Journal of Transportation and Logistics,
JTL 2025, 10 (2): 362–387

https://doi.org/10.26650/JTL.2025.1609360

Submitted 29.12.2024
Revision Requested 16.02.2025
Last Revision Received 23.03.2025
Accepted 01.05.2025
Published Online 28.07.2025

# Journal of Transportation and Logistics

Research Article

Open Access

# Adaptive Time-Based Clustering for Optimizing Bus Fleet Management: A Case Study on İzmir's Public Transportation System

Nurhan Dudaklı [1], Behice Meltem Kayhan [2] & Ümit Kuvvetli [3]

[1] İzmir Bakırçay University, Faculty of Economics and Administrative Science, Department of International Trade and Business, İzmir, Türkiye
[2] Manisa Celal Bayar University, Faculty of Business, Department of Business Administration, Manisa, Türkiye
[3] İzmir Bakırçay University, Faculty of Economics and Administrative Science, Department of Business Administration, İzmir, Türkiye

**Abstract** This study introduces an adaptive time-based clustering strategy to optimize bus fleet requirements in public transportation systems by leveraging passenger boarding data from İzmir's network. Addressing key challenges in scheduling and fleet sizing, the proposed method uses the K-Means clustering algorithm to segment boarding densities into optimally determined time intervals specific to each bus line and direction. By adapting the number and boundaries of time intervals to actual demand patterns across weekdays and weekends, the model offers a more responsive and efficient allocation of fleet resources. The results demonstrate that the adaptive clustering approach significantly outperforms the conventional fixed-interval strategy, reducing both maximum and average bus requirements while maintaining service quality. This improvement is especially notable for high-demand or highly variable lines, where resource flexibility is critical. While the study shows promising results, it also acknowledges limitations such as the exclusion of passenger waiting times and the diversity of the fleet composition. Future research may include integrating alternative clustering algorithms, incorporating alighting data, and developing multi-criteria operational planning models. These enhancements will further support the evolution of data-driven, adaptive planning tools for more efficient and sustainable urban transport systems.

**Keywords** Public Transportation Optimization · Adaptive Clustering · Frequency setting · Bus Fleet Management

**Author Note** This paper is an extended and full version of the study initially presented at the Conference of YAEM 2024.

Citation: Dudaklı, N., Kayhan, B. M. & Kuvvetli, Ü. (2025). Adaptive Time-Based clustering for optimizing bus fleet management: A case study on İzmir's public transportation system. *Journal of Transportation and Logistics*, *10*(2), 362-387. https://doi.org/10.26650/JTL.2025.1609360



Corresponding author: Nurhan Dudaklı nurhan.dudakli@bakircay.edu.tr

# Adaptive Time-Based Clustering for Optimizing Bus Fleet Management: A Case Study on İzmir's Public Transportation System

Urban public transportation, as a critical component of urban life, serves as a key indicator of a city's level of development. The rapid expansion of cities, population growth, and the increasing reliance on private vehicles contribute to a range of challenges, including traffic congestion, air pollution, noise, and excessive energy consumption. The most effective solution to mitigate these issues lies in the adoption of integrated and well-planned urban public transportation systems (Ibarra-Rojas et al., 2015; Mirza and Jain, 2025). However, the significant reliance on public resources in urban transportation systems necessitates equitable service delivery across all parts of the city and the efficient execution of these services. Addressing the ever-growing travel demand in a fair and efficient manner within the constraints of limited resources is particularly challenging in metropolitan areas. Although different transportation modes (rail system, ferry, suburban, etc.) are used to tackle this issue, buses often constitute the main backbone of the urban public transportation system in many cities. Consequently, it is essential to use buses in a way that ensures the highest quality service to the maximum number of passengers, given existing constraints such as the availability of buses and personnel. Achieving this goal is only possible through effective planning, which is referred to in the literature as Transit Network Planning (TNP).

The TNP problem consists of several subproblems, two of which are Transit Network Design (TND) and Frequency Setting (FS). TND focuses on designing efficient transit routes by optimizing stop locations, route structures, and terminal placements to enhance overall system coverage and efficiency. FS, on the other hand, involves determining the optimal number of vehicle trips per hour for each route, ensuring that service levels align with demand fluctuations across different times of the day and days of the week. In the literature, some studies address both TND and FS jointly, while others focus specifically on the FS problem. In this study, we focus on the FS problem and propose an adaptive strategy for adjusting service frequencies based on variations in passenger demand.

Previous research on the FS problem typically quantifies passenger flows based on peak demand periods, such as the number of passengers traveling during rush hour. This approach often assumes worst-case scenarios, focusing on an hour-long window of maximum demand. While effective for managing peak load conditions, it overlooks fluctuations in passenger demand throughout the day, potentially leading to inefficiencies during off-peak hours. Oliveira et al. (2024) highlighted this limitation and suggested that future research could explore FS strategies that dynamically adjust to varying demand across different time intervals.

To overcome these limitations, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful tools for optimizing public transportation systems. AI-driven methods have demonstrated significant potential for optimizing urban mobility, with projections estimating a 44% profitability increase in the transportation and storage sectors and a 27% rise in public services by 2035 (Purdy and Daugherty, 2017; UITP, 2020). Reflecting this trend, recent studies have leveraged ML techniques for various transportation applications, including travel time prediction (Tran et al., 2020; Yuan et al., 2020), understanding travel patterns (Shalit et al., 2020), public transit analysis (Mayaud et al., 2019; Sosnowska and Skibski, 2018), ticket price forecasting (Aditi et al., 2020; Branda et al., 2020), and passenger satisfaction assessment (dell'Olio et al., 2018; Garrido et al., 2014). However, despite these advancements, there remains a gap in research focusing on adapting service frequencies based on historical passenger demand patterns.

This study aims to bridge this gap by proposing an innovative ML-based adaptive strategy that adjusts service frequency according to variations in passenger demand throughout the day. Unlike conventional models that primarily focus on peak demand periods, this approach segments the day into distinct time intervals based on observed fluctuations in demand. By employing clustering algorithms to analyze boarding data, the proposed method uncovers demand patterns and enables more efficient and responsive scheduling.

Traditional FS methods often rely on peak demand periods, typically concentrating on managing rush hour traffic and overlooking the demand fluctuations that occur during off-peak hours. This limitation can result in inefficiency in service provision. The adaptive strategy introduced in this study adjusts service frequencies based on varying demand patterns across different times of the day, leading to better resource usage and improved network efficiency. By incorporating a data-driven, demand-responsive approach, this study represents a significant advancement in transit network planning, offering a more flexible and optimized solution for managing service frequencies

To validate the proposed approach, a case study was conducted in Izmir, the third most populous city in Türkiye. Passenger boarding data from key transit routes were analyzed using clustering techniques to optimize the number of buses and service frequency. This study demonstrates how ML techniques can automate processes that are still manually conducted in many public transportation systems. By minimizing human error and introducing continuous digital optimization, the study aims to elevate transit planning to its optimal state while paving the way for long-term system enhancements.

The contributions of this study can be summarized as follows:

- Unlike traditional fixed-strategy FS methods, this study introduces an adaptive approach that segments time periods based on daily passenger density patterns, ensuring more effective service planning.
- The use of ML algorithms in the study provides more accurate results by optimizing the fleet size of buses and service frequencies.
- By analyzing passenger boarding data, the study adopts a data-driven approach that supports more efficient resource allocation and demand-responsive service provision.
- The analyses conducted in the Izmir case provide a concrete example for the real-world application of the proposed method. The developed approach is adaptable to other cities and different transportation modes, offering broad-scale benefit potential.

The structure of the paper is as follows: Section 2 reviews the relevant literature and outlines the research contribution. Section 3 defines the problem and describes the methodology, including the general structure of the transit network problems. Section 4 presents the proposed adaptive approach. Section 5 discusses the case study and findings, including the data used, comparison of the fixed and adaptive approaches, and interpretation of the results. Finally, Section 6 provides the conclusion and outlines directions for future research.

## Literature Review and Research Contributions

Given its significance in transportation planning, the FS problem has been extensively explored in the literature from various perspectives. Researchers have employed various methods and focused on diverse objectives to address the challenges associated with FS. In this study, the relevant literature is reviewed in detail under several key categories: Problem Type, FS Strategy, Method, Objective, Case Study, and Case Location. A summary of the major recent studies is provided in Table 1.

The reviewed studies primarily focus on the FS problem, which aims to optimize service frequencies within transit networks. Additionally, some studies have explored the Transit Network Design and Frequency Setting (TNDFS) problem, which simultaneously addresses both network design and frequency optimization. A few works also delve into specialized problems such as Electric Transit Network Design and Frequency Setting (E-TNDFS) and Subline Frequency Setting (SFS), which consider factors like electric vehicle integration and subline services. These advancements reflect the growing complexity and innovation in public transportation planning.

All the reviewed studies adopted a fixed strategy, where service frequencies are predetermined and remain static, without dynamically adapting to fluctuations in passenger demand throughout the day. The primary objectives of these studies are focused on minimizing operational costs such as total travel time, user costs, and fleet size, with some studies also targeting the maximization of transfer efficiency or the reduction of waiting times.

**Table 1**

*A comparative summary table of recent studies*

| References | Problem Type | FS Strategy | Method | Objective | Case Study | Case Location |
|---|---|---|---|---|---|---|
| Gadepalli et al. (2024) | FS | Fixed | Mathematical Programming | Minimizing users' travel time and the total cost of public transportation operations | ✓ | Visakhapatnam/ India |
| Oliveira et al. (2024) | FS | Fixed | Metaheuristics | Minimizing the operator costs and total user costs | ✓ | Maceió/Brasil |
| Aksoy and Alver (2024) | E-TNDFS | Fixed | Metaheuristics | Minimizing the operator costs and total user costs | ✓ | Karaman/ Turkey |
| Aksoy and Mutlu (2024) | FS | Fixed | Metaheuristics | Minimizing the total user costs, fleet penalty, and passenger load penalty | | Mandl's Network |
| Benli and Akgün (2023) | TNDFS | Fixed | Mathematical Programming | Minimizing the total travel time and the fleet size | ✓ | Kayseri/Turkey |
| Jing et al. (2023) | TND | Fixed | Metaheuristics | Maximizing the transfer efficiency and minimizing the sum of the operating costs | | Mandl's Network |
| Gkiotsalitis et al. (2022) | SFS | Fixed | Mathematical Programming | Minimizing waiting times, running costs, and fleet size | ✓ | Eberbach/ Germany |
| Mutlu et al. (2022) | FS | Fixed | Metaheuristics | Minimizing the total travel cost | | Mandl's Network |
| Vlachopanagiotis et al. (2021) | TNDFS | Fixed | Metaheuristics | Minimizing user and operator costs | | Mandl's Network |
| Tian et al. (2021) | FS | Fixed | Mathematical Programming | Minimizing the total operating and | | The Sioux Falls Network |

| References | Problem Type | FS Strategy | Method | Objective | Case Study | Case Location |
|---|---|---|---|---|---|---|
| | | | | passenger transit costs | | |
| Mutlu et al. (2021) | FS | Fixed | Metaheuristics | Minimizing the total infection risk | | Mandl's Network |
| Bertsimas et al. (2020) | FS | Fixed | Mathematical Programming | Minimizing the waiting time and price | ✓ | Tokyo/Japan<br>Boston/USA |
| Wei et al. (2020) | FS | Fixed | Metaheuristics | Minimizing the total walking, driving, and waiting times for passengers | ✓ | Chongqing/China |
| Kim et al. (2019) | TNDFS | Fixed | Metaheuristics | Minimizing the user and operator costs | | Mandl's Network |
| Ibarra-Rojas et al. (2019) | FS | Fixed | Mathematical Programming | Minimizing the total passenger and operational costs | ✓ | Santiago/Chile |
| This study | FS | Adaptive | K-Means | Minimizing the fleet size | ✓ | İzmir/Turkey |

*E-TNDFS: Electric Transit Network Design and Frequency Setting; SFS: Subline Frequency Setting; TNDFS: Transit Network Design and Frequency Setting

Regarding the methods employed, several studies have used mathematical programming approaches to optimize vehicle frequency, reduce operational costs, and enhance service efficiency. For instance, Gadepalli et al. (2024) developed a novel framework to integrate paratransit with formal public transport systems, applying Integer Linear Programming (ILP) to minimize users' travel time and the total cost of public transport operations. Benli and Akgün (2023) proposed a multi-objective Mixed Integer Nonlinear Programming (MINLP) model for the TND and FS problem to meet passenger demand while incorporating real-life transit network features, such as travel time, transfers, waiting times, overcrowding, under-utilization of vehicles, and fleet size. Gkiotsalitis et al. (2022) described the SFS problem under uncertain passenger demand as a Mixed Integer Linear Programming (MILP) problem, maximizing resource utilization while minimizing operational and passenger waiting time costs. Tian et al. (2021) applied MILP to optimize transit frequency, aiming to minimize total operating and passenger transit costs while addressing congestion on common bus lines. Bertsimas et al. (2020) focused on jointly optimizing FS and pricing in multimodal transit networks, aiming to minimize wait times while considering constraints such as budget, capacity, and passenger preferences, achieving near-optimal solutions in a short time using data from Boston and Tokyo. Ibarra-Rojas et al. (2019) addressed the Bus Lines Synchronization Problem (BLSP) by integrating FS, timetable optimization, and passenger route assignments into a MINLP with time-indexed variables, capturing route choices and associated cost components.

In addition to mathematical programming, metaheuristic methods have been employed to address larger-scale problems, particularly in complex networks where exact mathematical programming approaches may be computationally expensive or time-consuming. Oliveira et al. (2024) proposed a methodology based on biased random-key genetic algorithms (BRKGA), aiming to minimize passengers' waiting times and operational costs, specifically the distance covered by buses. Applied to a real case study, their methodology showed an improvement of over 10% in system performance for both metrics compared to the current configuration. Aksoy and Alver (2024) used a bi-level optimization framework with MODEA to design electric bus networks with wireless charging, minimizing operator and user costs by optimizing routes, frequencies, and charging stops. Aksoy and Mutlu (2024) evaluated the performance of five metaheuristic algorithms, Artificial Bee Colony (ABC), Differential Evolution (DE), Firefly Algorithm (FA), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO), in solving the FS problem, highlighting the importance of algorithm selection and parameter tuning. Jing et al. (2023) proposed a multi-objective optimization model for the TND, incorporating the transfer efficiency and operational costs. They solved the problem using the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) on Mandl's benchmark network, demonstrating the trade-offs between improving the transfer efficiency and increasing the costs. Mutlu et al. (2022) developed a bi-level optimization model for the FS problem in bi-modal networks, considering car and bus flow interactions. Using the DE algorithm, they tested the model on Mandl's benchmark network, demonstrating its effectiveness in reducing transportation costs while consistently generating similar frequency sets across multiple runs. Vlachopanagiotis et al. (2021) proposed an alternating-objective GA to solve the TNDFS problem, aiming to achieve Pareto optimality between user and operator costs, achieving a 5%–6% improvement over previous methods. Wei et al. (2020) employed GA to minimize the total walking, driving, and waiting times for passengers in public transportation networks. The methodology was tested on a real-world case study in Chongqing, China, where it demonstrated improved performance in reducing travel time for passengers. Mutlu et al. (2021) and Kim et al. (2019) approached the problem from distinct perspectives, incorporating factors such as disease transmission risk and equity, which are often overlooked in traditional models. Mutlu et al. (2021) considered the risk of disease transmission in public transportation as a travel cost, developing a risk minimization problem within the FS context. They developed a bi-level optimization model to minimize cumulative infection risks at stops using the DE algorithm. In contrast, Kim et al. (2019) addressed sustainability in public transportation by incorporating modal and spatial equity into the TNDFS problem, developing a bi-level model with a GA to determine line frequencies, while the lower model calculates modal split and traffic assignments.

In contrast to the reviewed studies, this study introduces an adaptive strategy for FS, which analyzes historical passenger demand data to propose variable service frequencies throughout the day, rather than relying on static, fixed frequencies. By focusing on historical patterns, this approach aims to optimize fleet size and service efficiency, addressing the limitations of fixed strategy models that do not adapt to changes in demand. Unlike previous works that solely rely on mathematical programming or metaheuristics for optimization, this study employs a K-Means clustering algorithm to develop adaptive FS strategies, offering a dynamic yet non-real-time solution. This approach fills a notable gap in the existing literature, where the focus has primarily been on fixed strategies, and provides a more flexible alternative to traditional methods, ensuring a more efficient use of resources and a better alignment with passenger demand.

## Problem Description and Methodology

### Transit Network Planning (TNP) Problems

Public transportation systems in developing countries face persistent challenges, with one significant issue being the mismatch between the supply of vehicles and passenger demand on various routes. This imbalance arises from inaccurate estimations of passenger demand and insufficient planning regarding the number and size of vehicles required to meet demand (Pojani and Stead, 2015). In some regions, public transportation services operate without fixed timetables or schedules, often dispatching vehicles only when fully loaded. Such practices result in excessively long waiting times during off-peak periods and unmet demand during peak hours due to inadequate service capacity (Abuaisha and Abu-Eisheh, 2023).

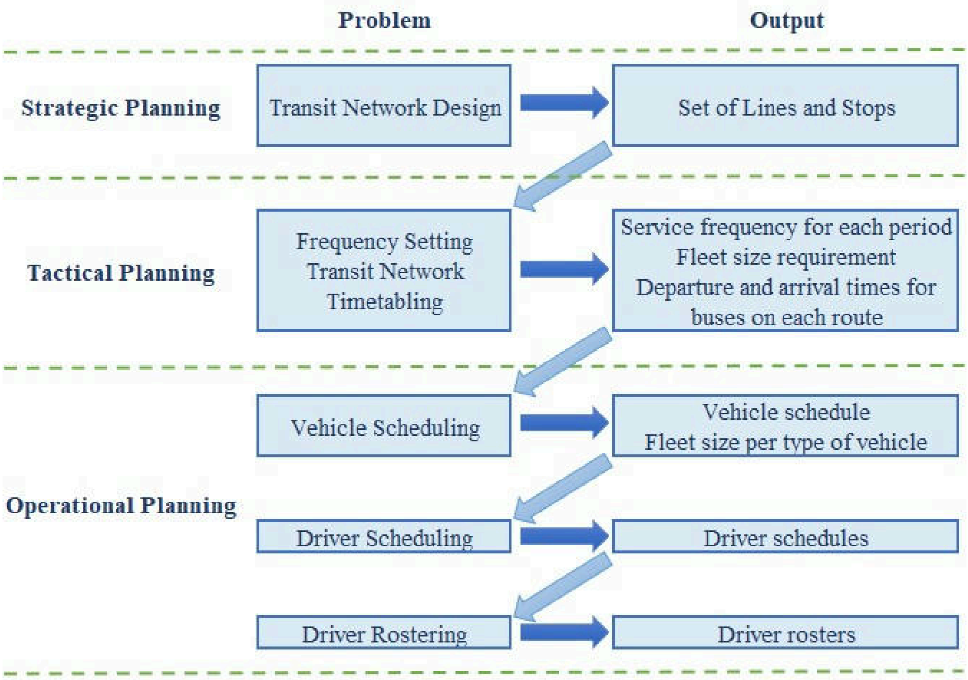
Addressing these inefficiencies requires a structured and systematic approach to transit planning, commonly referred to as the Transit Network Planning (TNP) problem. TNP plays a crucial role in designing efficient and sustainable urban transportation systems by optimizing various interconnected decision-making processes. This complex planning process has attracted significant attention from researchers across various fields. The TNP problem is typically divided into the following interconnected subproblems (Ceder, 2007; Ibarra-Rojas et al., 2015; Gadepalli et al., 2024):

- Transit Network Design (TND): Determining the optimal layout of transit lines and operational features, such as stop spacing, route structure, and terminal locations, to achieve specific objectives.
- Frequency Setting (FS): Determining the optimal number of vehicle trips per hour for each route in a transit system based on demand variations across different times of the day, days of the week, and day types.
- Transit Network Timetabling (TNT): Setting vehicle arrival and departure times at stops to align with frequencies, meet demand, optimize transfers, and minimize waiting times.
- Vehicle Scheduling (VS): Assigning vehicles to trips to ensure coverage of planned trips while minimizing operational costs.
- Driver Scheduling (DS): Assigning drivers to trips, ensuring coverage while complying with labor regulations and minimizing wage costs.
- Driver Rostering (DR): Organizing pre-defined daily duties into long-term work schedules (rosters) for drivers, ensuring compliance with labor regulations and cost efficiency.

These subproblems interact with each other and encompass tactical, strategic, and operational decisions within the planning process, as illustrated in Figure 1.

**Figure 1**

*Interaction Between the Stages of TNP (Ibarra-Rojas et al., 2015)*



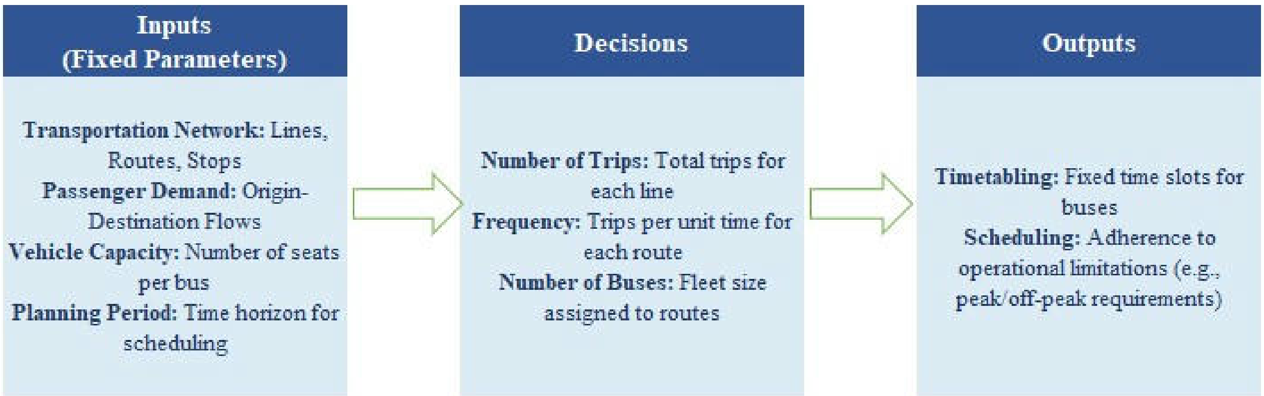

As seen in Figure 1, the FS problem is a key aspect of the tactical planning stage in the TNP. It involves deciding the number of vehicles to operate on each route during a specific time frame (e.g., hour, day) within a transit network to adequately meet passenger demand (Ceder, 2007; Oliveira et al., 2024). The FS directly affects fleet size requirements: higher frequencies necessitate a larger fleet, while an insufficient number of vehicles increases passenger waiting times, reducing service quality and satisfaction (Mete et al., 2022). Conversely, excessive vehicle deployment leads to inefficient resource use and higher operational costs (Ceder, 2007; Benli and Akgün, 2023). Thus, optimizing FS is critical for balancing these trade-offs, ensuring both operational efficiency and high-quality service.

## The General Structure of The Problem

The general problem structure of public transportation optimization encompasses two primary perspectives: (i) user-oriented factors, such as service quality, which is measured by trip frequency and headway, and (ii) service-provider-oriented factors, including ownership and operational costs, represented implicitly by the number of vehicles and trips. These factors are influenced by the parameters derived from the transport network and passenger flows between the origin–destination pairs. Key parameters include bus lines (defined by their length and stop sequences), bus capacities, and specified departure frequencies (minimum and maximum) for each line or route (Oliveira et al., 2024). The general problem structure and the classical input-output relationships are depicted in Figure 2.

**Figure 2**

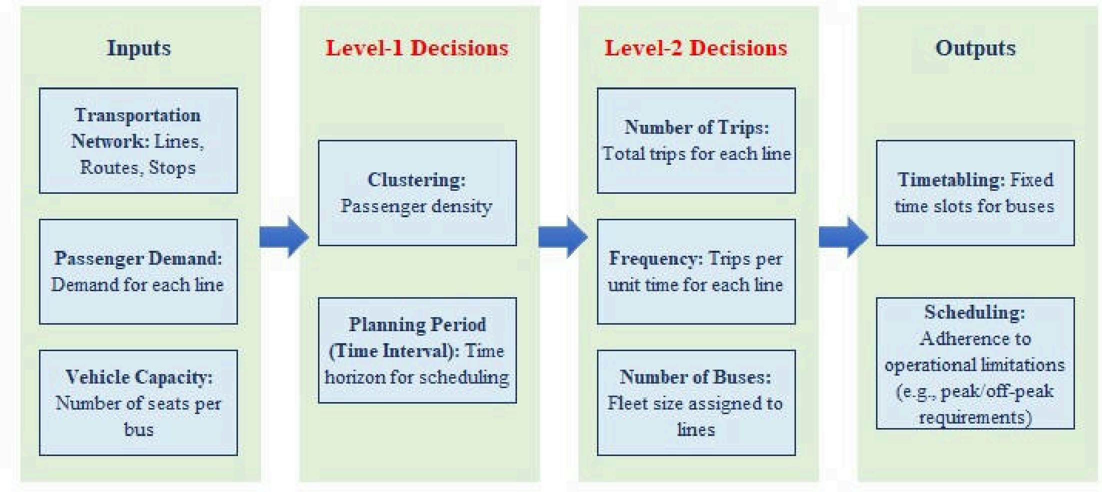
*Structure of the Frequency Setting and Fleet Sizing Problem*




A critical challenge in addressing this problem is determining the optimal frequency and fleet size to meet passenger demand, which fluctuates significantly throughout the day. These variations are most pronounced when comparing peak and off-peak periods, characterized by differences in hourly demand, trip distributions, average trip lengths, and traffic congestion. Traditional approaches often segment demand into predefined periods, such as peak and off-peak periods, under the assumption that this division sufficiently captures demand variability (Jara-Diaz et al., 2017). However, such fixed frameworks fail to account for the nuanced nature of daily demand patterns, which can vary significantly even within peak or off-peak windows. Using ML techniques, such as clustering, this method identifies patterns in boarding times, enabling a more precise and adaptive segmentation of time periods. This ensures a better alignment of fleet size and service frequency with real-world demand patterns, enhancing both operational efficiency and passenger satisfaction. The framework for the proposed approach is illustrated in Figure 3.

**Figure 3**

*Framework of the Proposed Adaptive Time Segmentation Approach for Optimizing Fleet Size and Service Frequency*




A key component of this approach is the adaptive definition of the planning period. By clustering passenger boarding data, the system adapts to daily demand variations, allowing for a more granular allocation of fleet size, trips, and frequencies. This flexibility transforms the planning period into an adaptive element of the optimization process, significantly improving the system performance.

The proposed framework begins with inputs such as the transportation network (lines, routes, stops), passenger demand (per bus line), and vehicle capacity (seating per bus). At the first decision level, clustering techniques segment the day into flexible time horizons that reflect demand fluctuations. This enables the model to recommend optimal frequencies and fleet sizes for each line and time period, ensuring efficient fleet usage and improved service quality. At the second decision level, the model optimizes the key operational variables: the total number of trips per line, trip frequencies, and fleet size allocation. Unlike traditional approaches, which often treat fleet size as a constraint rather than a decision variable, this study directly incorporates fleet size into the decision-making process, ensuring a more effective optimization. This process results in two main outputs: timetabling, which defines fixed time slots for buses, and scheduling, which adheres to operational constraints such as peak and off-peak requirements. Together, these elements ensure that system performance aligns with passenger demand while maintaining operational efficiency.

## Proposed Adaptive Clustering Approach

With the increasing adoption of automated fare collection systems, detailed onboard transaction data has become an asset for optimizing public transportation systems. Studies such as Pelletier et al. (2011) have highlighted the potential of smart card data for analyzing travel patterns and improving transit operations. Data mining techniques have proven effective in extracting actionable insights, such as identifying passenger behaviors (Ma et al., 2013). Building on these advancements, Carrel et al. (2015) introduced a high-resolution passenger tracking system that generates precise databases, including details such as boarding times, stations, fare types, and transit lines. Leveraging these technologies, this study employs data mining methods to analyze passenger boarding densities and travel patterns, with the goal of optimizing bus fleet sizes.

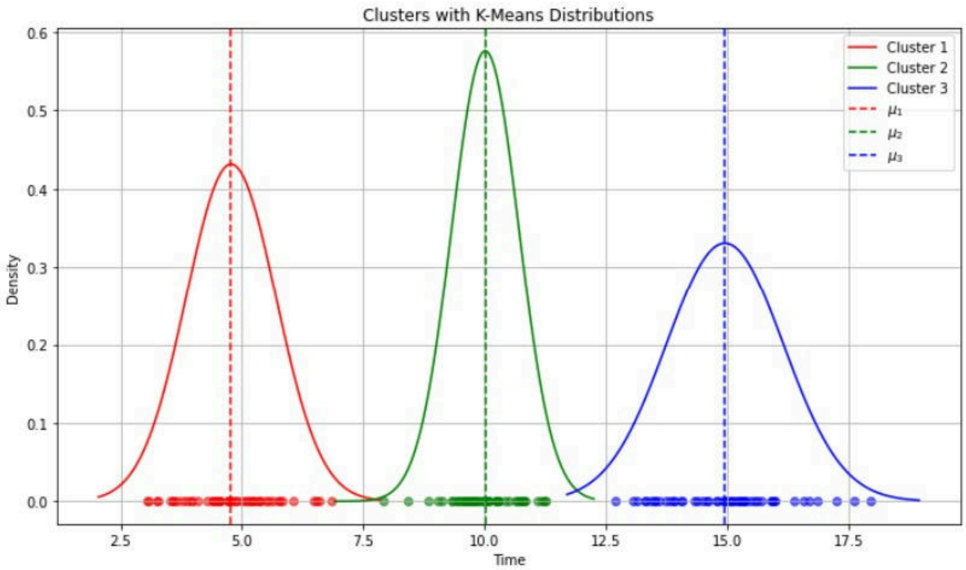
The proposed approach is practical, easy to implement, and requires minimal input data, making it ideal for optimizing public transportation services in resource-constrained settings. Its adaptable framework ensures applicability across various transportation modes, including road-based and rail-based systems. By integrating clustering algorithms, demand analysis, and resource allocation, it provides a comprehensive solution to enhance the efficiency and quality of public transportation systems.

Addressing two key challenges in public transportation planning (understanding dynamic, line-specific passenger demand patterns that vary throughout the day and utilizing real-world data for FS and fleet sizing), the approach employs clustering algorithms to group similar demand patterns. This enables time-specific planning and resource allocation. Traditionally, clustering has focused on analyzing demand density based on location for network design tasks, such as identifying high-demand regions, optimizing routes, and determining transportation modes and capacities (Charris et al., 2019; Xie et al., 2020; Chen & Shan, 2021; Lee et al., 2022; Pei et al., 2022; Bagheri et al., 2024).

In contrast, the proposed method applies clustering algorithms to classify passengers based on boarding times rather than locations, as illustrated in Figure 4. This time-based clustering uses the K-Means algorithm to group the boarding demand in distinct clusters, each represented by a Gaussian distribution curve. Individual data points, represented by scattered markers, are grouped into their assigned clusters. The vertical dashed lines mark the mean (μ) of each cluster, indicating the central tendency of demand within the respective time intervals.

**Figure 4**

*Example of Time-Based Demand Clustering Using K-Means*




This adaptive segmentation of the day into intervals, driven by fluctuations in passenger demand, enables the precise optimization of fleet size and trip frequencies tailored to specific time periods. By leveraging the efficiency of the K-Means algorithm with large datasets, this approach enhances the responsiveness of public transportation systems, aligning operations with real-world demand variations.

K-Means is a widely used unsupervised ML algorithm for clustering data into distinct groups. Its primary objective is to partition $n$ data points into $k$ clusters by minimizing the Within-Cluster Sum of Squares (WCSS), which quantifies the variance within each cluster. For a dataset divided into $k$ clusters, the WCSS is mathematically defined as in Eq. (1):

$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} \| x - \mu_i \|^2 \quad (1)$$

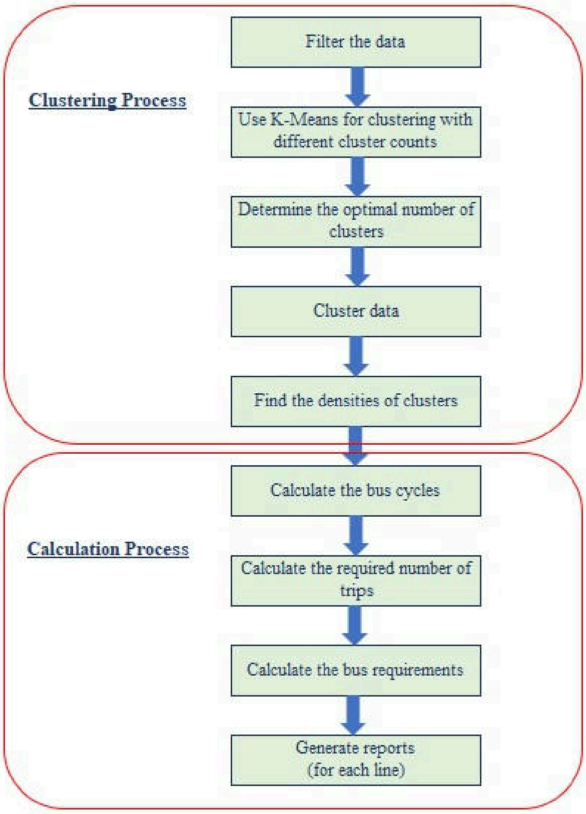
Where k is the number of clusters; $C_i$ is the i$^{th}$ cluster; $\mu_i$ is centroid of the i$^{th}$ cluster; $x$ is data point in the i$^{th}$ cluster; $\| x - \mu_i \|^2$ is squared Euclidean distance between $x$ and the centroid $\mu_i$. The optimal cluster count, or "elbow point," is identified where the WCSS curve shows a sharp reduction followed by a plateau, which is determined using the second derivative of the curve.

However, K-Means does not automatically determine the optimal number of clusters. To address this limitation, the Elbow Method is applied. This method involves evaluating various cluster counts and calculating the WCSS for each configuration. The optimal number of clusters is identified at the "elbow point," where the WCSS exhibits a significant decrease before plateauing. This method determines the cluster count at which compactness (measured by WCSS) achieves a balance with simplicity.

Once the clusters are formed, their time intervals are calculated based on the data points assigned to each cluster. This ensures that each cluster represents a distinct period with similar characteristics, enabling efficient planning and resource allocation. Based on the concept of time-based density clustering, Figure 5 illustrates the flow diagram of the proposed approach for optimizing bus fleet requirements. The methodology involves clustering hourly boarding data to identify time intervals with similar demand patterns. These clusters are then used to calculate the required number of trips, bus cycles, and ultimately the optimal number of buses for each cluster.

**Figure 5**

*Flow Diagram of the Proposed Approach*




The process begins with data preprocessing, which normalizes features to ensure that all attributes contribute equally to the clustering process. This step eliminates biases caused by varying feature magnitudes, making the data suitable for distance-based algorithms. The procedure then continues by clustering the day into distinct time intervals based on passenger density patterns. These clusters are subsequently used to define optimal scheduling periods, accurately reflecting demand variations.

**Table 2**

*Pseudo Code: Proposed Approach for Time-Based Clustering and Fleet Optimization*

| **Input:** |
|---|
| Passenger Data: Boarding times and demand for each line. |
| Transportation Network: Lines, routes, and stops. |
| Vehicle Capacity: Maximum number of passengers per bus. |
| Initial Parameters: Number of clusters (e.g., k for K-Means). |
| **Output:** |
| Time-based clusters representing passenger boarding density. |
| Optimal number of trips and fleet size for each time cluster |
| ***Step 1: Data Preprocessing*** |
| Load Passenger Data: |
| Input boarding times and demand data for all lines. |
| Normalize the Boarding Times: |
| Convert boarding times into numerical values (e.g., hours or minutes of the day). |
| Aggregate Demand by Time: |
| Group demand data by hour of the day for the initial analysis. |
| ***Step 2: Time-Based Clustering*** |
| Apply K-Means Clustering: |
| Define the number of clusters k. |
| Input: Boarding times as features. |
| Output: Clusters representing time intervals with similar demand density. |
| Assign Clusters to Time Periods: |
| Map each cluster to a specific time interval (e.g., morning peak, afternoon off-peak, evening peak). |
| ***Step 3: Calculate the Required Trips for Each Cluster*** |
| For each cluster: |
| Total Demand = Sum of passenger demand in the cluster. |
| Required Trips = Total Demand/Vehicle Capacity⌉ (round up to the nearest integer). |
| ***Step 4: Calculate the Fleet Size for Each Cluster*** |
| For each cluster: |
| Cluster Duration = Maximum boarding time Minimum boarding time in the cluster. |
| Bus Cycles = (Cluster Duration × 60) / Trip Time (in minutes). |
| Required Buses = Required Trips/Bus Cycles (round up to the nearest integer). |
| ***Step 5: Generate schedule and Schedule*** |
| Assign Trips to the Timetable: |
| Distribute trips across the cluster duration evenly. |
| Optimize Schedule: |
| Ensure that buses are dispatched to meet demand while minimizing waiting times. |
| ***Step 6: Validation*** |
| Simulate Demand Satisfaction: |
| Check if the assigned trips and fleet size meet the demand within each cluster. |

| |
|---|
| Adjust Parameters (if necessary): |
| Increase/decrease k or fleet size to improve the results. |
| ***Step 7: Output Results*** |
| Clusters: Time intervals with corresponding demand densities. |
| Trip and Fleet Requirements: |
| Number of trips and buses needed for each cluster. |
| Timetable: Optimized schedule for each line |

The pseudocode aligns with the steps outlined in Table 2, summarizing the clustering workflow and providing a detailed procedure for preprocessing data, identifying the optimal number of clusters using the Elbow Method, and performing K-Means clustering to effectively segment time-based demand. Once the time periods are established, the approach calculates the required number of trips for each line within these intervals, facilitating the precise determination of fleet size and line capacity to meet passenger demand while maintaining high service quality. Finally, the approach calculates the bus requirements for each cluster and generates detailed reports for each line and direction, supporting efficient transportation planning.

**Table 3**

*Table for Sets, Indices, Parameters and Variables*

| **Symbol** | **Description** |
|---|---|
| L | Set of all bus lines. |
| l | A specific bus line from the set L. |
| D | Set of directions for each line (forward and backward directions). |
| d | Direction of the line, where d=1 represents forward and d = 2 represents backward. |
| K | Set of time windows determined for each direction of each line. |
| k | A specific time window associated with a direction d of a line l. |
| B | Set of bus types available for operation. |
| b | A specific type of bus from set B. |
| **Variables** | **Description** |
| $RT_{ldk}$ | Required trip number for time window k of direction d of line l. |
| $RB_{ldk}$ | Required bus number for time window k of direction d of line l. |
| $CD_{ldk}$ | Cluster density (boarding demand) of time widow k of direction d of line l |
| $TL_{ldk}$ | Time length of the cluster (time interval) of time window k of direction d of line l |
| $CB_{lk}$ | Cycle number of each bus traveling on line l in time window k. |
| **Fixed Parameters** | |
| $CT_l$ | Cycle time for each bus operated on line l (e.g., time taken for a complete trip in minutes). |
| $cap_b$ | Capacity of bus b (e.g., maximum number of passengers per trip). |

After clustering the data, the density of each cluster was calculated to estimate the required number of bus trips. Based on the cluster density and duration, bus cycles and trip requirements were computed. The calculation process begins by using cluster densities $CD_{ldk}$ to estimate the total required trips $RT_{ldk}$, which are determined based on the bus capacity. Subsequently, the available time window (cluster's time length) $TL_{ldk}$ and the cycle time $CT_l$ are used to calculate the number of cycles $CB_{lk}$ that a bus can complete within the cluster's time interval. Finally, the required number of buses $RB_{ldk}$ is computed by dividing the required trips by the number of bus cycles, ensuring adequate allocation to meet passenger demand. The notations used are summarized in Table 3, and the detailed formulas are provided below for clarity and precision.

Required Trip Number $RT_{ldk}$ shows the total number of trips required to serve the demand in a specific time window *(k)* for a given direction *(d)* of line *(l)*. The ceiling function ensures that the value is rounded up to the nearest integer, ensuring that the total number of trips is sufficient to meet passenger demand without underestimating the requirement. $RT_{ldk}$ is calculated by the following formula.

$$RT_{ldk} = ceiling\left(C\frac{D_{ldk}}{c}ap_b\right) \forall l, d, k \tag{2}$$

Where $CD_{ldk}$ cluster density (boarding demand), $cap_b$ capacity of the bus *b* (fixed parameter). After calculating the required trips per time window of each direction of each line, we can calculate the required bus number $RB_{ldk}{}^{k}$ by the following formula.

$$RB_{ldk} = R\frac{T_{ldk}}{C}B_{lk} \forall l, d, k \tag{3}$$

Where $RT_{ldk}$ required trips, $CB_{lk}$ and bus cycles. The bus cycles related to each line depending on the travel times were calculated by the following equation. The number of bus cycles $CB_{lk}$ that represents the number of trips that can be done by a bus on line *(l)* for a given time window *(k)* is calculated as:

$$CB_{lk} = \max\left(T\frac{L_{ldk}}{C}T_l, 1\right) \forall l, k \tag{4}$$

Where $TL_{ldk}$ time length of the cluster (time interval in minutes), $CT_l$ is the cycle time for each bus operated on line *l* (fixed parameter), and the max ($TL_{ldk}/CT_l$, 1) ensures there is at least one cycle to avoid division by zero.

Equations (2-4) provide the foundation for calculating the line-specific and direction-specific bus requirements. However, to ensure an optimized fleet allocation, a mathematical model is necessary to minimize the total fleet size *F* while still meeting passenger demand across all lines and time clusters.

$$Min = F \tag{5}$$

*s.t.:*

$$F \geq \sum F_l \tag{6}$$

$$F_l \geq \max(RBC_{lk}) \forall l \tag{7}$$

$$RBC_{lk} \geq \max(RB_{ldk}) \forall l, k \tag{8}$$

By structuring fleet requirements hierarchically from individual clusters to lines and the entire network, the model achieves a balance between operational efficiency and passenger demand satisfaction. The goal is to minimize the total fleet size *F* required to operate the entire transportation network while maintaining service efficiency (Equation 5). The constraint given in Equations (6) ensures that the total fleet size *F* is at least the sum of the fleet sizes allocated to each line $F_l$. It guarantees that the number of buses assigned to the network is sufficient to meet the service demand of all bus lines. As given in Equation (7), the fleet size allocated to each line $F_l$ must be at least the maximum number of required buses per cluster $RBC_{lk}$, ensuring that the peak demand across all time clusters for a given line is adequately met. The final constraint (Equation 8) ensures that the required buses for each cluster $RBC_{lk}$, must be at least the maximum required buses $RB_{ldk}$ across both directions (*d=1,2*) of each line *(l)* for a given time cluster (*k*). It guarantees that the fleet allocation considers directional demand variations and peak periods.

A key challenge in the proposed approach is determining the optimal number of clusters for the entire line. Due to the bidirectional nature of transit lines, passenger density patterns often vary significantly between the two directions. This results in a situation where the optimal clustering algorithm may identify different numbers of clusters and corresponding time windows for each direction. To determine the optimal number of clusters for bidirectional transit lines, the approach considers the maximum passenger density ($CD_{ldk}$) associated with each line and cluster (time window). For each potential clustering configuration, the algorithm identifies the maximum density across all clusters for both directions of the line. The optimal clustering configuration is then selected by minimizing the maximum density across all clusters with the following condition.

$$K_{ld} = K_l \ \forall l, d, \text{where } K_l = \min_k \left( \max_d \left( CD_{ldk} \right) \right) \forall l \tag{9}$$

This method ensures that the clustering scheme accommodates the highest-density time periods while providing equitable time intervals that accurately reflect passenger demand patterns in both directions. By harmonizing the time windows and clusters through the minimization of the maximum density, the system achieves efficient scheduling and balanced resource allocation. This enhancement makes the optimization model robust and well-equipped to handle the inherent asymmetry of the bidirectional demand.

## Case Study and Findings

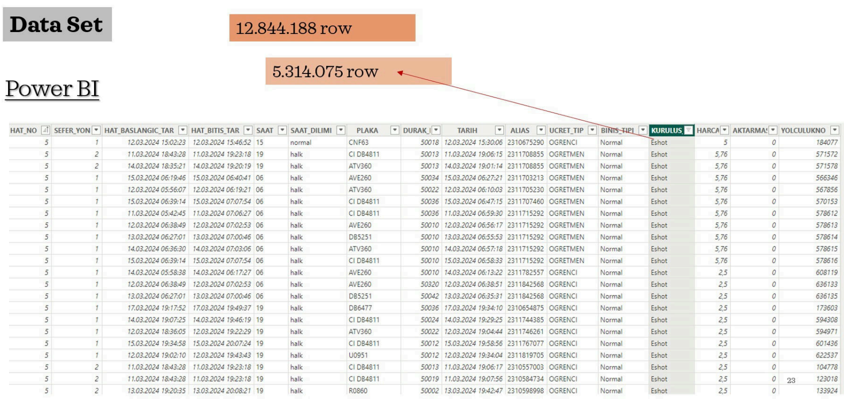
This study was conducted to test and validate the proposed approach using data from İzmir's public transportation system. By analyzing detailed boarding data over a one-week period, this study evaluates the effectiveness of the proposed method in optimizing fleet size and service frequency. The dataset, which includes boarding times, locations, travel directions, transportation modes, vehicle types, and passenger categories, serves as the basis for applying clustering algorithms to identify boarding density patterns. The results provide critical insights into high-demand times and locations, verifying the practicality and accuracy of the approach in enhancing operational efficiency and passenger satisfaction.

### The Case Data

To implement the proposed adaptive time-based clustering method, detailed data from İzmir's public transportation system were used. The data, collected via "İzmirim Kart" system and stored in Power BI, represents a comprehensive overview of one week of operations across the entire public transportation network. Figure 6 provides an overview of the dataset, highlighting its scale, with over 12.8 million rows of public transportation data, including 5.3 million rows specifically related to bus operations. The visualization underscores the dataset's comprehensiveness, containing critical fields such as boarding time, stop, passenger type, and fare details. This extensive dataset enables a comprehensive analysis of the bus frequency and fleet size optimization problem while offering opportunities for addressing other transportation-related challenges.

**Figure 6**

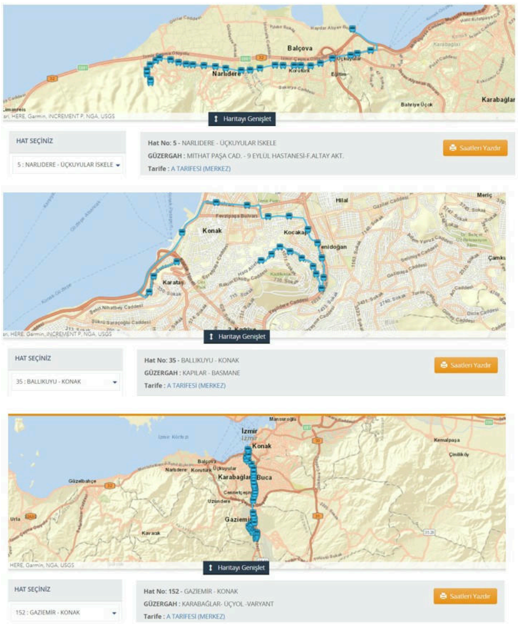
*Overview of the Dataset: Passenger Boarding Data and Attributes Stored in Power BI*

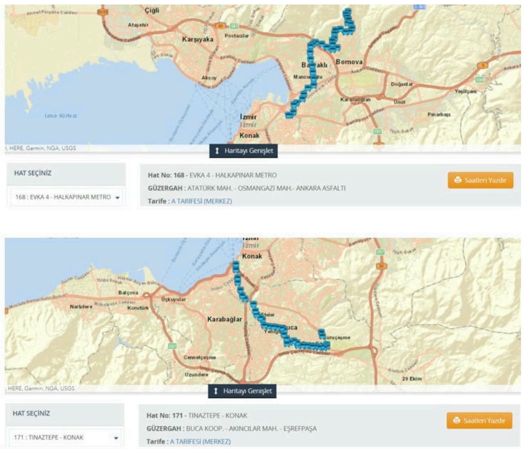


For this case study, five distinct bus lines from the network were selected based on input from system experts. These lines, 5, 35, 152, 168, and 171, were chosen to reflect diverse operational characteristics, including variations in route lengths, cycle times, bus types, and passenger patterns. The geographic routes of the selected lines are depicted in Figures 7, offering a visual representation of their spatial coverage within the transportation network. Additionally, the table in Figure 6 presents the detailed attributes of these lines, including their line IDs, bus types, passenger capacities, and cycle times. By focusing on these representative lines, the study underscores the flexibility and scalability of the proposed approach in addressing the varied contexts within public transportation systems.

**Figure 7**

*Geographic Coverage and Operational Attributes of Selected Bus Lines*





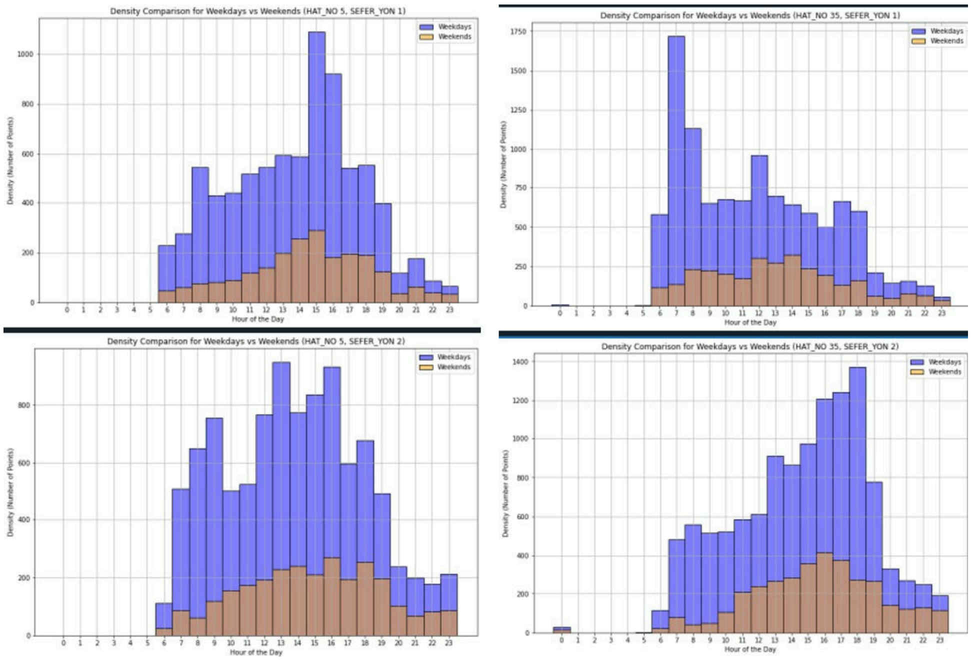
| LINE_NO | Bus Type | Capacity | Cycle Time |
|---|---|---|---|
| 5 | solo | 60 | 90 |
| 35 | midi | 40 | 60 |
| 152 | articulated | 100 | 130 |
| 168 | articulated | 100 | 90 |
| 171 | articulated | 100 | 120 |

Following the detailed presentation of the dataset and selected bus lines, a general analysis of passenger density was conducted to gain insights into the demand patterns across the selected lines. This analysis focused on identifying variations in passenger density between weekdays and weekends, as well as differences based on travel direction. Passenger demand in public transportation systems often varies significantly depending on the day of the week, reflecting differing commuter travel needs on weekdays versus weekends. To explore this, the dataset was segmented into weekday and weekend subsets, and the density of boardings for each line was analyzed separately. This differentiation is critical for understanding the temporal shifts in demand and optimizing the fleet size and frequency accordingly.

In addition, directional variations in passenger density were examined to assess whether travel patterns differed between the forward and backward directions of each line. These variations are often influenced by factors such as commuting flows, the distribution of residential and commercial areas, and the location of transit hubs. For example, higher density in one direction during morning hours may indicate commuter flows toward central business districts, while the reverse direction may experience greater density in the evening as passengers return home.

**Figure 8**

*Passenger Density Comparison by Hour, Line, Direction, and Day of the Week for the Selected Lines.*




This investigation reveals that passenger density and travel behavior vary significantly depending on the day of the week (weekdays vs. weekends) and the direction of travel. The variations highlight the dynamic nature of passenger demand, with distinct peaks during specific hours based on travel patterns. These differences are critical for planning optimal bus fleet sizes and schedules. As illustrated in Figure 8, the bar graphs compare passenger density for two selected lines, 5 and 35, in both directions (SEFER_YON 1=Direction 1 and SEFER_YON 2=Direction 2) across weekdays and weekends. The graphs clearly demonstrate variations in density patterns, underscoring the importance of incorporating day- and direction-specific factors into transit planning.

The results of this analysis provided essential insights into how passenger demand varies across different contexts, highlighting the need for adaptive and data-driven strategies in fleet optimization and service planning. By understanding these variations, the proposed approach can better address the unique operational needs of each line and direction, ensuring an efficient and balanced allocation of resources.

## The Comparison of Fixed and Adaptive Interval Approaches

This section compares two approaches to determine time intervals in managing the bus fleet and frequencies. The first approach, currently employed by ESHOT, divides the day into fixed time periods based on expert knowledge and historical patterns, reflecting traditional heuristic methods. The second approach proposed in this study introduces an adaptive, clustering-based methodology, where time intervals are analytically determined to adapt to the unique passenger density patterns of each line and direction. By contrasting these two approaches, this section highlights the limitations of the fixed interval method and the advantages of the adaptive clustering approach in achieving a more tailored and efficient public transportation system.

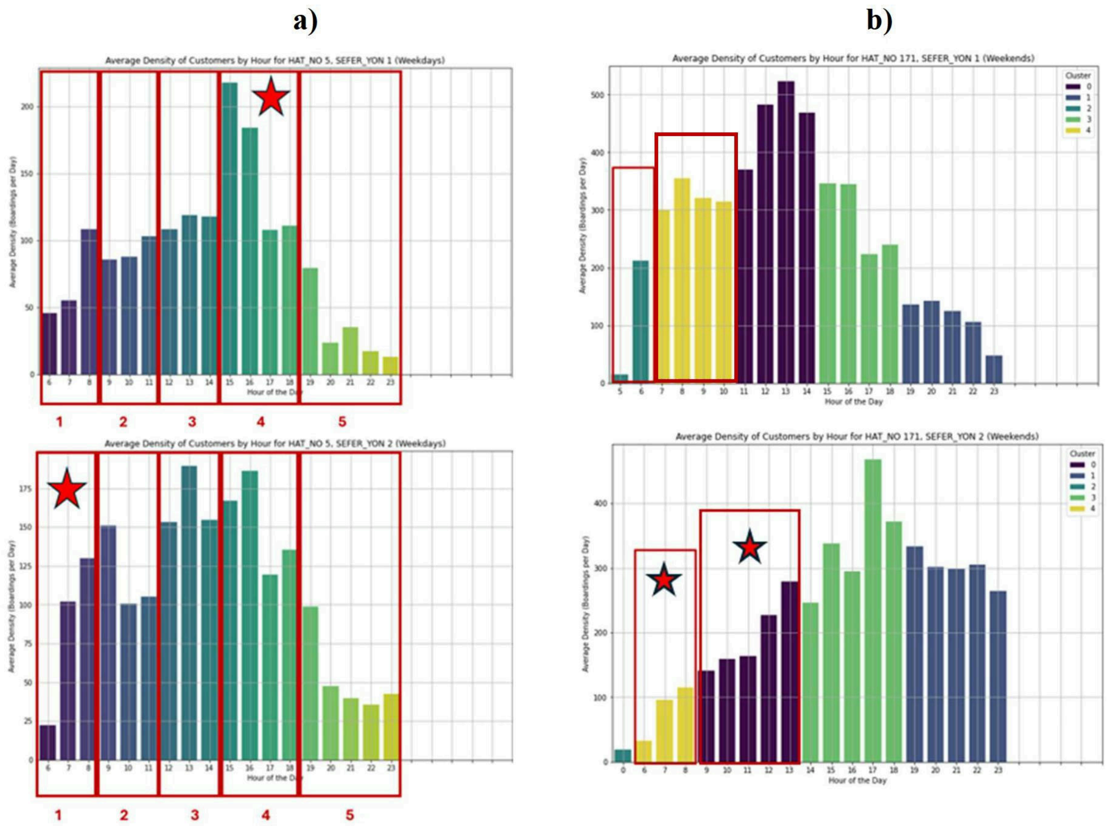
The first approach divides the day into five predefined periods based on operational experience: 1) 6:00–9:00, 2) 9:00–12:00, 3) 12:00–15:00, 4) 15:00–19:00, and 5) 19:00–24:00. Within each time window, bus frequencies and fleet requirements were determined using the maximum observed passenger density. Notably, the same time intervals are applied for both directions of each line, avoiding the need for directional customization. While this heuristic method is simple and aligns with traditional practices, it lacks the flexibility to accommodate specific demand patterns. Figure 9a illustrates the trip requirement calculation using the fixed-interval approach for line 5. The maximum density within each pre-defined time interval is used to determine the required number of trips, with intervals set uniformly across all directions and lines.

The adaptive clustering approach proposed in this study introduces two alternative clustering patterns to optimize time intervals and resource allocation in public transportation systems. Similar to the fixed-interval method, it calculates the required frequency and number of buses based on the maximum passenger density observed within each cluster (time interval). However, the adaptive approach offers enhanced flexibility by generating two cluster patterns for each line and direction. These patterns vary in the number and range of time intervals, enabling tailored segmentation based on specific passenger demand patterns.

The optimal cluster pattern is selected by minimizing the maximum passenger density across all clusters for the entire line. This strategy ensures a balanced distribution of passenger loads across time intervals while mitigating peak demand pressures. Figure 9b illustrates the trip requirement calculation using the adaptive interval approach for line 171, where the time intervals are determined separately for each line and direction. Required trips are calculated based on the minimum of the maximum densities across clusters, thereby optimizing fleet use. By adjusting the time intervals and service frequencies, this approach offers a more precise and efficient method for fleet optimization, outperforming both the fixed-interval method and the less flexible clustering techniques. This study highlights the benefits of tailored scheduling in addressing real-world transportation demands effectively.

**Figure 9**

*Comparison of Trip Requirement Calculation Between Fixed (Line 5) and Adaptive Interval (Line 171) Approaches*

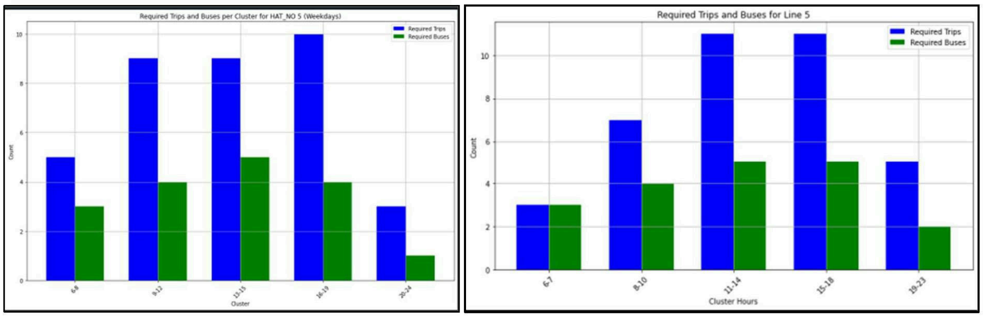



$CD_{lk} = max_d\ (CD_{ldk}) \quad \forall l, k$

$CD_{lk} = min_k\,(max_d\ (CD_{ldk}) \quad \forall l, k$

**Figure 10**

*Comparison of Required Trips and Buses for Fixed and Adaptive Clustering (K-Means) on Line 5*




The two bar charts compare the required trips and number of buses for Line 5 under the fixed clustering approach and the adaptive clustering approach utilizing the K-Means algorithm to segment time intervals based on passenger demand. In the fixed approach, uniform time intervals (e.g., 6-8, 9-12) are applied across all lines and directions, leading to inefficiencies such as overestimated resource allocation during non-peak hours and insufficient adaptability to actual demand patterns. For example, during peak periods like 13-15 and 16-19, the fixed approach requires 10 trips and 4-5 buses, but it fails to efficiently match demand during non-peak hours. In contrast, the adaptive clustering approach leverages the K-Means algorithm to create flexible time intervals (e.g., 6-7, 8-10) tailored to demand density, ensuring that resource allocation closely aligns with passenger needs. For instance, during peak hours (e.g., 11-14 and 15-18), required trips (11) and buses (5) are well-aligned with passenger density. Non-peak intervals (e.g., 6-7, 19-23) reflect reduced resource requirements, avoiding overallocation. This demonstrates that the adaptive clustering approach, powered by K-Means, enhances efficiency by adapting to actual demand patterns, optimizing resource utilization, and reducing over-allocation during low-demand periods. A detailed discussion of the results and findings will be presented in the following section.

## The Results and Discussion

The results and discussion section evaluates the effectiveness of the proposed adaptive clustering approach compared to the fixed time interval method currently employed in practice. By analyzing key metrics such as total trips, maximum and minimum buses, and average buses required for five selected bus lines, this section explores operational efficiency and resource use under both approaches. The comparison is made separately for weekdays and weekends to account for distinct passenger demand patterns. This detailed assessment highlights the benefits of a data-driven clustering strategy, emphasizing its capacity to optimize fleet allocation and service delivery. Table 4 summarizes the results of the fixed interval and adaptive clustering approaches for optimizing bus trips and fleet requirements on weekdays and weekends across the five selected bus lines.

**Table 4**

*Comparison of Fixed Intervals vs. Adaptive Clustering for Bus Fleet Opt. on Weekdays and Weekends*

| WEEKDAYS | LINE_NO | Total Trips | Max Buses | Min Buses | Avg Buses |
|---|---|---|---|---|---|
| Fixed Time Intervals | 5 | 66 | 5 | 1 | 3.1 |
| | 35 | 118 | 6 | 1 | 3.6 |
| | 152 | 62 | 6 | 2 | 4.2 |
| | 168 | 87 | 10 | 1 | 4.1 |
| | 171 | 127 | 13 | 2 | 7.4 |
| | Total | 460 | 40 | 7 | 22.4 |

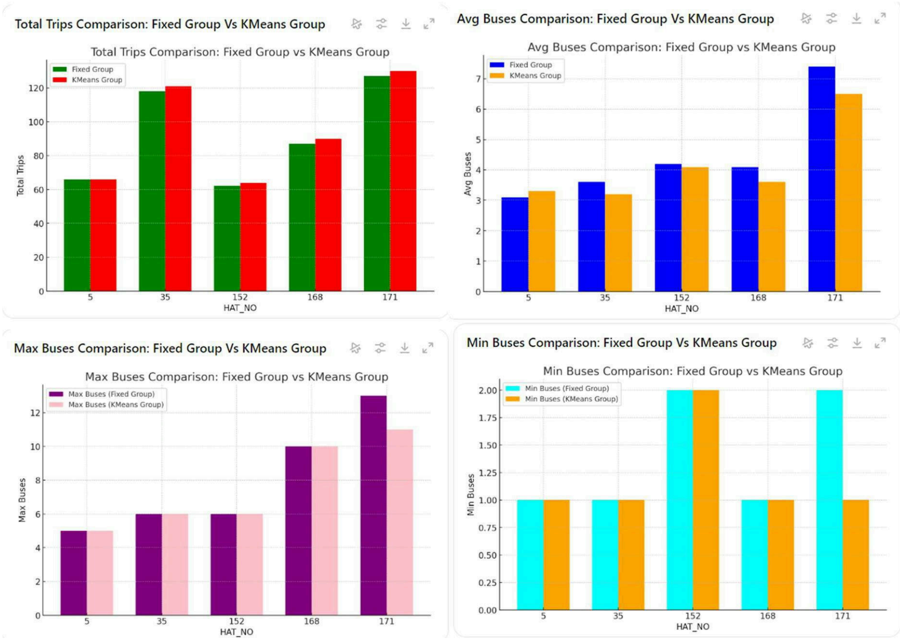
| WEEKDAYS | LINE_NO | Total Trips | Max Buses | Min Buses | Avg Buses |
|---|---|---|---|---|---|
| Dynamic Clustering Intervals (K-Means) | **LINE_NO** | **Total Trips** | **Max Buses** | **Min Buses** | **Avg Buses** |
| | 5 | 66 | 5 | 1 | 3.3 |
| | 35 | 121 | 6 | 1 | 3.2 |
| | 152 | 64 | 6 | 2 | 4.1 |
| | 168 | 90 | 10 | 1 | 3.6 |
| | 171 | 130 | 11 | 1 | 3.6 |
| | Total | 471 | 38 | 6 | 6.5 |
| | **% improvement** | 2 | 5 | 14 | 8 |
| **WEEKENDS** | **LINE_NO** | **Total Trips** | **Max Buses** | **Min Buses** | **Avg Buses** |
| Fixed Time Intervals | 5 | 46 | 4 | 1 | 2.2 |
| | 35 | 86 | 5 | 1 | 2.7 |
| | 152 | 54 | 6 | 2 | 9.7 |
| | 168 | 59 | 5 | 1 | 2.8 |
| | 171 | 99 | 10 | 2 | 5.6 |
| | Total | 344 | 30 | 7 | 17 |
| **WEEKENDS** | **LINE_NO** | **Total Trips** | **Max Buses** | **Min Buses** | **Avg Buses** |
| Dynamic Clustering Intervals (K-Means) | 5 | 45 | 3 | 1 | 2.1 |
| | 35 | 87 | 5 | 1 | 2.5 |
| | 152 | 55 | 5 | 2 | 3.4 |
| | 168 | 59 | 4 | 1 | 2.4 |
| | 171 | 101 | 10 | 2 | 5.2 |
| | Total | 347 | 27 | 7 | 15.6 |
| | **% improvement** | **1** | **10** | **0** | **8** |

**Weekdays**: For Line 171, the adaptive clustering approach reduced the maximum number of buses required from 13 (fixed interval) to 11, while the average number of buses decreased from 7.4 to 6.5. Total trips remained comparable between the two methods for most lines; however, the adaptive approach achieved greater efficiency by lowering peak fleet demands without compromising service levels. This highlights its ability to make tailored adjustments to passenger demand patterns, especially for lines with varying densities throughout the day.

**Weekends**: The benefits of the adaptive clustering approach are even more pronounced during weekends. The adaptive approach decreased the average number of buses required for Lines 5, 152, and 168 while maintaining similar total trip numbers. For instance, Line 5 required 4 buses with the fixed interval approach but only 3 with adaptive clustering. This reduction reflects the adaptive approach's capacity to adapt to fluctuating and typically lower weekend demand, ensuring optimized fleet allocation while preserving service quality.

**Figure 11**

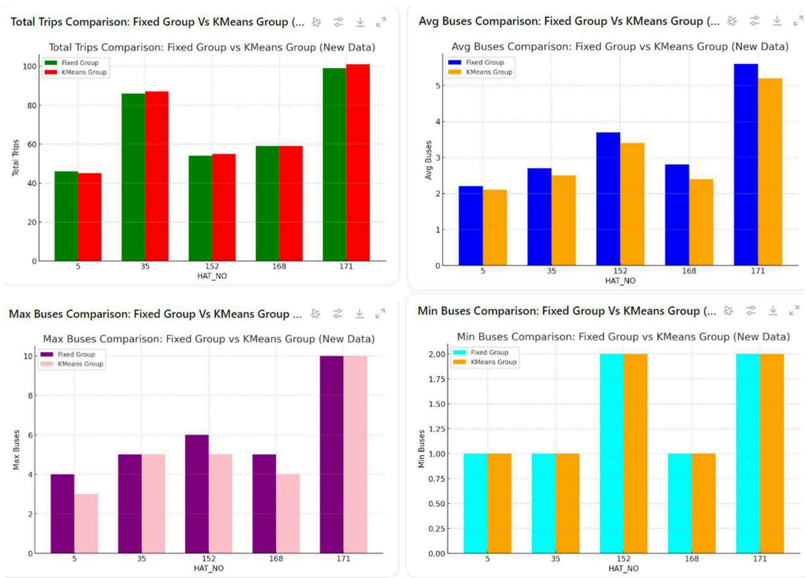
*Comparison of Fixed Interval and Adaptive Clustering Approaches for Bus Scheduling Metrics (weekdays)*




The comparison between the fixed interval and adaptive clustering (K-Means) approaches for weekdays, as shown in Figure 11, highlights several key differences in bus requirements. For minimum buses, the adaptive approach slightly reduces the requirements for line 171, while the results for lines 5, 35, 152 and 168 remain similar. Maximum bus requirements show significant improvements with the adaptive approach, particularly for line 171, where peak demand densities are better distributed, reducing the maximum number of buses needed. Average bus usage consistently improved with the adaptive clustering method across most lines; for instance, line 171 showed a decrease in average bus requirements, showcasing the efficiency of adaptively optimized intervals in balancing the fleet. Total trips remain comparable between the two approaches, demonstrating that both meet demand adequately, but the adaptive clustering approach achieves this with fewer buses overall, highlighting its superior resource optimization for weekday operations.

**Figure 12**

*Comparative Analysis of Fixed Interval and Adaptive Clustering Approaches for Key Bus Metrics (weekends)*

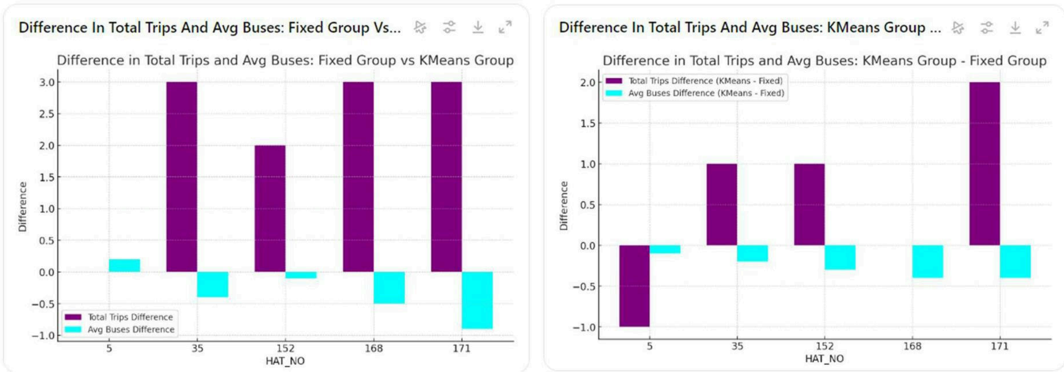


These graphs illustrate the performance metrics for the fixed interval and adaptive clustering approaches during weekends, as presented in Figure 12. The total trips and average number of buses required demonstrate that the adaptive clustering approach achieves superior resource utilization by adapting clusters to specific weekend demand patterns. The maximum number of buses required shows a noticeable reduction under the adaptive clustering method, indicating a more effective allocation of resources during high-demand periods. Meanwhile, the minimum number of buses required remains consistent between both approaches, reflecting the adaptive method's ability to maintain efficiency in low-demand scenarios. Overall, the analysis highlights adaptive clustering as a more flexible and efficient solution for weekend transit planning.

Figure 13 compares the total trips and average buses between the adaptive clustering (K-Means) approach and the fixed interval method for both weekdays (Figure 13a) and weekends (Figure 13b). The vertical axis represents the difference between the two approaches, where positive values indicate higher values in the K-Means method, while negative values show reductions compared to the fixed interval approach. This visualization highlights how demand-responsive clustering impacts fleet operations and resource allocation.

**Figure 13**

*Difference Trips and Average Buses: (K-Means values-Fixed Interval values)*




a) b)

In Figure 13a (weekdays), the adaptive clustering approach leads to an increase in the total trips for lines 35, 152, 168, and 171, with the most significant difference observed in line 35. This suggests that the K-Means method slightly increases the number of trips to better align with the fluctuating passenger demand. Meanwhile, the average bus requirement remains nearly the same or slightly decreases, demonstrating that adaptive clustering distributes resources more efficiently while maintaining or enhancing service levels. A similar trend is evident in Figure 13b (weekends) although the differences are less pronounced. The adaptive approach increases total trips more noticeably for lines 168 and 171, while also reducing the number of average buses. This indicates that the data-driven clustering method effectively captures demand variations, allowing for a more precise allocation of fleet resources and minimizing inefficiencies caused by static scheduling.

One of the key implications of these findings is that the increase in total trips, while maintaining or even reducing the number of required buses, suggests improved resource utilization. The K-Means approach enables higher service frequency without additional fleet resources, ensuring that bus operations are more closely aligned with real-time passenger demand. This optimization enhances operational efficiency by enabling a more balanced, demand-driven allocation of resources.

The results further indicate that the proposed adaptive clustering approach is particularly beneficial for lines with high passenger density (e.g., line 171) or significant demand fluctuations (e.g., line 168). By tailoring time intervals and fleet allocation based on actual demand patterns, this approach minimizes the inefficiencies associated with static scheduling. Specifically, for lines with substantial demand variability, the data-driven method enables more precise scheduling and fleet distribution, reducing resource wastage during off-peak hours while ensuring sufficient capacity during peak periods.

Ultimately, the adaptive clustering approach offers a more demand-responsive scheduling framework, effectively reducing peak fleet requirements while maintaining service reliability. Although the total number of trips slightly increased, this reflected improved service availability without a proportional rise in bus deployment, indicating better resource usage rather than inefficiency.

## Conclusion and Future Work

This study proposed a data-driven, adaptive approach to optimize bus fleet requirements across different lines and directions in İzmir's public transportation system. By applying K-Means clustering to passenger boarding data and identifying optimal time-based clusters using the Elbow Method, the model enables more responsive and efficient resource planning tailored to real demand patterns. Unlike traditional fixed interval methods, which apply uniform time segments regardless of demand variability, the proposed adaptive (FS) strategy allows for flexible time intervals and cluster numbers specific to each line and direction.

The results highlight that the adaptive clustering approach improves fleet efficiency by aligning service frequency with dynamic passenger density, especially for lines with fluctuating or high-volume demand. The model demonstrated that similar or improved service levels could be maintained, and in some cases enhanced, with fewer buses by redistributing trips more effectively. Notably, increased total trip numbers in the adaptive model were achieved without increasing the number of required buses, suggesting better vehicle use and enhanced service responsiveness.

The study also has modeling limitations. The current model does not account for passenger waiting times, operational costs (e.g., fuel, maintenance, labor), or service reliability metrics. It assumes a homogeneous fleet and relies solely on K-Means clustering, which may not be optimal for all data structures. Additionally, while weekday and weekend demand differences are considered, alternative clustering criteria or operational constraints could be explored through scenario-based analysis. On this basis, the future research directions can be summarized as follows.

- Integrated Operational Planning: Future studies could extend the proposed approach to include detailed scheduling and route planning for synchronized optimization of bus frequency, dispatching, and fleet allocation.
- Inclusion of Operational Costs: Factoring in costs such as fuel consumption, driver wages, and maintenance would support more financially sustainable planning alongside fleet efficiency.
- Passenger Waiting Time Optimization: Including waiting time as a decision variable could improve both service quality and customer satisfaction.
- Fleet Composition Diversification: Incorporating multiple bus types (e.g., solo, articulated) would enable dynamic vehicle allocation based on time-specific demand.
- Exploring Alternative Clustering Techniques: Testing other clustering algorithms (e.g., DBSCAN, hierarchical clustering) may reveal better segmentation for complex or irregular demand patterns.

By addressing these future directions, the proposed model has the potential to evolve into a comprehensive decision-support system for public transportation agencies, enabling more adaptive, cost-effective, and passenger-centered fleet management. Applying the model to the entire transit network in future studies would allow for a more thorough evaluation of its broader operational impacts. This expanded application would help validate the scalability and real-world feasibility of the approach, while also demonstrating its potential to support system-wide planning and policymaking for public transit authorities.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- N.D., N.M.K., Ü.K.; Data Acquisition- N.D., N.M.K., Ü.K.; Data Analysis/Interpretation- N.D., N.M.K., Ü.K.; Drafting Manuscript- N.D., N.M.K., Ü.K.; Critical Revision of Manuscript- N.D., N.M.K., Ü.K.; Final Approval and Accountability- N.D., N.M.K., Ü.K.

Conflict of Interest: Authors declared no conflict of interest.

Acknowledgement: The authors express their gratitude to ESHOT General Directorate, Department of Transportation Planning, and Statistics Branch Directorate for their support and provision of data used in this research.

Author Details

**Nurhan Dudaklı (Asst. Prof)**

[1] İzmir Bakırçay University, Faculty of Economics and Administrative Science, Department of International Trade and Business, İzmir, Türkiye

0000-0002-5593-5335 nurhan.dudakli@bakircay.edu.tr

**Behice Meltem Kayhan (Asst. Prof)**

[2] Manisa Celal Bayar University, Faculty of Business, Department of Business Administration, Manisa, Türkiye

0000-0001-6881-2580 behice.kayhan@cbu.edu.tr

**Ümit Kuvvetli (Asst. Prof)**

[3] İzmir Bakırçay University, Faculty of Economics and Administrative Science, Department of Business Administration, İzmir, Türkiye

0000-0002-9567-3675 umit.kuvvetli@bakircay.edu.tr

## References

Aditi, A. D., Dureja, A., Abrol, S., & Dureja, A. (2020). Prediction of ticket prices for public transport using linear regression and random forest regression methods: A practical approach using machine learning. In U. Batra, N. Roy, & B. Panda (Eds.), *Data science and analytics* (pp. 140–150). Springer. https://doi.org/10.1007/978-981-15-5827-6_12

Aksoy, İ. C., & Alver, Y. (2024). Addressing electric transit network design frequency setting problem with dynamic transit assignment. *Transportmetrica B: Transport Dynamics, 12*(1), 2318566. https://doi.org/10.1080/21680566.2023.2318566

Aksoy, İ. C., & Mutlu, M. M. (2024). Comparing the performance of metaheuristics on the transit network frequency setting problem. *Journal of Intelligent Transportation Systems*, 1–20.

Bagheri, K., Samany, N. N., Toomanian, A., Jelokhani-Niaraki, M., & Hajibabai, L. (2024). A planar graph cluster-routing approach for optimizing medical waste collection based on spatial constraint. *Transactions in GIS*.

Benli, A., & Akgün, İ. (2023). A multi-objective mathematical programming model for transit network design and frequency setting problem. *Mathematics, 11*(21), 4488. https://doi.org/10.3390/math11214488

Bertsimas, D., Ng, Y. S., & Yan, J. (2020). Joint frequency-setting and pricing optimization on multimodal transit networks at scale. *Transportation Science, 54*(3), 839–853. https://doi.org/10.1287/trsc.2019.0939

Branda, F., Marozzo, F., & Talia, D. (2020). Ticket sales prediction and dynamic pricing strategies in public transport. *Big Data and Cognitive Computing,* 4(4), 36. https://doi.org/10.3390/bdcc4040036

Carrel, A., Lau, P. S., Mishalani, R. G., Sengupta, R., & Walker, J. L. (2015). Quantifying transit travel experiences from the users' perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses. *Transportation Research Part C: Emerging Technologies, 58*, 224–239.

Ceder, A. (2007). *Public transit planning and operation: Theory, modeling and practice*. Elsevier, Butterworth-Heinemann.

Charris, E. L. S., Montoya-Torres, J. R., & Guerrero-Rueda, W. (2019). A decision support system for technician routing with time windows: A case study of a Colombian public utility company. *Academia Revista Latinoamericana de Administración, 32*(2), 138–158.

Chen, Y., & Shan, T. (2021). Research on flexible public transportation planning based on node importance clustering. In *CICTP 2021* (pp. 1800–1807).

dell'Olio, L., Ibeas, A., de Oña, J., & de Oña, R. (2018). Chapter 9 – Data mining approaches. In L. dell'Olio, A. Ibeas, J. de Oña, & R. de Oña (Eds.), *Public transportation quality of service* (pp. 155–179). Elsevier. http://dx.doi.org/10.1016/B978-0-08-102080-7.00009-4

Gadepalli, R., Bansal, P., Tiwari, G., & Bolia, N. (2024). A tactical planning framework to integrate paratransit with formal public transport systems. *Transportation Research Part D: Transport and Environment, 136*, 104438.

Garrido, C., de Oña, R., & de Oña, J. (2014). Neural networks for analyzing service quality in public transportation. *Expert Systems with Applications, 41*, 6830–6838. http://dx.doi.org/10.1016/j.eswa.2014.04.045

Gkiotsalitis, K., Schmidt, M., & van der Hurk, E. (2022). Subline frequency setting for autonomous minibusses under demand uncertainty. *Transportation Research Part C: Emerging Technologies, 135*, 103492.

Ibarra-Rojas, O. J., Delgado, F., Giesen, R., & Muñoz, J. C. (2015). Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological, 77*, 38–75.

Ibarra-Rojas, O. J., Muñoz, J. C., Giesen, R., & Knapp, P. (2019). Integrating frequency setting, timetabling, and route assignment to synchronize transit lines. *Journal of Advanced Transportation, 2019*(1), 9408595.

Jara-Díaz, S., Fielbaum, A., & Gschwender, A. (2017). Optimal fleet size, frequencies and vehicle capacities considering peak and off-peak periods in public transport. *Transportation Research Part A: Policy and Practice, 106*, 65–74.

Jing, D., Yao, E., Chen, R., & Sun, X. (2023). Optimal design method of public transit network considering transfer efficiency. *IET Intelligent Transport Systems, 17*(6), 1118–1136.

Kim, M., Kho, S. Y., & Kim, D. K. (2019). A transit route network design problem considering equity. *Sustainability, 11*(13), 3527.

Lee, I., Cho, S. H., Kim, K., Kho, S. Y., & Kim, D. K. (2022). Travel pattern-based bus trip origin-destination estimation using smart card data. *PLOS ONE, 17*(6), e0270346.

Ma, X., Wu, Y. J., Wang, Y., Chen, F., & Liu, J. (2013). Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies, 36*, 1–12.

Mayaud, J. R., Tran, M., & Nuttall, R. (2019). An urban data framework for assessing equity in cities: Comparing accessibility to healthcare facilities in Cascadia. *Computers, Environment and Urban Systems, 78*, 101401. https://doi.org/10.1016/j.compenvurbsys.2019.101401

Mete, S., Çelik, E., & Gül, M. (2022). Predicting the time of bus arrival for public transportation by time series models. *Journal of Transportation and Logistics, 7*(2), 541–555.

Mirza, A. M., & Jain, R. K. (2025). Review of public transportation integration and modeling strategies: Toward seamless urban mobility. *Multidisciplinary Reviews, 8*(1), 2025018–2025018.

Mutlu, M. M., Aksoy, İ. C., & Alver, Y. (2021). COVID-19 transmission risk minimization at public transportation stops using Differential Evolution algorithm. *[Manuscript submitted for publication]*.

Mutlu, M. M., Aksoy, İ. C., & Alver, Y. (2022). Transit frequency optimization in bi-modal networks using differential evolution algorithm. *Teknik Dergi, 33*(5), 12601–12616.

Oliveira, J. L. A., Aquino, A. L., Pinheiro, R. G., & Nogueira, B. (2024). Optimizing public transport system using biased random-key genetic algorithm. *Applied Soft Computing, 158*, 111578.

Pei, J., Zhong, K., Li, J., & Yu, Z. (2022). PAC: Partial area clustering for re-adjusting the layout of traffic stations in city's public transport. *IEEE Transactions on Intelligent Transportation Systems, 24*(1), 1251–1260.

Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies, 19*(4), 557–568.

Pojani, D., & Stead, D. (2015). Sustainable urban transport in the developing world: Beyond megacities. *Sustainability, 7*(6), 7784–7805.

Purdy, M., & Daugherty, P. (2017). *How AI boosts industry profits and innovation*. Accenture Ltd.

Shalit, N., Fire, M., & Ben-Elia, E. (2020). Imputation of missing boarding stop information in smart card data with machine learning methods. In *Intelligent Data Engineering and Automated Learning – IDEAL 2020* (pp. 17–27). https://doi.org/10.1007/978-3-030-62362-3_3

Sosnowska, J., & Skibski, O. (2018). Path evaluation and centralities in weighted graphs—An axiomatic approach. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* (pp. 3856–3862).

Tian, Q., Wang, D. Z., & Lin, Y. H. (2021). Service operation design in a transit network with congested common lines. *Transportation Research Part B: Methodological, 144*, 81–102.

Tran, L., Mun, M., Lim, M., Yamato, J., Huh, N., & Shahabi, C. (2020). DeepTRANS: A deep learning system for public bus travel time estimation using traffic forecasting. *Proceedings of the VLDB Endowment, 13*, 2957–2960. https://doi.org/10.14778/3415478.3415518

UITP Asia Pacific Centre for Transport Excellence (CTE). (2020). *Artificial intelligence in mass public transport: Executive summary*. https://cms.uitp.org/wp/wp-content/uploads/2020/08/UITP-AP-CTE-AI-in-PT-Executive-Summary-Dec-2018_0.pdf

Vlachopanagiotis, T., Grizos, K., Georgiadis, G., & Politis, I. (2021). Public transportation network design and frequency setting: Pareto optimality through alternating-objective genetic algorithms. *Future Transportation, 1*(2), 248–267.

Wei, M., Liu, T., Sun, B., & Jing, B. (2020). Optimal integrated model for feeder transit route design and frequency-setting problem with stop selection. *Journal of Advanced Transportation, 2020*(1), 6517248.

Xie, Y., Guo, Y., Zhou, T., Mi, Z., Yang, Y., Sadoun, B., & Obaidat, M. S. (2020). A strategy to alleviate rush hour traffics in urban areas based on school-trip commute information. *IEEE Systems Journal, 15*(2), 2630–2641.

Yuan, Y., Shao, C., Cao, Z., He, Z., Zhu, C., Wang, Y., & Jang, V. (2020). Bus dynamic travel time prediction: Using a deep feature extraction framework based on RNN and DNN. *Electronics, 9*(11), 1876. https://doi.org/10.3390/electronics9111876