafter, the researchers conducted the performed tests for homoscedasticity, multicollinearity, and linearity on the variables that passed the normality test.

To do a homoscedasticity test, the scatterplot relationship between the predicted values of the dependent variable (ZPRED) and their corresponding residuals (SPRESID) was looked at. The anticipated outcome is an indiscriminate and non-patterned scatterplot,

Table 2. Initial and Final Result for Predictor and Criterion Variables

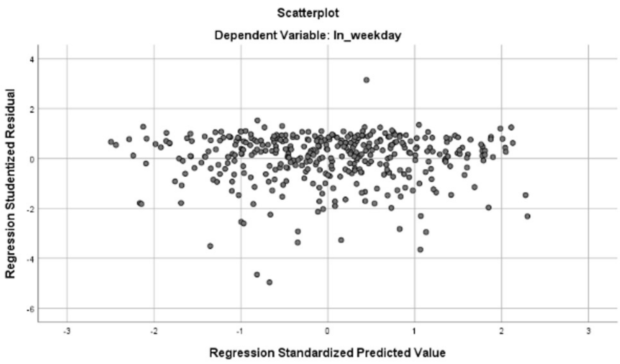
Variables	Initial Result					Final Result				
	N	Skewness		Kurtosis		N	Skewness		Kurtosis	
		z-score	α 0.05 ±1.96	z-score	α 0.05 ±1.96		z-score	α 0.05 ±1.96	z-score	α 0.05 ±1.96
y1	416	16.99	x	21.62	x	377	-1.21	normal	-0.79	normal
y2	416	17.47	x	20.46	x	377	-1.71	normal	-0.81	normal
x1	416	14.33	x	12.91	x	377	-0.28	normal	-1.89	normal
x2	416	14.88	x	12.57	x	377	0.23	normal	-1.74	normal
x3	416	3.13	x	-1.92	normal	377	0.04	normal	-6.81	x
x4	416	4.95	x	-4.05	x	377	3.60	x	-6.78	x
x5	416	15.88	x	6.78	x	377	14.26	x	4.85	x
x6	416	-18.43	x	50.16	x	377	-31.58	x	128.96	x
x7	416	0.88	normal	7.54	x	377	-12.76	x	16.57	x

as depicted in Figure 3. However, the observed scatterplot exhibited uncertainty and displayed bias, which posed challenges for subjective evaluation. Consequently, to substantiate the evaluation, the Glejser Test was employed to identify indications of heteroscedasticity within the model. This was achieved by regressing the independent variable against its absolute residuals. The Glejser test computations determined that the significance values of all variables exceeded 0.05 (see Table 3 for the weekday model and Table 4 for the weekend model). This result confirms the model's absence of heteroscedasticity symptoms.

The researcher examined the tolerance and variance inflation factor (VIF) values to assess the presence of multicollinearity within the model. Analysis of the collinearity statistics reveals that all independent variables exhibit tolerance values exceeding 0.1, while VIF values remain below 10 (see Table 5). These findings collectively indicate there is an insignificant correlation among the independent variables.

The examination of linearity between the independent variables and the dependent variable was conducted through the utilization of scatterplots. The graphical outcomes portray a clear positive linear correlation between variable Y and X1, while conversely revealing a negative linear association between variable Y and X2.

After checking for normality, linearity, multicollinearity, and homoscedasticity in the earlier steps of the process, the adjusted R square values for the weekday and weekend models were found to be 0.819 and 0.763, respectively (see Table 6). These models yielded estimated standard errors of 0.402 and 0.485, respectively. Notably, the models' viability is further reinforced by



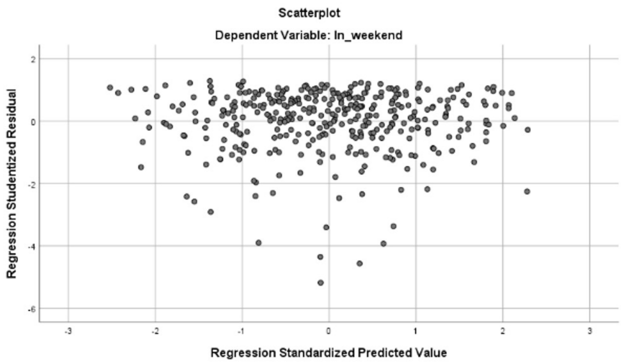


Figure 3. Homoscedasticity Test: Scatterplot

E ISSN 2442-3149 | P ISSN 2355-472X
https://journal.itltrisakti.ac.id/index.php/jmtranslog

Table 3. Homoscedasticity Test: Glejser test Weekday Model

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.322	0.156		2.067	0.039
	X1	-0.005	0.018	-0.016	-0.286	0.775
	X2	0.001	0.015	0.003	0.045	0.964

a. Dependent Variable: ABS_RES1

Table 4. Homoscedasticity Test: Glejser test Weekend Model

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.631	0.189		3.343	0.001
	X1	-0.027	0.022	-0.072	-1.245	0.214
	X2	-0.023	0.018	-0.071	-1.232	0.219

a. Dependent Variable: ABS_RES2

the significant F-values, registering at 0.000 (<0.005) for both weekday and weekend models. These F-values, measuring 851.779 for the weekday model and 605.681 for the weekend model, were determined from an extensive dataset comprising 377 entries spanning 20 sub-districts within the city of Bandung. The subsequent analytical stride entails a comprehensive investigation into the outcomes of the ANOVA and coefficient analyses (see Table 7). This pursuit aims to ascertain the partial relationships between the designated independent variables and the prevailing dependent variables.

The assessment of autocorrelation is another crucial step in validating assumptions. This test seeks to ensure the absence of autocorrelation within the linear regression model, with the Dubbin-Watson statistic serving as the diagnostic measure. The analysis revealed noteworthy results: the variable x1, which is building area, demonstrated statistical significance with a p-value of <0.05, whereas the "x2" variable exhibited a p-value exceeding 0.05. This divergence indicates a robust correlation between the building area

Table 5. Multicollinearity Test Result

Model		Weekday	
		Tolerance	VIF
1	(Constant)		
	X1	0.805	1.243
	X2	0.805	1.243

Table 6. Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
Weekday[b]	.906[a]	.820	.819	.40276	.820	851.774	2	374	.000
Weekend[c]	.874[a]	.764	.763	.48557	.764	605.681	2	374	.000

a. Predictors: (Constant), ln_hubdist, ln_building_area
b. Dependent Variable: ln_weekday
c. Dependent Variable: ln_weekend

Table 7. ANOVA[a] Weekday

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	276.345	2	138.172	851.774	.000[b]
	Residual	60.669	374	.162		
	Total	337.014	376			

a. Dependent Variable: ln_weekday

b. Predictors: (Constant), ln_hubdist, ln_building_area

Table 8. ANOVA[a] Weekend

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	285.612	2	142.806	605.681	.000[b]
	Residual	88.181	374	.236		
	Total	373.793	376			

a. Dependent Variable: ln_weekend

b. Predictors: (Constant), ln_hubdist, ln_building_area

variable and its influence on cafe visitors, in contrast to the "x2" variable, which appears to lack a substantial correlation with weekday visitor patterns.

The significance of the "x1" variable suggests that a larger activity area, in this case, the coffee shop space, holds the potential to attract a higher number of visitors and employees during weekdays. An intriguing observation is that cafes situated in proximity to Bandung's public transportation network do not exhibit a significant relationship. This observation aligns with on-site findings, where cafes boasting distinctive features like scenic views, comfort, and diverse menus attract a substantial number of visitors. However, these attributes are not captured within the model devised by the author.

Consequently, the resultant regression function for coffee shop trip attractions in Bandung is formulated as follows:

y1 (weekday model) = 0.992 (x1) - 0.013 (x2) - 1.998

Deconstructing the regression equation yields the following insights: a. The positive regression coefficient of the building area variable (x1) signifies that the nature of one's occupation exerts an influence on the escalation of trips or the frequency of visitor arrivals at the designated cafes in Bandung. For example, for every unit increase in the building area, there are 0.992 more visitors during the week. Also, the regression coefficient for the "hub-distance" variable (x2) has a negative value, which means that for every unit increase in hub-distance, there are fewer trips or fewer visitors arriving at the chosen cafes in Bandung. This decrease corresponds to a 0.013 reduction in weekday visitor arrivals. In essence, these interpretations underline the differential impact of the examined variables on the weekday visitor patterns at the coffee shops in Bandung.

The analysis for modelling trip attraction to coffee shops in Bandung involved the utilisation of mathematical modelling, specifically focusing on the correlation between the criterion and predictor variables. This analysis was conducted through the application of multiple linear regressions employing the enter method. The results are depicted in Table 8, revealing that the regression model associated with weekdays yielded a noteworthy F value of 605.681 with a highly significant p-value of 0.000. This outcome highlights the robustness and

Table 9. Coefficients Weekend

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.243	.280		-8.009	.000
	ln_x1	1.020	.033	.878	31.356	.000
	ln_x2	.009	.027	.009	.313	.755

a. Dependent Variable: ln_weekend

significance of the model in explaining the variability in weekday trip attraction to coffee shops within Bandung area.

The assessment of autocorrelation also constitutes a pivotal assumption test. This evaluation aims to establish a linear regression model devoid of autocorrelation, gauged through the examination of the Dubbin-Watson statistic. Through the conducted analysis, distinct patterns emerged: the x1 variable exhibited a significance value of <0.05, whereas the x2 variable did not manifest a significance value exceeding >0.05 (see Table 9). This disparity underscores that the x1 maintains a robust correlation concerning the impact of cafe visitors, while the x2's correlation with visitor impact is comparatively weaker.

The x1 variable's significance implies that the expansiveness of the activity area—specifically, the cafe—holds substantial potential to attract visitors and employees on a nearly daily basis. Another intriguing revelation pertains to the proximity of cafes to Bandung's public transportation network, which does not yield a significant relationship. This observation aligns with on-site findings, suggesting that the coffee shop most frequented by a substantial number of visitors possesses its own distinctive allure, such as scenic views, comfort, and a diverse menu. Unfortunately, these elements are not incorporated into the model developed in this research.

Consequently, the resulting regression function for cafe tourism attractions in Bandung is formulated as follows:

y2 (weekend model) = 1.020 (x1) + 0.009 (x2) - 2.243

The equation yields the following insights: a. The regression coefficient for building area (x1) exhibits a positive value, signifying that the nature of one's profession influences the increase in trips or the frequency of visitor arrivals at the selected café in Bandung, with a magnitude of 1.020 during weekends. b. The regression coefficient for hub-distance (x2) demonstrates a negative value, indicating that each additional vehicle load results in a reduction of trips or the frequency of visitor arrivals at the selected café in Bandung by 0.009 during weekends.

2) TAR for smart transportation planning

Instead of using a multiple linear regression approach, the TAR can also be calculated by dividing the number of people by the predictor variable for all coffee shops (N = 377) and disaggregating the data for weekdays and weekends, as well as considering the standard deviation. The grand average of TAR at peak hour was found to be 0.13 person/m^2/hour on weekdays and 0.14 person/m^2/hour on weekends. As can be seen in Table 10, the range of values on weekdays was between 0.02 and 0.42 person/m^2/peak hour with a standard deviation (S.D.) of 0.04 (person/m^2/peak hour), and the range of values on weekends was between 0.01 and 0.22 person/m^2/peak hour with a S.D. of 0.05 (person/m^2/peak hour).

This estimation is considered more dynamic compared to the use of standard trip rate values that are only confined to peak hours. Historically, the ITE guidelines have served

Table 10. Trip Attraction Rate by Weekday/Weekend

	person/m²/peak hour on weekdays	**person/m²/peak hour on weekends**
TAR Estimation	0.13	0.14
Standard Deviation	0.04	0.05

as a prominent framework for forecasting trip rates based on activity patterns. These standards have yet to effectively differentiate between the allure of peak hours and that of non-peak hours. It is noteworthy that these two time frames hold distinct attractions; nonetheless, it is still needed to conduct a comprehensive examination of foot traffic during standard hours. Furthermore, the developed model reveals variations in trip rates between weekdays and weekends.

The trip rate of coffee shops during peak periods holds a pivotal role in the distribution of trips based on land use per square meter. By way of comparison, a coffee or donut shop without a drive-through window, as a service-based activity, yields 43.38 trips per 1,000 square feet, equivalent to 0.467 trips per square metre (Institute of Transportation Engineers, 2017). Moreover, when contrasting these trip rates against the outcomes of travel surveys conducted by Greene and Kannan (2011) in Western New York cities, the travel attraction rate (TAR) varies between the city of Bandung and the other sampled locations. These surveys were carried out manually, encompassing 13 coffee shops as samples to derive the TAR during peak hours. The divergent rates identified are presented and juxtaposed in Table 11, with the mean value being employed for analysis.

The utilisation of ITE standards, which reference cities in the United States, may not necessarily be relevant when applied in cities across Indonesia. As evidenced by Greene & Kannan (2011), the areas therein alone exhibit varied TAR. The various states in the USA also provide different average TARs, as seen in California (D. E., 2013). Several previous studies related to trip rates in aggregate transportation modelling in Indonesia utilised ITE standards to calibrate the obtained values (Rahayu et al., 2019; Maulana & Alvinsyah, 2018). However, the user characteristics between the United States and other countries tend to differ (Goh et al., 2023), influencing

Table 11. Coffee Shop Trip Rates Comparison in peak hours

Country	**Precedent**	**Average TAR**	**Sample Size (Coffee Shop)**	**References**
New York, USA	Rochester	0.110	1	(Greene & Kannan, 2011)
	Brighton	0.053	2	
	Victor	0.054	2	
	Geneseo	0.037	2	
	Irondequoit	0.040	2	
	Henrietta	0.045	3	
	Greece	0.066	1	
California, USA	San Jose	1.178	6	(Jacob, 2013)
USA	ITE guidelines	0.467	8	(Institute of Transportation Engineers, 2017)
This study (Indonesia, Bandung)		**0.140**	**377**	**Google Maps Popular Times**

their motivations to visit coffee shops. Google Maps' popular times could more accurately depict the actual attractions present in a city compared to utilising ITE standards.

Utilising open data for estimating TAR offers an alternative approach to modelling trip attractions in transportation planning. Activity-based travel demand models originate from individuals' motivations to partake in specific activities (Adenaw & Bachmeier, 2022). The complexity emerges from the need to plan transportation networks that accommodate the dynamic and distinctive behavioural patterns of populations across various cities and countries. Likewise, visitor rates exhibit disparities between weekdays and weekends. Our developed model adeptly accommodates these intricacies by discerning TAR estimates across temporal dimensions.

The number of samples used is significantly larger compared to other methods, thereby rendering the TAR more representative in the city of Bandung. Employing open data to estimate TAR from popular times offers the advantage of capturing a substantial amount of data with less effort. In comparison to conventional methods, such as direct field surveys, web scraping can efficiently gather objects and minimise human error. Although the TAR estimates from popular times still require calibration for further use, in this study, our intention is to demonstrate the efficiency of open data utilisation in calculating TAR in the city of Bandung.

This approach can facilitate the establishment of an intelligent system that serves as a comprehensive database for smart transportation planning (Karami & Kashef, 2020). Open data presents an expansive and tailored dataset within the confines of Bandung. Nevertheless, the benchmarks offered by ITE pertain to U.S. cities, and the deduced TAR notably deviates from these values. Each city could potentially exhibit distinct TAR values relative to others. Consequently, the creation of an intelligent system for sophisticated transportation planning, particularly in the realm of activity-based trip attraction modeling, and the measurement of other activity types are required. Google Maps' search outcomes provide these categories, grouping them automatically even with varying search terms. This alternative avenue has the potential to propel forward smart transportation planning.

The approach of estimating TAR using open data holds the potential to provide a data source for smart transportation management systems. The promotion of smart transportation planning in Indonesia gained momentum with the enactment of Minister of Transportation Regulation of the Republic of Indonesia No. PM 76 of 2021 concerning Smart Transportation Management Systems in Traffic and Road Transport. This system is directed, among other goals, at supporting effective and efficient road planning. The presence of open data as a dynamic dataset brings the advantage of being self-sufficient (Iliashenko et al., 2021), thus contributing to the efficiency of smart transportation management systems.

D. Conclusion

The result of the trip attraction model for cafés in Bandung on weekdays is y = 0.992 x 1 - 0.013 x 2 - 1.998, with an R2 value of 0.8819. If seen from the value of the R2 value, it shows the proportion of the effect of building area and distance to access public transport to cafés on weekdays to the arrival frequency variable of 88%. The building area influences the increase in trip frequency of 0.992; each distance to public transport decrease will decrease trip frequency by 0.013. Factors influencing the attraction of coffee shops in Bandung are the building area and the distance to access public transport.

Just as the obtained results reveal, the TAR of coffee shops in Bandung and other references vary. This diversity indicates that these standards may not necessarily be suitable for estimating trip rates in other cities. The limitations of resources and time to estimate TAR for a specific activity in a certain city have been reasons for utilizing these standards. This study substantiates the employment of popular times as an estimate of TAR, which can mitigate the constraints of time and resources. Open data can curtail the necessity for field surveys, providing a level of confidence akin to the historical data offered by ITE. This approach also presents opportunities to support the development of smart transportation management systems in Indonesia. Self-sufficient open data can serve as an efficient data source. This efficiency extends beyond the computation of TAR and encompasses estimations for other components as well. Although this study does not primarily focus on other applications, their validation can be undertaken in future research endeavors.

This study has several limitations that should be considered for future research endeavors. First, the study employed web scraping relying on Google Maps' search results with the keyword "coffee shop" or any relevant keywords, leading to uncertainty in object classification as cafes due to voluntary input from the place owner. Manual elimination of self-proclaimed coffee shops was performed to address this issue. Second, the research lacked identification of dependent variables from cafe visitor preferences, such as convenience, cleanliness, facility completeness, menu availability, and proximity to residences, which could influence visitors's choices. Third, the research lacked calibration for converting Google Popular Times data into visitor counts, introducing risks of overestimation and underestimation. Despite this, the estimation was rationalized by considering coffee shop internal space requirements to mitigate these risks.

E. References

Adenaw, L., & Bachmeier, Q. (2022). Generating Activity-Based Mobility Plans from Trip-Based Models and Mobility Surveys. *Applied Sciences*, *12*(17), 8456. https://doi.org/10.3390/app12178456

Basuki, Y., Rahayu, S., & Rahmawati, N. W. (2020). Analysis of Backward Methods for Determining Trip Attraction Model on Commercial and Service Area in Sukun Raya Road, Banyumanik, Semarang. *IOP Conference Series: Earth and Environmental Science*, *409*(1), 012018. https://doi.org/10.1088/1755-1315/409/1/012018

BPS Kota Bandung. (2023). *Pertumbuhan Ekonomi Kota Bandung Tahun 2022*.

Cats, O., & Ferranti, F. (2022). Voting with one's feet: Unraveling urban centers attraction using visiting frequency. *Cities*, *127*, 103773. https://doi.org/https://doi.org/10.1016/j.cities.2022.103773

D. E., J. (2013). *Smart Growth Trip Generation of Coffee Shops*. California State University.

Ferreira, J., Ferreira, C., & Bos, E. (2021). Spaces of consumption, connection, and community: Exploring the role of the coffee shop in urban lives. *Geoforum*, *119*, 21–29. https://doi.org/10.1016/j.geoforum.2020.12.024

Goh, B. H., Yuen, C. W., & Onn, C. C. (2023). Improvement of trip generation rates for mixed-use development in Klang Valley, Malaysia. *Scientific Reports*, *13*(1), 2699. https://doi.org/10.1038/s41598-023-29748-w

Greene, C., & Kannan, V. (2011). A Trip Generation Study of Coffee/Donut Shops in Western New York. *Institute of Transportation Engineers (ITE) Journal*, *81*(6), 40–45.

Hariandini, T. B., & Maharani, Y. (2021). *Coffee Shop Place Attachment during Covid-19 Pandemic in Bandung*. https://doi.org/10.2991/assehr.k.211228.077

Iliashenko, O., Iliashenko, V., & Lukyanchenko, E. (2021). Big Data in Transport Modelling and Planning. *Transportation Research Procedia*, *54*, 900–908. https://doi.org/10.1016/j.trpro.2021.02.145

Ismiyati, I., & Hermawan, F. (2018). Lifestyle as an Influential Factor to Urban Mobility

Transport: a Case Study of Semarang City, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *123*, 012020. https://doi.org/10.1088/1755-1315/123/1/012020

Javed, S. A., Debnath, M., Nadim, Md. H., Anwar, Md. A., & Chowdhury, S. (2020). Estimation of Trip Attraction Rates and Models for Shopping Centers in Dhaka City. *Journal of Transportation Systems*, *5*(1). https://doi.org/10.5281/zenodo.3733088

Karami, Z., & Kashef, R. (2020). Smart transportation planning: Data, models, and algorithms. *Transportation Engineering*, *2*, 100013. https://doi.org/10.1016/j.treng.2020.100013

Klinkhardt, C., Woerle, T., Briem, L., Heilig, M., Kagerbauer, M., & Vortisch, P. (2021). Using OpenStreetMap as a Data Source for Attractiveness in Travel Demand Models. *Transportation Research Record: Journal of the Transportation Research Board*, *2675*(8), 294–303. https://doi.org/10.1177/0361198121997415

Krotov, V., Johnson, L., & Silva, L. (2020). Legality and Ethics of Web Scraping. *Communications of the Association for Information Systems*, *47*, 539–563. https://doi.org/10.17705/1CAIS.04724

Maulana, N., & Alvinsyah. (2018). *Pengembangan Model Bangkitan Perjalanan Berbasis Metoda ITE (Kasus: Guna Lahan Tipe Rumah Sakit)* (13).

Ortúzar, J. de D., & Willumsen, L. G. (2011). *Modelling Transport* (4th ed.). John Wiley & Sons, ltd.

Ponnuswamy, S., & Victor, D. J. (2012). *Urban Transportation: Planning, Operation and Management* (1st Edition). McGraw-Hill Education. https://www.accessengineeringlibrary.com/content/book/9781259002731

Rahayu, S., Basuki, Y., & Gritanarum, M. (2019). TRIP RATE DAN POLA PERGERAKAN PADA TRANSMART SETIABUDI SEMARANG SEBAGAI SALAH SATU PUSAT PERBELANJAAN BERKONSEP MULTI-AKTIVITAS. *Jurnal Pengembangan Kota*, *7*(1), 77. https://doi.org/10.14710/jpk.7.1.77-82

Rizki, M., Joewono, T. B., Belgiawan, P. F., & Irawan, M. Z. (2021). The travel behaviour of ride-sourcing users, and their perception of the usefulness of ride-sourcing based on the users' previous modes of transport: A case study in Bandung City, Indonesia. *IATSS Research*, *45*(2), 267–276. https://doi.org/10.1016/j.iatssr.2020.11.005

Setiati, G., Santosa, I., & Syarief, A. (2015). GENDER DAN PLACE ATTACHMENT PADA COFFEE SHOP DI BANDUNG. *Jurnal Sosioteknologi*, *14*(3), 298–310. https://doi.org/10.5614/sostek.itbj.2015.14.3.8

Williams, M. N., Grajales, C. A. G., & Kurkiewicz, D. (2013). Assumptions of Multiple Regression: Correcting Two Misconceptions. *Practical Assessment, Research, and Evaluation*, *18*(11).