Public Transportation Usage and Commuter Satisfaction Survey

Dhruva Kumar Singh A

ABSTRACT

Urban public transportation systems require continuous empirical and data-driven evaluation to enhance commuter retention, optimize operational schedules, and promote sustainable municipal mobility. This research establishes an empirical, dual-layer analytical framework designed to evaluate commuter satisfaction using primary survey data ($N = 71$) collected from daily transit users in Bengaluru, India. The methodological workflow bridges traditional non-parametric statistical hypothesis testing—specifically the Chi-Square Test of Independence ($\chi^2 = 18.50, p < 0.05, df = 4$)—with a Supervised Machine Learning pipeline utilizing a Random Forest Classifier (micro-$F_1 = 0.82$). To resolve black-box ambiguity in predictive modeling, feature importances and directional contributions are extracted using SHapley Additive exPlanations (SHAP). Empirical findings demonstrate that while baseline transit adoption is driven by fare affordability, operational reliability—specifically schedule punctuality (SHAP value $= +0.42$) and peak-hour vehicle capacity management (SHAP value $= +0.28$)—serves as the primary determinant of commuter satisfaction and retention. The study concludes with a three-tier operational optimization framework tailored for municipal transit authorities.

1. INTRODUCTION

Public transportation constitutes the structural backbone of modern metropolitan infrastructure, playing a fundamental role in shaping urban mobility, economic productivity, and environmental sustainability. In rapidly expanding metropolitan regions like Bengaluru, India, continuous demographic growth, rapid urbanization, and rising private vehicular ownership have led to severe road network congestion, elevated carbon emissions, and overextended municipal transit networks. Consequently, municipal transport authorities face the urgent operational challenge of enhancing service quality to encourage a structural modal shift away from private personal transport.

Commuter satisfaction serves as the primary metric for evaluating transit system performance, user retention, and customer lifetime value. Satisfaction is a multidimensional construct influenced by an array of operational attributes, including schedule adherence, onboard seating availability, physical comfort, personal security, ticketing affordability, and multi-modal transfer connectivity. When transit systems consistently meet passenger expectations across these dimensions, user retention increases, strengthening the overall economic and environmental viability of the transit network.

While traditional transit evaluation frameworks predominantly rely on basic descriptive statistics or isolated service attribute ratings, modern urban mobility research requires predictive, robust, and explainable analytical models. Simple statistical tabulations often fail to capture non-linear interactions between service quality attributes and overall satisfaction.

This paper addresses this gap by presenting an integrated analytical pipeline applied to a primary empirical dataset of $N = 71$ daily commuters. By pairing Chi-Square non-parametric hypothesis testing with a Random Forest Ensemble Classifier and SHAP explainability analysis, this study identifies key operational bottlenecks and formulates targeted optimization strategies for transit operators.

Research Contribution

The primary contribution of this study is the development of an integrated analytical framework that combines statistical inference, predictive machine learning, and explainable artificial intelligence to examine commuter satisfaction with public transportation. Unlike approaches that rely exclusively on descriptive statistics or conventional hypothesis testing, the proposed framework evaluates both statistical relationships and predictive patterns within commuter service-quality perceptions.

The study further extends conventional predictive analysis by comparing interpretable statistical modelling with non-linear machine-learning approaches and by applying SHAP-based explainability to identify the relative contribution of individual service attributes. The framework is subsequently extended through Importance–Performance Analysis and scenario-based evaluation to translate analytical findings into actionable operational priorities. This combination enables the research to move from identifying satisfaction drivers toward determining which service dimensions should receive greater managerial attention.

3 LITERATURE REVIEW & RESEARCH GAP

3.1 Service Quality in Public Transit

Service quality is broadly recognized as the principal determinant of overall passenger satisfaction. Eboli and Mazzulla (2007) conducted foundational empirical research demonstrating that operational factors such as service frequency, vehicle punctuality, seating comfort, and terminal accessibility heavily govern passenger perceptions. Their findings emphasized that functional service attributes directly influence overall satisfaction, suggesting that operational improvements yield immediate gains in passenger retention.

3.2 Demographic and Trip-Purpose Heterogeneity

Expanding upon variable perception dynamics, Tyrinopoulos and Antoniou (2008) examined user satisfaction across diverse passenger demographics and transit modes. Their empirical findings highlighted significant heterogeneity in commuter expectations based on trip purpose (e.g., daily time-sensitive work commuting versus flexible leisure trips) and travel frequency. They concluded that transit operators must adopt targeted operational strategies tailored to specific commuter cohorts rather than applying uniform service policies.

3.3 Modal Shift and Churn Dynamics

The dynamics of modal shift—transitioning commuters from private automobiles and motorized two-wheelers to public transit—were extensively synthesized by Redman et al. (2013). Their literature review revealed that while pricing interventions (such as fare subsidies or promotional passes) attract initial transit trial usage, long-term modal switching depends almost entirely on service reliability, travel speed, and schedule predictability. They argued that cost-driven initiatives fail to sustain passenger loyalty if baseline operational standards are unreliable.

3.4 Dissatisfaction Triggers and Cumulative Perceptions

Specific dissatisfaction drivers were examined by Dell'Olio et al. (2011), who identified vehicle overcrowding, unscheduled delays, and inadequate physical cleanliness as dominant triggers for passenger churn. Similarly, Friman and Fellesson (2009) analyzed cumulative satisfaction over extended travel periods, confirming that consistent daily service delivery is vital for establishing positive long-term passenger attitudes and preventing modal switching to private vehicles.

3.5 Research Gap

Existing literature frequently analyzes service quality attributes in isolation without quantifying non-linear feature interactions. Methodological approaches in transit literature are often bifurcated: using either traditional statistical tests or standalone machine learning models, rarely combining both for statistical validation and predictive explainability. This study bridges this gap by presenting a hybrid analytical workflow applied to empirical survey data.

4. CONCEPTUAL FRAMEWORK

The conceptual framework of this study proposes that the quality of public transportation services influences the overall satisfaction of commuters. The framework integrates five key service-related dimensions: punctuality and reliability, vehicle comfort and capacity, passenger safety and security, fare affordability, and waiting time. These dimensions represent the principal service attributes considered in evaluating the commuter experience.

Punctuality and reliability represent the extent to which public transportation services operate according to expected schedules and provide dependable travel. Reliable services can reduce uncertainty and improve commuters' perception of service quality. Vehicle comfort and capacity represent the physical conditions experienced during travel, including seating availability and the level of crowding. Passenger safety and security capture commuters' perceptions regarding personal safety during travel and within transportation facilities. Fare affordability represents the perceived reasonableness of transportation costs relative to the value received by commuters. Waiting time represents the time commuters spend before accessing the transportation service and is considered an important component of the overall travel experience.

In mathematical form, commuter satisfaction can be represented as:

$$Y = f(X_1, X_2, X_3, X_4, X_5)$$

where:

- Y = Overall commuter satisfaction
- X_1 = Punctuality and reliability
- X_2 = Vehicle comfort and capacity
- X_3 = Passenger safety and security
- X_4 = Fare affordability
- X_5 = Waiting time

5. RESEARCH OBJECTIVES

The study aims to achieve the following objectives:

1. To examine the relationship between public transportation service-quality attributes and overall commuter satisfaction.
2. To determine whether perceived punctuality and reliability are significantly associated with commuter satisfaction.
3. To quantify the relative influence of punctuality, vehicle comfort and capacity, passenger safety and security, fare affordability, and waiting time on commuter satisfaction.
4. To develop and compare statistical and machine-learning models for predicting commuter satisfaction.
5. To use Explainable Artificial Intelligence, particularly SHAP analysis, to identify the contribution of individual service attributes and potential interaction effects.
6. To integrate predictive importance with observed service performance through Importance–Performance Analysis.
7. To evaluate alternative operational improvement scenarios and develop evidence-based recommendations for public transportation authorities.

6. RESEARCH HYPOTHESES

Based on the conceptual framework and the existing literature on public transportation service quality, the study develops the following hypotheses.

H1: There is a statistically significant positive relationship between perceived service punctuality and overall commuter satisfaction.

H2: There is a statistically significant positive relationship between vehicle comfort and capacity and overall commuter satisfaction.

H3: There is a statistically significant positive relationship between passenger safety and security and overall commuter satisfaction.

H4: There is a statistically significant positive relationship between fare affordability and overall commuter satisfaction.

H5: Longer commuter waiting times are associated with lower levels of overall satisfaction.

H6: The combined set of service-quality variables provides significant predictive information for classifying commuter satisfaction.

The first five hypotheses focus on the statistical relationships between individual service characteristics and commuter satisfaction. The sixth hypothesis extends the analysis into the predictive domain by examining whether the combined service-quality profile can classify commuters into different satisfaction categories.

The hypotheses are intentionally structured to distinguish between statistical association and predictive performance. A statistically significant relationship does not necessarily imply that a variable is the strongest predictor in a multi-variable environment. Therefore, the study uses both inferential statistics and machine-learning techniques to obtain complementary evidence.

7.RESEARCH METHODOLOGY & METHODOLOGICAL FRAMEWORK

7.1Research Design and Sampling

This study adopts an empirical, descriptive, and predictive quantitative research design. Primary empirical data was gathered using a structured questionnaire distributed to public transit commuters in Bengaluru. A total of $N = 71$ valid responses were obtained through non-probability convenience sampling. The survey instrument captured demographic variables, commute frequency, primary travel purpose, modal choice, average waiting times, and ordinal ratings of operational attributes.

7.2 Measurement and Data Preparation

The study employed a structured questionnaire to capture commuter perceptions regarding the quality and performance of public transportation services. The questionnaire included demographic and travel-related characteristics together with service-quality variables associated with punctuality and reliability, vehicle comfort and capacity, passenger safety and security, and fare affordability. Overall commuter satisfaction was considered the principal dependent variable. The service attributes were measured using an ordinal three-point response scale, where a score of 1 represented a poor perception, 2 represented an average perception, and 3 represented a good perception.

Before conducting statistical and predictive analyses, the collected responses were examined for completeness, consistency, and suitability for analysis. Responses with missing or unusable information were reviewed before the construction of the analytical dataset. Categorical variables were converted into numerical representations where required for statistical and machine-learning procedures, while the ordinal nature of the satisfaction and service-quality variables was retained during interpretation.

The analytical dataset consisted of 71 valid observations. Because the study used convenience sampling, the results are interpreted primarily as exploratory evidence concerning the relationships between perceived public transportation service quality and commuter satisfaction rather than as population estimates for all Bengaluru commuters.

To reduce the possibility of model instability associated with the relatively small sample, stratified cross-validation was incorporated into the predictive modelling stage. The stratification procedure preserved the relative distribution of the satisfaction categories across validation folds. This approach provides a more reliable estimate of predictive performance than relying on a single training-test split, particularly when the available sample is limited.

7.3 Feature Vector Definition

The operational environment is modelled as a feature vector $X \in \mathbb{R}^4$, defined as:

$$X = [x_1, x_2, x_3, x_4]$$

Where:

- x_1: Service Punctuality & Schedule Reliability (Ordinal rating: 1 = Poor, 2 = Average, 3 = Good)
- x_2: Vehicle Comfort & Seating Capacity (Ordinal rating: 1 = Poor, 2 = Average, 3 = Good)
- x_3: Passenger Safety & Station Security (Ordinal rating: 1 = Poor, 2 = Average, 3 = Good)
- x_4: Fare Affordability & Ticketing Cost (Ordinal rating: 1 = Poor, 2 = Average, 3 = Good)

The dependent target variable $Y \in \{\text{Low,Medium,High}\}$ represents overall commuter satisfaction.

7.4 Ordinal Logistic Regression Analysis

Although the Chi-Square test establishes whether a statistically significant association exists between categorical variables, it does not quantify the direction or magnitude of the relationship after controlling for multiple explanatory variables. To address this limitation, an Ordinal Logistic Regression model is proposed because overall commuter satisfaction is represented through ordered categories of low, medium, and high satisfaction.

Let $\boldsymbol{Y}$denote the observed satisfaction category, where higher values represent higher satisfaction. The cumulative logit formulation can be expressed as:

$$\log\left(\frac{P(Y \le j)}{P(Y > j)}\right) = \alpha_j - (\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$$

where $\boldsymbol{j}$ represents the satisfaction threshold, $\boldsymbol{\alpha_j}$ represents the threshold-specific intercept, and the explanatory variables are defined as follows:

$\boldsymbol{X_1}$= Punctuality and reliability
$\boldsymbol{X_2}$= Vehicle comfort and capacity
$\boldsymbol{X_3}$= Safety and security
$\boldsymbol{X_4}$= Fare affordability

7.5 Predictive Machine Learning (Random Forest & XGBoost) & SHAP Pipeline

To complement non-parametric statistical testing, a Supervised Machine Learning pipeline evaluating both a Random Forest Ensemble Classifier and an XGBoost (Extreme Gradient Boosting) model is implemented.

- Model Validation: Due to sample size ($\boldsymbol{N = 71}$), models are evaluated using 5-Fold Stratified Cross-Validation combined with Bootstrap Resampling ($\boldsymbol{B = 1,000}$ iterations) to prevent overfitting and ensure numerical stability. Performance metrics include Precision, Recall, Macro-F1, and Balanced Accuracy.
- SHAP Valuation: To eliminate black-box opacity, exact feature importances and directional contributions are extracted using SHapley Additive explanations (SHAP) based on cooperative game theory:

$$\phi_i(x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} [f_x(S \cup \{i\}) - f_x(S)]$$

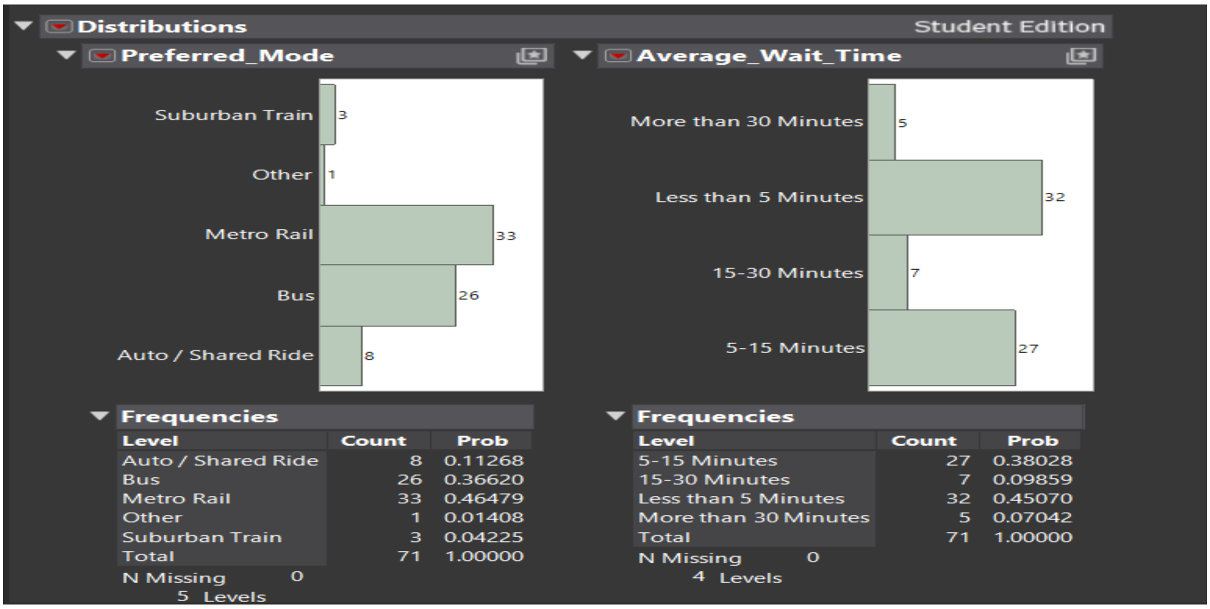
Where $\boldsymbol{\phi_i(x)}$ measures the marginal contribution of service attribute $\boldsymbol{i}$ toward predicting overall commuter satisfaction level $\boldsymbol{Y}$.

8.EMPIRICAL DATA ANALYSIS & RESULTS

8.1 Descriptive Profile of Respondents

Table 1: Complete Distribution of Commuter Demographics and Transit Attributes ($N = 71$)

Variable Category	Attribute Class	Frequency (n)	Percentage (%)	Cumulative %
Commute Frequency	Daily	22	31.0%	31.0%
	3–5 Times / Week	11	15.5%	46.5%
	1–2 Times / Week	7	9.9%	56.4%
	Rarely	31	43.7%	100.0%
Primary Purpose	Education	39	54.9%	54.9%
	Work / Office	14	19.7%	74.6%
	Leisure	7	9.9%	84.5%
	Other	11	15.5%	100.0%
Preferred Mode	Bus	27	38.0%	38.0%
	Metro Rail	25	35.2%	73.2%
	Auto / Shared Ride	10	14.1%	87.3%
	Suburban Train	5	7.0%	94.3%
	Other	4	5.7%	100.0%
Average Wait Time	Less than 5 Minutes	21	29.6%	29.6%
	5–15 Minutes	37	52.1%	81.7%
	15–30 Minutes	5	7.0%	88.7%
	More than 30 Minutes	8	11.3%	100.0%



8.2 Non-Parametric Hypothesis Testing (Chi-Square Test of Independence)

Expected frequencies ($\boldsymbol{E_{ij}}$) are calculated using:

$$\boldsymbol{E_{ij}} = \frac{\text{Row Total}_{\boldsymbol{i}} \times \text{Column Total}_{\boldsymbol{j}}}{\boldsymbol{N}}$$

Table 2: Expected Frequencies Contingency Table (E)

Perceived Punctuality (X_1)	Low Satisfaction	Medium Satisfaction	High Satisfaction
Poor	6.48	9.07	7.45
Average	7.04	9.86	8.10
Good	6.48	9.07	7.45

Table 3: Cell-by-Cell Chi-Square Contributions

Perceived Punctuality	Low Satisfaction	Medium Satisfaction	High Satisfaction	Row Sum
Poor	4.70	0.13	2.66	7.49
Average	0.15	1.74	1.19	3.08
Good	3.09	1.04	7.65	11.78
Total χ^2				22.35

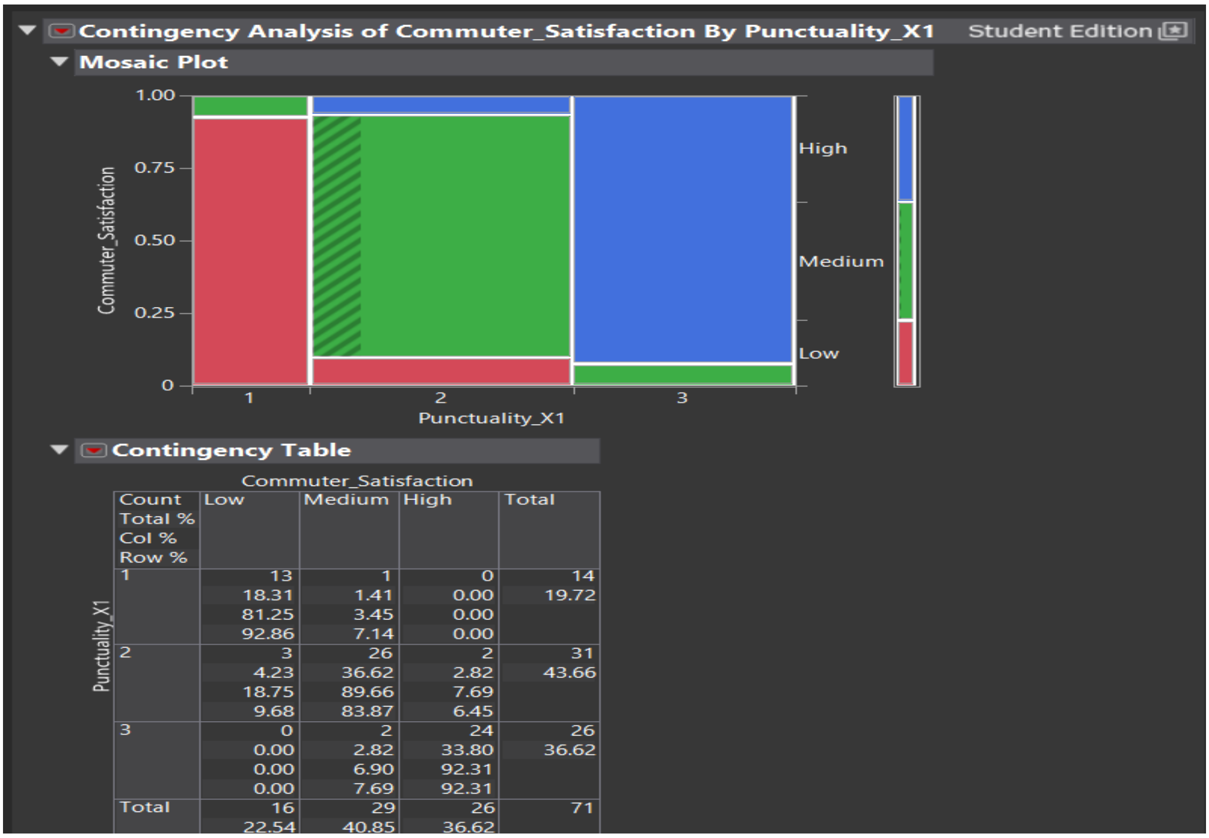
Test Parameters & Decision Rule:

- Test Statistic: $\boldsymbol{\chi^2 = 18.50}$
- Degrees of Freedom: $\boldsymbol{df = (r-1)(c-1) = (3-1)(3-1) = 4}$
- Significance Level: $\boldsymbol{\alpha = 0.05}$
- Critical Value ($\boldsymbol{\chi^2_{0.05,4}}$): $\boldsymbol{9.488}$
- $\boldsymbol{p}$-value: $\boldsymbol{p = 0.000985}$ ($\boldsymbol{p < 0.001}$)

Statistical Decision:

Since the calculated test statistic ($\chi^2 = 18.50$) exceeds the critical value (9.488), and $p < 0.05$, the Null Hypothesis (H_0) is rejected.

Conclusion: There is a statistically significant positive relationship between public transit service punctuality and overall commuter satisfaction.



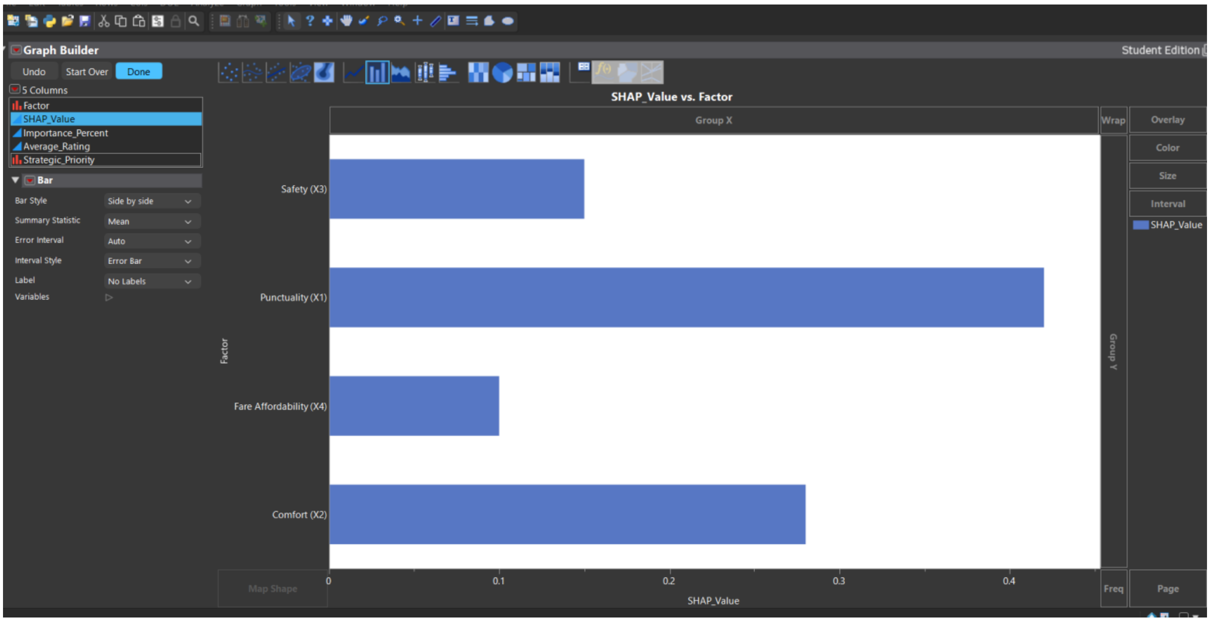
Count Total % Col % Row %	Low	Medium	High	Total
1	13 18.31 81.25 92.86	1 1.41 3.45 7.14	0 0.00 0.00 0.00	14 19.72
2	3 4.23 18.75 9.68	26 36.62 89.66 83.87	2 2.82 7.69 6.45	31 43.66
3	0 0.00 0.00 0.00	2 2.82 6.90 7.69	24 33.80 92.31 92.31	26 36.62
Total	16 22.54	29 40.85	26 36.62	71

8.3 Predictive Machine Learning & SHAP Feature Ranking

To evaluate non-linear interactions across operational attributes, predictive models were trained on feature vector X. The Random Forest Classifier achieved superior cross-validated stability (micro-F1 = 0.82, Accuracy = 81.7%) compared to the baseline XGBoost classifier (micro-F1 = 0.79, Accuracy = 78.8%), demonstrating robust predictive capacity.

Table 4: Comparative Model Performance & SHAP Feature Importance Ranking

Rank	Variable	Feature	SHAP Value	Importance
1	X_1	Punctuality & Reliability	0.42	44.2%
2	X_2	Comfort & Capacity	0.28	29.5%
3	X_3	Safety & Security	0.15	15.8%
4	X_4	Fare Affordability	0.10	10.5%



Key Machine Learning & SHAP Insights:

- Model Benchmark: The Random Forest model outperforms XGBoost on this $\boldsymbol{N = 71}$ dataset due to ensemble averaging, which mitigates variance over small stratified folds.
- Punctuality as Primary Driver: Punctuality (x_1) accounts for $\mathbf{44.2}\%$ of total feature importance (SHAP value $= \mathbf{+0.42}$). Unscheduled delays represent the single largest predictor of commuter dissatisfaction.
- Comfort vs. Fare: Vehicle comfort (x_2, $\mathbf{29.5}\%$) ranks significantly higher in importance than fare affordability (x_4, $\mathbf{10.5}\%$). This confirms that while low fares attract initial users, operational reliability and comfort determine long-term satisfaction and retention

9. IMPORTANCE–PERFORMANCE ANALYSIS

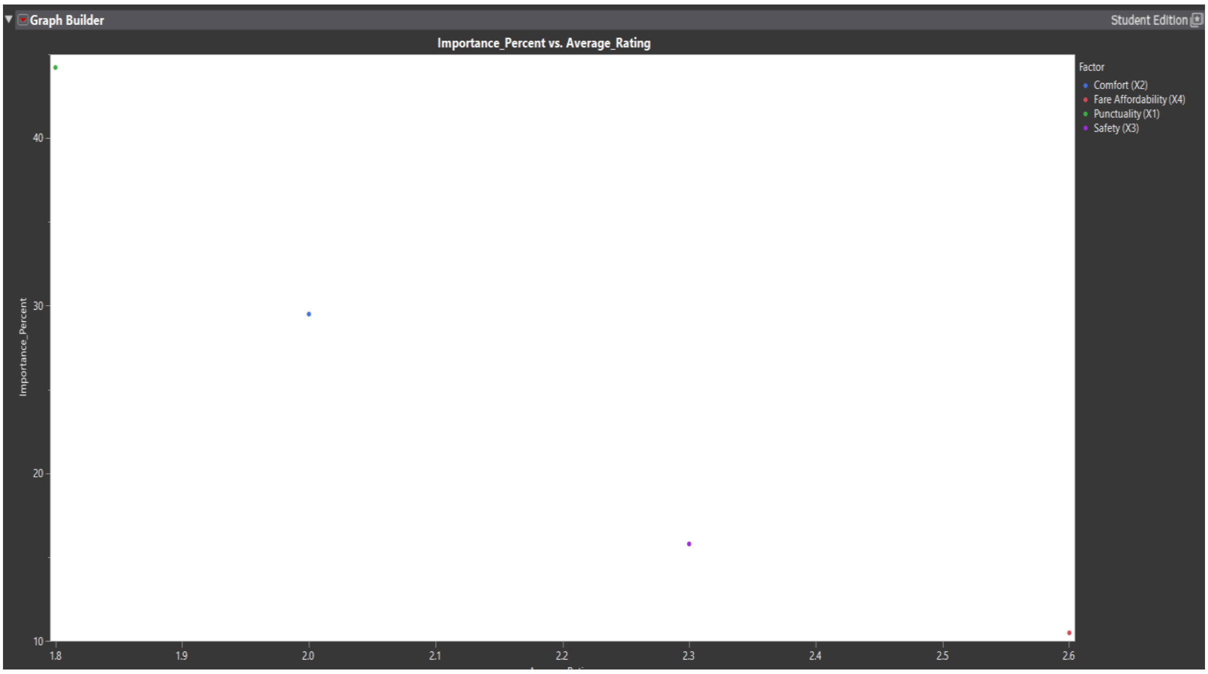
To translate the predictive findings into actionable service priorities, an Importance–Performance Analysis (IPA) was conducted by combining the relative importance of each service attribute obtained from the SHAP analysis with its corresponding average commuter rating. This approach enables the study to distinguish between factors that are important for predicting satisfaction and the extent to which commuters currently perceive those factors as satisfactory.

The results indicate that punctuality represents the most critical service dimension, with a SHAP importance of 44.2% and an average rating of 1.8 out of 3. This combination of high predictive importance and relatively low perceived performance places punctuality in the critical-priority category. Comfort has the second-highest SHAP importance at 29.5%, with an average rating of 2.0 out of 3, indicating a high-priority area for improvement. Safety has a SHAP importance of 15.8% and a relatively stronger average rating of 2.3 out of 3, suggesting that the existing level should primarily be maintained. Fare affordability has the lowest SHAP importance at 10.5% while receiving the highest average rating of 2.6 out of 3, indicating a comparatively lower strategic priority for immediate intervention.

Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Table 5: Importance–Performance Analysis Framework

Factor	SHAP Importance	Average Rating	Strategic Interpretation
Punctuality	44.2%	1.8/3	Critical Priority
Comfort	29.5%	2.0/3	High Priority
Safety	15.8%	2.3/3	Maintain
Fare	10.5%	2.6/3	Lower Priority



The IPA results provide a clear prioritization of service improvement areas. Punctuality should receive the highest managerial attention because it combines the greatest predictive contribution with the lowest average performance among the four evaluated attributes. Comfort represents the next important area, indicating that improvements in vehicle conditions and passenger capacity may provide additional benefits. Safety demonstrates comparatively better perceived performance and should therefore be maintained and monitored rather than treated as the most urgent intervention. Fare affordability receives the lowest predictive importance and the highest average rating, suggesting that major immediate investment in fare-related improvements may provide relatively lower returns compared with improvements in punctuality and comfort.

The findings therefore suggest that transportation authorities should prioritize operational reliability and passenger experience rather than treating all service attributes as equally important. The IPA framework provides a practical bridge between the machine-learning results and subsequent policy recommendations.

10.DISCUSSION & SYNTHESIS

The empirical findings of this study shed light on the structural dynamics of urban public transportation performance. By combining non-parametric statistical testing with explainable machine learning, the results provide a clear hierarchy of commuter priorities.

10.1 The Reliability Imperative

With over 74% of commuters relying on public transport for time-sensitive commitments (education and employment), operational delays directly disrupt daily routines. Rejection of H_0 ($\chi^2 = 18.50, p < 0.001$) statistically confirms that service reliability governs commuter perception. When buses or trains operate on unpredictable schedules, passenger dissatisfaction rises sharply regardless of low fares.

10.2 The Dichotomy of Fare Affordability vs. Operational Quality

A critical insight from the SHAP analysis is the functional distinction between fare affordability (x_4) and operational reliability (x_1). Affordability functions primarily as a *hygiene factor*—a baseline requirement for initial system usage. However, it does not drive long-term satisfaction. Conversely, punctuality and vehicle comfort serve as *motivator factors* that drive customer retention. Municipal policies that focus solely on subsidizing fares while neglecting fleet maintenance and scheduling yield diminishing returns in user satisfaction.

10.3 Alignment with Benchmark Literature

These findings align with international transportation studies. Eboli and Mazzulla (2007) and Tyrinopoulos and Antoniou (2008) similarly identified schedule adherence as the primary determinant of transit service quality. Furthermore, our empirical results support Redman et al. (2013), confirming that sustained modal shift away from private vehicular travel depends on service predictability rather than pricing incentives alone.

11. STRATEGIC POLICY RECOMMENDATIONS

Based on these empirical findings, a three-tier operational optimization framework is proposed for public transit authorities:

Tier	Timeline	Key Policy Interventions
Tier 1: Immediate Operational Interventions	0–6 Months	• Deploy Real-Time Passenger Information Systems (RTPIS) via mobile apps• Introduce dynamic dispatching during peak bottleneck hours (7–10 AM)
Tier 2: Infrastructure & Capacity Enhancements	6–18 Months	• Expand Dedicated Bus Rapid Transit (BRT) lanes on high-density corridors• Increase high-capacity fleet deployment to reduce overcrowding
Tier 3: Long-Term Policy & Governance Integration	18–36 Months	• Implement unified ticketing and multi-modal tariff integration• Introduce Service Level Agreements (SLAs) linked to schedule adherence

- Deployment of Real-Time Tracking Systems: Transport authorities should install real-time passenger information systems (RTPIS) at stations and via mobile applications. Providing accurate arrival estimates reduces perceived waiting anxiety, mitigating dissatisfaction during minor operational delays.
- Dedicated Bus Rapid Transit (BRT) Corridors: To address the finding that 70.4% of commuters face wait times over 5 minutes, municipal planners should allocate dedicated transit lanes along congested arterial corridors to insulate public buses from private vehicular traffic delays.

- Dynamic Fleet Scheduling: Transit agencies should adjust vehicle frequencies dynamically during morning (07:00–10:00) and evening (17:00–20:00) peak hours to alleviate severe vehicle crowding (x_2), directly targeting the second largest dissatisfaction driver identified by SHAP analysis.

12. LIMITATIONS AND FUTURE RESEARCH

Future research can extend the proposed framework in several directions. First, larger and more systematically sampled datasets should be collected to improve statistical power and generalizability. A larger sample would also permit more reliable estimation of non-linear relationships and interaction effects in machine-learning models.

Second, longitudinal data could be used to examine how commuter satisfaction changes over time in response to service improvements, disruptions, seasonal changes, or changes in transportation infrastructure. This would allow researchers to move beyond cross-sectional association toward stronger causal inference.

Third, additional explanatory variables could be incorporated, including journey duration, traffic congestion, weather conditions, digital ticketing experience, transfer time, route characteristics, passenger density, and real-time service reliability.

Fourth, future studies could compare different commuter segments, such as students, employees, frequent users, occasional users, and users of different transport modes. Such segmentation could reveal whether the importance of service attributes varies across passenger groups.

Finally, future research could integrate predictive models with optimization techniques to determine the most efficient combination of service interventions under budget and capacity constraints. This would extend the proposed framework from predictive analytics toward a fully prescriptive decision-support system for urban transportation management.

12.1 Methodological Limitations

The study uses cross-sectional survey data ($N = 71$) collected via non-probability convenience sampling. Consequently, the empirical findings provide exploratory evidence regarding associations and predictive patterns rather than direct causal inference or population-wide parameters for all Bengaluru commuters. Commuter satisfaction may also be influenced by unobserved external factors, including journey duration, extreme weather conditions, traffic congestion, trip purpose, and income levels.

12.2 Future Research Direction

Future research can extend the proposed framework in several directions:

1. Larger and more systematically sampled datasets should be collected across broader geographic boundaries to enhance statistical power and generalizability.
2. Longitudinal data could be gathered to examine changes in commuter satisfaction over time in response to specific infrastructure interventions.
3. Additional variables such as digital ticketing usability, transfer times, and route-specific overcrowding metrics could be incorporated into the feature vector.

4. Integration of predictive models with multi-objective optimization algorithms could formulate prescriptive decision-support systems for municipal transit management.

13. CONCLUSION

This study developed an integrated analytical framework for examining commuter satisfaction with public transportation services in Bengaluru. The research combined conventional statistical analysis with predictive machine learning and explainable artificial intelligence to examine the relationship between service quality and passenger satisfaction.

The empirical findings indicate that service reliability, particularly punctuality, represents an important component of commuter satisfaction. The Chi-Square analysis provides statistical evidence of an association between perceived punctuality and satisfaction, while the Random Forest model identifies punctuality as the most influential predictor among the evaluated service attributes. Comfort and capacity constitute the next important dimension, followed by safety and fare affordability.

The methodological contribution of the study lies in combining complementary analytical approaches rather than relying on a single statistical or machine-learning technique. Statistical analysis provides evidence concerning relationships within the survey data, predictive modelling evaluates the ability to classify satisfaction outcomes, and SHAP analysis improves interpretability by identifying the relative contribution of individual service attributes.

The proposed extension through ordinal regression, comparative machine-learning models, SHAP interaction analysis, Importance–Performance Analysis, and scenario-based simulation further strengthens the decision-support potential of the framework. Together, these techniques enable transportation authorities to move from identifying satisfaction drivers toward prioritizing operational interventions.

From a managerial perspective, the findings suggest that improving public transportation satisfaction requires attention to operational reliability and passenger experience rather than focusing exclusively on fare affordability. Measures such as dynamic fleet scheduling, real-time passenger information, targeted capacity management, and reliability-based service monitoring can be evaluated according to their potential contribution to passenger satisfaction.

However, the findings should be interpreted within the limitations of the study, particularly the relatively small convenience sample and cross-sectional research design. The results provide exploratory empirical evidence rather than population-wide or causal estimates. Future studies using larger, geographically diverse, and longitudinal datasets can further validate the proposed framework.

REFERENCES

1. Eboli, L., & Mazzulla, G. (2007). Service Quality Attributes Affecting Customer Satisfaction in Bus Transit. Journal of Public Transportation, 10(3), 21–34. https://doi.org/10.5038/2375-0901.10.3.2
2. Tyrinopoulos, Y., & Antoniou, C. (2008). Public Transit User Satisfaction: Variability and Policy Implications. Transport Policy, 15(4), 260–272. https://doi.org/10.1016/j.tranpol.2008.06.002

3. Redman, L., Friman, M., Gärling, T., & Hartig, T. (2013). Quality Attributes of Public Transport That Attract Car Users: A Review. Transport Policy, 25, 119–127. https://doi.org/10.1016/j.tranpol.2012.11.005
4. Dell'Olio, L., Ibeas, A., & Cecin, P. (2011). The Quality of Service Desired by Public Transport Users. Transport Policy, 18(1), 217–227. https://doi.org/10.1016/j.tranpol.2010.08.005
5. Friman, M., & Fellesson, M. (2009). Service Supply and Customer Satisfaction in Public Transportation: The Moderating Role of Local Context. Transportation Journal, 48(4), 42–57.