



Data Article

Psychological determinants of transport mode choice: A comprehensive dataset from Pabna Municipality, Bangladesh



Md. Nayem Hossain^{a,*}, Nakib Aman^b, Md. Rakibul Islam Sabbir^c,
Md. Rashedul Haque^a, MD. Tahidul Islam^b

^a Department of Civil Engineering, Pabna University of Science and Technology, Pabna-6600, Bangladesh

^b Department of Computer Science and Engineering, Pabna University of Science and Technology, Pabna-6600, Bangladesh

^c Department of Urban and Regional Planning, Pabna University of Science and Technology, Pabna-6600, Bangladesh

ARTICLE INFO

Article history:

Received 25 October 2025

Revised 23 December 2025

Accepted 3 February 2026

Available online 13 February 2026

Dataset link: [A Comprehensive Dataset of Transportation Mode Choice \(Original data\)](#)

Keywords:

Urban transport

Psychological factors

Transport mode choice

Sustainable transport planning

Machine learning

ABSTRACT

This data article introduces a detailed and structured dataset derived from a survey exploring the psychological factors that influence transport mode choices in Pabna Municipality, northwest of Bangladesh. This dataset covers a wide range of socio-demographic factors, i.e., income, occupation level, and age, as well as the factors related to the choice of the mode of transportation, money expenditure, and a set of psychological determinants, i.e., the perceived safety, financial constraint, and environmental consciousness. These aspects are often under-represented in existing research. The data were collected using a carefully structured questionnaire survey of 348 samples conducted over a three weeks survey of urban residents on different points. The dataset is an important resource to researchers working on urban planning, transportation behaviour analysis, environmental psychology and public health as it provides a more detailed understanding of how psychological factors interact with mode choice in transportation. This will help in policy-making, to contribute to the broader field of urban transport planning. To validate the reliability and predictive capability of the dataset, several supervised machine learning models were applied,

* Corresponding author.

E-mail address: nayem.191101@s.pust.ac.bd (Md. N. Hossain).

Social media: [@MdNayemHos64119](#) (Md. N. Hossain)

demonstrating strong classification performance for transport mode choice. The dataset can be used in future research on inclusion-based, user-friendly transport systems especially in Bangladesh as well as other rapidly urbanising regions, as it is in line with global sustainability.

© 2026 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>)

Specifications Table

Subject	Urban Transportation.
Specific subject area	Transport planning, Transportation safety, Public health, Sustainable mobility.
Type of data	Table and raw, Cleaned, well-structured data.
Data collection	Data were collected through a structured questionnaire survey involving 348 participants. Data were recorded in a field notebook during face-to-face interview.
Data source location	Data collected in Municipality Area, which lies within the Pabna District Under Rajshahi Division, Bangladesh.
Data accessibility	Repository name: Mendeley Data Data identification number: https://doi.org/10.17632/9kgzcvwm7.1 Direct URL to data: A Comprehensive Dataset of Transportation Mode Choice - Mendeley Data
Related research article	None

1. Value of the Data

- The data source provides a detailed analysis of the degree to which psychological and socio-demographical variables such as perceptions of safety, financial limitations, and environmental concerns influence transport mode choices among residents living in an urban area.
- It can provide constructive ideas to urban planners and policymakers to design a transport system that is beneficial for different perspectives, enhancing transport performance and the quality of life of its populations.
- It has the potential to be used by public health researchers to investigate the linkage between transport mode choice and mental well-being, to allow a more hybrid focus on transport policies and health in urban areas.
- The data covers a wide range of psychological and practical factors that have an influence on the selection of transport modes. The data can enable the analysis of transport behaviour to be considered in a more multi-dimensional way by considering emotional or mental aspects (e.g. safety perceptions, anxiety, past negative experiences) as well as conventional aspects (e.g. cost, distance).
- The dataset exists as an optimized structure that supports both statistical analysis and machine learning modelling operations. It has already been applied to train both machine learning and Artificial Neural Network with high accuracy through this data which has been successfully applied in training processes. This demonstrates the dataset's utility for advanced AI analysis methods.
- Researchers in other regions can implement the data collection process as a benchmark for creating comparable valuable databases particularly in low and middle budget contexts. The dataset may form a basis for comparative future research in similar urban contexts, and it may develop interdisciplinary research on urban development, transportation economics, and behavioural science on mode selection.

2. Background

Travel-mode choice is one of the key determinants and critical factor in shaping urban mobility and traffic flows distribution, and such variables significantly influence the mode choice [1]. The correct choice of mode could significantly decrease the time waste, and increase productivity rates in the most densely populated surroundings [2]. There are individual, social, and infrastructural factors affecting urban mode choice; among them, the individual motivation and the sense of personal control are the key factors [3]. Factors influencing mode choice for work trips include travel distance, time, availability, health benefits, comfort, safety, cost, and environmental impact [4]. Convenience, safety, and affordability are some of the elements that can determine the travel mode of families in urban areas [5]. Most research on transport in Bangladesh tends to concentrate on the major cities, leaving smaller municipalities largely unexplored. The transport mode choice is determined by the individual, social, and infrastructural reasons with the top factors being safety, comfort, convenience, and length of travel, and the walking distance is considered as a random parameter [6–8]. Moreover, the application of the dataset goes well beyond Pabna Municipality. Researchers can use it in developing more sophisticated models of transport, and policymakers can develop transport solutions.

3. Data Description

This dataset describes the complexity of the factors that determine transport mode choice and effects it has on individuals. It is structured in some categories that include socio-demographic, transport preferences, psychological factors, and mental or emotional well-being. This survey was conducted over three weeks in 2025 in English for the questionnaire, while explaining the audience instruction in Bangla, their native language to maximizing participation. The essential different variables are to be gathered by the questionnaire. The types of variables contained in the dataset are as illustrated below:

- **Socio-Demographic Variables:** Age, Gender, Occupation, Income Source, Monthly Income, Origin, Destination.
- **Transport Mode Variables:** Mode of Transport, Average Expenditure for Transportation, Travel Objective, Distance Influence on Mode Choice, Importance of Safety in Mode Choice, Mobility Barrier on Mode Choice, Accident Memories on Mode Choice, Environmental Impact on Mode Choice, Usually Selected Specific Transport Mode, Usually Avoided Specific Transport Mode.
- **Psychological Factors:** Importance of Cost in Mode Choice, Effect of Cost on Mode Choice, Daily Activity Influence on Mode Choice, Perceived Safety in Bus, Perceived Safety in Unperceived Safety in Easybike, Perceived Safety in Rickshaw, Perceived Safety in Bike, Perceived Safety in Car, Perceived Safety in Cycle.
- **Mental and Emotional Well-being:** Depression, Happiness, Mental Stress, Urgency, Weather Condition.

The data is organized within a folder structure that includes different categories of data files, as shown below:

Main File.zip:

1. Survey Data Folder:

- **Raw Data (.xlsx):** This file contains the unprocessed survey responses from the initial data collection from the survey.
- **Questionnaire (.docx):** The document containing the actual survey questions that were asked of participants during the survey.
- **Cleaned Data (.xlsx):** The file contains the cleaned version of survey data with missing values handled and prepared for use in machine learning and other statistical analysis.

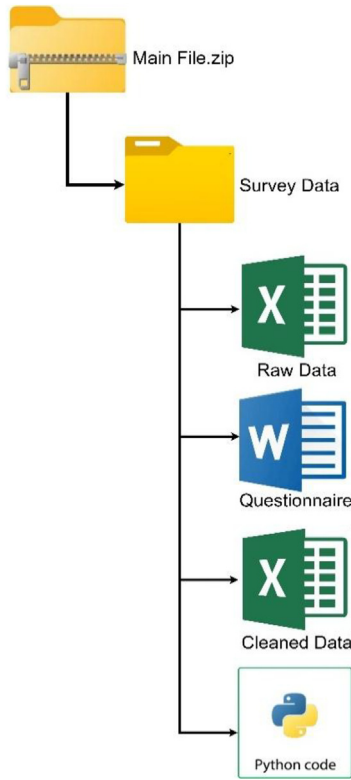


Fig. 1. Structure of the main dataset folder.

- **Python Code (.ipynb):** The jupyter notebooks file contain Python code that was used in executing machine learning operations.

The architecture of the dataset comprises the Main File.zip, which contains the major directories of Survey Data, which include the main Survey Data. In this folder, files related to Raw Data, Questionnaire, Cleaned Data, and Python Code will be found as shown in the diagram below Fig. 1.

The dataset represents a balanced and diverse urban population. Among the respondents, 59.5 % were male and 40.5 % female, with ages ranging from 13 to 67 years and a mean age of 29.4 years ($SD = 11.2$). Respondents were distributed across all 15 municipal wards (origin) and reported trip destinations spanning 19 different ward locations, confirming adequate geographic dispersion. Occupational composition included students (48.3 %), businesspersons (12.9 %), housewives (12.4 %), service employees (11.5 %), teachers (3.7 %), and a small fraction of others (11.2 %) such as retired or unemployed individuals. All collected responses were entered and cleaned in Microsoft Excel, followed by verification through SPSS. Missing values (<2 %) were treated using mean substitution for continuous variables and mode imputation for categorical variables. The dataset was further formatted and validated for consistency before conversion into machine learning-ready form. No weighting adjustments or calibration weights were applied, as the sample proportions closely reflected the municipal census distribution. The following tables provide a detailed analysis of the survey data collected from respondents across Pabna Municipality (Tables 1–5).

This information is useful to learn more about trust rates in the safety of vehicles in different transportation modes, can give better insights into how mental processes that are related to

Table 1
Descriptive statistics of dataset.

Category	Values (Frequency)
Gender	Male (207), Female (141).
Occupation	Advocate (4), Business (45), Doctor (3), Driver (9), Farmer (1), House wife (43), Job (40), Labourer (11), Student (168), Tailor (5), Teacher (13), Unemployment (6).
Income source	Yes (157), No (191).
Travel Objective	Education (174), Medical/Recreation (38), Shopping (50), Work (86).
Origin Ward No.	Ward-01 (29), Ward-02 (22), Ward-03 (27), Ward-04 (24), Ward-05 (24), Ward-06 (22), Ward-07 (20), Ward-08 (22), Ward-09 (24), Ward-10 (26), Ward-11 (21), Ward-12 (19), Ward-13 (24), Ward-14 (20), Ward-15 (24).
Destination Ward No.	Ward-01 (15), Ward-02 (64), Ward-03 (25), Ward-04 (3), Ward-05 (12), Ward-06 (4), Ward-07 (39), Ward-08 (9), Ward-09 (14), Ward-10 (11), Ward-11 (66), Ward-12 (11), Ward-13 (19), Ward-14 (11), Ward-15 (3), Outside East (22), Outside South (10), Outside West (8), Outside North (2).
Usually selected Transport Mode	Bike (33), Bus (60), Car (13), CNG (10), Cycle (4), Easybike (171), Rickshaw (57).
Reason of Transport Mode	Availability (55), Comfort (4), Cost effectiveness (48), Health benefit (9), Job (16), Safety (74), Time savings (134), Weather condition (8).
Usually Avoided specific Transport Mode	Bike (60), Bus (40), Car (8), CNG (27), Cycle (14), Easybike (9), Rickshaw (5), None (185).
Why you avoid this transport mode	Health issue (24), Safety (86), Time saving (16), Weather Condition (6), Others (20), None (196).
Average Expenditure for transportation	500tk (89), 1000tk (137), 2000tk (72), 3000tk (50).
Effect of Cost on Mode Choice	Yes (160), No (188).
Daily Activity Influence Mode choice	Yes (116), No (232).
Daily Activity Influence Mode choice	Yes (155), No (193).
Importance of Cost in Mode Choice	Extreme (18), High (85), Medium (171), Low (48), No-effect (26).
Importance of Safety in Mode Choice	Extreme (17), High (84), Medium (184), Low (45), No-effect (18).
Mobility Barrier on Mode Choice	Yes (227), No (121).
Accident Memories on Mode Choice	Yes (102), No (246).
Environmental Impact on Mode Choice	Extreme (41), High (146), Medium (70), Low (55), No-effect (36).
Trust level based on safety in Bus	Extreme (44), High (96), Medium (153), Low (49), No-effect (6).
Trust level based on safety in CNG	Extreme (17), High (49), Medium (143), Low (134), No-effect (5).
Trust level based on safety in Easybike	Extreme (36), High (75), Medium (178), Low (57), No-effect (2).
Trust level based on safety in Rickshaw	Extreme (49), High (101), Medium (136), Low (59), No-effect (3).
Trust level based on safety in Bike	Extreme (8), High (59), Medium (111), Low (146), No-effect (24).
Trust level based on safety in Car	Extreme (71), High (153), Medium (77), Low (19), No-effect (28).
Trust level based on safety in Cycle	Extreme (119), High (41), Medium (54), Low (108), No-effect (26).
Depression	Bike (75), Bus (34), Car (22), CNG (6), Cycle (9), Easybike (31), Rickshaw (171).
Happiness	Bike (103), Bus (39), Car (59), CNG (4), Cycle (7), Easybike (24), Rickshaw (112).
Mental stress	Bike (68), Bus (45), Car (30), CNG (9), Cycle (19), Easybike (46), Rickshaw (131).
Urgency	Bike (77), Bus (37), Car (79), CNG (106), Easybike (22), Rickshaw (27).
Weather Condition	Bike (46), Bus (43), Car (38), CNG (19), Cycle (4), Easybike (72), Rickshaw (126).

Table 2

Trust level based on safety across different transport modes.

Statistics Measures	Trust level based on safety in Bus	Trust level based on safety in CNG	Trust level based on safety in Easybike	Trust level based on safety in Rickshaw	Trust level based on safety in Bike	Trust level based on safety in Car	Trust level based on safety in Cycle
Valid Responses(N)	348	348	348	348	348	348	348
Missing Responses	0	0	0	0	0	0	0
Mean	2.65	3.18	2.75	2.61	3.34	2.37	2.66
Std. Error of Mean	0.05	0.046	0.047	0.051	0.049	0.06	0.075
Median	3	3	3	3	3	2	3
Std. Deviation	0.932	0.866	0.87	0.955	0.918	1.112	1.408
Variance	0.869	0.75	0.757	0.912	0.842	1.236	1.984
Sum	921	1105	958	910	1163	824	925

Table 3

Selected and avoidance of specific transport modes.

	Bike	Bus	Car	CNG	Cycle	Easybike	None	Rickshaw
Selected specific transport mode	9.5 %	17.2 %	3.7 %	2.9 %	1.1 %	49.1 %	0.0 %	16.4 %
Avoided specific transport mode	17.2 %	11.5 %	2.3 %	7.8 %	4.0 %	2.6 %	53.2 %	1.4 %

psychology in the urban transportation decision-making process using the perception of safety as an example.

To assess the predictive capability of the dataset, a set of supervised machine learning algorithms was applied to classify transport mode choice. The performance comparison in [Table 6](#) shows that ensemble-based models consistently outperformed traditional classifiers. The Extra Trees Classifier achieved the highest performance with an accuracy of 0.8887 and an AUC of 0.9837, closely followed by the Random Forest Classifier and Light Gradient Boosting Machine, indicating strong discriminative power and robust generalization. Gradient boosting-based methods such as XGBoost and Gradient Boosting Classifier also showed competitive results, while distance-based (K-Nearest Neighbors) and rule-based (Decision Tree) models demonstrated moderate performance. In contrast, simpler probabilistic models like Naive Bayes and the Dummy Classifier yielded significantly lower accuracy, reflecting the non-linear and multi-dimensional nature of transport decision behavior.

This data can be of crucial value to an urban planner or a policymaker who develops a transportation process that shall consider safety of the people and also prepare efficient sustainable mobility. Furthermore, the data can be used as a benchmark in comparative studies of the other developing regions, which helps to present general research on the transport behaviour, safety, or mental well-being.

A correlation matrix was generated to examine the relationships among the socio-demographic factors, travel characteristics, and mode choice determinants included in the dataset. As shown in [Fig. 2](#), most variable pairs exhibit weak to moderate correlations, indicating low multicollinearity and suggesting that the variables capture distinct behavioral dimensions. Notably, income source shows a moderate positive correlation with destination (ward no.) (0.59), implying that higher-income individuals tend to travel toward specific wards. Average expenditure for transportation is moderately associated with importance of cost in mode choice (0.41), reflecting that individuals who spend more on travel are more cost-conscious when selecting transport modes. Additionally, distance influence on mode choice correlates positively with effect of cost on mode choice (0.34) and daily activity influence (0.24), indicating that longer travel distances amplify both cost sensitivity and trip-purpose relevance in transport decisions. Conversely, some negative associations were observed, such as between occupation and travel objective (-0.57), suggesting differing trip purposes across occupational groups. Overall,

Table 4

Descriptive statistics for trust level.

Trust level of Different Vehicles	Selecting vehicles	Percentages	Coefficient (Mean, SD)
Trust level based on safety in Bus	Extreme	12.6 %	Mean: 2.65 Standard Deviation: 0.932
	High	27.6 %	
	Medium	44.0 %	
	Low	14.1 %	
	No effect	1.7 %	
Trust level based on safety in CNG	Extreme	4.9 %	Mean: 3.18 Standard Deviation: 0.866
	High	14.1 %	
	Medium	41.1 %	
	Low	38.5 %	
	No effect	1.4 %	
Trust level based on safety in Easy bike	Extreme	10.3 %	Mean: 2.75 Standard Deviation: 0.87
	High	21.6 %	
	Medium	51.1 %	
	Low	16.4 %	
	No effect	0.6 %	
Trust level based on safety in Rickshaw	Extreme	14.1 %	Mean: 2.61 Standard Deviation: 0.955
	High	29.0 %	
	Medium	39.1 %	
	Low	17.0 %	
	No effect	0.9 %	
Trust level based on safety in Bike	Extreme	2.3 %	Mean: 3.34 Standard Deviation: 0.918
	High	17.0 %	
	Medium	31.9 %	
	Low	42.0 %	
	No effect	6.9 %	
Trust level based on safety in Car	Extreme	20.4 %	Mean: 2.37 Standard Deviation: 1.112
	High	44.0 %	
	Medium	22.1 %	
	Low	5.5 %	
	No effect	8.0 %	
Trust level based on safety in Cycle	Extreme	34.2 %	Mean: 2.66 Standard Deviation: 1.408
	High	11.8 %	
	Medium	15.5 %	
	Low	31.0 %	
	No effect	7.5 %	

Table 5

Impact of vehicle types on mental and emotional well-being.

Vehicles Types	Depression	Happiness	Mental stress	Urgency	Weather condition
Bike	21.5 %	29.6 %	19.5 %	22.1 %	13.2 %
Bus	9.8 %	11.2 %	12.9 %	10.6 %	12.4 %
Car	6.3 %	17.0 %	8.6 %	22.7 %	10.9 %
CNG	1.7 %	1.1 %	2.6 %	30.5 %	5.5 %
Cycle	2.6 %	2.0 %	5.5 %	0.0 %	1.1 %
Easybike	8.9 %	6.9 %	13.2 %	6.3 %	20.7 %
Rickshaw	49.1 %	32.2 %	37.6 %	7.8 %	36.2 %

the correlation structure supports the diversity and independence of variables included in the dataset, making it suitable for further multivariate and predictive modeling.

Table 6
Dataset validation and multiclass emotion classification using dataset.

Model	Classifier	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
et	Extra Trees Classifier	0.8887	0.9837	0.8887	0.8916	0.8878	0.8664	0.8674
rf	Random Forest Classifier	0.8811	0.9823	0.8811	0.8839	0.8798	0.8573	0.8584
lightgbm	Light Gradient Boosting Machine	0.8689	0.9802	0.8689	0.8731	0.8683	0.8427	0.8436
xgboost	Extreme Gradient Boosting	0.8643	0.9793	0.8643	0.8677	0.8632	0.8372	0.8384
gbc	Gradient Boosting Classifier	0.8475	0.9625	0.8475	0.8560	0.8463	0.8171	0.8190
knn	K Neighbors Classifier	0.8003	0.9546	0.8003	0.8011	0.7925	0.7604	0.7631
dt	Decision Tree Classifier	0.7972	0.8783	0.7972	0.7980	0.7948	0.7567	0.7580
nb	Naive Bayes	0.4298	0.8239	0.4298	0.5875	0.3764	0.3161	0.3657
dummy	Dummy Classifier	0.1662	0.5000	0.1662	0.0276	0.0474	0.0000	0.0000

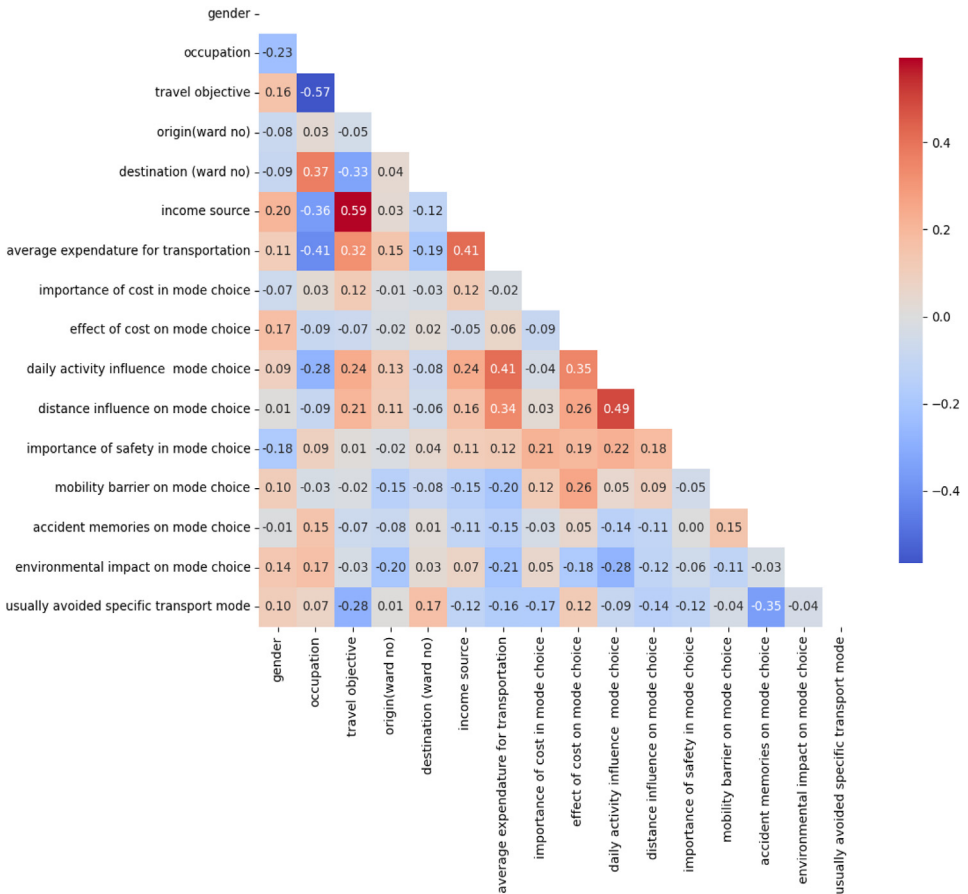


Fig. 2. Correlation of the selected features.

4. Experimental Design, Materials and Methods

Pabna being an area of diverse land use, increased urbanization, and developing transport infrastructure (road, rail, and river) is a dynamic case when thinking of transport mode choice. In this study, the municipality of Pabna in the Bangladesh region is taken as the study area. It was established in 1876 and promoted to municipality type in 1989 to type "A". Pabna municipality refers to a local administrative unit in the Pabna district of the Rajshahi division that is found between the latitudes of 23°59'15"N and 24°02'15"N, and the longitudes of 89°12'00"E and 89°15'45"E (Fig. 3).

The flowchart in Fig. 4 outlines the systematic process followed for data collection and analysis in this study, specifically during the survey conducted.

The survey methodology for this study follows a structured process, starting with the establishment of research objectives to guide the data collection and analysis. The target group has been specified as the residents of the Pabna Municipality. The sampled space was carefully selected and the survey was conducted across a period of three weeks in June 2025. The survey employed Judgment-Based Sampling, where participants were selected based on predefined criteria, including occupation, income, and transportation behaviours, to provide context-specific insights aligned with the research objectives. A total of 400 structured questionnaires were distributed across various residential and commercial zones within all 15 administrative wards, with an emphasis on capturing respondents from different occupational, income, and age groups. A total of 348 valid responses were received, yielding a response rate of approximately 87 %, while 32 partially completed forms (8 %) were discarded during the data-cleaning stage. The survey sample of 348 respondents in municipality type 'A' who had been finally chosen randomly with a broad representation of various ages, occupations, and socioeconomic backgrounds, capturing a broad spectrum of travellers' patterns. The survey was administered face-to-face (in-person) and each questionnaire required approximately 12–15 min to complete. Participants were informed about the study's purpose and confidentiality assurance, and informed consent was obtained before participation.

The questionnaire included both closed and open-ended questions, which enabled the participants to express their preferences and experiences in details. The questions were designed to capture both objective factors (e.g. transport spending, usage rate) and subjective factors (e.g. perceptions of safety, sense of well-being). So, the data set encompasses numerous variables that can affect the choice of transport mode one way or another, including psychological and practical ones. The emotional and mental aspects (e.g., safety perceptions, anxiety, past negative experiences) and the conventional aspects (e.g., cost, distance) were found, which, when combined with the conventional aspects of transport behaviour (e.g., cost, distance) can provide a finer view of transport behaviour through the data. The survey was designed in such a way that in addition to the mode choice preferences during each trip, also the psychological determinants, which are associated with certain origin-destination pairs, could be retrieved.

A pilot test was conducted in an initial study to perfect the questionnaire, thus making it empirically effective and easy to understand. To eliminate selection bias and ensure demographic diversity, participants were picked randomly in the demarcated survey boundaries. The administration of the surveys was done on a face-to-face basis. The data collection process took place over a period of three weeks; therefore, the final dataset was exposed to exact validation measures to ensure the data was complete and accurate.

The data was collected and then cleaned and coded, and where necessary, missing values and imputed. The data was subjected to preliminary analysis for personal information and travel related information. The data was analyzed through SPSS (Statistical Package for Social Science) where multinomial logistic regression is employed, and Pycrate for Machine Learning (ML).

The map shows in Fig. 5 the pre-selected area of the survey covered under Pabna Municipality with a 100-meter buffer zone around each sampling location was taken to reduce the spatial contaminations and maintain methodological integrity. The buffer zone ensures that these survey points not only capture direct responses but also consider the influence of nearby features,

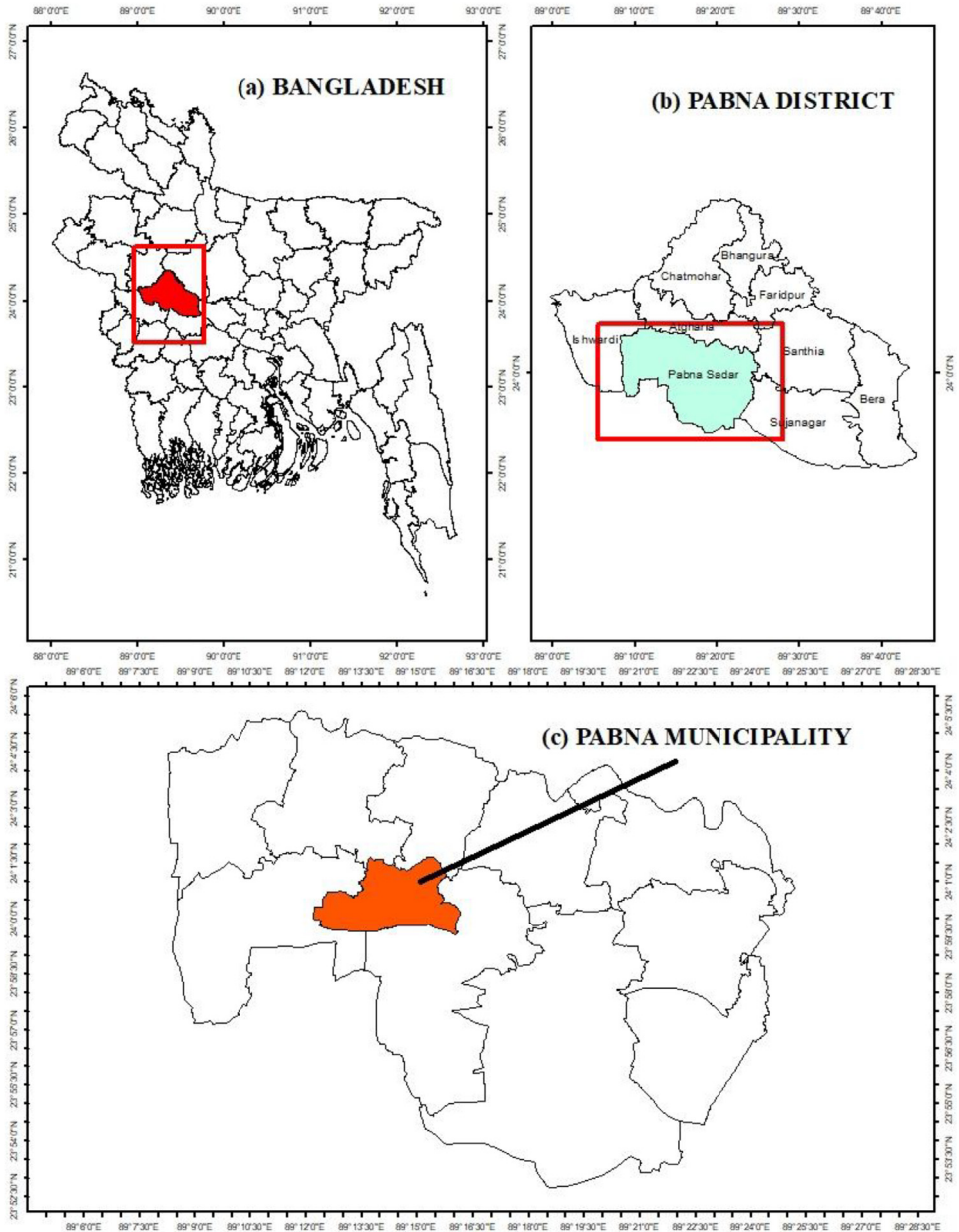


Fig. 3. Study area map for data collection.

improving the robustness of the data. The points were chosen carefully so as to ensure that different neighbourhoods and wards in the municipality were exhaustively covered. Field survey used was conducted face-to-face and at the whole buffer area to obtain a more diverse respondent sample. In every survey area, the respondents were randomly contacted both before and immediately after their mode of transport decision. The sample included students, working professionals and retirees, and thus age was diversified in the final dataset. This process reduced

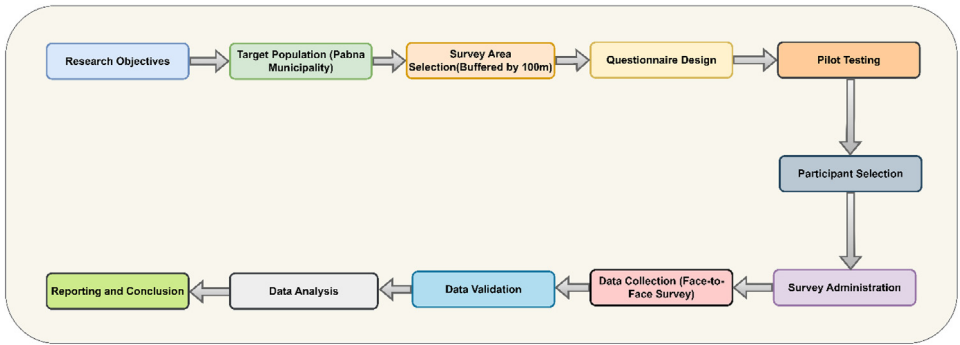


Fig. 4. Overview of survey design.

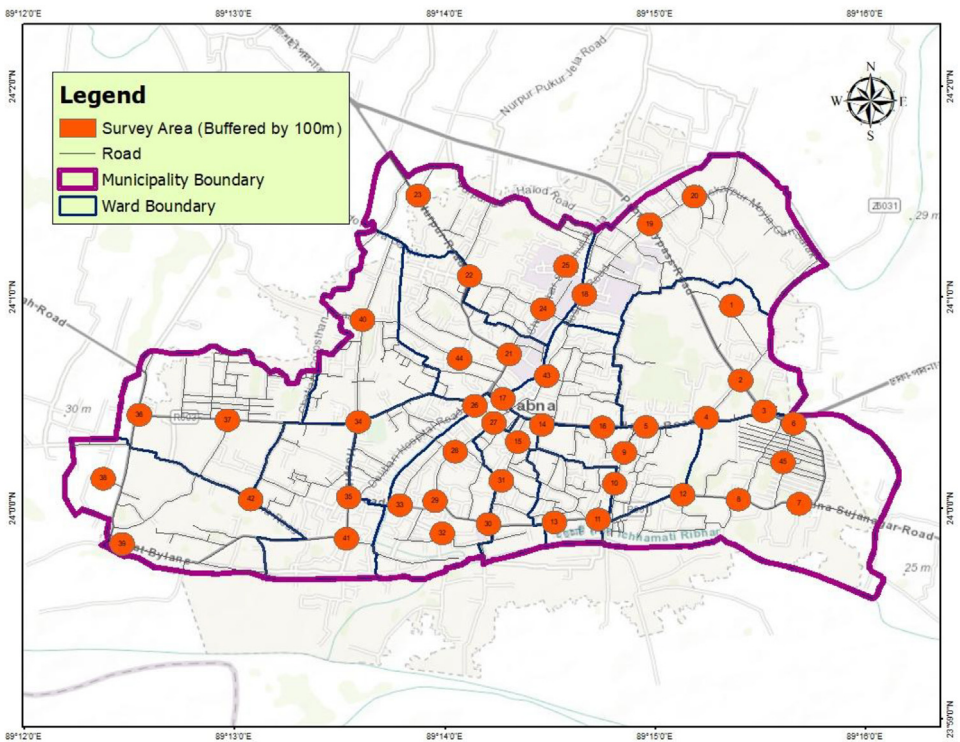


Fig. 5. Survey area with 100 m buffer zone for data collection across Pabna Municipality.

selection bias as it ensured that there was random selection in the buffer zone as opposed to being dependent on one fixed location.

4.1. Design of questionnaire

Work place interview was adopted for the study, since the study focuses on the transport mode choice factors. The method of data collection adopted for this study was face-to-face interview technique. Questionnaire was developed based on the factors influencing the mode choice

identified from the literature review. Questionnaire was designed to collect, personal information and travel related information, in addition to the data on psychological factors. The questionnaire was structured into four major sections like Socio-demographic characteristics, Travel behavior, psychological determinants and Emotional and mental well-being. This questionnaire was specifically designed for this particular study (ad hoc) to capture the psychological, behavioral, and demographic determinants of transport mode choice among residents of Pabna Municipality.

4.2. Sampling method

The Judgment-Based Sampling method was employed for this survey. Judgment-Based Sampling is a non-probability-based technique where participants are selected based on predefined criteria that are relevant to the study's objectives. In this case, participants were chosen based on factors such as occupation, income, transportation behaviours, and mental health variables. This method allowed for targeted selection of individuals who could provide the most relevant insights into the transportation behaviours. These targeted selections ensure a well-defined representation from key socio-demographic groups, aligning with the study's objectives to explore transportation choices and socio-economic factors. Our goal was to conduct survey a total of 400 participants, ensuring a balanced distribution to achieve representative coverage of Pabna Municipality. Among them a total of 348 respondents were included after cleaning, providing adequate statistical power to detect meaningful differences in transport mode preferences. The target population for this study was residents of Pabna Municipality, a dynamic urban area in Bangladesh, known for its diverse land use and developing transport infrastructure. Surveys are administered using structured questionnaires, with the collection process being randomized by Face-to-Face Interviews.

Participants were selected based on specific transportation behaviours and demographic factors to understand the factors that influence individuals make transportation decisions. The study targeted who use public transportation, private cars, bicycles, and those who primarily walk. Additionally, participants were chosen from various age groups, income levels, and occupational backgrounds to reflect diverse perspectives. Participants and experience were required to use their selected mode of transportation frequently. Recruitment took place in the field, focusing on locations where individuals from the identified groups were naturally found, such as transportation hubs, road junction, business districts, suburban areas, and residential zones. The recruitment process was planned to be respectful, clear, and informative. Participants were approached during peak hours at busy locations. The researcher provided a brief overview of the study, emphasizing that participation was voluntary, and offered a small incentive (e.g., entry into a raffle for a gift card) to encourage participation. Informed consent was obtained, ensuring that participants understood the confidentiality of their responses. The sampling covered multiple wards within Pabna Municipality and fewer surveys per survey area with buffer zone. If certain groups are underrepresented (such as minorities), oversampling is applied in those strata to ensure that each group's data adequately contributes to the analysis.

The sampling design also incorporated psychological factors including trust in transport modes, happiness, depression, and mental stress to capture their influence on transport decisions. The sampling approach ensured that all major transport modes used in the municipality (e.g., rickshaw, bike, car, easy bike) were approximately represented. Income variation was captured by including participants from low-, middle-, and high-income brackets, enabling analysis of economic influences on transport choices. Since there is a wide range of differences in transport behaviour based on age, income, and geographic location, the study also takes into account average expenditure on transportation to assess how financial constraints influence transport choices. The dataset contains a diverse range of responses related to mental stress, happiness, urgency level, and transport choices, enriching the behavioural analysis. Considering socio-demographic, psychological, environmental, and economic dimensions strengthen the validity of the sampling by incorporating all major determinants of transport behaviour. The targeted use of Judgment-Based Sampling based on predefined socio-demographic, psychologi-

Table 7

Power analysis table.

	Power ^b	Predictors		Test Assumptions		
		Total	Test	N	Partial ^c	Sig.
Type III F-test ^a	0.995	8	8	348	0.3	0.05

a. Intercept term is included.
b. Predictors are assumed to be fixed.
c. Multiple partial correlation coefficient.

cal, and transportation-related criteria ensures a diverse and context-specific sample, minimizing sampling bias and improving the relevance and representativeness of the dataset.

4.3. Cochran sample-size calculation

The data were collected through a man-to-man questionnaire. This face-to-face method not only served to come up with answers on sensitive issues that are normally difficult to obtain via self-administered surveys, but also helped to clarify ambiguous answers in real time. Cochran's Sample Size Formula was used to determine sample size using the following (Eq. 1):

$$n = \frac{Z^2 \times p \times (1 - p)}{E^2} \quad (1)$$

$$n = \frac{1.96^2 \times 0.5 \times (1 - 0.5)}{0.0525^2} \approx 348.02$$

where,

n = Sample size; $Z = 1.96$ (for 95 % confidence level); $P = 0.5$ (max variability assumption); $E = 0.0525$ (5.25 % margin of error).

4.4. Alpha (margin-of-error) calculation and power analysis

The margin of error (ME) was calculated using the following Eq. (2):

$$ME = Z \sqrt{\frac{p \times q}{n}} \quad (2)$$

$$ME = 1.96 \sqrt{\frac{0.5 \times 0.5}{348}} \approx 0.0525 (\pm 5.25\%)$$

This indicates that the survey estimates are reliable within ± 5.25 % at the 95 % confidence level.

Where, sample size $n = 348$; Confidence level: for 95 % α is 0.05; Z-value for 95 % confidence: $Z = 1.96$; Conservative population proportion: $p = 0.5$; Complement: $q = 1 - p = 0.5$ (Table 7).

The power analysis using Type III F-test indicated a high statistical power of 0.995 for the model with 8 predictors (independent variables) and 348 observations. The partial correlation coefficient was 0.3, and the significance level was set at 0.05, confirming the adequacy of the sample size for detecting effects. Thus, the required initial sample size of infinite population is about 348. In transient populations, effective population size correction or stratified sampling allows to achieve selection of the effective sample and sufficient representation without much

change in the calculation of the sample size. Infinite-population assumption is commonly utilized as a simplifying assumption that remains a reasonable estimate. All the entries were carefully read and cross-examined to ensure an accurate and consistent data was interpolated.

Moreover, the findings of the study can be used to influence future transport policy in a way that aligns with the psychological needs of the passengers. One of the biggest advantages of this data is the focus on the complexity of transport decision-making.

Limitations

The data was taken only in Pabna Municipality and might not wholly represent the transport behavior of people in other developing countries with different urbanization and cultural contexts. The short time survey duration may not capture seasonal mode variation or long-term effects in transport behavior. The survey is still limited to fully understanding the deep psychological influence among different age groups. The survey data does not fully reflect the actual environmental impact of the different transport modes.

Ethics Statement

All participants were provided with detailed information about this survey, and informed consent was obtained prior to their participation. Their personal information in the survey remained confidential. Also, any respondents who experienced discomfort from the survey were right to withdraw at any time. The survey was conducted in the municipality with the university's approval, and no additional municipal authorization was required for conducting the survey in this area. For all participants under eighteen years old, informed consent was taken from their parents or the legal guardians, who were present during the survey and gave consent on behalf of the minor.

CRedit Author Statement

Md. Nayem Hossain: Conceptualization, Methodology, Writing (Original draft preparation, Reviewing and Editing), Data Collection, Data review, Investigation. **Nakib Aman:** Supervision, Data review. **Md. Rakibul Islam Sabbir:** Data Collection, Investigation. **Md. Rashedul Haque:** Supervision, Data review. **MD. Tahidul Islam:** Data review, Data Collection.

Data Availability

[A Comprehensive Dataset of Transportation Mode Choice \(Original data\)](#) (Mendeley Data)

Acknowledgments

The authors would like to express their deepest gratitude towards Department of Urban and Regional Planning for their contributions to the research. We would like to thank Pabna Municipality authorities for permission to conduct the survey and supporting our survey activities. We also thank the residents of these municipalities area for participating this survey by answering questions from survey team members. The data sets mentioned in this article are publicly available; it is still obligatory that researchers must follow the rules of citation is invaluable.

Declaration of Competing Interest

The authors report no financial or personal conflicts of interest between themselves and others.

References

- [1] L. Chen, Y. Zhao, Z. Liu, X. Yang, Construction of commuters' Multi-mode choice model based on public transport operation data, *Sustainability* vol. 14 (22) (Nov. 2022) 15455 2022, Vol. 14, Page 15455, doi:[10.3390/SU142215455](https://doi.org/10.3390/SU142215455).
- [2] M.B.H. Taş, K. Özkan, İ. Sarıççek, A. Yazici, Transportation mode selection using reinforcement learning in simulation of urban mobility, *Appl. Sci.* vol. 15 (no. 2) (Jan. 2025) 806 2025, Vol. 15, Page 806, doi:[10.3390/APP15020806](https://doi.org/10.3390/APP15020806).
- [3] A. Javaid, F. Creutzig, S. Bamberg, Determinants of low-carbon transport mode adoption: systematic review of reviews, *Environ. Res. Lett.* 15 (10) (Sep. 2020) 103002, doi:[10.1088/1748-9326/ABA032](https://doi.org/10.1088/1748-9326/ABA032).
- [4] S. Kelela, Y.M. Emagnu, K.K. Berta, Assessing the determinant factors influencing transport mode choice: a case of Debre Berhan City, *J Adv Transp* 2025 (1) (Jan. 2025) 2393859, doi:[10.1155/ATR/2393859](https://doi.org/10.1155/ATR/2393859).
- [5] L. McCarthy, A. Delbosc, G. Currie, A. Molloy, Factors influencing travel mode choice among families with young children (aged 0–4): a review of the literature, *Transp Rev* 37 (6) (Nov. 2017) 767–781, doi:[10.1080/01441647.2017.1354942](https://doi.org/10.1080/01441647.2017.1354942);JOURNAL:JOURNAL:TTRV20;REQUESTEDJOURNAL:JOURNAL:TTRV20;WGROU:STRING:PUBLICATION.
- [6] A. Javaid, F. Creutzig, S. Bamberg, Determinants of low-carbon transport mode adoption: systematic review of reviews, *Environ. Res. Lett.* 15 (10) (Sep. 2020) 103002, doi:[10.1088/1748-9326/ABA032](https://doi.org/10.1088/1748-9326/ABA032).
- [7] W. Yang, Q. Chen, J. Yang, Factors affecting travel mode choice between high-speed railway and road passenger transport—Evidence from China, *Sustain.* 2022 vol. 14 (no. 23) (Nov. 2022) 15745 Vol. 14, Page 15745, doi:[10.3390/SU142315745](https://doi.org/10.3390/SU142315745).
- [8] L. Chen, Y. Zhao, Z. Liu, X. Yang, Construction of commuters' Multi-mode choice model based on public transport operation data, *Sustain.* 2022 vol. 14 (22) (Nov. 2022) 15455 Vol. 14, Page 15455, doi:[10.3390/SU142215455](https://doi.org/10.3390/SU142215455).