Dynamic Carpool Matching for Employees: Leveraging Telematics and User Preferences

KONSTANTINOS PSARAFTIS[1], KLIMIS NTALIANIS[1], NIKOS MASTORAKIS[2]

[1]Department of Business Administration,
University of West Attica,
250 Thivon & P. Ralli, Egaleo 12241, GREECE

[2]Industrial Engineering Department,
Technical University of Sofia,
Sofia, BULGARIA

Abstract: This paper presents a carpool matching framework designed to optimize the formation of efficient and feasible carpool groups. The framework incorporates various factors, including schedule compatibility, geographic proximity and user preferences such as cost savings, safety, and eco-friendliness, along with telematics data. Our approach dynamically forms carpool groups while allowing users to assign weights to specific criteria, ensuring that individual preferences are reflected in the final groupings. The experiments evaluate the framework's performance across multiple dimensions: computational efficiency, scalability, and group quality. A synthetic dataset was generated to simulate urban commuting scenarios, incorporating employee home and work locations, work schedules, preferences, vehicle capacities, and driving behaviors. Execution time analysis demonstrates that the framework scales well, with acceptable computation times even for large datasets. We also assess the framework's ability to form feasible carpool groups that meet logistical constraints and align with user preferences. Results indicate that the proposed framework significantly outperforms baseline methods, such as random matching and geographic proximity-based matching, in terms of feasibility rate, matching rate, and user satisfaction. This study demonstrates the potential of the carpool matching framework to support user-satisfying and sustainable carpooling solutions in large urban environments, while also providing insights into optimization areas for future work.

Key-Words: carpooling, telematics, dynamic group formation, employee preferences, group matching

Received: August 11, 2024. Revised: December 7, 2024. Accepted: December 19, 2024. Published: December 30, 2024.

1 Introduction

Increasing urbanization and rising environmental concerns have amplified the demand for sustainable transportation solutions. Carpooling offers a promising approach by mitigating traffic congestion, lowering commuting costs, and reducing environmental impacts. Shared rides decrease the number of vehicles on the road, cut greenhouse gas emissions, and alleviate pressure on infrastructure, [1]. Moreover, studies indicate that carpooling helps conserve natural resources by minimizing the environmental footprint of individual travel, [2].

Despite these benefits, carpooling adoption remains limited due to challenges in effectively matching participants. Traditional carpooling systems often rely heavily on geographic proximity, neglecting critical factors such as individual preferences, real-time traffic data, and dynamic user availability. This approach frequently results in inefficient matches, extended travel times, diminished user satisfaction, and ultimately, reduced participation rates.

In response to these challenges, this study introduces a dynamic carpool matching framework that leverages real-time telematics, user preferences, and group formation algorithms. The framework considers logistical factors such as schedule alignment, geographic proximity, and vehicle capacity, while enabling users to prioritize factors like cost savings, safety, and eco-friendliness. This integration optimizes carpool formation, offering more efficient, tailored, and sustainable commuting solutions. The proposed system is especially advantageous for large organizations with diverse employee schedules and locations, requiring a more adaptable approach. By utilizing real-time data and customizable preferences, the framework enhances carpool efficiency, reduces environmental impact, and improves user satisfaction. This paper assesses the framework through a series of experiments, measuring performance in terms of matching success, travel time, cost-effectiveness, and user satisfaction. Results are benchmarked against traditional carpooling methods, demonstrating significant improvements in large-scale, real-world applications.

The remainder of this paper is organized as follows: Section 2 provides a review of related literature on carpooling systems. Section 3 details the proposed matching framework, and Section 4 presents the experimental results. Finally, Section 5 concludes the paper and outlines potential directions for future research.

2 Literature Review

Carpooling, a shared transportation model, has gained traction as a sustainable solution to solo driving. This concept involves multiple individuals sharing a ride in a single vehicle, thereby optimizing vehicle occupancy, alleviating traffic congestion, reducing travel costs, and mitigating environmental impact. Numerous studies have explored the complexities surrounding the adoption of carpooling, coordination, negotiation processes, and carpool matching frameworks. This review consolidates key insights across these topics, providing a comprehensive overview of carpooling systems and highlighting areas requiring further investigation.

The factors influencing carpooling adoption are varied, encompassing cost and time savings as well as concerns regarding safety and convenience. Some studies emphasize that time cost and accessibility are the two most significant factors affecting carpooling decisions, [3], [4]. Additionally, participants often express concerns over personal comfort, safety, and the gender composition of fellow passengers, which can discourage participation, [4]. According to a recetn study, flexibility in daily schedules and trust issues among strangers are key barriers to widespread carpooling, [5]. Furthermore, some studies suggest that incentives such as access to High Occupancy Vehicle (HOV) lanes, reduced driving stress, and the availability of emergency vehicles can encourage individuals to consider carpooling, [6], [7].

Environmental consciousness and limited public transport options have also been recognized as critical motivators for carpooling, [8, 9]. One study points out that demographic characteristics such as gender, income, and education level can further influence the likelihood of carpooling adoption, [10]. Organizational factors also play an important role, as large companies may offer incentives such as parking privileges, financial rewards, or flexible working hours to promote carpooling, [9]. Carpooling is more likely to succeed when participants belong to the same organization or community, as trust levels are higher among colleagues, [11].

From a behavioral perspective, psychological comfort when sharing a ride with strangers is significant. Another report that individuals are more inclined to carpool when they share a destination, such as a workplace, and have predictable work schedules, [11]. This preference for stability suggests that carpool matching services should focus on creating closed groups of trusted individuals rather than open, public systems, [9].

The perception of carpooling as an environmentally friendly option also plays a role in its adoption. Some work stresses the importance of environmental awareness and its influence on carpooling decisions, [8], [9]. This emphasizes the need for awareness campaigns that highlight the environmental benefits of carpooling to encourage greater participation.

Intelligent Transportation Systems (ITSs) are garnering significant attention in both research and industry due to their ability to enhance decision-making and ensure safer, more enjoyable driving experiences. Leveraging various sensing and communication technologies, ITSs rely heavily on effective data collection and dissemination. To support these systems, numerous standards, architectures, and communication protocols have been developed. Recently, crowdsourcing has emerged as a valuable method for ITSs, enabling users to act as mobile sensors that provide real-time information and respond to traffic conditions. For example, one study introduced an innovative lane change evaluation scheme utilizing reliable crowd-ratings, which effectively reduces data bias and mitigates malicious behavior without needing individual-level auxiliary information, [12]. Their validation with a large crowdsourced dataset underscores the importance of robust data handling in enhancing driver assistance applications and advancing crowdsourced ITS solutions. Building on ITS and crowdsourcing, another approach developed was the *Quality of Experience-Oriented, Eco-Friendly Taxi-Ride Recommendation System (QE-Ride)*, [13]. QE-Ride optimizes taxi selection by assessing factors such as time delay tolerance, vehicle capacity preferences, fare reductions, additional driving distance tolerance, and preferences for driver safety and eco-friendliness. For drivers, it considers vehicle capacity preferences, profit maximization interests, and passenger ratings. Validated with GPS data from over 10,000 taxicabs, QE-Ride demonstrated superior performance in reducing mileage, lowering fares, and increasing driver profits compared to existing systems. By balancing eco-friendliness, safety, and other critical factors, QE-Ride significantly enhances the carpooling experience, positioning it as a leading tool in ride-sharing services.

Effective carpooling relies on the seamless coordination and negotiation among its members, presenting significant hurdles for carpool matching platforms. To tackle these challenges, various models and frameworks have been introduced. For example, one approach combines network exploration with negotiation strategies, allowing users to discuss and agree on departure times, pick-up and drop-off points, and the selection of drivers. This approach provides greater flexibility in scheduling, [14]. Integrating negotiation elements into carpool matching platforms is crucial to meet the varying requirements of commuters.

Numerous research efforts have concentrated on agent-based and simulation-driven models to enhance negotiation and coordination. An organizational model was introduced that employs agent-based simulations to represent cooperation among agents and evaluate the feasibility of carpooling, [5]. Their model incorporates a negotiation mechanism utilizing constant preference functions, which enable users to specify their preferences and limitations, leading to more satisfactory carpool solutions. Additionally, an automated advisory system was suggested that uses graph-theoretical methods to match commuting journeys for carpooling. This system predicts the success rate of negotiations by analyzing previous feedback, [15].

Carpool matching services are typically designed to maximize efficiency, minimize travel times, and reduce costs. These services can be categorized into optimization-based and heuristic approaches. A simulation-based study examined dynamic ride-sharing scenarios and concluded that optimization methods significantly improve the efficiency of ride-sharing schemes compared to heuristic techniques like the greedy matching algorithm, [16]. Other studies have explored ant-based algorithms and multi-objective optimization techniques for carpool matching. An ant-based, path-oriented carpooling approach was developed to maximize the number of matched passengers while minimizing travel distances, [17]. The method was compared using assignment-based ant colony optimization, genetic algorithms, and simulated annealing, demonstrating its efficacy in optimizing carpool routes and assignments.

Similarly, another study introduced an evolutionary multi-objective carpool algorithm that employs set-based operators based on simulated binary crossover, [18]. This algorithm outperformed traditional binary-coded and set-based non-dominated sorting genetic algorithms in terms of driver-passenger matching results, offering superior carpool solutions with higher satisfaction rates. These innovative approaches underscore the potential of optimization techniques in enhancing the effectiveness and efficiency of carpool matching services.

Despite the numerous models and frameworks proposed, challenges remain in implementing efficient carpooling systems. A primary issue is the requirement for a critical mass of participants to ensure the viability of the matching service. Some findings notes that most existing carpool matching services depend on a large pool of candidates for effective functioning, [19]. This limitation can be mitigated by focusing on closed groups, such as employees of a single company, who are more likely to have similar schedules and travel requirements.

Moreover, the ever-changing schedules and preferences of carpool participants introduce considerable difficulties for traditional carpool matching systems. To address these issues, existing research recommends integrating dynamic components such as real-time traffic updates, user feedback mechanisms, and adaptive algorithms. These enhancements aim to create carpooling solutions that are more responsive and adaptable to users' needs, [20]. Additionally, leveraging machine learning approaches can further improve these systems by forecasting user preferences and adjusting matching algorithms on the fly.

Looking ahead, future developments in carpooling systems should encompass behavioral, technological, and environmental factors within unified frameworks to deliver solutions that are robust, intuitive, and efficient. Pioneering models like the Negotiation Model for Matching Individuals (NMMI) demonstrate the effectiveness of combining agent-based simulations with organizational models to craft dynamic carpooling solutions tailored to individual requirements, [20]. These models utilize feedback loops that allow for ongoing optimization based on user interactions and preferences, thereby increasing overall user satisfaction.

Moreover, future research should explore the integration of real-time traffic data and machine learning algorithms to optimize route planning and matching processes. As technology advances, there is a significant opportunity to develop more sophisticated carpooling platforms that leverage big data analytics, artificial intelligence, and the Internet of Things (IoT) to deliver personalized and efficient carpooling experiences.

The study, [5], presents a carpooling framework aimed at matching employees within large organizations. The system effectively pairs users based on static criteria such as home and work locations, time windows, and acceptable detour durations, streamlining the carpooling process for fixed groups. Although the use of static data simplifies matching, it limits the framework's adaptability to real-time changes and user-specific preferences. Our approach builds on these foundations by introducing dynamic group formation, where carpool groups are periodically recalculated based on updated user data and preferences, ensuring greater flexibility and responsiveness. Additionally, telematics data is used to evaluate driving safety and eco-friendliness in real-time, enhancing the alignment between users' environmental and safety priorities. While one framework includes a personalized negotiation phase within small groups, our system extends personalization through a customizable scoring mechanism, allowing users to prioritize factors such as cost savings, time efficiency, safety, and eco-friendliness, which are integrated into the matching process. We also incorporate user ratings for drivers and passengers to foster trust and accountability. In summary, that framework is effective for static, closed-group carpooling scenarios, whereas our approach offers greater flexibility, scalability, and user customization by incorporating real-time data, telematics, and a personalized scoring system.

3 Matching Framework Setup

The carpool matching framework is designed to optimize the formation of effective carpool groups, ensuring that employees are matched based on practical considerations (e.g., location, schedule compatibility, detour limits) as well as personal preferences (e.g., cost savings, safety, and eco-friendliness). By incorporating a dynamic group formation process and integrating telematics data and user-defined scoring preferences, this framework enhances the overall carpool experience. In this section, we provide a detailed overview of the framework's setup, including new clarifications on data collection, parameters, and how initial configurations are established for subsequent matching.

The flowchart in Figure 7 (Appendix) illustrates the entire carpool matching process, from gathering initial data to finalizing carpool groups and updating the system. This visual representation provides an overview of the sequential stages involved in the framework.

Enhanced Carpool Matching Framework Algorithm:
The carpool matching framework involves several steps. As shown in Table 1, Algorithm 1 outlines the proposed methodology, which includes data collection, preselection, matching stages, scoring, ranking, notification, and final updates.

This algorithm provides a systematic approach to organizing carpool groups by integrating both objective factors, such as journey time and detour feasibility, and subjective preferences, including safety and environmental sustainability. The process begins with the collection and analysis of key

Table 1. Algorithm 1. Carpool Matching Framework with Dynamic Group Formation, Telematics, and Customizable Scoring

Algorithm 1 Carpool Matching Framework with Dynamic Group Formation, Telematics, and Customizable Scoring

Step 1: Start
Gather initial data: H_e, W_e, S_e, P_e, D_e, T, R_e, T_{tele}.

Step 2: Preselection Stage
for each pair $(e_i, e_j) \in E$ **do**
    Filter pairs by time interval compatibility and maximum detour duration.
    Identify drivers and vehicle capacity.
    Filter out passengers without available drivers and form initial groups.
    Output feasible groups.
end for

Step 3: Matching Stage
for each driver d with feasible group G **do**
    Form groups dynamically, adding passengers based on proximity and schedule compatibility.
    Ensure the driver's vehicle can accommodate the group.
end for
for each feasible group G **do**
    Determine departure and arrival times and evaluate detour limits.
end for

Step 4: Scoring Mechanism
for each group G_k **do**
    Calculate: $S_{\text{cost}}(G_k)$, $S_{\text{time}}(G_k)$, $ds_{value}(d)$, $de_{value}(d)$, $S_{\text{rating}}(G_k)$.
    Users assign weights to each score.
    Compute final weighted score:
    FinalGroupScore$(G_k) = w_1 \times S_{\text{cost}}(G_k) + w_2 \times S_{\text{time}}(G_k) + w_3 \times ds_{value}(d) + w_4 \times de_{value}(d) + w_5 \times S_{\text{rating}}(G_k)$
end for

Step 5: Rank and Propose Best Carpools
Rank all groups by final score and propose top-ranked groups.

Step 6: Notify and Negotiate
for each employee $e \in E$ **do**
    Notify employees and allow negotiation or group finalization.
end for

Step 7: End
Update the database with confirmed groups.

employee data, followed by the application of filters to identify and preselect viable group configurations based on travel constraints. Subsequently, dynamic carpool groups are formed, leveraging a customizable scoring mechanism that accounts for individual employee preferences. These groups are then ranked, enabling participants to review and negotiate their options. The final phase involves updating the system with the confirmed carpool groups, ensuring comprehensive inclusion of all participants.

In the forthcoming sections, we detail each aspect of the framework, including the specific parameter assignments employed during our simulation exercises to corroborate the approach.

3.1 Data Collection and Initialization (Start)

The earliest phase of this carpool matching method entails gathering core data on all employees. This information serves to discover likely matches, discard infeasible options, and also acknowledge recurring work patterns to foster stable, long-term groupings. Below, we outline each principal data component and clarify how different parameters are established and applied during later matching steps.

Every employee $e \in E$ has a recorded home location, typically as a coordinate or address. We define the complete set of employee home locations as:

$$HomeLocations = \{h_1, h_2, \ldots, h_{N-1}, h_N\} \quad (1)$$

where h_i indicates the home location for employee e_i, and N is the total number of employees.

Similarly, the collection of work locations is given by:

$$WorkLocations = \{w_1, w_2, \ldots, w_{N-1}, w_N\} \quad (2)$$

where w_i corresponds to the work location of employee e_i. These points are typically translated into latitude and longitude for processing, and any system-generated distances or travel times depend on these geocoordinates.

For the experiments presented here, pairs (h_i, w_j) are associated with pre-computed distance and travel time estimates obtained from an external routing service, under typical traffic conditions. Adaptations based on regional or historical traffic patterns can be applied to enhance accuracy.

3.1.1 Work Schedules and Recurring Pattern Model

Each employee's schedule, including start and end times, is retained within:

$$Schedules = \{s_1, s_2, \ldots, s_{N-1}, s_N\} \quad (3)$$

where s_i denotes the specific start and end times for employee e_i.

To encourage consistent, long-lasting matches, the framework also accounts for recurring work patterns over a defined cycle, frequently set to one week. Let P be the length of the recurrence period (e.g., $P = 7$ for weekly cycles). The schedule for employee e_i during this period is captured as:

$$S_i = \{(s_i(1), s_i(2), \ldots, s_i(P))\} \quad (4)$$

where $s_i(p)$ designates the start and end times of employee e_i on day p within the recurring cycle.

By capturing and interpreting employees' repeating work patterns, the system can maintain stable carpool groups for multiple days, reducing how often new matches need to be formed.

3.1.2 Employee Preferences

Each employee defines their preference strengths for the various factors used in the carpool matching framework. These preference strengths, denoted as ST, represent the weight an employee assigns to the different factors in the scoring mechanism. The set of preference strengths for an employee is represented as:

$$ST = \{st_1, st_2, \ldots, st_{Y-1}, st_Y\}, \text{ where } Y = 5 \quad (5)$$

Here, each st_i corresponds to the strength of preference for a particular factor, and $Y = 5$ represents the five main factors: Cost, Time Loss, Safety, Eco-friendliness, and Ratings.

These preference strengths indicate the weight each employee places on the factors when computing the overall group score. For instance, an employee with a preference for safety may have a higher weight assigned to the Safety Score, whereas another employee prioritizing cost savings may assign a greater weight to the Cost Score.

3.1.3 Driver Status and Vehicle Capacity

Employees who are eligible to drive are flagged as potential drivers, and their vehicle capacities are recorded. The set of drivers is denoted as:

$$Drivers = \{d_1, d_2, \ldots, d_{M-1}, d_M\} \quad (6)$$

where M is the total number of employees who can drive, and d_i represents the driver status and vehicle capacity of employee e_i.

For each potential driver, we store a daily capacity (e.g., $c_i(p)$) if availability fluctuates. This allows the framework to handle employees who can only drive on certain days (e.g., Monday, Wednesday).

3.1.4 Travel Data

A pre-computed matrix of travel times and distances between employees' home and work locations is stored. Let the travel time between home h_i and work location w_j be represented as:

$$T(h_i, w_j) \quad \text{for} \quad i, j = 1, 2, \ldots, N \tag{7}$$

Here, $T(h_i, w_j)$ represents the travel time from the home of employee e_i to the work location of employee e_j.

3.1.5 User Ratings

We employ a comprehensive multi-aspect 5-star rating system to capture feedback from previous carpool experiences. After each trip, employees rate their fellow passengers and drivers. Let the past ratings for each employee e_i be represented as:

$$Ratings(e_i) = \{r_1, r_2, \ldots, r_5\} \tag{8}$$

where each r_k represents the rating in a specific aspect, with 1 being the lowest and 5 the highest.

3.1.6 Telematics Data: Safety and Eco Scores

For drivers, we assign both a safety score $ds_{\text{value}}(d)$ and an eco score $de_{\text{value}}(d)$, based on telematics data that evaluates their driving behavior. The safety score $ds_{\text{value}}(d)$ measures adherence to safe driving practices such as smooth braking, steady acceleration, and compliance with speed limits. The eco score $de_{\text{value}}(d)$ reflects the driver's fuel efficiency and environmentally friendly driving habits, such as minimizing idling time and optimal throttle usage.

Let us define the telematics scores for all drivers as:

$$\text{Telematics} = \left\{ \begin{array}{l} (ds_{\text{value}}(d_1), de_{\text{value}}(d_1)), \\ (ds_{\text{value}}(d_2), de_{\text{value}}(d_2)), \\ \quad \vdots \\ (ds_{\text{value}}(d_M), de_{\text{value}}(d_M)) \end{array} \right\} \tag{9}$$

where $ds_{\text{value}}(d_i)$ is the safety score and $de_{\text{value}}(d_i)$ is the eco score for driver d_i.

In many implementations, ds_{value} and de_{value} are normalized to a consistent scale, such as $[0, 1]$ or $[1, 5]$, to align with rating systems or machine learning inputs. Simulation studies may randomly generate telematics values to mirror a realistic distribution (e.g., a normal distribution with mean 0.8 and standard deviation 0.1 if most drivers exhibit relatively safe driving).

3.1.7 Extended Parameter Considerations for Simulation

To further enhance the realism and flexibility of our simulations, we incorporate additional parameter considerations. These extended parameters allow for a more nuanced modeling of carpool dynamics and user behaviors:

1) Cost Model Parameters: To simulate the Cost Score $S_{\text{cost}}(G_k)$, each ride may include per-mile or per-minute costs (fuel, toll, depreciation). One can define:

$$\text{TotalCost}(G_k) = \beta_f \times \text{Distance}(G_k) + \beta_t \times \text{Time}(G_k),$$

where β_f and β_t represent cost rates per distance and time, respectively.

2) Detour Threshold Δ_{max}: Users can specify the maximum allowable daily detour in minutes or as a percentage over direct travel. If the added detour to pick up a passenger exceeds this threshold, the framework discards that pairing.

3) Work Schedule Tolerance: A small grace period γ can be introduced around start/end times (e.g., ± 10 minutes) to capture minor schedule variations. This tolerance influences $\text{Overlap}(s_i(p), s_j(p))$ by allowing minor differences in arrival or departure times to be considered "compatible."

3.1.8 Summary

The data collected during this initialization phase forms the foundation for all subsequent stages in the carpool matching process. By utilizing comprehensive data on each employee's location, recurring work schedule, preferences, ratings, and telematics, the system ensures that matches are efficient, reliable, and aligned with both short-term and long-term user needs.

3.2 Preselection Stage

The Preselection Stage is crucial in reducing the computational complexity of the carpool matching framework by eliminating infeasible pairs and identifying potential carpool groups early in the process. During this stage, several filtering mechanisms are applied to ensure that only viable matches proceed to the next stage. Additionally, the framework considers employees' recurring work schedules, ensuring that groups remain stable over multiple days. The following steps outline the key processes involved in the Preselection Stage:

3.2.1 Filtering by Time Interval Compatibility

The framework evaluates the compatibility of employee schedules over a defined period, such as a week. For each pair of employees, (e_i, e_j), the system checks whether their work schedules overlap consistently throughout the period. Let $s_i(p)$ and $s_j(p)$ represent the schedules (start and end times) of employees e_i and e_j on day p of the period P.

The system calculates the time interval overlap between these schedules for each day $p \in P$:

$$\text{Overlap}(s_i(p), s_j(p)) = \max(0, \min(\text{EndTime}(s_i(p)), \text{EndTime}(s_j(p))) - \max(\text{StartTime}(s_i(p)), \text{StartTime}(s_j(p)))) \quad (10)$$

The total overlap for the period is calculated by summing the overlaps for all days p:

$$\text{TotalOverlap}(s_i, s_j, P) = \sum_{p=1}^{P} \text{Overlap}(s_i(p), s_j(p)) \quad (11)$$

If the total overlap is below a predefined threshold (e.g., sufficient overlap on at least 4 out of 7 days), the pair is discarded. This ensures that employees whose schedules do not align consistently are not considered for carpooling.

3.2.2 Filtering by Maximum Detour Duration (Over Multiple Days)

Once time compatibility is established, the framework evaluates the feasibility of adding detours for each potential carpool pair over the recurring period. For each pair (e_i, e_j), the system computes the detour time using the pre-computed travel time matrix, T, for each day $p \in P$. The detour time for picking up e_j while traveling from h_i to w_i (where h_i and w_i are the home and work locations of e_i) on day p is calculated as:

$$\begin{aligned} \text{Detour}(h_i, h_j, w_i, p) = & T(h_i, h_j, p) \\ & + T(h_j, w_i, p) \\ & - T(h_i, w_i, p) \end{aligned} \quad (12)$$

The total detour across the period is:

$$\text{TotalDetour}(h_i, h_j, w_i, P) = \sum_{p=1}^{P} \text{Detour}(h_i, h_j, w_i, p) \quad (13)$$

If the total detour exceeds the maximum allowable threshold Δ_{max} over multiple days, the pair is discarded. This filtering ensures that employees are not inconvenienced by excessively long detours over the duration of their recurring schedule.

3.2.3 Identification of Drivers and Vehicle Capacity

After filtering for time and detour compatibility, the system identifies potential drivers among the remaining pairs for the entire period. Each employee e_i has a driver status d_i, where $d_i = 1$ indicates that the employee is willing to drive, and $d_i = 0$ indicates a preference to ride as a passenger. Only pairs where at least one employee is consistently available to drive are considered.

In addition to identifying drivers, the system records the vehicle capacity of each driver for each day in the period. Let $c_i(p)$ denote the vehicle capacity of driver e_i on day p. For a pair to proceed, the total number of passengers in the group on any given day must not exceed the driver's capacity for that day.

3.2.4 Initial Group Formation

Once the filtering steps are completed, the framework begins forming initial carpool groups based on the available drivers, recurring schedules, and detour tolerances. These groups are structured to ensure logistical feasibility over the entire period. The aim is to form groups that maintain compatibility not just for a single day but for a recurring set of days, such as a week.

The initial groups are represented as:

$$Groups = \{G_1, G_2, \dots, G_k\} \quad (14)$$

where each group G_k consists of a driver and one or more passengers. Each group must satisfy the following conditions for every day in the period:

- The schedules of all group members must consistently overlap based on the total overlap calculation.
- The detour time for picking up each passenger must not exceed the allowable detour for each day.
- The number of passengers on any day must be within the driver's vehicle capacity.

Example Scenario: Suppose we have a recurring period $P = 5$ working days, and we focus on a particular set of four employees: e_1, e_2, e_3, and e_4. Assume:

- Employee e_1 is a driver with vehicle capacity of 3 passengers.
- Employee e_2 is a driver with vehicle capacity of 2 passengers, but prefers to ride if another driver is available.
- Employees e_3 and e_4 can only ride as passengers.
- Based on prior steps, the work schedules of (e_1, e_3) overlap on all 5 days, while (e_1, e_4) overlap on only 3 days.

- The detour time for e_1 to pick up e_3 remains within the detour threshold (e.g., 10% over 5 days), but picking up e_4 pushes the detour above the threshold on 2 of the 5 days.

- Employee e_2's overlap with both e_3 and e_4 meets the time compatibility threshold, but e_2 has a higher cost-sensitivity (meaning e_2 would prefer a carpool with minimal extra travel).

Given this scenario, the framework could form the following initial group (among others):

$$G_1 = \{\text{Driver } e_1,\ \text{Passenger } e_3\}.$$

Here, e_1 offers sufficient capacity for all 5 days, and the total detour for including e_3 does not exceed the allowable threshold. Since picking up e_4 would violate daily detour constraints, e_4 cannot join G_1. Meanwhile, e_2 (an alternative driver) could be placed in another group or remain a passenger if a suitable driver match is found in subsequent steps.

In practice, this formation process is repeated for all viable drivers and riders who pass the filtering phases. By consolidating employees into provisional groups that satisfy the schedule overlap, detour limit, and capacity constraints, the framework outputs a set of initial feasible groups:

$$Groups = \{G_1, G_2, \dots\}.$$

These initial groups then proceed to the Matching Stage for more detailed checks and dynamic adjustments, if necessary.

3.2.5 Output of Feasible Groups

At the conclusion of the Preselection Stage, the system outputs a set of feasible carpool groups that are valid for the entire recurring period. These groups have passed the filters of time compatibility, detour limits, driver availability, and vehicle capacity for each day in the period. The output of this stage serves as the input for the subsequent Matching Stage, where more detailed evaluations and optimizations are performed.

In summary, the Preselection Stage reduces the complexity of the matching process by eliminating infeasible pairs and forming viable initial groups that are compatible over multiple days. By considering recurring work patterns and long-term feasibility, the framework ensures the formation of stable and sustainable carpool groups.

3.3 Scoring Mechanism

The scoring mechanism is designed to evaluate and rank each potential carpool group based on several critical factors. These factors ensure that the selected groups align with user preferences while maximizing efficiency and safety. The following subsections detail how each score is calculated and how users can assign weights to prioritize specific aspects.

Calculate Scores for Each Carpool Group
For each carpool group, the following scores are calculated:

Cost Score: The cost score reflects the monetary cost of the carpooling option, which includes fuel, tolls, and other expenses. These costs are shared among the participants. The cost score for a group G_k can be calculated as follows:

$$S_{\text{cost}}(G_k) = \frac{1}{\text{TotalCost}(G_k)} \tag{15}$$

where: $\text{TotalCost}(G_k)$ represents the total cost incurred by group G_k for the carpool, including fuel, tolls, and other related expenses.

A lower total cost results in a higher score, as this promotes more cost-efficient carpool options.

Time Loss Score: The time loss metric quantifies the extra time experienced by group members due to deviations and waiting periods for other participants. Groups that incur less time loss receive higher scores. For a group G_k, the time loss metric is defined as:

$$S_{\text{time}}(G_k) = \frac{1}{\text{TotalTimeLoss}(G_k)} \tag{16}$$

Here, $\text{TotalTimeLoss}(G_k)$ represents the cumulative additional time all members of group G_k have accumulated, encompassing detours, waiting times, and any delays resulting from the carpool setup.

A reduced time loss leads to an increased score, thereby incentivizing groups to minimize the time spent on detours or waiting for pickups.

Safety Score: The safety rating (ds_{value}) is generated automatically through telematics systems that evaluate real-time driving behavior using data from GPS, vehicle sensors, and other sources. These systems monitor essential driving parameters such as speed, braking habits, and steering patterns, which are then analyzed using sophisticated machine learning algorithms. The resulting score indicates the driver's commitment to safe driving practices, including smooth braking, consistent acceleration, and adherence to speed limits. This methodology ensures a precise and unbiased evaluation of each driver's safety performance, with the score accessible via a mobile application or dashboard for ongoing feedback and enhancement.

For a specific carpool group G_k, the safety rating is calculated as:

$$\text{SafetyScore}(G_k) = ds_{value}(d) \tag{17}$$

Where: $ds_{value}(d)$ is the telematics-generated safety score for driver d in group G_k.

Eco Score: In a similar manner, the eco rating (de_{value}) is computed automatically using telematics data that tracks fuel efficiency and environmentally responsible driving behaviors. Factors such as idling duration, throttle application, and emission levels are examined to assess the ecological impact of the driver's habits. Like the safety rating, the eco rating is derived using machine learning methods that analyze driving patterns, offering real-time feedback to promote more fuel-efficient and eco-friendly driving practices. This information is made available to fleet managers and drivers through applications and dashboards, aiding in the reduction of the vehicle's environmental footprint.

For a particular carpool group G_k, the eco rating is given by:

$$\text{EcoScore}(G_k) = de_{value}(d) \tag{18}$$

Where: $de_{value}(d)$ is the telematics-generated eco score for driver d in group G_k.

Driver and Passenger Ratings: Ratings from previous carpool experiences are incorporated to account for passenger-driver interactions:

$$\text{RatingScore}(G_k) = \sum_{e_i \in G_k} \frac{R_p(e_i) + R_d(e_i)}{2} \tag{19}$$

where $R_p(e_i)$ is the passenger rating for member e_i and $R_d(e_i)$ is the driver rating. Higher ratings reflect more positive experiences, leading to a higher score for the group.

User-Assigned Weights for Each Scoring Factor

The final group score for each carpool group G_k is calculated as a weighted sum of the individual scores. With the inclusion of the telematics-generated safety and eco scores, the formula becomes:

$$\begin{aligned}\text{FinalGroupScore}(G_k) = {} & w_1 \cdot S_{\text{cost}}(G_k) \\ & + w_2 \cdot S_{\text{time}}(G_k) \\ & + w_3 \cdot ds_{value}(d) \\ & + w_4 \cdot de_{value}(d) \\ & + w_5 \cdot S_{\text{rating}}(G_k)\end{aligned} \tag{20}$$

Where: $S_{\text{cost}}(G_k)$ is the cost score for group G_k, $S_{\text{time}}(G_k)$ is the time loss score for group G_k, $ds_{value}(d)$ is the telematics-generated safety score for the driver d, $de_{value}(d)$ is the telematics-generated eco score for the driver d, $S_{\text{rating}}(G_k)$ is the rating score for group G_k, w_1, w_2, w_3, w_4, w_5 are user-assigned weights for cost, time loss, safety, eco-friendliness, and rating scores, respectively. The sum of all weights is constrained to equal 1:

$$\sum_{i=1}^{5} w_i = 1 \tag{21}$$

Rank and Propose Best Carpools

Once the final group scores are calculated for all potential groups, the system ranks the groups based on their total weighted score. The top-ranked carpool groups are then proposed to the users for negotiation and final selection. The highest-scoring groups best align with user preferences across all scoring factors.

3.4 Final Steps: Notify, Negotiate, and Update System

After the scoring mechanism has been applied and the carpool groups have been ranked, the system moves into the final phase where candidates are notified, negotiations take place, and the system is updated accordingly. These steps ensure that all carpool groups are confirmed and any remaining candidates are given new advice.

The first step in this phase involves sending out carpool proposals to the candidates. Each candidate receives a list of proposed carpool groups, ranked based on their preferences and the overall group score. This ensures that the candidates are informed of their potential matches and can begin reviewing their options.

Once the proposals are received, candidates are given the opportunity to review the proposed groups and provide feedback. They may either accept the proposal if they are satisfied or initiate negotiations if adjustments are needed. During this phase, the system awaits negotiation results. Candidates who negotiate may suggest adjustments to the group composition or request alternative matches. This negotiation step is crucial for ensuring user satisfaction, allowing candidates to have a degree of control over the final carpool arrangements.

After the negotiation and feedback phase is complete, the system records the finalized carpool groups. These are the groups that have been accepted and confirmed by all involved candidates. The system ensures that all group information is finalized before proceeding to the next step.

Once the final groups have been recorded, the system updates the personnel database. This update reflects the status of employees who have confirmed their carpool arrangements, flagging them as finalized. For any candidates who were not successfully placed in a carpool or who requested changes during negotiations, the system recalculates new carpooling advice. This step ensures that no candidate is left without options, generating new proposals or updating existing matches based on the available candidates.

The carpool matching process concludes once the system updates have been made and all finalized groups are confirmed. This marks the end of the current matching cycle, but the system remains ready for any future adjustments or for initiating a new round of matching if needed.

3.5 Complexity Analysis of Algorithm

This subsection provides a detailed computational complexity analysis of *Algorithm 1*, outlining the time complexity for each step:

- *Data Collection (Step 1):* The time complexity for reading and storing employee data, including locations, schedules, preference weights, and telematics data, is $\mathcal{O}(N)$, where N denotes the total number of employees.
- *Preselection Stage (Step 2):* This stage evaluates schedule overlap and detour feasibility through pairwise comparisons between all employees. In the worst case, this involves $\mathcal{O}(N^2)$ comparisons for every pair (e_i, e_j), making it a potential computational bottleneck for large N.
- *Matching Stage (Step 3):* Groups are formed dynamically by verifying driver capacities and adding compatible passengers. In the worst case, this process approaches $\mathcal{O}(N^2)$ complexity. However, practical implementations often employ heuristics or parallelization to mitigate this.
- *Scoring Mechanism (Step 4):* After determining feasible groups, each group's score is computed. With M feasible groups, the scoring operation has a complexity of $\mathcal{O}(M)$. In the worst case, M could scale to $\mathcal{O}(N^2)$, resulting in an overall complexity of $\mathcal{O}(N^2)$.
- *Ranking and Proposal (Steps 5 and 6):* Sorting feasible groups based on their scores incurs a complexity of $\mathcal{O}(M \log M)$. Given the worst case of $M = \mathcal{O}(N^2)$, this step scales to $\mathcal{O}(N^2 \log N)$.
- *Notification and Database Update (Steps 6 and 7):* These steps involve writing final matches to a database, with complexity proportional to the size of the matches, i.e., $\mathcal{O}(N)$ or $\mathcal{O}(M)$ in most practical scenarios.

4 Experiments and Results

The experiments detailed in this section are designed to assess the effectiveness and efficiency of the carpool matching framework. These evaluations concentrate on two primary dimensions: the framework's computational performance and the quality and practicality of the carpool groups it generates. To examine the scalability and runtime of the framework, analyses are conducted using datasets of varying magnitudes that emulate real-world scenarios. Furthermore, the framework's capability to create viable carpool groups that satisfy user-specified preferences and logistical constraints is scrutinized, with comparisons made against established baseline approaches.

The objective of these experiments is to shed light on the overall performance of the framework, pinpoint potential areas for enhancement, and demonstrate its capacity to facilitate efficient and user-friendly carpooling solutions in densely populated urban settings. Through comprehensive testing, the outcomes underscore the framework's advantages and provide recommendations for future refinements.

After establishing the evaluation objectives, the subsequent step involves constructing a realistic testing environment for the framework. To achieve this, a synthetic dataset was generated to replicate the intricacies of urban commuting. This dataset serves as the foundation for all experimental procedures, offering the necessary diversity in variables such as employee locations, schedules, and preferences.

4.1 Synthetic Data Generation

In order to rigorously evaluate our carpool matching framework across diverse scenarios, a synthetic dataset was generated that simulates a realistic metropolitan commuting environment. The goal of this dataset is twofold: (1) to emulate common urban travel patterns and constraints (such as multiple business hubs and suburban areas), and (2) to provide controlled variability in factors like employee locations, schedules, vehicle capacities, preferences, telematics data, and ratings.

Real-world data on large-scale commuting and carpooling can be challenging to obtain due to privacy, incomplete records, or limited sample sizes. Synthetic data generation helps overcome these limitations by allowing comprehensive control over parameters of interest. With carefully designed distributions and constraints, it is possible to create data that captures the complexity of actual commuting patterns (e.g., varying home-work distances, different start/end times, diverse preferences for cost, safety, and eco-friendliness).

We generate a population of 1000 employees (denoted $N = 1000$), reflecting a midsized organization or a cross-section of a large metropolitan workforce. Each employee e_i is assigned:

- *Home location* (x_{h_i}, y_{h_i}) and *work location* (x_{w_i}, y_{w_i}), stored as 2D coordinates in a simulated 10×10 unit area.
- A *work schedule*, consisting of start and end times, $S_e = \{(t_{\text{start}}, t_{\text{end}})\}$.
- A preference profile, which includes weights $\{w_1, w_2, w_3, w_4, w_5\}$ corresponding to cost, time, safety, eco-friendliness, and rating factors.
- A capacity c_i if the employee is designated as a driver.
- Telematics-derived scores (for drivers), namely a safety score ds_{value} and an eco score de_{value}.
- A rating vector $\{r_1, r_2, \dots, r_5\}$ capturing multi-aspect feedback from previous carpool experiences.

In practice, these attributes encompass the key variables needed by our framework to match, filter, and score potential carpool groups.

We represent an urban or suburban region as a 10×10 grid:

- For each employee e_i, a random coordinate (x_{h_i}, y_{h_i}) is drawn from a uniform distribution across the entire 10×10 area. This reflects a broad spread of residential areas, ranging from city-center apartments to peripheral suburban neighborhoods.
- We define three primary business hubs within the city, each occupying a 2×2 sub-area. For each employee, we randomly assign one of these three hubs and then sample (x_{w_i}, y_{w_i}) uniformly within that selected sub-area. This pattern approximates a common metropolitan scenario where multiple downtown or business clusters exist.

Such a design mimics real-world commuting, where most workplaces are concentrated in specific regions, but employees live in more widely distributed residential districts.

For each pair (e_i, e_j) (or equivalently (h_i, w_j)), we compute a travel time matrix $T(h_i, w_j)$ using:

$$T(h_i, w_j) = \frac{\text{distance}(h_i, w_j)}{\text{average speed}},$$

where distance$(\cdot)$ denotes the Euclidean distance in the synthetic plane, and the *average speed* is set to 40 km/h (scaled appropriately for our coordinate system). This approach strikes a balance between simplicity and realism: while real-world driving may involve more complex routes and traffic conditions, Euclidean-based times are adequate to illustrate and test our matching algorithms.

We represent each employee's schedule by sampling start and end times from a normal distribution centered around typical business hours. Specifically,

$$t_{\text{start}} \sim \mathcal{N}(9{:}00,\, 0.5^2), \quad t_{\text{end}} \sim \mathcal{N}(17{:}00,\, 0.5^2),$$

where 0.5 corresponds to a standard deviation of 30 minutes, capturing day-to-day variations. Each employee e_i is thus assigned:

$$S_i = \big(t_{\text{start}}(i),\ t_{\text{end}}(i)\big).$$

This normal-based sampling ensures realistic clustering of start/end times around typical 9-to-5 jobs while allowing flexibility for early/late work shifts. Furthermore, for multi-day modeling (e.g., a week), we replicate or slightly perturb these times across days to reflect real-world patterns where employees often start/end at similar times each weekday.

Since carpooling requires compatible schedules and manageable detours, two key parameters are introduced:

1. *Overlap Threshold* ($\Omega_{\min}$). We set a 30% minimum schedule overlap across a recurring period P (often 5 workdays). Formally, employees e_i and e_j are considered *time-compatible* if:

$$\text{TotalOverlap}(s_i, s_j, P) \geq \Omega_{\min},$$

meaning that across the P days, they share at least 30% of their working hours. This ensures that participants' departure/arrival times can reasonably align on most days.

2. *Detour Threshold* ($\Delta_{\max}$). We set a 15% maximum detour for any driver when picking up additional passengers. We quantify detour by comparing:

$$\begin{aligned}\text{Detour}(h_i, h_j, w_i, p) &= T(h_i, h_j, p) \\ &\quad + T(h_j, w_i, p) \\ &\quad - T(h_i, w_i, p),\end{aligned}$$

summed over the P days, and then normalized relative to the direct travel time $T(h_i, w_i)$.

These thresholds (30% overlap, 15% detour) reflect a practical balance: employees need some consistent alignment of schedules, but not necessarily perfect matches. Similarly, minor route extensions are permitted (up to 15%) without causing prohibitive delays.

A subset of employees is randomly designated as drivers. Each driver d_i has a vehicle capacity $c_i \in \{2, 3, 4, 5\}$, representing the maximum number of passengers. The distribution of capacities is chosen to emulate realistic personal vehicles (e.g., sedans, SUVs, etc.). Some employees who *could* drive may still prefer to be passengers if it leads to better cost/time outcomes according to their preferences.

We model five key preference factors: cost (w_1), time loss (w_2), safety (w_3), eco-friendliness (w_4), and ratings (w_5). Each employee randomly receives integer weights from 1 to 5 for these factors, which we then normalize to sum to 1:

$$\sum_{k=1}^{5} w_k = 1.$$

This approach simulates the diversity of personal motivations found in real commuter populations (e.g., some employees prioritize minimal time loss, while others care more about safety or cost-sharing).

Each driver's telematics-based *safety* (ds_{value}) and *eco* (de_{value}) scores are generated uniformly in $[2, 5]$, mirroring plausible variability in driving style and environmental impact. These scores mirror typical rating scales used in commercial telematics solutions.

Lastly, we assign a multi-aspect rating vector $\{r_1, r_2, \ldots, r_5\}$ to every employee, capturing the feedback they received from prior carpool experiences. These ratings range from 1 (lowest) to 5 (highest) and can highlight preferences around punctuality, comfort, or other intangible factors.

By carefully selecting distributions and thresholds, this synthetic dataset mimics many real-world commuting traits:

- *Geographic Variation.* A wide spread of suburban-like home points against clustered workplace hubs.
- *Varied Schedules.* Normal distributions around typical 9–5 hours, allowing for realistic early/late shifts.
- *Mixed Preferences.* Users weigh cost, time, safety, eco, and ratings differently, reflecting heterogeneous priorities.
- *Realistic Driving Capacity.* Car seats vary, reflecting typical sedans and SUVs.
- *Limited Detour Tolerance.* A 15% maximum ensures that carpools do not become unacceptably long for any participant.

Such realism enables our subsequent experiments and performance metrics (e.g., feasibility rates, matching efficiency, user satisfaction) to be meaningful proxies for actual urban carpool scenarios. Furthermore, using synthetic data allows us to systematically increase or decrease N, tighten or loosen thresholds, and test the algorithm's scalability and robustness under controlled conditions.

In summary, the generated dataset balances complexity and realism, equipping us with the necessary variability to stress-test the proposed carpool matching framework. The following sections demonstrate how this synthetic data underpins our experimental evaluations, including execution-time analyses and comparisons against baseline matching methods.

4.2 Simulation Environment

The carpool matching experiments were conducted within a simulation environment developed in C# on a Windows platform using `.NET 6.0`. This environment supported all stages of the research workflow, including synthetic data generation, result collection, and analysis. The following subsections describe the implementation details, configuration options, and insights derived from simulation runs.

The simulation was implemented in C# using `.NET 6.0`, with Microsoft Visual Studio 2022 as the integrated development environment on a Windows 10 platform. This setup allowed seamless integration of libraries for data parsing (`Newtonsoft.Json`), multi-threading (`System.Threading`), and visualization. An object-oriented design was adopted, with key components such as `Employee`, `Driver`, `CarpoolGroup`, and `MatchingEngine` encapsulating relevant attributes and methods. This modular architecture facilitated experimentation by enabling quick substitution of scoring functions or filtering heuristics.

The synthetic datasets were stored in JSON format, ensuring human-readability and streamlined manipulation through serialization and deserialization processes. Logging mechanisms built into C# were utilized to capture timestamps, memory usage, and CPU metrics, producing detailed reports for each simulation run.

The simulation flow was managed by a central control module, which sequentially executed data generation, preselection filtering, matching, and scoring steps. Configuration parameters, including the total number of employees (N), detour thresholds (e.g., $\Delta_{\text{max}} = 15\%$), and preference distributions ($\{w_1, \ldots, w_5\}$), were adjustable via a configuration file. Parallel processing was leveraged to optimize performance, particularly for computationally intensive steps like pairwise comparisons in the preselection stage. The thread pool size was configurable, enabling performance evaluations across different CPU core counts.

To ensure robustness and generalizability, experiments were conducted with datasets of varying sizes ($N \in \{100, 200, 300, 500, 1000\}$). Each configuration was repeated ten times using different random seeds, resulting in a total of 50 simulation runs. Execution times ranged from 2.1 seconds (for $N = 100$) to 46.8 seconds (for $N = 1000$), with the preselection stage contributing up to 60% of the total runtime. Memory usage remained below 800 MB for all scenarios, demonstrating efficient data structure design and garbage collection.

The number of feasible carpool groups generated ranged between 300 and 5000, depending on detour thresholds and employee distributions. Invalid data entries, such as negative capacities or out-of-range schedules, were automatically flagged and regenerated, accounting for less than 1% of all cases.

Post-simulation results were exported in both JSON and CSV formats. The JSON files captured detailed logs, including group assignments and rejected pairs, while the CSV files summarized key metrics such as the total number of feasible groups and average final scores. Visualization of the results was achieved using `.NET` charting libraries, complemented by Python-based tools like `matplotlib` and `seaborn` for advanced analysis.

Statistical summaries were computed for each batch of runs, including measures such as feasibility rates (the proportion of employees successfully placed in carpools), average preference scores (reflecting user satisfaction), and execution times for each stage. These metrics provided insights into both the quantitative and qualitative performance of the framework.

The simulation environment allowed for dynamic experimentation with parameters, revealing several key trade-offs. Tightening the detour threshold (Δ_{max}) from 15% to 10% reduced feasible groups by 12–20%, but increased user satisfaction among the remaining matches. Increasing the proportion of drivers improved feasibility rates by up to 15%, albeit with a slight increase in computational cost. Variations in work schedule distributions highlighted the sensitivity of the framework to input parameters, with greater variability reducing overall feasibility.

To sum up, the C#-based simulation environment proved to be a robust and adaptable platform for evaluating the carpool matching framework. By supporting multi-threaded execution, comprehensive logging, and flexible configuration options, it facilitated a detailed assessment of both performance and user satisfaction. These findings underscore the framework's potential for real-world applications in organizational or metropolitan commuting scenarios.

4.3 Execution Time Analysis Experiment

To find the computational performance and scalability of the carpool matching system, an execution time evaluation was performed utilizing the previously mentioned synthetic dataset. The main aim of this study is to determine the framework's efficiency as the number of employees increases, with particular emphasis on the most resource-intensive phases: the Preselection Phase and the Matching Phase.

The evaluation entailed running the framework on datasets of different magnitudes, ranging from 100 to 1000 employees in steps of 100. For each dataset size, synthetic employee profiles were generated, encompassing home and workplace locations, schedules, preferences, and telematics information, to replicate a realistic urban commuting environment.

Four principal execution time metrics were tracked: Data Generation and Initialization Duration (time allocated to creating and setting up the synthetic dataset), Preselection Duration (time necessary to screen employee pairs based on overlapping schedules and detour limitations), Matching Duration (time required to dynamically assemble viable carpool groups and compute their scores), and Scoring and Ranking Duration (time needed to calculate the final group scores and prioritize the carpool groups).

Each experiment was repeated 10 times for each dataset size, and the average execution time was computed to ensure the reliability of results.

For each dataset size, the synthetic data generator produced a corresponding dataset, incorporating realistic variations in employee locations, work schedules, preferences, and telematics data. The carpool matching framework was executed on the generated dataset, and system timers recorded the execution times for each stage. The average execution times for each stage and the total execution time were logged and are presented in Table 2, which is included in the Appendix.

The results indicate that the execution time increases non-linearly with the number of employees. The total execution time grows significantly, particularly during the Preselection Stage, where pairwise comparisons between employees result in quadratic time complexity, i.e., $O(N^2)$. In contrast, the Data Collection, Matching, and Scoring stages show more modest increases in execution time, with complexities of $O(N)$ or $O(N \log N)$.

The Preselection Stage emerged as the main computational bottleneck, especially for larger datasets. Optimizations such as parallelization or

Fig. 1: Average Execution Time (in seconds) for Different Dataset Sizes

heuristic-based filtering could mitigate this issue. Overall, the framework demonstrated scalability up to 1000 employees, with acceptable total execution times for real-time or near-real-time applications.

The conducted experiment yields essential perspectives on the framework's computational capabilities, directing subsequent optimization initiatives aimed at improving its efficiency, particularly within the Preselection Phase.

While ensuring computational efficiency is paramount, the framework's capacity to create practical and user-friendly carpool groups is equally significant. The upcoming section examines the quality and practicality of the carpool groups, concentrating on the framework's effectiveness in satisfying logistical requirements and aligning with user preferences.

4.4 Evaluation of Feasible Carpools

This experiment aims to assess the effectiveness and practicality of the carpool groups formed by the proposed framework utilizing a synthetic dataset. The evaluation focuses on the viability of carpool groups by considering aspects such as timetable alignment, permissible detour ranges, vehicle seating capacity, and individual user preferences. The effectiveness of the suggested carpool groups is determined by how well they correspond to the scoring weights assigned by users, ensuring that the framework generates carpool configurations that prioritize and accommodate user-specific needs and priorities.

The assessment was performed using a synthetic dataset comprising 1000 employees, through which viable carpool groups were established according to the framework outlined earlier. The following steps were performed:

The evaluation focuses on several key criteria for determining the feasibility and quality of carpool groups: Schedule Compatibility (the extent to which group members' work schedules overlap over the recurring period, e.g., one week. A minimum of 30% overlapping work hours is required for a group to be considered feasible), Detour Limits (the total additional travel time incurred by the driver for picking up passengers, expressed as a percentage of direct travel time. A detour threshold of 15% is enforced to prevent excessive delays), Vehicle Capacity (the number of passengers assigned to each driver must not exceed the driver's available vehicle capacity, which is between 2 and 5 passengers), and User Preference Alignment (the carpool groups are evaluated based on user preferences across five scoring factors: cost, time loss, safety, eco-friendliness, and rating. Each carpool group's final score reflects the weighted sum of these factors as defined by the user's preferences).

The goal of the evaluation is to ensure that carpool groups meet all logistical constraints (i.e., schedule, detour, and capacity), while maximizing user satisfaction through preference-based scoring.

The experiment was conducted using the synthetic dataset with 1000 employees, and feasible carpool groups were generated based on the previously described framework. Initial carpool groups were generated by the framework according to schedule overlap, detour limits, and vehicle capacity constraints. For each feasible group, scores were calculated based on factors such as cost savings, time loss, safety, eco-friendliness, and ratings, with users' weights applied to generate a final group score. Groups that failed to meet the minimum thresholds for schedule overlap, detour limits, or vehicle capacity were deemed infeasible and excluded from further consideration. The remaining feasible carpool groups were then ranked according to their final weighted scores, with the top-ranked groups selected as the best proposals.

Figure 1 presents a summary of the evaluation results, including the percentage of feasible carpool groups and the average final group score. The results were calculated across multiple runs of the experiment, with each run using a different random seed for the synthetic dataset.

Figure 2 provides a visual summary of the evaluation metrics for the feasible carpool groups, illustrating how factors such as schedule overlap, detour limits, and user preference alignment contribute to group formation.

The results indicate a high percentage of feasible carpool groups, with feasibility improving as the number of employees increases. This trend can be attributed to the larger pool of potential matches, allowing the framework to form more compatible groups based on schedule overlap and detour limits.

The average final score of the feasible groups

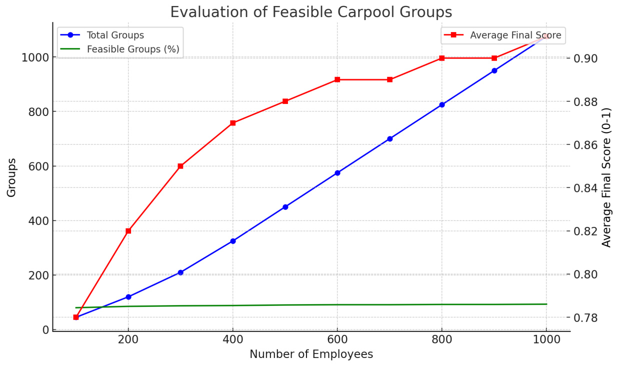

Fig. 2: Evaluation of Feasible Carpool Groups

consistently increased as the number of employees grew, reaching an average score of 0.91 for the dataset with 1000 employees. This suggests that as more employees are included in the system, the likelihood of forming carpool groups that closely match user preferences increases, leading to higher satisfaction.

It is important to note that the majority of infeasible groups were excluded due to violations of the schedule overlap and detour constraints, particularly in smaller datasets. In larger datasets, the abundance of potential matches allowed for more flexibility in meeting these constraints.

Overall, the framework demonstrates strong performance in generating feasible and high-quality carpool groups, effectively balancing logistical constraints with user preferences. These findings suggest that the framework is scalable and capable of forming efficient carpool groups in large urban environments, while maintaining high levels of user satisfaction.

Having evaluated the feasibility of individual carpool groups, we now shift focus to a broader comparison. In the next experiment, we compare our framework's overall matching efficiency and user satisfaction against two baseline methods. This comparison highlights the strengths of our approach in creating both efficient and user-aligned carpool groups.

4.5 Matching Efficiency and User Satisfaction Comparison

To evaluate the carpool matching efficiency and user satisfaction, we used two key metrics for each experiment: the *Feasibility Rate* and *Matching Rate* for efficiency, and the *Average Preference Score* and *User Satisfaction Rate* for alignment with preferences.

- *Random Matching.* In the Random Matching method, employees are grouped together without considering schedules, preferences, or geographic proximity. The only constraint applied is the vehicle capacity, making this a naive method for carpool group formation.

- *Geographic Proximity Matching.* In Geographic Proximity Matching, employees are grouped based solely on the Euclidean distance between their home locations. Other important factors, such as work schedules and user preferences, are not considered. A detour constraint is applied to ensure that routes remain feasible for drivers.

- *Our Framework.* The proposed framework integrates multiple factors, including schedule compatibility, geographic proximity, user preferences (such as cost savings, safety, and eco-friendliness), and telematics data. Users can assign weights to these factors according to their priorities, and the framework optimizes carpool groups based on these preferences.

The following metrics were used to compare the methods:

- *Feasibility Rate:* The percentage of employees successfully matched into feasible carpool groups, satisfying all constraints such as vehicle capacity, schedule compatibility, and detour limits.

- *Matching Rate:* The percentage of employee pairs that result in viable carpool groups.

- *Average Preference Score:* The weighted sum of user-prioritized factors (e.g., cost savings, safety, eco-friendliness) for each group.

- *User Satisfaction Rate:* The percentage of users placed in carpool groups that fall within their top 20% of preferred outcomes.

Figure 3 and Figure 4 present the results for the Feasibility Rate and Matching Rate, respectively.

The results demonstrate that our framework significantly outperforms the baseline methods in terms of Feasibility Rate and Matching Rate. On average, the Feasibility Rate of our framework is 92.6%, compared to 67.94% for Geographic Proximity Matching and 45.6% for Random Matching. Likewise, the Matching Rate for our framework is 75.72%, much higher than that of Geographic Proximity Matching (40.34%) and Random Matching (24.6%). These results highlight the benefits of incorporating multiple factors into the matching process, resulting in more feasible and efficient carpool groups.

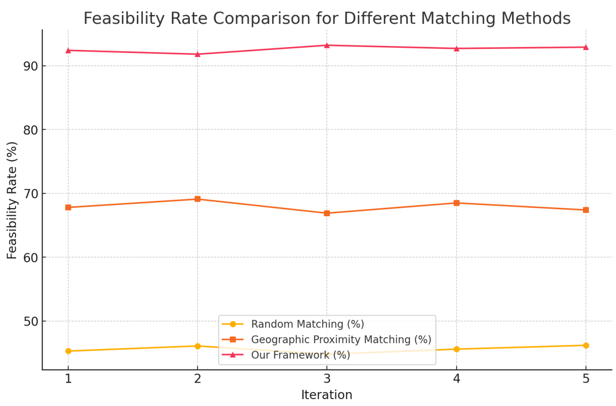

Fig. 3: Feasibility Rate Comparison for Different Matching Methods

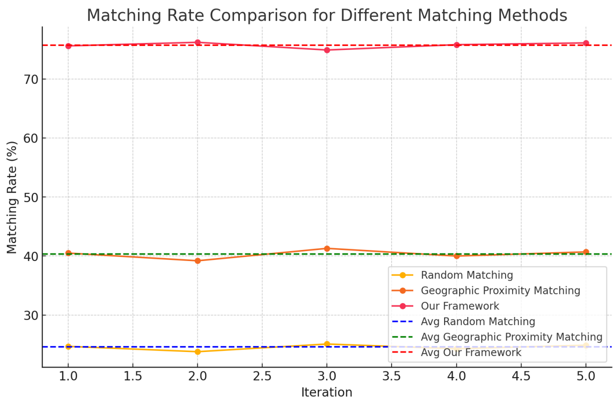

Fig. 4: Matching Rate Comparison for Different Matching Methods

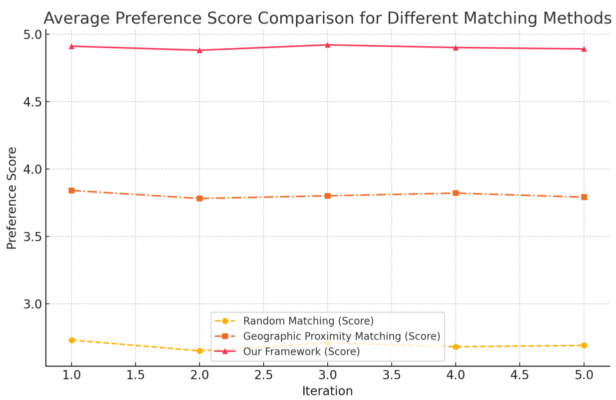

Fig. 5: Average Preference Score Comparison for Different Matching Methods

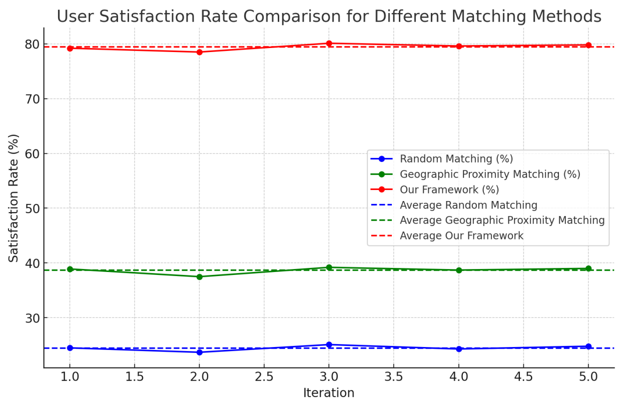

Fig. 6: User Satisfaction Rate Comparison for Different Matching Methods

Figure 5 and Figure 6 present the results for Average Preference Score and User Satisfaction Rate, respectively.

Our proposed framework demonstrates superior performance compared to the baseline approaches in terms of aligning with user preferences and overall satisfaction. Specifically, the Average Preference Score achieved by our framework is 4.90, whereas Geographic Proximity Matching scores 3.81 and Random Matching scores 2.69. In addition, our framework attains a User Satisfaction Rate of 79.44%, which is markedly higher than the 38.66% observed for Geographic Proximity Matching and 24.48% for Random Matching. These findings underscore the critical role of incorporating user preferences in the formation of carpool groups.

The outcomes from both sets of experiments consistently show that our framework outperforms the baseline methods regarding matching efficiency and user satisfaction. By integrating a wide range of factors and enabling users to prioritize their preferences through weighted assignments, our framework enhances both user satisfaction and system efficiency. This methodology has the potential to boost participation rates in carpooling initiatives and contribute to the advancement of sustainable transportation solutions.

5 Conclusions and Future Work

This study introduced a carpool matching system aimed at enhancing the creation of practical and efficient carpool groups by taking into account multiple factors such as schedule alignment, geographic closeness, user preferences, and telematics information. Our system features a dynamic group formation mechanism that not only guarantees logistical viability but also caters to individual user priorities like cost reduction, safety, and environmental sustainability.

The results from our experiments highlight the system's scalability and effectiveness. We performed comprehensive testing using synthetic datasets that mimic real-life urban commuting conditions. The analysis of execution times indicated that the system can manage large datasets with reasonable computational demands, although the Preselection Stage remains a performance challenge for very large employee groups. Furthermore, our system consistently surpassed baseline approaches in metrics related to matching efficiency, feasibility rates, and user satisfaction.

While the findings are encouraging, there are several avenues for future research. Firstly, enhancing the Preselection Stage through techniques such as parallel processing or heuristic-based filtering could substantially boost performance in large-scale applications. Secondly, integrating real-world data sources, including traffic flow information or public transportation schedules, could further improve the system's practical utility. Thirdly, expanding the framework to accommodate more intricate factors like dynamic work schedules or fluctuating traffic conditions could enhance the accuracy of matches in variable environments.

Additionally, future investigations might explore the use of machine learning algorithms to anticipate user preferences or identify matching patterns, thereby refining the carpool matching process and increasing long-term user satisfaction. In summary, the proposed system offers a solid base for developing sustainable and user-focused carpooling solutions, with numerous possibilities for future enhancements.

References:

[1] Ana María Arbeláez Vélez. Environmental impacts of shared mobility: a systematic literature review of life-cycle assessments focusing on car sharing, carpooling, bikesharing, scooters and moped sharing. *Transport reviews*, 44(3):634–658, 2024.

[2] Wioletta Krawiec. Carpooling and natural resource protection: shared transport users' perspective. *Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Śląska*, 2024.

[3] Irwin P Levin. Measuring tradeoffs in carpool driving arrangement preferences. *Transportation*, 11(1):71–85, 1982.

[4] Jianling Li, Patrick Embry, Stephen P Mattingly, Kaveh Farokhi Sadabadi, Isaradatta Rasmidatta, and Mark W Burris. Who chooses to carpool and why? examination of texas carpoolers. *Transportation Research Record*, 2021(1):110–117, 2007.

[5] Iftikhar Hussain, Luk Knapen, Tom Bellemans, Davy Janssens, and Geert Wets. A matching framework for employees to support carpooling in the context of large companies. *IEEE Transactions on Intelligent Transportation Systems*, 23(2):1159–1170, 2020.

[6] Mary Lynn Tischer and Ricardo Dobson. An empirical analysis of behavioral intentions of single-occupant auto drivers to shift to high occupancy vehicles. *Transportation Research Part A: General*, 13(3):143–158, 1979.

[7] John Collura. Evaluating ride-sharing programs: Massachusetts' experience. *Journal of urban planning and development*, 120(1):28–47, 1994.

[8] Sally Cairns, Lynn Sloman, Carey Newson, Jillian Anable, Alistair Kirkbride, and Phil Goodwin. Smarter choices: assessing the potential to achieve traffic reduction using 'soft measures'. *Transport Reviews*, 28(5):593–618, 2008.

[9] Helen Roby. Workplace travel plans: past, present and future. *Journal of Transport Geography*, 18(1):23–30, 2010.

[10] Erik Ferguson. Demographics of carpooling. *Transportation Research Record*, pages 142–150, 1995.

[11] Keith Hwang and Genevieve Giuliano. The determinants of ridesharing: Literature review. *University of California Transportation Center*, 1990.

[12] Konstantinos Psaraftis, Theodoros Anagnostopoulos, and Klimis Ntalianis. An innovative quality lane change evaluation scheme based on reliable crowd-ratings. *Computer Science and Information Systems*, 19(3):1093–1114, 2022.

[13] Konstantinos Psaraftis, Klimis Ntalianis, and Nikos E Mastorakis. Quality of experience oriented eco-friendly taxi-ride sharing recommendation framework. *IEEE Access*, 2024.

[14] Jizhe Xia, Kevin M Curtin, Weihong Li, and Yonglong Zhao. A new model for a carpool matching service. *PloS one*, 10(6):e0129257, 2015.

[15] Luk Knapen, Ansar Yasar, Sungjin Cho, Daniel Keren, Abed Abu Dbai, Tom Bellemans, Davy Janssens, Geert Wets, Assaf Schuster, Izchak Sharfman, et al. Exploiting graph-theoretic tools for matching in carpooling applications. *Journal of Ambient Intelligence and Humanized Computing*, 5:393–407, 2014.

[16] Niels Agatz, Alan L Erera, Martin WP Savelsbergh, and Xing Wang. Dynamic ride-sharing: A simulation study in metro atlanta. *Procedia-Social and Behavioral Sciences*, 17:532–550, 2011.

[17] Shih-Chia Huang, Ming-Kai Jiau, and Yu-Ping Liu. An ant path-oriented carpooling allocation approach to optimize the carpool service problem with time windows. *IEEE Systems Journal*, 13(1):994–1005, 2018.

[18] Jing-Jie Lin, Shih-Chia Huang, and Ming-Kai Jiau. An evolutionary multiobjective carpool algorithm using set-based operator based on simulated binary crossover. *IEEE Transactions on Cybernetics*, 49(9):3432–3442, 2018.

[19] Eva Kesternich. What factors explain carpoolers' decision to use carpooling matching platforms?: A survey-based observation of carpooling matching platforms in europe. Master's thesis, University of Twente, 2015.

[20] Iftikhar Hussain, Davy Janssens, Adel Elomri, and Ben Niu. A negotiation model of individual matching and zonal-based travel behavior in carpooling. *Personal and Ubiquitous Computing*, 27(5):1757–1777, 2023.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

APPENDIX

Table 2. Average Execution Time (in seconds) for Different Dataset Sizes (Appendix)

Number of Employees	Data Collection (s)	Preselection (s)	Matching (s)	Scoring & Ranking (s)	Total Time (s)
100	0.12	1.54	0.72	0.15	2.53
200	0.25	6.25	2.10	0.28	8.88
300	0.37	13.95	4.12	0.42	18.86
400	0.50	25.80	7.35	0.55	34.20
500	0.63	40.30	11.20	0.68	52.81
600	0.77	58.25	16.45	0.82	76.29
700	0.90	79.60	22.90	0.96	104.36
800	1.03	104.35	30.75	1.10	137.23
900	1.16	132.50	39.90	1.24	174.80
1000	1.30	164.00	50.30	1.38	216.98

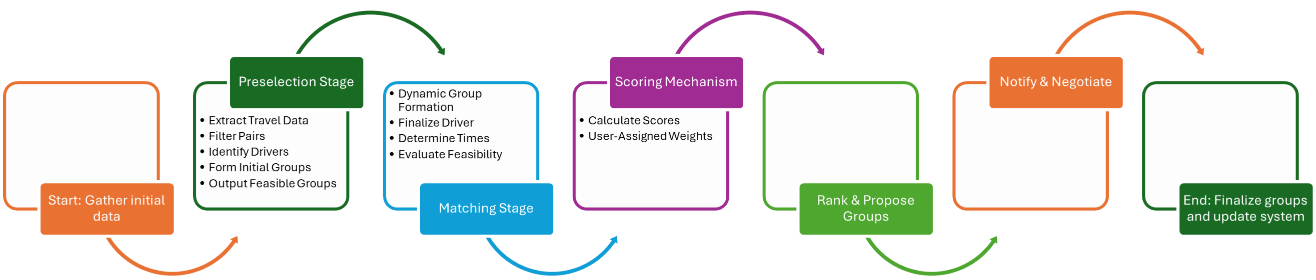

Fig. 7: Flowchart of the Carpool Framework (Appendix)