

**REGRESSION MODEL FOR FORECASTING TRIP GENERATION AT
TOLL ROAD REST AREAS IN INDONESIA (A CASE STUDY: BATANG-
SURAKARTA TOLL ROAD SECTION)**

Joko Sutrisno^{1*}, Pratikso², Alfa Narendra³

^{1,2}Doctoral Program in Civil Engineering, Faculty of Engineering, Universitas Sultan Agung Semarang, Jl. Raya Kaligawe Km. 4, Central Java, Semarang, Jawa Tengah, 50112, Indonesia

³Faculty of Engineering, Universitas Negeri Semarang, UNNES Campus Sekaran Gunungpati, Jawa Tengah, Semarang, 50229, Indonesia

Correspondence: 102023300010@std.unissula.ac.id^{1*}, pratikso@unissula.ac.id²,
Alfa.narendra@mail.unnes.ac.id³

Abstract

This study develops a trip generation prediction model at toll road rest areas in Indonesia using an integrated stochastic-deterministic approach, an innovation that has not been comprehensively applied in the local context. Primary data were collected from eight rest areas along the Trans Java toll corridor and analyzed using a multivariate regression model. Based on traffic condition differences, the study develops two separate models for the Lebaran and non-Lebaran periods. During the Lebaran period, the regression model obtained is: $Y = 345.60 - 81.50X_1 + 241.67X_2$, where X_1 represents the Gross Floor Area (GFA,) and X_2 is the number of fuel pumps, with an accuracy of $R^2 = 0.9989$ dan adjusted $R^2 = 0.9978$. For the non-Lebaran period, the developed model is: $Y = -56668.5 - 439.3X_1 - 3128.0X_2 + 18662.1X_3 + 434.2X_4 + 384.0X_5 - 120.6X_6$, with predictor variables including GFA (X_1), number of fuel pumps (X_2), number of facilities (X_3), types of additional facilities (X_4), connectivity (X_5), and travel direction toward Jakarta (X_6), with an accuracy of $R^2 = 0.9711$ and adjusted $R^2 = 0.9277$. Both models demonstrate a very high capability in explaining data variability, with significantly improved accuracy compared to the models of Mohd Syafie (2021) and MTGM (2010), which only consider GFA and the number of fuel pumps with lower accuracy. The novelty of this research resides in integrating stochastic and deterministic methods with more comprehensive and locally relevant predictor variables, resulting in models with substantially higher accuracy. These findings indicate that the developed models are more suited to the traffic characteristics of Indonesian rest areas and can serve as a basis for standardizing trip generation methodologies, which previously lacked established standards, and support policymakers and infrastructure planners in optimizing rest area capacity.

Keywords: Rest area, Multivariate Regression, Trip Generation, Dwell time, Stochastic Model, Deterministic Model.

Introduction

Indonesia has experienced a significant rise in motorization over the past decade, culminating in more than 159 million registered vehicles by 2023, 16.39% of which were passenger cars. (Polri, 2024). This figure surpasses half of the country's total population, which reached 275.80 million in the same year. (Badan Pusat Statistik, 2023), reflecting a rapidly increasing vehicle ownership rate and intensifying pressure on national road infrastructure. The sharp growth in vehicle volume has been particularly evident along key transport corridors, including toll roads, necessitating the development of complementary infrastructure such as rest areas (Tempat Istirahat dan Pelayanan, or TIP) to support

user needs, reduce accident risk due to fatigue, and enhance transport system efficiency. (Darmawan et al., 2021).

Rest areas serve an essential function within the toll road ecosystem, contributing to road safety, travel efficiency, and user comfort. (Menteri Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia, 2021). These facilities are not only designated for rest and refueling but also offer a variety of services, including dining, retail, and sanitation. (Fredri Saputro et al., 2022) (Laskara, 2021). Equitably distributed rest areas are instrumental in reducing fatigue-related accidents and promoting regional economic integration. (Purboyo & Santoso, 2016). Research suggests that optimal rest area spacing ranges between 20 and 40 kilometers to maintain accessibility and functional effectiveness. (Pérez-Acebo & Romo-Martin, 2019).

Despite the increasing demand and strategic importance of rest areas, Indonesia currently lacks standardized models for predicting trip generation at toll road rest areas, resulting in planning practices that are still largely empirical and site-specific. In contrast, developed countries employ standardized models to estimate trip generation rates. For example, the ITE Trip Generation Manual (2017) uses land use codes such as "convenience store with gas station" (code 945), relying primarily on variables like Gross Floor Area (GFA) to predict trip ends (Institute of Transportation Engineers, 2017). Similarly, the Malaysian Trip Generation Manual (MTGM, 2010) utilizes facility categories such as commercial areas and petrol stations with convenience stores (code 0707), incorporating Thousand Square Feet (TSF) and number of fuel dispensers as predictor variables (Ministry of Works Malaysia, 2010).

However, several studies have demonstrated that trip generation models relying solely on GFA and the number of fuel pumps tend to underestimate peak-hour traffic volumes, particularly in dynamic or multifunctional rest area environments. (Mohd Shafie et al., 2021). These shortcomings underscore the need for localized models that reflect Indonesia's unique traffic patterns, road user behavior, and the evolving role of rest areas not merely as stopover points but as emerging activity hubs and tourist destinations.

In response, the Indonesian Ministry of Transportation currently acknowledges rest areas as traffic-generating activity centers. However, calculation methods remain unstandardized, often relying on a mix of variables such as fuel dispenser count, floor area, or site-specific traffic volumes. (Kementerian Perhubungan, 2021). Within the classical four-step transportation modeling framework, trip generation is the foundational stage, serving as the basis for estimating travel demand originating from or attracted to a facility. (Shirly Wunas, 2011).

As of 2024, Indonesia has developed over 2,800 kilometers of toll roads and established 132 operational rest areas. (Badan Pusat Statistik, 2023); (Satuan Kerja Sekretariat Pengatur Jalan Tol, 2023). While this growth reflects infrastructure expansion, the transformation of rest areas into multifunctional service complexes has not been matched by the development of accurate, observation-based predictive models. The absence of such models hinders evidence-based planning and may result in suboptimal investments and service provision. Despite this development, rest areas have transformed beyond simple rest stops into multifunctional service centers and tourist destinations. (Menteri Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia, 2021) Yet accurate visitor prediction models remain unavailable. This lack of standardized models and the ongoing transformation highlight the urgent need for observation-based predictive models to optimize future rest area planning and management. (PT. Trans Marga Jawa Tengah, 2019).

Accordingly, this study aims to develop a trip generation prediction model for rest areas using 2024 primary data collected from eight rest areas across Indonesia. A multivariate regression approach is

employed to identify key variables such as GFA, number of fuel pumps, facility types, and inter-rest area distances while incorporating deterministic factors (e.g., facility quantity) and stochastic elements (e.g., travel uncertainty) to create a model that better reflects Indonesia’s local traffic characteristics.

Method

Data Collection Method

This study was conducted in 2024 using a primary survey method at eight rest areas along the Trans-Java toll road corridor (between Batang Regency and Surakarta City). The survey aimed to collect data on vehicle counts, dwell times, and patterns of facility usage at rest areas during the Lebaran (peak holiday period) and non-Lebaran (off-peak period).

Data collection was carried out using 360° cameras installed at the entrance, exit, and parking areas of each rest area. This technology enabled real-time 24-hour recording, license plate recognition, monitoring of vehicle entries and exits (trip ends), and observation of vehicle movement patterns within the parking areas.

The recorded data were processed using media players and Google Sheets to document vehicle types, arrival and departure times, dwell times, and the facilities used. Subsequently, the data were analyzed using R Studio and SPSS software for data cleansing, grouping/binning (with 15-minute intervals), descriptive statistical analysis, probability density function (PDF) analysis, correlation tests, and regression modeling to predict visitor numbers and dwell times

To ensure data accuracy, validation was conducted through cleansing and grouping methods. (Ridzuan & Wan Zainon, 2019) In accordance with the Indonesian Road Capacity Guidelines 2023 (Direktorat Jenderal Bina Marga, 2023) To eliminate outliers (e.g., due to adverse weather conditions or recording errors). Meanwhile, data grouping is performed by categorizing data into 15-minute time intervals. This process aims to ensure that the data used and analyzed accurately represent actual conditions or observational data.

Sample selection for the rest areas was conducted through purposive sampling (Lenaini, 2021), based on classifications of Types A, B, and C as stipulated in Permen PUPR No. 28/2021. The sampling selection criteria, which ensure representativeness particularly along the Trans-Java Toll Road (spanning 1,060 km with 132 rest areas), are illustrated in Figure 1, covering 72% of the total rest area population in Indonesia. The study locations focus on rest areas within the Central Java Province cluster, covering 195 km, which are considered strategic stopover points and gathering hubs for homecoming traffic, as well as having the highest service performance ratings in Indonesia (Sutrisno, 2025a). The selection of rest area survey samples is presented in Table 1 and Figure 2, including details on location, travel direction, rest area type, and management authority. The collected data were utilized to develop predictive models for trip generation and dwell times, as well as to provide recommendations for transportation system arrangements and sustainable urban land use planning (Shirly Wunas, 2011).

Table 1 Survey location data for rest areas

| No. | Location | City | Direction | Type | Toll Road Operator |
|-----|-------------------|-------------|-----------|------|----------------------------|
| 1 | KM 379, Gringsing | Kab. Batang | A | A | Jasa Marga Batang Semarang |
| 2 | KM 389, Jatirejo | Kab. Kendal | B | A | |

| No. | Location | City | Direction | Type | Toll Road Operator |
|-----|-------------------|---------------|-----------|------|-----------------------|
| 3 | KM 391, Rowoaking | Kab. Kendal | A | B | Trans Jateng Marga |
| 4 | KM 429, Ungaran | Kab. Semarang | A | A | |
| 5 | KM 456, Pabelan | Kota Salatiga | B | B | |
| 6 | KM 456, Pabelan | Kota Salatiga | A | B | |
| 7 | KM 487, Gumukrejo | Kab. Boyolali | B | B | |
| 8 | KM 487, Gumukrejo | Kab. Boyolali | A | B | |

Description: Direction A refers to traffic leaving (from) Jakarta, while Direction B refers to traffic heading (to) Jakarta; TIP = Tempat Istirahat dan Pelayanan (Rest and Service Area); BUJT = Badan Usaha Jalan Tol (Toll Road Business Entity); Source: Research Data, 2024.



Fig. 1. Trans-Java Toll Road Corridor (Jakarta – Banyuwangi)

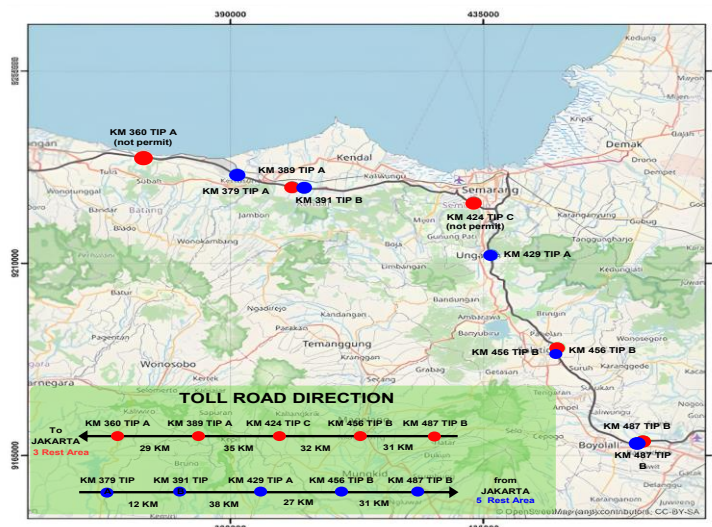


Fig. 2. Study Area Map of Batang – Semarang Toll Road Corridor

Processing Data

Data processing in this study involved the analysis of visitor vehicle counts and dwell times at eight toll road rest areas. The analysis was conducted using SPSS and R Studio software through a comprehensive sequence of statistical procedures, including normality tests, comparative tests of data characteristics between the Lebaran and non-Lebaran periods, probability density function (PDF) analysis, and multivariate regression for developing trip generation prediction models.

To determine the best-fitting data distribution representing the observed patterns, goodness-of-fit tests were performed using the Chi-square goodness-of-fit test, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The regression models were developed based on significant predictor variables, such as gross floor area (GFA), the number of fuel pumps, rest area types (A, B, and C), distances between rest areas, and the number of additional facilities.

Variable validation was carried out concerning previous studies, technical standards from the Institute of Transportation Engineers (ITE) and the Malaysia Trip Generation Manual (MTGM), as well as through Focus Group Discussions (FGDs) with transportation experts, toll road operators, and rest area management teams to ensure the models accurately represent real-world conditions and possess a high level of accuracy.

Normality Test

The normality test was conducted to verify whether the data approximated a normal distribution. Normality was assessed using the Kolmogorov–Smirnov test. (Bluman, 2014) and the Shapiro–Wilk test (Ghasemi & Zahediasl, 2012), along with visual inspections through Quantile–Quantile (Q–Q) plots (Field, 2024), skewness–kurtosis values, and normal curve histograms (Hair & Hair, 2010).

Difference Test

The difference test was conducted to evaluate variations in data characteristics between the Lebaran and non-Lebaran periods (considering both weekday and weekend conditions). This analysis aimed to identify significant differences in vehicle counts across these categories, providing a foundation for developing trip generation prediction models. The statistical methods applied in the difference test were determined based on the normality test results. If the data were not normally distributed, the Wilcoxon Rank Sum and Signed Rank Tests were employed for paired data. (Ghozali, 2017). The difference test results were further supported by visualizations using boxplots and histogram graphs, offering a clearer depiction of data distribution and inter-category differences.

Stochastic Model

A stochastic approach was employed to capture the uncertainty in trip generation patterns at rest areas by considering the probability distribution of empirical data (Yulianto et al., 2020). This model enables probabilistic estimation of vehicle volumes by accounting for random factors influencing vehicle arrivals and movements. The analysis involved three main stages: data distribution identification, probability density function (PDF) selection, and model validation.

Since the data were not normally distributed, a goodness-of-fit test was conducted to determine the best probabilistic model. The tested distributions included the Geometric, Poisson, Log-Normal, and Uniform PDFs as references for vehicle counts within specific time intervals. (Ross, 2019) (McHugh, 2013). The selection criteria for the optimal PDF were based on the results of the One-Sample Kolmogorov–Smirnov Test, Chi-Square Goodness-of-Fit Test. (McHugh, 2013). And the lowest values

of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), considering model complexity. (Burnham & Anderson, 2004).

The stochastic modeling process produced equations that describe vehicle counts at 15-minute intervals based on the most suitable distribution. Model validation was performed by comparing empirical and theoretical distributions using the Root Mean Square Error (RMSE) and Kolmogorov–Smirnov Distance (K–S Distance) to assess model accuracy. The model with the lowest AIC and BIC values was selected as the best fit, enabling probabilistic predictions that account for dynamic traffic scenarios and increased vehicle volumes during peak periods.

Deterministic Model

A deterministic approach was utilized to develop a trip generation prediction model for rest areas by precisely identifying significant variables that influence the number of visitor vehicles. (Saltelli et al., 2000). For this purpose, a multivariate regression analysis was conducted to examine the relationships between independent variables and the number of vehicles visiting rest areas. (Hair & Hair, 2010). Out of the initial 14 variables, only those that significantly influence the dependent variable, whether they can be controlled or not, were selected. (Ghozali, 2017).

In the context of decision-making for new rest area development, variables deemed crucial are those that potential investors can influence, such as gross floor area (GFA), the number of fuel pumps, and available facilities. The selection of independent variables was also guided by previous similar studies, focus group discussions (FGDs) with toll road concessionaires (BUJT), the Toll Road Regulatory Agency (BPJT), transportation experts, and questionnaires distributed to toll road users. Variabel independen yang paling signifikan dalam memprediksi volume kendaraan, meliputi: (1) GFA; (2) Number of fuel pumps (fuel stations or SPBU); (3) Number of standard facilities; (4) Number of additional facilities; (5) Connectivity; (6) Jakarta-bound direction indicator; (7) Distance to the preceding rest and service area; (8) Distance to the subsequent rest and service area; and (9) Parking capacity. After identifying the significant variables, sensitivity and multicollinearity tests were conducted to evaluate the impact of variable changes on the optimal model, using R Studio and SPSS software. The regression analysis was performed separately for the Lebaran and non-Lebaran periods to ensure that the resulting models accurately reflect traffic condition differences across these periods.

Data Analysis

1. Normality Test

The average number of visitor vehicles at each rest area within 15-minute intervals is illustrated in [Figure 3](#) (scatterplot) and [Figure 4](#) (boxplot). The rest area at KM 389 (Type A), in the Jakarta-bound direction, recorded the highest number of visitor vehicles. This is likely due to the more comprehensive range of facilities compared to other rest areas and the data collection occurring during the Lebaran homecoming traffic surge. The Focus Group Discussion (FGD) findings further emphasized that the availability of subsidized fuel stations (SPBU) at Type A rest areas is a key factor attracting toll road users. (Sutrisno, 2025a). Further analysis revealed a positive correlation between the completeness of rest area facilities and visitation rates, reinforcing the hypothesis that facility-related factors significantly contribute to the volume of vehicles stopping at rest areas.

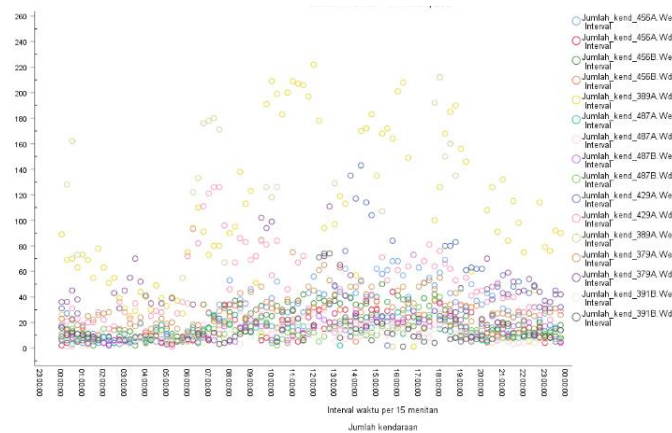


Fig. 3. Scatterplot of vehicle counts per 15-minute interval at rest areas during peak and off-peak periods.

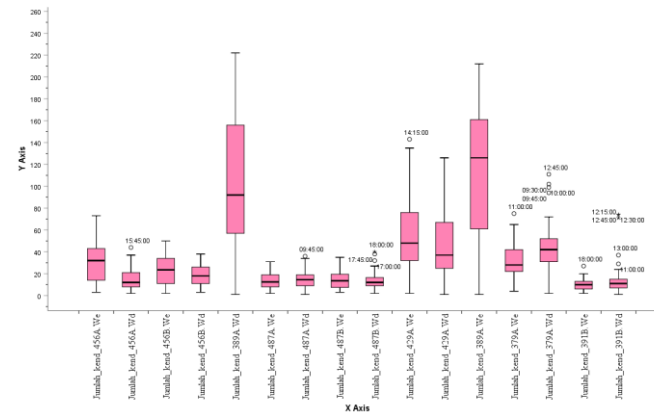


Fig. 4. Boxplot diagram of vehicle counts at each rest area.

The normality test results (Sutrisno, 2025b), conducted using the Kolmogorov–Smirnov and Shapiro–Wilk tests, indicated that the data did not follow a normal distribution ($p < 0.05$). This finding was corroborated by the skewness–kurtosis values and Q–Q plot visualizations, which showed deviations from the diagonal reference line. Consequently, subsequent analyses employed non-parametric approaches and alternative probability distributions, such as the Log-Normal and Uniform distributions, to provide a more accurate representation of traffic characteristics at the rest areas.

2. Difference Test

The results of the difference test using the Wilcoxon rank sum and signed rank tests are presented in

Table 2. Difference test results.

| Method: Wilcoxon rank sum exact test | Nilai |
|---|-----------|
| Conditions | p-value |
| Periode Lebaran dibandingkan Lebaran | Non 0.913 |
| Periode Lebaran dibandingkan Lebaran Kondisi Awal Pekan | Non 0.857 |

Periode Lebaran dibandingkan Non 0.714
 Lebaran Kondisi Akhir Pekan

Based on Table 2, the comparison of vehicle counts between the Lebaran and non-Lebaran periods, both during weekdays and weekends, yielded p -values > 0.05 . Thus, the null hypothesis is accepted, indicating no statistically significant difference in vehicle count patterns between the two periods. To support these findings, a summary of the research data is provided in Table 3 and visualized through an interval plot (Fig. 5), a boxplot (Fig. 6), and a histogram (Fig. 7).

Table 3. Summary of research data count

| Period | Categori | Data Count |
|-------------|----------|------------|
| Lebaran | Weekend | 3 |
| Lebaran | Weekdays | 2 |
| Non Lebaran | Weekend | 5 |
| Non Lebaran | Weekdays | 6 |

Analysis in Table 3 shows that the available dataset is relatively limited and generally does not follow a normal distribution. The survey was conducted 16 times across 8 rest areas, with two observations at each location. Observations during the Lebaran weekday period at two locations, “KM 379, Gringsing” and “KM 391, Rowoaking” are detailed in Tables 2 and 3. The Wilcoxon signed rank test results on the median data yielded p -values > 0.05 , indicating no significant differences in vehicle counts across the Lebaran and non-Lebaran periods, including during weekdays, weekends, and non-holiday periods. Consequently, the median patterns for these categories are statistically considered similar.

Although the p -values exceed 0.05 and the median data appear similar, visual analyses via interval plots, boxplots, and histograms reveal distribution differences between the periods. The limited dataset contributes to the model’s lower representativeness when further subdividing data into smaller groups, such as early-week and weekend categories during the Lebaran period.

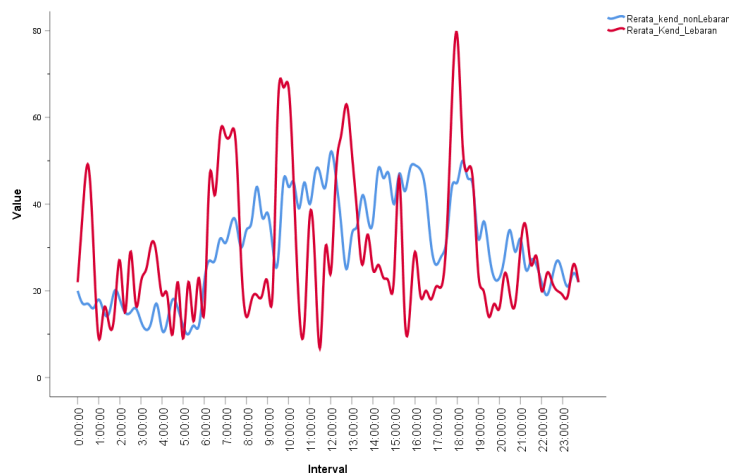


Fig. 5. Visitor vehicle counts during the Lebaran and non-Lebaran periods per 15-minute interval.

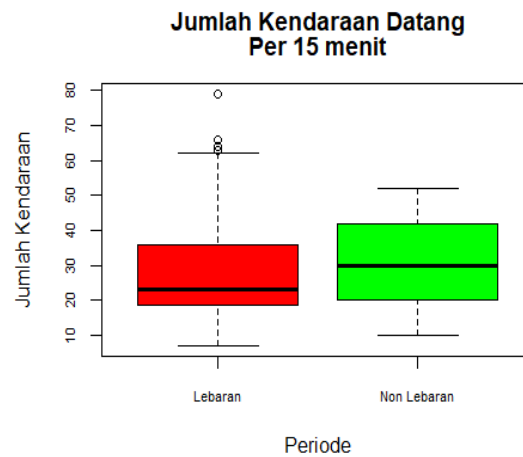


Fig. 6. Boxplot diagram of vehicle counts during the Lebaran and non-Lebaran periods

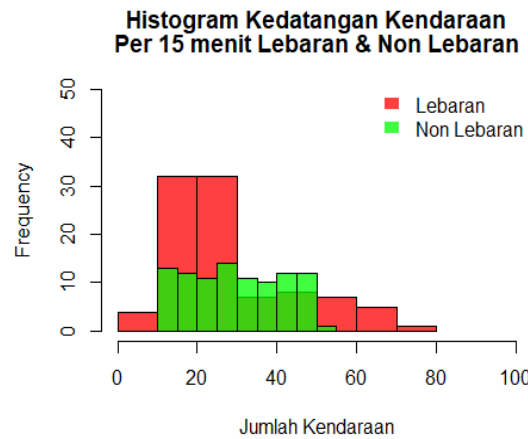


Fig. 7. Histogram of vehicle counts during the Lebaran and non-Lebaran periods

Results and Discussion

Probability Density Function (PDF) Equations

The data distribution analysis indicates that the data do not follow a normal distribution. The log-normal distribution is the best-fitting model for the Eid holiday period, while the uniform distribution provides a better fit for the non-Eid period. This difference suggests that the pattern of vehicle volumes at rest areas varies between the Lebaran and non-Lebaran periods.

The PDF equation for the Lebaran period follows a log-normal distribution, formulated as follows:

$$f(x) = \frac{1}{x \times 0.5081197\sqrt{2\pi}} \exp\left(\frac{(\ln x - 3.2285987)^2}{2 \times (0.5081197)^2}\right)$$

Meanwhile, during the non-lebaran period, the PDF adheres to a uniform distribution with the formulation:

$$f(x; 10; 52) = \begin{cases} \frac{1}{52 - 10}, & 10 \leq x \leq 52 \\ 0, & \text{otherwise} \end{cases}$$

Regression Model during the Lebaran Period

Based on the multivariate regression analysis presented in [Table 12](#), the significant variables influencing the vehicle volume of rest area visitors during the Lebaran period are the Gross Floor Area (GFA) and the number of fuel pump units (SPBU). The resulting deterministic model can be formulated as:

$$Y = 345.60 - 81.50X_1 + 241.67X_2$$

Where Y represents the number of vehicles visiting the rest area per day, X_1 denotes the Gross Floor Area (GFA), and X_2 refers to the number of fuel dispensers at the gas station.

The analysis results in [Table 12](#) indicate a very strong relationship between the number of vehicles and the two independent variables, with a high level of accuracy reflected by the R-squared value (coefficient of determination) of 0.9989. This suggests that 99.89% of the variability in the daily vehicle count can be explained by the model, demonstrating its excellent ability to capture data variability. The remaining 0.11% is influenced by other factors not examined in this study. The Adjusted R^2 value of 0.9978, which closely approximates the R^2 , further confirms the relevance and suitability of the selected variables, indicating that the model is sufficiently efficient for further development.

The interpretation of the model equation shows that when all independent variables are zero, the predicted vehicle volume is 345.6 vehicles per day during the Lebaran period. The GFA variable has a negative coefficient (-81.50), indicating that a 1-hectare increase in land area leads to a decrease in the daily vehicle count at the rest area. This result is statistically significant ($p = 0.05443$). Conversely, the number of fuel dispensers exhibits a positive and significant coefficient (241.67), implying that an increase in fuel pumps reduces vehicle queues, thereby encouraging drivers to visit the rest area, which significantly increases vehicle volume ($p = 0.0375$) without any multicollinearity issues.

Based on [Table 13](#), the deterministic model for predicting dwell time at the rest area during the Lebaran period is also influenced by GFA and the number of fuel dispensers. This deterministic model can be formulated as:

$$Y = 7809.3 - 81.61X_1 + 7033.6X_2$$

Where Y represents the daily dwell time (in seconds), X_1 denotes the Gross Floor Area (GFA), and X_2 refers to the number of fuel dispensers. The regression analysis results for the dwell time model at the rest area during the Lebaran period show an R-squared value of 0.9922. This indicates that the selected independent variables are relevant and appropriate in explaining the dependent variable, and the regression model is sufficiently efficient.

The interpretation of the dwell time model equation during the Lebaran period indicates that when all independent variables are zero, the predicted dwell time is 7,809.3 seconds per day. The GFA variable has a negative coefficient (-8,161.5), suggesting that a 1-hectare increase in land area reduces the daily dwell time. This result is statistically significant ($p = 0.05563$). Conversely, the number of fuel dispensers (X_2) is significant with a positive coefficient (7,033.6), indicating that an increase in fuel pumps reduces vehicle queues, allowing drivers to spend more time at the rest area. This, in turn, significantly increases vehicle volume ($p = 0.0375$) without any multicollinearity issues.

Regression Model during the non-Lebaran Period

Based on [Table 14](#), the model used to predict the daily number of vehicles at rest areas during the non-Lebaran period incorporates the following influential variables: Gross Floor Area (GFA), number of fuel dispensers, number of standard facilities, number of additional facilities, connectivity, and direction toward Jakarta. The model is formulated as follows:

$$Y = -56668.5 - 439.3X_1 - 3128.0X_2 + 18662.1X_3 + 434.2X_4 + 384.0X_5 - 120.6X_6$$

Where Y represents the daily number of vehicles visiting the rest area during the non-lebaran period, with X_1 as the GFA variable, X_2 as the number of fuel dispensers, X_3 as the number of standard facilities, X_4 as the number of additional facilities, X_5 as connectivity, and X_6 as the direction toward Jakarta.

Overall, this model demonstrates excellent performance in predicting the daily number of vehicles at rest areas, with an R-squared value of 0.9711, indicating that 97.11% of the variability in vehicle volume can be explained by the variables included in the model. Variables such as GFA, number of fuel dispensers, number of standard facilities, number of additional facilities, connectivity, and direction toward Jakarta contribute significantly to predicting vehicle volume, as evidenced by a p-value of 0.0048 (< 0.05). The adjusted R-squared value is 0.9277 (92.77%), which is very close to the R-squared value, confirming that the model is robust and the independent variables are relevant and appropriate in explaining the dependent variable, making the regression model efficient.

The model interpretation indicates that during the non-Lebaran period, the GFA has a negative coefficient of -439.3, suggesting that an increase of 1 hectare in gross floor area tends to reduce the daily vehicle volume, although this relationship is not statistically significant ($p = 0.4629$). Conversely, the number of fuel dispensers has a significant negative coefficient (-241.67, $p = 0.0107$), indicating that an increase in fuel dispensers is associated with a decrease in vehicle volume—possibly due to enhanced traffic management capacity that reduces congestion. In addition, the number of standard facilities (X_3) has a significant positive effect ($p = 0.0129$), while additional facilities (X_4) and connectivity (X_5) show positive but statistically insignificant effects. The direction toward Jakarta (X_6) has a negative coefficient (-120.6), suggesting a tendency for decreased vehicle volume in that direction, though the effect is not statistically significant ($p = 0.7921$).

To predict the daily dwell time at rest areas during the non-Lebaran period, the regression model (Table 15) is formulated as follows:

$$Y = -292640.84 - 1006.93X_1 - 17315.32X_2 + 98141.27X_3 + 827.23X_4 + 68.72X_5 - 1483.46X_6$$

Where Y represents the daily dwell time (in seconds), and X_1 to X_6 are defined as previously described. The model exhibits an R^2 value of 0.9788 and an adjusted R^2 of 0.9471, indicating that the model explains 97.88% of the variability in daily dwell time.

The model interpretation reveals that the Gross Floor Area (GFA) has a negative coefficient of -1006.93, suggesting that each additional hectare of land area tends to decrease the daily dwell time. However, this effect is not statistically significant ($p = 0.62232$). Conversely, the number of fuel pumps shows a significant negative coefficient of -17315.32 ($p = 0.002$), indicating that increasing the number of pumps is associated with shorter dwell times, possibly due to improved traffic management and faster refueling processes. The number of standard facilities (X_3) has a significant positive contribution ($p = 0.0029$), emphasizing that more core amenities (e.g., restrooms, dining areas) lead to longer dwell times. In contrast, additional facilities (X_4) and connectivity (X_5) display positive but statistically insignificant effects, suggesting that while they may enhance user experience, their impact on dwell time is less pronounced. Finally, the direction toward Jakarta (X_6) shows a negative coefficient of -1483.46, implying that rest areas serving vehicles traveling away from Jakarta generally experience shorter dwell times. Nonetheless, this relationship is not statistically significant ($p = 0.3766$).

To determine the influence level of each variable, a ranking analysis was conducted based on the results presented in Tables 14 and 15. During the Lebaran period, the number of fuel pumps ranked first with the most significant positive influence on vehicle volume, while the Gross Floor Area (GFA) held the second rank with a negative impact. In contrast, for the non-Lebaran period, the vehicle volume model

ranking reveals that the number of standard facilities exerts the strongest positive influence (rank 1), followed by additional facilities (rank 2) and connectivity (rank 3). Meanwhile, both GFA and the number of fuel pumps exhibit negative impacts on vehicle volume. Regarding the dwell time model, the number of standard facilities also dominates as the most influential positive factor, underscoring the importance of core amenities in prolonging stay durations. Conversely, both GFA and the number of fuel pumps demonstrate negative effects on dwell time.

Regarding the variable rankings for the visitor volume model and dwell time model based on [Tables 16 and 17](#), the analysis for the non-Lebaran period reveals the following findings: For the visitor volume model at rest and service areas (RSAs): The number of standard facilities holds the highest positive influence (rank 1), underscoring the significance of essential amenities such as toilets and dining areas in attracting visitors. Additional facilities (rank 2) and connectivity (rank 3) also exhibit positive effects, indicating that the provision of supplementary amenities and improved accessibility can increase vehicle volume. The Jakarta-bound direction variable shows a negative influence (rank 4), suggesting that drivers traveling toward Jakarta are less likely to stop at RSAs. Both the Gross Floor Area (GFA) (rank 5) and the number of fuel pump nozzles (rank 6) have negative effects on the number of visitors.

For the dwell time model at rest and service areas (RSAs) during the non-lebaran period, the number of standard facilities exhibits the strongest positive influence (rank 1), reinforcing the importance of core amenities in prolonging visitor dwell time. Both the number of additional facilities (rank 2) and connectivity (rank 3) show positive effects, though their impact is less significant compared to standard facilities. In contrast, the Jakarta-bound direction (rank 5) and the number of fuel pump nozzles (rank 6) exert negative influences on dwell time. Thus, during the non-lebaran period, standard facilities emerge as the most positively influential factor in extending visitors' stay, while GFA and the number of fuel pump nozzles are associated with a decrease in dwell time.

Regression Model Comparison

The model comparison was conducted to validate the developed mathematical models by comparing them with observed data and previous models. The observed vehicle count data from this study were compared against The mathematical models generated using R Studio and SPSS, the Mohd. Syafie model, and the MTGM (2010) model. The ITE TGM (2017) model could not be included in the comparison due to data limitations that hindered its development. The comparison results between the observed data and the four models are presented in Table 16, demonstrating that the model developed in this study exhibits higher accuracy and better representation of traffic conditions on Indonesian toll roads.

Table 4 Model Comparison Results

| Location | Periods | Model | | | | |
|----------|----------|--------------------------|-----------------------|-------------------|---------------------------|-------------------|
| | | Observation ⁰ | R Studio ¹ | SPSS ² | Mohd. Syafie ³ | MTGM ⁴ |
| KM 389B | Weekend | 2257 | 2257 | 2258 | 2194 | 1669 |
| KM 389B | Weekdays | 5123 | 5123 | 5124 | 2194 | 1669 |
| KM 429A | Weekend | 1395 | 1822 | 1823 | 2330 | 1719 |
| KM 429A | Weekdays | 2249 | 1822 | 1823 | 2330 | 1719 |

| | | | | | | |
|---------|----------|------|------|------|------|------|
| KM 456A | Weekend | 1480 | 1287 | 1287 | 834 | 1165 |
| KM 456A | Weekdays | 709 | 903 | 903 | 834 | 1165 |
| KM 487A | Weekend | 546 | 421 | 421 | 834 | 1165 |
| KM 487A | Weekdays | 294 | 421 | 421 | 834 | 1165 |
| KM 379A | Weekend | 1718 | 1685 | 1685 | 1922 | 1568 |
| KM 379A | Weekdays | 1650 | 1685 | 1685 | 1922 | 1568 |
| KM 391A | Weekend | 337 | 313 | 314 | 834 | 1165 |
| KM 391A | Weekdays | 289 | 313 | 314 | 834 | 1165 |
| KM 456B | Weekend | 729 | 536 | 536 | 834 | 1165 |
| KM 456B | Weekdays | 726 | 920 | 920 | 834 | 1165 |
| KM 487B | Weekend | 390 | 484 | 484 | 834 | 1165 |
| KM 487B | Weekdays | 577 | 484 | 484 | 834 | 1165 |

The formulations underlying the models presented in Table 5 are detailed in Table 6, which includes: Model 0: Observed data, Model 1: Output from R Studio, Model 2: Output from SPSS, Model 3: Derived from Mohd. Syafie’s study, and Model 4: Standardized MTGM (2010) model. Model Classification: Models 1, 2, and 3: Based on Indonesian toll road rest area categories. Model 3: Uses a petrol station with a restaurant category. Model 4 (MTGM 2010): Applies commercial areas, petrol service stations, and petrol stations with convenience stores classifications (code 0707). The model comparison involved eight rest areas during both Lebaran and non-Lebaran periods to assess model performance across different traffic scenarios.

Table 5 Equation of models

| Model ⁰ | Observation | |
|--------------------|------------------------------|--|
| Model ¹ | Lebaran (vehicle/day) | $345.6 - 81.5 \text{ gfa} + 241.67 \text{ jml_nozzle_spbu}$ |
| | Non Lebaran (vehicle/day) | $-56668.5 - 439.3 \text{ gfa} - 3128.0 \text{ jml_nozzle_spbu} + 18662.1 \text{ jml_fasilitas} + 434.2 \text{ jml_fasilitas_tambahan} + 384.0 \text{ connectivity} - 120.6 \text{ ke_jkt}$ |
| Model ² | Lebaran (vehicle/day) | $345.602 - 81.504 X_1 + 241.673 X_2$ |
| | Non Lebaran (vehicle/day) | $-56668.518 - 439.283 X_1 - 3128.0 X_2 + 18662.125 X_3 + 434.179 X_4 + 384.0 X_5 - 120.607 X_6$ |
| Model ³ | Evening peak (pcu/hour) | $22.660 \text{ (pump)} + 138.934$ |

Model⁴ Evening peak 8.401(pump) + 194.007
(pcu/hour)

1 hectare \approx 107.64 TSF (thousand square feet)

1 PCU/peak hour x 2 hours x 6 DEF \approx 12 kendaraan/hari

If there are 1,500 PCU/hour during the 2-hour peak period with a Daily Expansion Factor (DEF) of 6, the converted volume results in 18,000 vehicles per day

Based on Tables 16 and 17, along with the visualized model comparison results presented in Figures 11 and 12, it can be concluded that the mathematical model developed in this study demonstrates higher accuracy compared to previous models. Furthermore, it provides a better representation of the toll road conditions in Indonesia.

Based on Figures 11 and 12, it can be concluded that the observed data model for predicting the number of vehicles visiting rest areas, whether using SPSS or R Studio software, exhibits a very small and consistent difference. The model accuracy comparison presented in Table 18 shows that the rest area trip generation model developed in this study possesses very high mathematical accuracy and is well-suited to the variables relevant to Indonesia.

Model 1 and Model 2 successfully incorporate a broader range of contextual variables, including GFA, the number of fuel dispensers, standard facilities, additional facilities, connectivity, and travel direction, all of which are relevant to traffic patterns on Indonesian toll roads. In contrast, Model 3 and Model 4 rely solely on the number of fuel dispensers as the primary variable, making them overly simplistic and less representative of the complexity of rest areas in Indonesia.

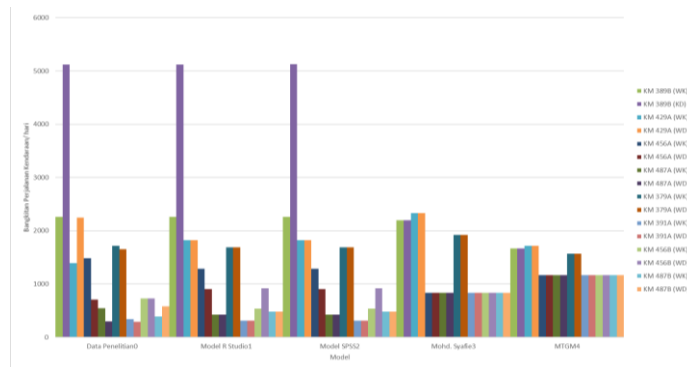


Figure 8 Model Comparison Diagram

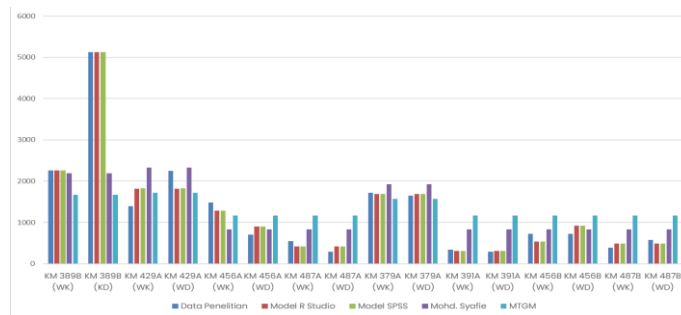


Figure 9 Model Comparison Diagram by Rest Area

Table 6 Model Accuracy Comparison

| Model | RMSE (Root Square Error) | MAE (Mean Absolute Error) | R ² (Determination coefficient) |
|------------------------------------|-----------------------------|------------------------------|---|
| 1. Model 0 (Data observasi) | - | - | 1.000 |
| 2. Model 1 (Regresi R Studio) | 58.3 | 32.1 | 0.9989 |
| 3. Model 2 (Regresi SPSS) | 60.1 | 34.0 | 0.9980 |
| 4. Model 3 (Mohd. Syafie, 2021) | 121.7 | 85.6 | 0.9200 |
| 5. Model 4 (MTGM, 2010) | 198.3 | 142.1 | 0.8900 |

Overall, the deterministic model developed in this study demonstrates high prediction accuracy for both vehicle volume and dwell time at rest areas, effectively explaining a significant portion of data variability. These findings make an important contribution to the development of trip generation standards in Indonesia and can serve as a foundation for planning and managing rest area infrastructure to anticipate future increases in vehicle volume.

Conclusion

This study successfully developed a predictive trip generation model for rest areas along Indonesian toll roads using multivariate regression analysis. The results indicate two regression models for daily vehicle volume at rest areas, applicable to both Lebaran and non-Lebaran periods. During the Lebaran period, the significant predictor variables affecting both daily vehicle volume and dwell time at rest areas are the Gross Floor Area (GFA) and the number of fuel dispensers (nozzles) at gas stations. The deterministic model for vehicle volume achieved an R² of 0.9989, explaining 99.89% of the variance in rest area visitor vehicle volume, with an adjusted R² of 0.9978. For dwell time, the deterministic model recorded an R² of 0.9922, indicating it explains 99.22% of dwell time variance, with an adjusted R² of 0.9844.

In the non-Lebaran period, significant predictor variables include GFA, number of fuel dispensers, number of standard facilities, additional facilities, connectivity, and direction toward Jakarta. The deterministic model for vehicle volume yielded an R² of 0.9711, capturing 97.11% of the variance with an adjusted R² of 0.9277, while the dwell time model achieved an R² of 0.9788 and an adjusted R² of 0.9471. These findings make a significant contribution to the development of trip generation standards in Indonesia, where no standardized guidelines currently exist. The deterministic models serve as valuable tools for planning and managing rest areas, particularly in anticipating future vehicle volume

increases and designing capacity for parking and non-standard facilities. Model interpretation based on a Focus Group Discussion (FGD) with transportation experts and rest area operators led to recommendations for including variables such as travel time, availability of subsidized and non-subsidized fuel, and rest area layout design. However, travel time did not exhibit a significant influence in this study, as it was represented by the distance before and after rest areas (within a 20-40 km range), suggesting that its impact may be more prominent over longer distances. The availability of subsidized fuel is represented through rest area types: Type A offers subsidized fuels (Pertalite and Solar), whereas Type B provides only non-subsidized options (Pertamax, Dexlite, etc.). Furthermore, discussions with experts indicate that rest area layout design influences visitor interest. Complex or convoluted vehicle circulation can reduce comfort and decrease the number of visitors. This finding correlates with the GFA variable, which exhibits collinearity with parking capacity, suggesting that an excessively large area may lead to inefficient circulation pathways. In this study, the terms *lebaran* and *non-lebaran* can be generalized into long weekends, weekends, and weekdays, aligning with a more common approach in transportation analysis. In the context of Indonesia, the Eid holiday period represents the largest long weekend travel pattern of the year. By analyzing dwell time during the Eid and non-Eid periods, rest area operators can optimize the allocation of parking capacity, standard facilities (e.g., fuel stations, restrooms), and additional amenities (e.g., cafés, shopping areas, and meeting rooms) based on visitor needs.

Future research is recommended to incorporate additional variables, such as travel time and rest area layout design, to enhance the predictive model. Furthermore, model validation using additional datasets from other toll road segments is necessary to ensure its accuracy across different traffic conditions.

Policy implications include adopting the model for standardized trip generation regulations, improving rest area infrastructure planning with more accurate demand predictions, and enhancing regional economic activity through the integration of commercial facilities.

Declaration of competing interests

The authors declare that there are no financial or personal conflicts of interest that could have influenced the research presented in this study.

Declaration of AI Use

During the preparation of this work, the authors utilized AI assistance to enhance the language quality and readability. Following the use of this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Reference

1. Badan Pusat Statistik. (2023). *Statistik Indonesia (Statistical Yearbook of Indonesia) 2023* (Direktorat Diseminasi Statistik, Ed.; 03200.2303, Vol. 11010001).
2. Bluman, A. (2014). *Elementary Statistics: A step-by-step approach*, 9e. McGraw-Hill. [https://books.google.co.id/books?hl=id&lr=&id=EIZvEAAAQBAJ&oi=fnd&pg=PP1&dq=Bluman,+A.+G.+\(2013\).+Elementary+Statistics:+A+Step+by+Step+Approach.+McGraw-Hill+Education.&ots=239FxnAHs&sig=COMrLvGXZz_68nJe_Ehdvs8-NVQ&redir_esc=y#v=onepage&q&f=false](https://books.google.co.id/books?hl=id&lr=&id=EIZvEAAAQBAJ&oi=fnd&pg=PP1&dq=Bluman,+A.+G.+(2013).+Elementary+Statistics:+A+Step+by+Step+Approach.+McGraw-Hill+Education.&ots=239FxnAHs&sig=COMrLvGXZz_68nJe_Ehdvs8-NVQ&redir_esc=y#v=onepage&q&f=false)

3. Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304.
4. Darmawan, E., Sari, S. R., Werdiningsih, H., & Enis, A. S. (2021). Rest Area Development in Central Java (Case Study: Rest Area on Semarang-Solo Toll Road). *International Journal of Sustainable Transportation Technology*, 4(1), 1–8.
5. Direktorat Jenderal Bina Marga, P. (2023). Pedoman Kapasitas Jalan Indonesia No. 09/ P/ BM 2023 (Vol. 1, Issue 021).
6. Field, A. (2024). *Discovering Statistics Using IBM SPSS Statistics* (J. Seaman, Ed.; 6th ed.). Sage.
7. Fredi Saputro, R., Dian Susanti, A., & Wardianto, G. (2022). ARRANGEMENT OF SPATIAL RELATIONSHIPS IN TOLL ROAD REST AREA Case Study: Rest Area km. In *ARSIP Jurnal Arsitektur* (Vol. 2, Issue 1).
8. Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. In *the International journal of endocrinology and metabolism* (Vol. 10, Issue 2). Brieflands. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3693611/>
9. Ghazali, I. (2017). *Aplikasi Analisis Multivariate Dengan Program IBM SPSS 25*, Edisi 9 (9th ed., Vol. 9). Badan Penerbit Universitas Diponegoro.
10. Hair, J. F., & Hair, J. F. (Jr). (2010). *Multivariate Data Analysis* (7th ed.). Prentice Hall. <https://books.google.co.id/books?id=JIRaAAAAAYAAJ>
11. Institute of Transportation Engineers. (2017). *Trip Generation Manual 10th Edition • Volume 1: Desk Reference Institute of Transportation Engineers* (Vol. 1).
12. Kementerian Perhubungan, R. I. (2021). Peraturan Menteri Perhubungan Republik Indonesia Nomor PM 17 Tahun 2021 Tentang Penyelenggaraan Analisis Dampak Lalu Lintas. www.peraturan.go.id
13. Lenaini, I. (2021). Teknik Pengambilan Sampel Purposive dan Snowball Sampling. 6(1), 33–39. <https://doi.org/10.31764/historis.vXiY.4075>
14. McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica*, 23(2), 143–149.
15. Menteri Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia. (2021). Peraturan Menteri Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia Nomor 28 Tahun 2021 Tentang Tempat Istirahat dan Pelayanan Pada Jalan Tol.
16. Ministry of Works Malaysia, H. P. U. (2010). *Malaysian-Trip-Generation-Manual-2010* (Vol. 1).
17. Mohd Shafie, S. A., Leong, L. V., & Mohd Sadullah, A. F. (2021). A Trip Generation Model for a Petrol Station with a Convenience Store and a Fast-Food Restaurant. *Sustainability*, 13(22), 12815. <https://doi.org/10.3390/su132212815>
18. Pérez-Acebo, H., & Romo-Martin, A. (2019). Service and Rest Areas in Toll Motorways in Poland: Study of Distribution and Facilities. *Transport Problems*, 14(2), 155–164. <https://doi.org/10.20858/tp.2019.14.2.14>
19. Polri, K. (2024). Electronic Registration Identification. [Http://Rc.Korlantas.Polri.Go.Id:8900/Eri2017/Index.Php](http://Rc.Korlantas.Polri.Go.Id:8900/Eri2017/Index.Php).
20. PT. Trans Marga Jawa Tengah. (2019). Rencana Pembangunan Tempat Istirahat dan Pelayanan Jalan Tol Semarang - Solo.
21. Purboyo, H., & Santoso, E. B. (2016). The growth of the rest area at the development axis in the southern part of West Java. *Procedia-Social and Behavioral Sciences*. <https://www.sciencedirect.com/science/article/pii/S1877042816308126>
22. Ridzuan, F., & Wan Zainon, W. M. N. (2019). A Review of Data Cleansing Methods for Big Data. *Procedia Computer Science*, 161, 731–738. <https://doi.org/https://doi.org/10.1016/j.procs.2019.11.177>
23. Ross, S. M. (2019). *A first course in probability* (Tenth edition). Pearson. https://api.pageplace.de/preview/DT0400.9781292269238_A39514643/preview-9781292269238_A39514643.pdf
24. Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modeling. *Statistical Science*, 377–395.
25. Satuan Kerja Sekretariat Pengatur Jalan Tol, D. J. B. M. P. (2023). Standar Pelayanan Minimum Jalan Tol 2023.

26. Shirly Wunas. (2011). Kota Humanis Integrasi Guna Lahan & Transportasi di Wilayah Suburban: Vol. viv (A. Wijaya, Ed.; 1st ed.). Brilian Internasional.
27. Sutrisno, J. (2025a). Hasil FGD Model Bangkitan Perjalanan Rest Area di Jalan Tol Indonesia (Studi Kasus: Ruas Tol Batang - Surakarta).
28. Sutrisno, J. (2025b). Model Bangkitan Perjalanan Rest and Service Area di Jalan Tol Indonesia (Studi Kasus: Ruas Tol Batang - Surakarta).
29. Yulianto, B., Setiono, Sugiyarto, Purnomo, S., & Prasetyo, R. A. (2020). Study of Standard Trip Attraction Models of Various Land Use in Surakarta City. *Journal of Physics: Conference Series*, 1625(1), 012037. <https://doi.org/10.1088/1742-6596/1625/1/012037>