Final Report

Smart Rideshare Matching – Feasibility of Utilizing Personalized Preferences

Shaguna Gupta[1], Shi Feng[2], and B. Brian Park[2]
[1]Trinity College Dublin, [2]University of Virginia
Email: guptas2@tcd.ie, {sf5jh, bp6v}@virginia.edu

Date
December 2024

Prepared for the Sustainable Mobility and Accessibility Regional Transportation Equity Research Center,
Morgan State University, CBEIS 327, 1700 E. Coldspring Lane, Baltimore, MD 21251

ACKNOWLEDGMENT

This research was supported by the Sustainable Mobility and Accessibility Regional Transportation Equity Research Center at Morgan State University and the University Transportation Center(s) Program of the U.S. Department of Transportation.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The report is funded, partially or entirely, under grant number 69A3552348303 from the U.S. Department of Transportation's University Transportation Centers Program. A version of this report was accepted for presentation and publication at the 11th International Conference on Vehicle Technology and Intelligent Transportation Systems 2025 in Porto, Portugal.

1. Report No. SM13	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Smart Rideshare Matching – Feasibility of Utilizing Personalized Preferences		**5. Report Date** December 2024
		6. Performing Organization Code
7. Author(s) *Include ORCID #* Shaguna Gupta, Shi Feng, and B. Brian Park (0000-0003-4597-6368)		**8. Performing Organization Report No.**
9. Performing Organization Name and Address University of Virginia		**10. Work Unit No.**
		11. Contract or Grant No. 69A3552348303
12. Sponsoring Agency Name and Address US Department of Transportation Office of the Secretary-Research UTC Program, RDT-30 1200 New Jersey Ave., SE Washington, DC 20590		**13. Type of Report and Period Covered** Final, September 2023 -December 2024
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

We investigated the feasibility of utilizing vehicular telematics data for ride-share matching. Our focus was to optimize ride-share matching for users by utilizing personal preferences, such as home and workplace locations, as well as departure and arrival times. A case study was conducted using participants' vehicular telematics data to analyze their commuting patterns between their homes and the University of Virginia (UVA) campus. Using vehicle trip, departure, and arrival data from April 2022, this research analyzed vehicle trips over two weeks to identify individuals with similar commuting schedule preferences. By clustering vehicles based on proximity and timing, we proposed a framework for matching individuals who share similar arrival and departure schedule preferences and live close to each other, thereby facilitating coordinated ride-sharing opportunities. The findings are presented through visualizations illustrating ride-matching potential, particularly during peak commuting hours. Ride-share matching would offer a convenient solution to UVA commuters while maintaining their commuting flexibility. This approach could also offer a more sustainable transportation solution that enhances travel efficiency, lowers environmental impact, and supports the broad adoption of ride-sharing within academic and urban settings. The proposed framework provides a scalable model for systematic ride-sharing implementation and could guide future research and policy development for sustainable campus mobility solutions.

17. Key Words: Rideshare matching, Big data analytics, Vehicular telematics data	**18. Distribution Statement**

19. Security Classif. (of this report) : Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 23	**22. Price**

Abstract

We investigated the feasibility of utilizing vehicular telematics data for ride-share matching. Our focus was to optimize ride-share matching for users by utilizing personal preferences, such as home and workplace locations, as well as departure and arrival times. A case study was conducted using participants' vehicular telematics data to analyze their commuting patterns between their homes and the University of Virginia (UVA) campus. Using vehicle trip, departure, and arrival data from April 2022, this research analyzed vehicle trips over two weeks to identify individuals with similar commuting schedule preferences. By clustering vehicles based on proximity and timing, we proposed a framework for matching individuals who share similar arrival and departure schedule preferences and live close to each other, thereby facilitating coordinated ride-sharing opportunities. The findings are presented through visualizations illustrating ride-matching potential, particularly during peak commuting hours. Ride-share matching would offer a convenient solution to UVA commuters while maintaining their commuting flexibility. This approach could also offer a more sustainable transportation solution that enhances travel efficiency, lowers environmental impact, and supports the broad adoption of ride-sharing within academic and urban settings. The proposed framework provides a scalable model for systematic ride-sharing implementation and could guide future research and policy development for sustainable campus mobility solutions.

Keywords: Rideshare matching, Big data analytics, Vehicular telematics data

1 INTRODUCTION

Current transport-share systems or carpooling typically rely on users to actively request or offer a ride to passengers and to coordinate the time and pickup location. Services such as Lyft and Uber have addressed this problem by using location data to provide ride services that are convenient and on-demand. These aspects of ride-share services might also explain the widespread use of personal cars, as they allow drivers to combine commutes with other activities (e.g., picking up kids to and from school, running errands, going to off-campus meetings, etc.). This convenience, however, comes at a great personal and societal cost, including increased traffic congestion and parking demand. Despite various agencies' incentives and discounts for ride-sharing, this kind of service has not been widely used in daily commuting for the reasons mentioned above as well as coordination challenges, scheduling requirements, commitment, and having to actively request or offer rides.

Ride-sharing can be an efficient and sustainable mode of transport for work colleagues or people who live in close proximity, especially in areas with limited public transport, walking, or cycling options. Ride-sharing or car-pooling can be beneficial for organizations and individuals in terms of reduced travel costs, improved parking efficiency, decreased traffic congestion, and positive environmental impacts.

In this research, we attempted to increase engagement with ride-sharing in a university community by analyzing vehicular telematics data made available by Wejo, a private connected vehicle company based in the United Kingdom. The analyses focused on evaluating potential ride-share matches based commuters' departure and arrival data to and from the University of Virginia (UVA). UVA is based in Charlottesville and faces challenges in managing traffic flow due to its limited parking spots. Because of the ever-increasing parking demands on the campus, there is an opportunity to optimize traffic flow using ride-sharing initiatives.

This research presents an analysis of vehicle trips to and from UVA over a two-week period in April 2022, with the goal of identifying potential ride-sharing opportunities based on personalized commuting arrival/departure time preferences. The key objectives of this analysis are to recommend ride-sharing matches based on timing and proximity. By identifying individuals who arrive at UVA at similar times and leave UVA at similar departure times, a matching framework can coordinate ride-sharing opportunities that align with their schedules. Furthermore, the study suggests ride-sharing options for individuals commuting to and from nearby locations relative to each other to enhance convenience. These criteria are used to group vehicles based on their arrival and departure locations, both at UVA and at their home locations before and after UVA trips.

This research details the methodology used to identify these ride-sharing opportunities, evaluates the results obtained from the analysis, and presents visualizations to support the findings. The insights gained from this study could be instrumental in developing a systematic ride-sharing program that enhances the commuting experience for UVA students, faculty, and staff. We expect that the approach used in this research is applicable to many similarly situated communities and institutions where optimizing transportation systems can significantly lower environmental impact and improve overall travel efficiency. Tailored to different settings, this framework can serve as a versatile solution for fostering sustainable and convenient commuting practices in urban and academic environments alike. The remainder of this paper is organized as follows. Section 2 summarizes the status of the existing literature related to car sharing, followed by section 3 with data and approaches used to analyze the vehicular data used in this study. Section 4 presents feasibility evaluation and results, followed by section 5 – discussion, section 6 – conclusions, and section 7 – future work.

2 LITERATURE REVIEW

Interest in analyzing ride-sharing as a solution to traffic congestion, parking issues, and saving energy in cities, companies, and college campuses has led to extensive research in optimizing ride-sharing models and road networks. Our research has reviewed relevant contemporary research papers on carpooling models, carpooling preferences, and ways to improve ride-sharing systems.

Many researchers have proposed models and algorithms aimed at enhancing car-sharing systems to benefit both users and car-sharing companies. Focusing on the user perspective, Narman et al. presented a model that employs a two-layer matching system [Narman et al., 2021], and Hussain et al. proposed a system specifically designed to optimize car sharing frameworks for employees in large organizations [Hussain et al., 2022], while Masoud and Jayakrishnan introduced a real-time algorithm to address the ride-matching problem within a flexible ride-sharing system [Masoud and Jayakrishnan, 2017]. In the two-layer model that Narman et al. developed, the first layer matches riders based on personal characteristics, such as safety, punctuality, and comfort. The second layer limits waiting times by allowing riders and drivers to set a personalized threshold. Using a machine learning-based recommendation system, the model achieved a 90 percent accuracy rate in predicting rider preferences, leading to an increase in successful matches and completed trips. Considering car sharing in large companies, Hussain et al. developed a framework that considers factors like home location, target destination, time windows, and personal behavior to optimize carpooling groups. The system updates schedules in real-time, allowing for flexible carpooling solutions. The study shows that the proposed framework can efficiently generate optimal carpooling solutions in real-time, emphasizing its potential in managing recurrent travel demand among managed groups, such as company employees. The flexible system that Masoud et al. proposed allows for dynamic, real-time matching and multi-hop rides, considering users' preferences and minimizing waiting times. Their algorithm is also able to solve large-scale ride-matching problems quickly, providing comfort to riders through optimal routing and reducing the number of transfers. These studies developed innovative models and used algorithmic approaches to enhance user experiences in car-sharing systems, ultimately fostering greater participation and computational efficiency in car sharing systems.

Some researchers have also concentrated on optimizing road networks, particularly in relation to road congestion and capacity. De Palma et al. focused on the impact of dynamic congestion on carpool matching in their paper [de Palma et al., 2022]. The study incorporated scheduling preferences and dynamic bottleneck congestion in its ride- sharing framework. The theoretical results suggested that, if the only inconvenience is a potential detour, the optimal matching occurs when drivers and passengers are sequenced based on their location. In addition, the authors also discovered that when people have different desired arrival times, matching becomes more complicated, and penalties for tardiness may arise. Xingyuan Li et al. presented a ride sharing trip-assignment model using a bi-level programming approach for optimizing the reserve capacity of road networks. The programming was then formatted into a single-layer optimization problem [Li et al., 2024]. They concluded that subsidizing ride-sharing drivers can significantly improve the reserve capacity of road networks, almost equating to the effects of expanding road capacity without ride-sharing. These studies integrate dynamic congestion and user preferences when optimizing road networks, demonstrating that effective ride-sharing strategies can significantly enhance road capacity while mitigating congestion.

Other researchers have examined the social aspects of carpooling. Limited participation in carpooling can be attributed to specific strategies employed by ride-hailing companies and concerns expressed by passengers. Naumov and Keith focused on the economic and environmental impacts of ride-hailing [Naumov and Keith, 2023]. The study finds that most ride-hailing trips are unprofitable without subsidies, especially for pooled rides. The study suggested that ride-hailing platforms can improve revenue and reduce environmental impact by adjusting pricing strategies. Specifically, increasing the price difference between individual and pooled rides could decrease total vehicle miles traveled, benefiting ride-hailing companies and urban environment. Linchao Li et al. explored what influences college students' views on carpooling by implementing a multinomial logit model based on a survey-based data [Li et al., 2023]. The study concluded that individual attributes like sex and age do not significantly affect attitudes. However, travel habits, such as whether students use public transportation or have previous carpooling experience, could have a crucial impact on students' views on carpooling. Concerns about safety and cost are key reasons why carpooling is not more popular among students. Safety and cost are primary concerns, a reliable carpooling information platform specifically for students could therefore increase support for carpooling. These studies highlight the social challenges facing carpooling, including economic and environmental impacts, passenger concerns, and social attitudes.

These studies provided various insights from diverse perspectives on how ride-sharing systems can be modeled and optimized, as well as the factors influencing users' preferences.

Despite the valuable findings provided by the reviewed studies, there exists a notable gap in the research: few studies have utilized real-world vehicle trajectory data that specifically capture drivers' departure and arrival time preferences in their daily commuting trips. Most existing ride-sharing models rely on simulations or survey data, which may not fully reflect the dynamic nature of actual commuting patterns. In addition, the integration of vehicle telematics data has been underexplored in the context of optimizing ride-sharing on university campuses. To address this gap, our study proposes an innovative approach that leverages real-world vehicular telematics data to analyze commuting patterns at the University of Virginia (UVA), and the potential for matching based on their arrival and departure times at UVA, which generates private route options for users. This research allows for more accurate identification of potential ride-sharing opportunities based on the timing and availability of potential vehicle trips on campus, which aims to reduce traffic, improve efficiency, and create a more sustainable transportation system in the city of Charlottesville.

3 MATERIALS AND METHODS

As noted, our research focuses on identifying patterns where individual commuters arriving at and departing from the university campus at similar times or from similar locations could benefit from shared transportation. Leveraging trip data, including Trip IDs, Vehicle IDs, ignition status, and geographical coordinates, we can cluster vehicles based on their arrival and departure times and locations. This allows us to propose ride-sharing opportunities that could reduce traffic congestion, lower transportation costs, and contribute to a more sustainable campus environment.

3.1 Vehicular Telematics Data

This research used vehicular telematics data made available to the research team by the Virginia Department of Transportation (VDOT). VDOT purchased the data from the Wejo company. Out of 26 weeks of data, this research utilized vehicle trips from ten working days in April 2022, focusing on Trip IDs, Vehicle IDs, ignition status, and geographical coordinates (latitude and longitude) to identify potential ride-sharing clusters. These particular weekdays were selected because other months were affected by COVID-19 (some still teleworking during Fall 2021) and because of irregularities caused by typical University academic activities (e.g., Spring recess in March, graduation in May). The data comprises the following key attributes as shown in Table 1.

Table 1: Data Format and Column Descriptions

Column Name	**Column Description**
dataPointId	Unique ID for each data point.
journeyId	ID for vehicle's movements to ignition
off. capturedTimestamp	Timestamp for data point (1s precision).
locationl_atitude	Latitude to 6 decimal places.
location_longitude	Longitude to 6 decimal places.
location_geohash	Geohash for Earth's surface square of data point.
location_postalCpde	Zip or postal code of data point.
location_regionCode	Region/state code of data point.
location_countryCode	Country code of data point.
metrics_speed	Vehicle speed in km/h at data point.
metrics_heading	Heading from 0 (North) to 359.
vehicle_wejoVehicleTypeId	ID for vehicle's specific characteristics.
vehicle_status_ignitionStatus	Ignition status.
dt	Date from captured timestamp.
hr	Hour from captured timestamp.
tripID	Unique int ID for journey based on start time.
edge index	Edge index in matched points list (if matched).
way_id	OpenStreetMap way ID.
ulid	ULID for individual journeyId.

3.2 VA Boundary

To accurately determine which trips were associated with UVA, a boundary polygon was created to represent the UVA campus, including its parking lots [University of Virginia, 2023]. This polygon was used to filter trips that either started or ended within the UVA boundary. Only trips that had their ignition turned off (KEY_OFF) within this boundary were considered as arriving at UVA, while trips that had their ignition turned on (KEY_ON) within the boundary were considered as departures.

Figure 1 shows the UVA campus boundary on OpenStreetMap [OpenStreetMap contributors, 2017] used to filter relevant trips and this is being utilized to understand the arrivals and departures in the subsequent sections.

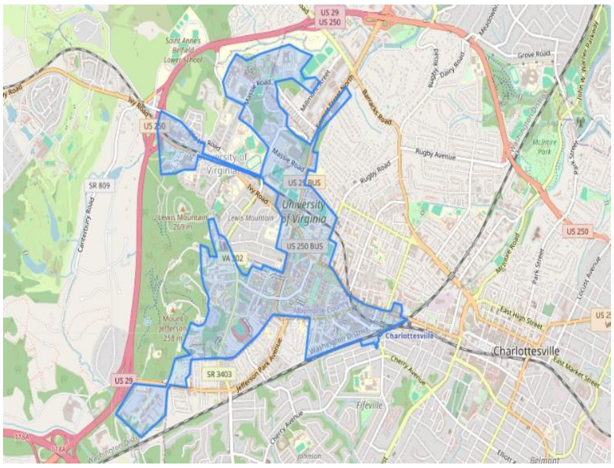

Figure 1: UVA Polygon Boundary

3.3 Vehicle Operational Schedule Understanding

The initial phase in the methodology involved loading and transforming the data. Timestamps recorded in Coordinated Universal Time (UTC) were converted to Eastern Time (EDT) to align with the local time zone at UVA. The data was then filtered to include only relevant trips, specifically those that occurred between 5 AM and 8 PM on weekdays, and within a predefined set of valid postal codes within the 25-mile radius of the UVA area.

3.3.1 Arrival and Departure Schedule

Based on the UVA polygon boundary, the filtered data was further processed to identify arrivals and departures at UVA. The study specifically analyzed trips that ended or started within the UVA polygon area, considering these as the first trip of the day to UVA and the last trip of the day from UVA consecutively. Based on this, we finalized the trips that arrived or departed.

- Arrivals Extraction Algorithm

  This algorithm 1 is designed to find the first trip each vehicle makes to UVA on any given day. A trip is classified as an arrival at UVA when the vehicle's ignition is turned off, represented by the KEY OFF ignition status. The algorithm processes data related to vehicle trips and extracts the earliest instance of a vehicle arriving at UVA on each day within the dataset. This approach filters the dataset to include only trips where the vehicle's ignition was turned off (i.e., arrival events) and then sorts the data by vehicle and timestamp.

Algorithm 1 UVA First Trip Arrival Extraction Algorithm

```
Input: DataFrame df, Polygon uva_polygon
Output: DataFrame first_rows_of_first_trip
Convert capturedTimestamp_est to trip date (date only).
Filter rows where vehicle_status_ignitionStatus is KEY OFF (vehicle arrived).
Sort DataFrame by vehicle wejoVehicleTypeId and capturedTimestamp_est.
Group by vehicle_wejoVehicleTypeId and trip date to get the first tripID of the
    day.
Join the dataset with the first tripID to get the full details of the first trip.
Convert the DataFrame to Pandas format for spatial filtering.
Convert back to Polars DataFrame.
Group by vehicle wejoVehicleTypeId and trip date to keep only the first rows of
    the first trip.
Drop unnecessary columns.
Return the cleaned DataFrame.
```

- Departures Extraction Algorithm
  This algorithm 2 identifies the last trip for each vehicle that departs from UVA on a given day. The trip is considered a departure if the vehicle's ignition is turned on, represented by the KEY ON ignition status. The algorithm focuses on identifying the last trip that departs from UVA for each vehicle on that day. This is designed to primarily extract the first row of the last recorded trip for vehicle departures from a defined location based on its timestamp.

Algorithm 2 UVA Last Trip Departure Extraction Algorithm

```
Input: DataFrame df, Polygon uva_polygon
Output: DataFrame containing first row of the last trip for each vehicle
Create a new column trip_date by extracting the date from capturedTimestamp_est.
Filter df where vehicle_status_ignitionStatus is KEY_ON.
Sort df by vehicle wejoVehicleTypeId and capturedTimestamp_est.
Group df by vehicle_wejoVehicleTypeId and trip_date.
    For each group, select the last tripID.
Join the original df with last_trip_ids_df to get all rows of the last trip.
Convert the resulting DataFrame to Pandas for spatial filtering.
Perform any spatial filter using uva_polygon (optional step).
Convert the filtered DataFrame back to Polars.
For each vehicle and day, extract the first row of the last trip.
Drop unnecessary columns.
Return the resulting DataFrame.
```

Both algorithms identify the first trip and last trip of each vehicle arriving at and departing from UVA on a given day. The key steps involve filtering for relevant trips (ignition off/on), sorting, grouping by vehicle and date, extracting the first trip, and cleaning the dataset to remove unnecessary information. The algorithms also utilize Pandas and Polars for data manipulation and transformation.

3.3.2 Start and End Trip Schedule

- Start and End Point Extraction Algorithm

  This code processes trip data of vehicles by identifying start and end points based on key events (key-on and key-off) before and after vehicle arrivals and departures at UVA as mentioned in algorithm 3. The process focuses on merging, filtering, and ranking key-on and key-off events to capture possible home locations by calculating the distances between start and end trip locations.

Algorithm 3 Compute Start and End Trip Points and Distances

1: **Input:** Arrival data, Departure data, Key-on event data, Key-off event data, Geographical boundary for UVA parking lot

2: **Output:** Excel file with start and end trip points, arrival and departure times, and calculated distances.

3: Load Data

Load arrival, departure, key-on, and key-off data from Parquet files into Pandas DataFrames.

Convert *capturedTimestamp_est* to timezone-naive datetime format. Extract *date* from *capturedTimestamp_est*.

4: Preprocess Data

Select columns: *vehicle_wejoVehicleTypeId*, *capturedTimestamp_est*, *date*, *location latitude*, *location longitude*.

Remove duplicate records to retain unique trip events.

5: Merge Arrival Data with Key-On Events

Merge *arrival_df* with key-on events (*all_trips_df_key_on*) on *vehicle_wejoVehicleTypeId* and *date*.

Retain the first three key-on events before the arrival for each vehicle and date.

6: Merge Departure Data with Key-Off Events

Merge *departure df* with key-off events (*all trips df key off*) on *vehicle wejoVehicleTypeId* and *date*.

Retain the first three key-off events after the departure for each vehicle and date.

7: Spatial Filtering

Apply spatial filtering to remove trip points that fall within the UVA parking lot boundary.

8: Calculate Distances

Use the Haversine formula to calculate:

- Distance between start and end trip points.
- Distance between arrival and departure points.

9: Final Data Preparation

10: Merge cleaned arrivals and departures into a final DataFrame with vehicle ID, dates, trip times, and distances.

11: Save to Excel

Save the final DataFrames as Excel files: Example (Master_trip_3LL.xlsx).

3.4 Clustering Approach

The filtered data was further processed to identify potential ride-sharing opportunities. The analysis involved visualizing and clustering vehicles based on spatial and temporal proximity at various stages of their trips [Jain and Dubes, 1988]. Three distinct clusters were defined to gain insights into the ride-sharing potential. The logic for this clustering involved grouping vehicles based on the following criteria:

- Arrival Points: Vehicles arriving at UVA over the duration of 10 days composes our arrival points. This will lead to the potential for detecting vehicles that arrive around the same time. The arrivals spread across UVA boundary over 10 days can be visualized as following. The visualizations in Figure 2 and Figure 3 illustrate the density of the vehicles around UVA parking lots.

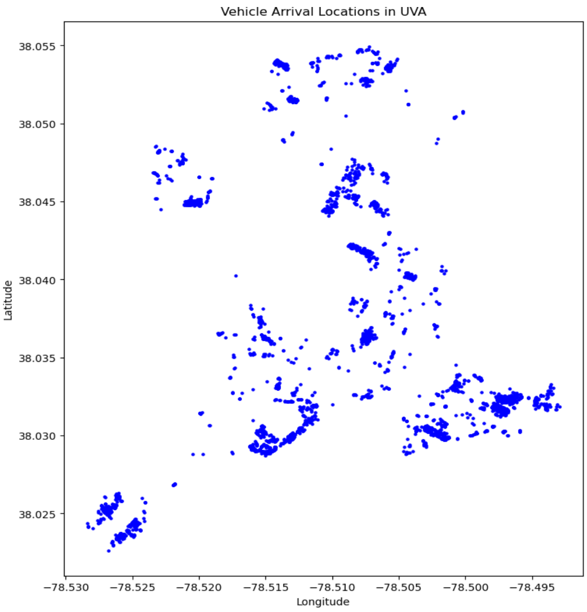

Figure 2: Arrivals spread at UVA boundary

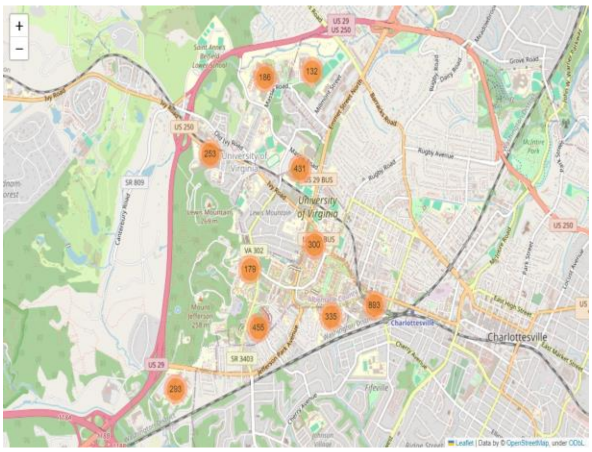

Figure 3: Arrivals at UVA

Departure Points: The location of vehicles leaving UVA over the duration of 10 days captures the departure points. Again, departure points will help identify the vehicles departing around the same time, offering a chance to cluster and optimize ride-sharing for return trips. The departures spread across UVA boundary can be visualized in Figure 4 and Figure 5 as follows:

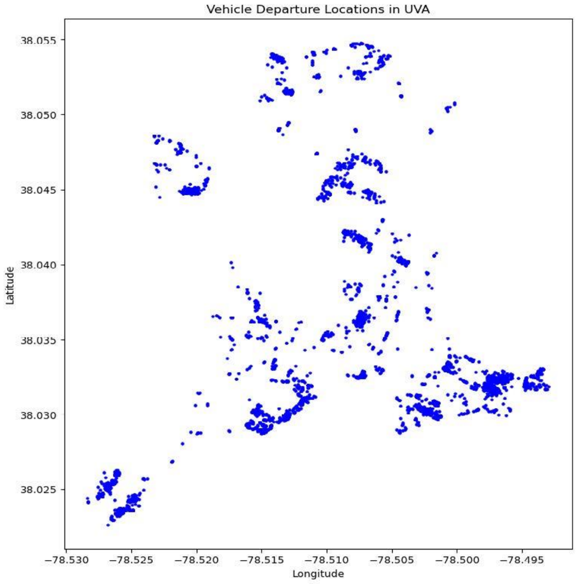

Figure 4: Departures spread at UVA boundary

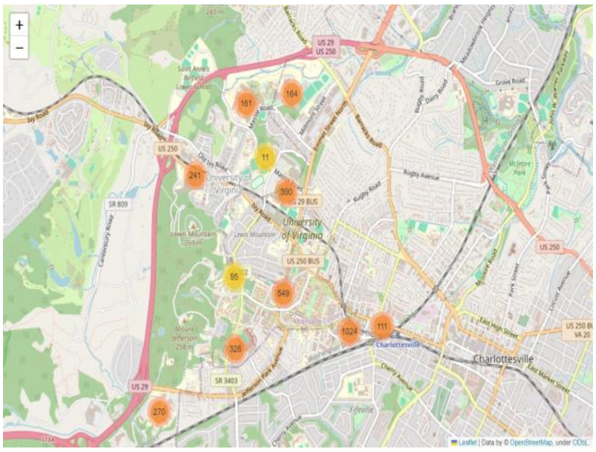

Figure 5: Departures from UVA

- Home-to-UVA Points: These points represent our Home-to-UVA start locations, highlighting where vehicles begin their trips. This information is crucial for identifying potential ride-sharing clusters by analyzing the spatial proximity of vehicles departing for UVA.

- UVA-to-Home Points: Vehicles departing from UVA to return to home locations will represent the UVA-to-Home end locations. This will allows for the identification of vehicles returning to similar home locations at similar times, which enables the coordination of ride-shares for return trips.

3.4.1 Defining Clusters

1. Arrival Time Clustering: Vehicles arriving at UVA within a similar time window (e.g., 15–30-minute intervals) were grouped. This temporal clustering helps identify vehicles that could potentially share rides to reduce the number of trips to UVA.

2. Departure Time Clustering: Similarly, vehicles departing from UVA within the same time windows were grouped to facilitate ride-sharing for the return journey. This clustering helps identify vehicles that leave UVA at similar times, enabling ride-sharing for the trip home.

3. Home Location Clustering: This clustering method identified start and final locations of vehicles. Vehicles that started their trips from nearby locations (e.g., within 2 miles of each other) could be identified as potential ride-sharing candidates. This was done for both the trip to UVA and the return trip home. Therefore, vehicles that begin their trips from nearby locations heading toward UVA or leaving UVA for nearby locations were grouped together, suggesting ride-sharing routes for both legs of the journey.

This analysis allowed us to group vehicles that start and end their trips from and to nearby locations respectively within a predefined radius (e.g., 2 miles) or using a clustering algorithm such as K-means and arrive at UVA and depart from UVA within the same time window [Li and Chung, 2020].

4 EVALUATION AND RESULTS

The evaluation of the clustering process was conducted by analyzing the number of vehicles that were successfully grouped based on the defined criteria. The results indicate several significant clusters of vehicles that could benefit from ride-sharing, especially during peak arrival and departure times. The analysis identified several clusters of vehicles that could potentially benefit from ride-sharing.

4.1 Evaluation Metric 1: Number of vehicles in each day within the 10-day bracket

The provided graphs show the daily number of unique vehicle arrivals and departures at UVA over a 10-day period in April 2022. In Figure 6, vehicle arrivals ranged from a low of 297 on April 15 to a high of 372 on April 22, with some fluctuation observed throughout the period. In Figure 7, vehicle departures ranged from a minimum of 285 on April 15 to a maximum of 365 on April 13, indicating a similar level of day-to-day variability. Overall, both arrivals and departures exhibit peaks and troughs, but both the distribution of the unique vehicle arrivals and departures follow a similar pattern in the researched period.

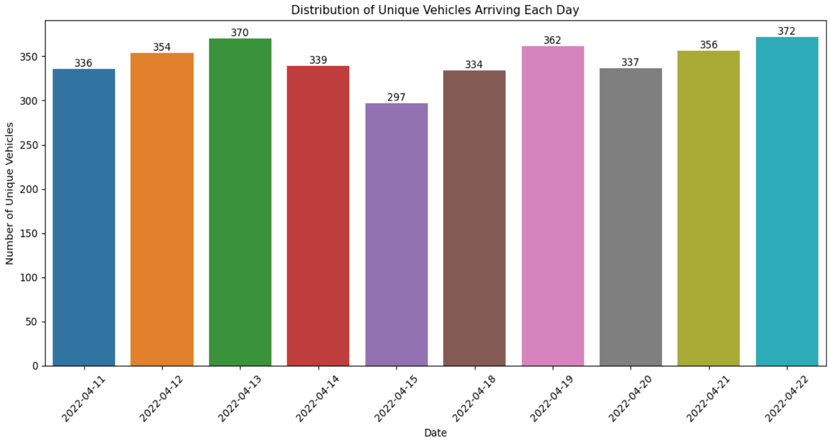

Figure 6: UVA Unique Arrivals Each Day

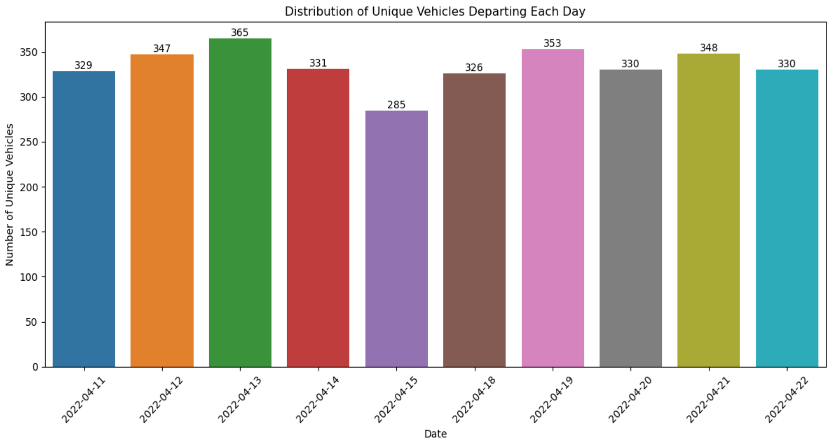

Figure 7: UVA Unique Departures Each Day

4.2 Evaluation Metric 2: Number of vehicles in each hour for 10 days

The following figures illustrate the hourly distribution of vehicle arrivals and departures at UVA over the researched 10-day period.

In Figure 8, arrivals peak during the morning hours, particularly between 7 AM and 9 AM, with the highest number of unique vehicles observed around 9 AM. After this peak, the number of arrivals gradually decreases throughout the day until the evening. Figure 9 shows departures, which exhibit a different pattern with peaks occurring in the late afternoon, particularly between 3 PM and 5 PM, with a peak at around 4 PM. Departures steadily increase from late morning until reaching their maximum, after which they decline towards the evening hours. The data indicates that arrivals are concentrated in the morning while departures peak in the late afternoon, suggesting typical commuting behavior.

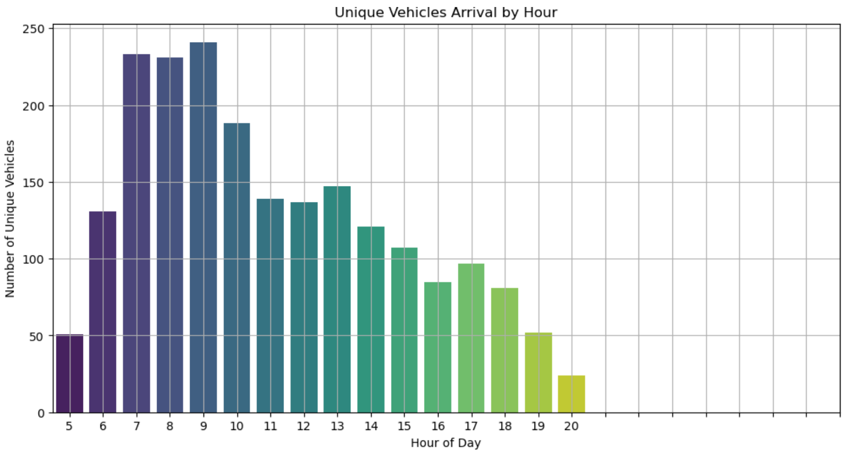

Figure 8: UVA Unique Arrivals per hour for 10 Days

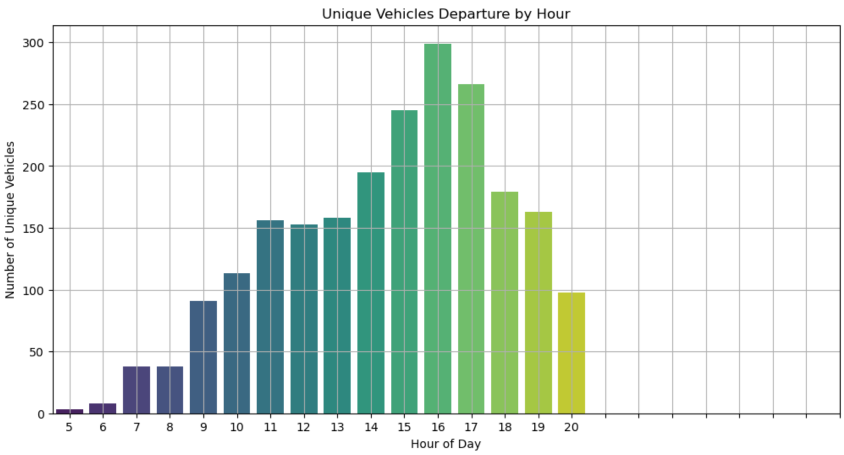

Figure 9: UVA Unique Departures per hour for 10 Days

4.3 Evaluation Metric 3: Number of vehicles in every 30-minute window for 10 days

The following figures contain the number of unique vehicle IDs arriving and departing at UVA in 30-minute intervals across the 10-day research period.

In Figure 10, the arrivals show a clear peak during the morning hours, especially between 7:30 AM and 9:30 AM, with the highest number reaching 164 around 7:30 AM. After this peak, the number of arrivals steadily declines throughout the day, indicating that most arrivals happen in the early part of the day. In Figure 11, the departures reveal a different trend. Departures increase gradually throughout the day, peaking between 4:00 PM and 5:30 PM, with the highest number reaching 182 unique vehicle IDs at 4:30 PM. After this peak, the number of departures decreases toward the evening hours. This pattern suggests a typical commuting behavior where vehicles arrive in the morning and depart in the late afternoon.

Arrival and Departure Clusters at UVA: Significant clustering was observed for vehicles arriving and departing within 30-minute windows, indicating a potential for coordinated ride-sharing as shown in the figures.

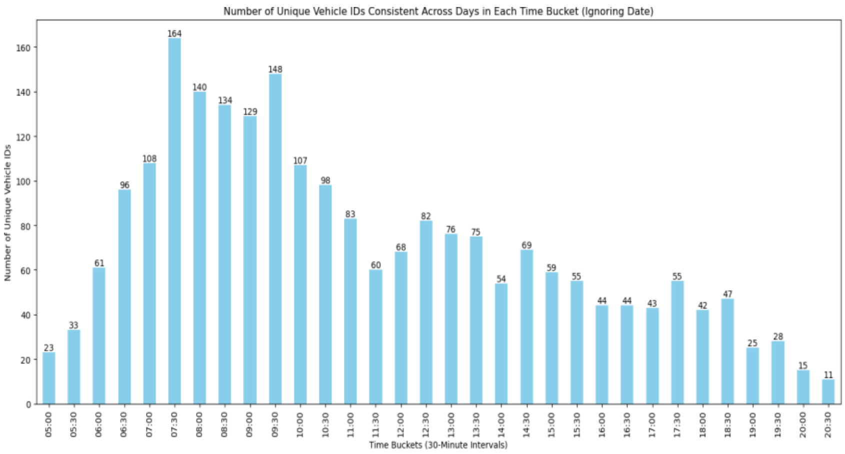

Figure 10: UVA Unique Arrivals for 10 Days

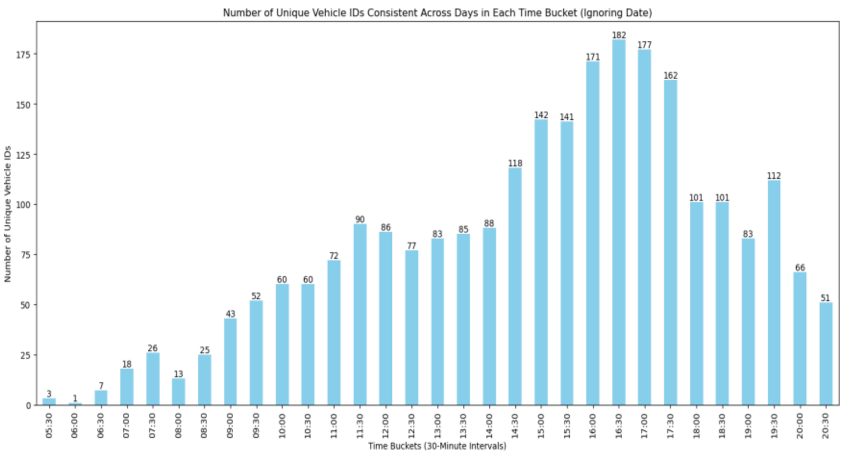

Figure 11: UVA Unique Departures for 10 Days

4.4 Evaluation Metric 4: Vehicles distribution in total 10 days

Table 2 shows the distribution of vehicle counts for the different number of days starting from 10 days to 9 days and so on, with each number of day's absolute count and corresponding percentage of the total. Row with number of vehicles coming only 1 day out of a total 10 days has the highest count at 375, making up 38.38% of the total. This may be due to an influx of drivers traveling to the hospital or visiting the university for other purposes. As the total number of days increases, their respective counts and percentages generally decrease, with 10 total days showing the lowest count at 76, which accounts for 7.78% of the total percentage of vehicles. This results in 977 total unique vehicle entries to UVA over 10 days. A similar pattern is seen for vehicles exiting UVA over the same time period.

Table 2: Distribution of Arrival Counts by Vehicle ID and Days

Total Days	**Count**	**Percentage**
10	76	7.78%
9	46	4.71%
8	35	3.58%
7	36	3.68%
6	46	4.71%
5	46	4.71%
4	72	7.37%
3	92	9.42%
2	153	15.66%
1	375	38.38%
Grand Total	**977**	**100.00%**

4.5 Evaluation Metric 5: Vehicle distribution for proximity analysis

4.5.1 Elbow Plot

Figure 12 illustrates the process of determining the optimal number of clusters for a dataset using the Elbow Method. The x-axis represents the number of clusters, while the y-axis represents the distortion, which is typically the sum of squared distances from each point to its assigned cluster center. As the number of clusters increases, the distortion decreases because the points are divided into more clusters, thus reducing the average distance to the cluster centers.

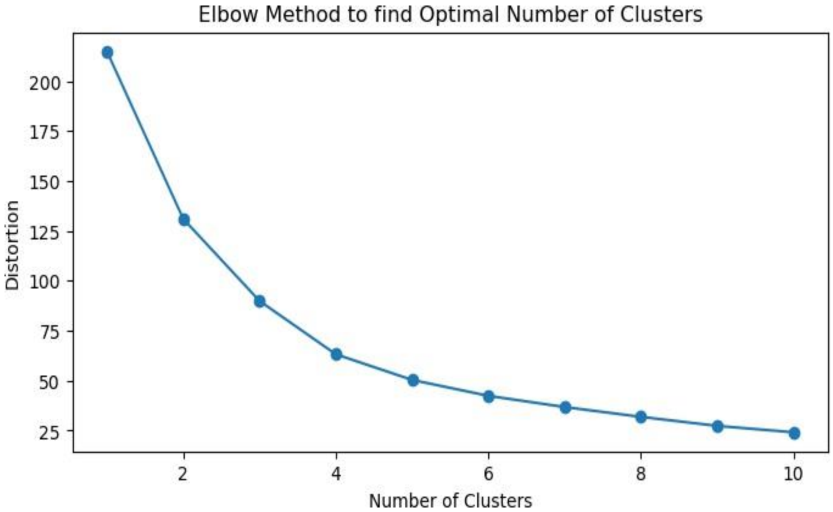

Figure 12: Elbow Plot with 10 clusters

Initially, when the number of clusters is small, adding more clusters significantly reduces the distortion, as seen by the steep decline from 1 to 4 clusters. However, after around 4 clusters, the reduction in distortion becomes less pronounced, as seen by the flattening of the curve. The point where this flattening occurs is considered the "elbow," which is typically regarded as the optimal number of clusters. In this case, the elbow appears to be around 4 clusters, suggesting that this is a good choice for the number of clusters in the data representing vehicle start locations. Beyond this point, increasing the number of clusters can offer diminishing returns in terms of reducing distortion. In the following sections, clusters are visualized for different sets such as a set of 4 clusters and a set of 5 clusters.

4.5.2 Clustering Results Visualizations

The following cluster characteristics are used to optimally divide the start locations into zones.

1. Cluster Centroids: Each cluster will have a central point (centroid) representing the mean of all latitude and longitude values within the cluster. The centroids allow us to understand the general geographic location of each cluster.

2. Cluster Spread: We calculate the minimum, maximum, and mean latitude and longitude values for each cluster, which gives us an idea of how wide or narrow each cluster is geographically.

3. Cluster Density: We count the number of start locations within each cluster and compare that to the geographic spread to determine how densely packed the points are. Higher densities indicate tightly concentrated clusters, while lower densities show a more dispersed cluster.

The tables and summary provided below contain the results of clustering the trip start locations using two sets of cluster configurations: one with 4 clusters as shown in Figure 13 and the other with 5 clusters as shown in Figure 14. The silhouette scores and statistical summaries for each clustering set are reported.

- **First Set: 4 Clusters**

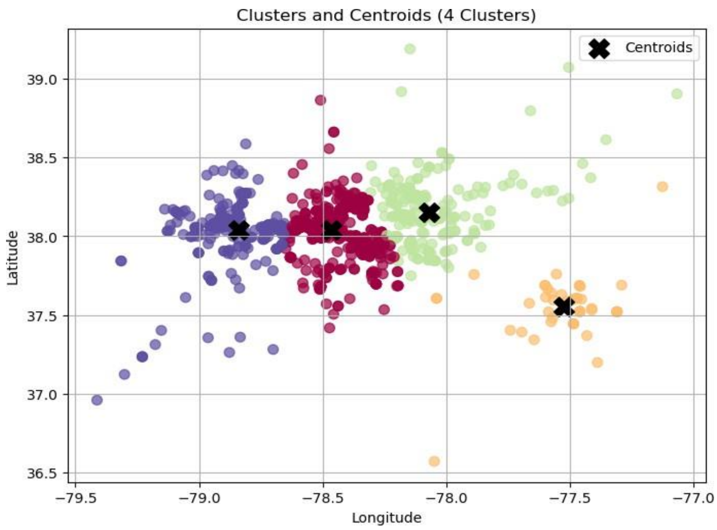

Figure 13: Start Locations Clusters with Centroid

- **Second Set: 5 Clusters**

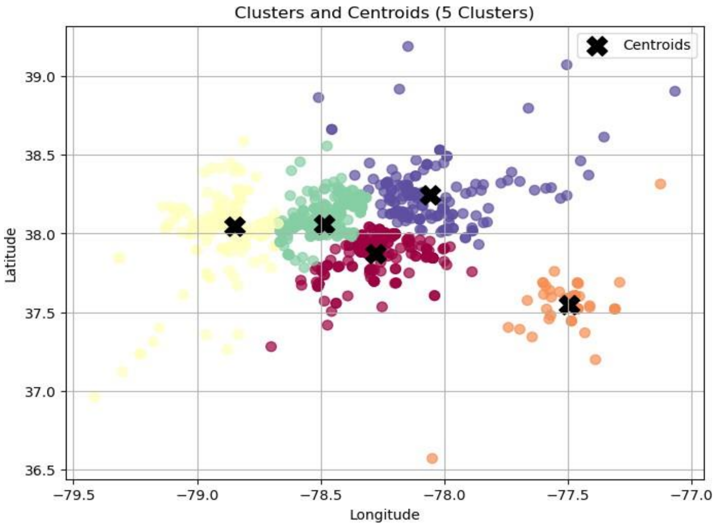

Figure 14: Start Locations Clusters with Centroid

4.5.3 Interpreting Densities:

The density of a cluster in Table 3 represents how closely packed the data points are within each cluster. Higher density indicates that the points are closer to each other, whereas lower density indicates more spread-out points.

Table 3: Comparison of Densities (points per unit area) of 4 clusters vs 5 clusters

Cluster	Densities_4_clusters	Densities_5_clusters
0:	3569.27	806.88
1:	63.31	106.57
2:	237.77	533.83
3:	501.96	6896.55
4:	-	167.27

Density Distributions: In the 4-cluster set, cluster 0 has the highest density (3569.27), while cluster 1 has the lowest (63.31). This means that the data points are tightly packed in cluster 0 and more spread in cluster 1. In the 5-cluster set, cluster 3 has a very high density (6896.56), while cluster 1 has the lowest (106.57). This suggest that after splitting into 5 clusters, more data points are tightly grouped.

4.5.4 Interpreting Summary:

- **Silhouette Score for Cluster sets 1 and 2:** The silhouette score is a measure used to evaluate the quality of clustering by calculating how similar an object is to its own cluster (cohesion) compared to other clusters (separation). It ranges from -1 to +1, with a score closer to 1 indicating well-defined clusters and values closer to -1 suggesting that samples may have been assigned to the wrong clusters.

  In this case, the Silhouette Score for the first set of clusters, which consists of 4 clusters, is 0.5578. This indicates relatively good separation between the clusters but suggests there might be room for improvement. When increasing the number of clusters to 5, the Silhouette Score improves slightly to 0.5630, indicating a better fit. The higher score in the second set with 5 clusters shows that the clustering structure is more well-defined compared to the first set. However, since the improvement is marginal, it suggests that adding more clusters may only slightly enhance the overall clustering quality in this case. Therefore, while the second set with 5 clusters performs better, the difference between the two scores is small, indicating that both 4 and 5 clusters are reasonable options.

- **Cluster Spread Summary:** The Cluster Spread Summary provides an overview of the statistics for the trip starting points (latitude and longitude) across different clusters. Two sets of clusters are presented: Set 1, with four clusters provided in Table 4, and Set 2, with five clusters provided in Table 5. For each cluster, the mean, minimum, and maximum values for the starting latitude and longitude are provided, along with the count of data points in each cluster.

  Set 1 features four clusters with the following characteristics:

  Cluster 0 has the highest count (1369) and represents trips starting with a latitude range from 37.42 to 38.87 and a longitude range from -78.65 to -78.19. Cluster 1 has a much smaller count (49) with latitudes ranging from 36.57 to 38.32 and longitudes from -78.05 to -77.12, showing a more compact range. Cluster 2 and Cluster 3 also show distinct geographical spreads with varying counts, 254 and 383, respectively. Set 2 increases the number of clusters to five, providing a more detailed view:

  Cluster 0 is reduced in size compared to the previous set, with a count of 245 and latitudes from 37.28 to 38.04. Cluster 3 becomes the dominant cluster with 1212 trips, indicating a higher density of trips starting in the given range of latitude and longitude. The additional Cluster 4 introduces a new cluster with a mean latitude of 38.24 and longitude of -78.06, accounting for 189 trips. In both sets, the clusters cover a wide geographical range, and the differences in mean, minimum, and maximum values reflect varying concentrations of trips across the region. Set 2, with its five clusters, provides a more granular view, offering insights into smaller, more specific regions of the trip starting points.

Table 4: Trip Start Latitude and Longitude Statistics by Cluster for Set 1

Cluster	Trip start latitude			Trip start longitude			Count
	Mean	Min	Max	Mean	Min	Max	Count
0	38.037342	37.416632	38.865326	-78.466910	-78.649063	-78.194489	1369
1	37.551325	36.568424	38.315376	-77.523632	-78.048103	-77.124741	49
2	38.152525	37.749020	39.190842	-78.071578	-78.306068	-77.066010	254
3	38.041136	36.958107	38.587189	-78.840288	-79.412605	-78.655846	383

Table 5: Trip Start Latitude and Longitude Statistics by Cluster for Set 2

Cluster	Trip start latitude			Trip start longitude			Count
	Mean	Min	Max	Mean	Min	Max	Count
0	37.871467	37.280174	38.042206	-78.279540	-78.699982	-77.887306	245
1	37.544520	36.568424	38.315376	-77.493320	-78.048103	-77.124741	46
2	38.044617	36.958107	38.587189	-78.849947	-79.412605	-78.671181	363
3	38.060848	37.782543	38.557156	-78.487998	-78.669220	-78.316055	1212
4	38.244091	37.930283	39.190842	-78.056641	-78.508377	-77.066010	189

5 DISCUSSION

While each day has around 300 vehicles arriving to and departing from the UVA campus, it is noted that UVA currently employs about 16,000 faculty and staff in addition to 7,000 UVA health employees. Given the Wejo data is only available from newer vehicles made after 2015 by selected automakers, the study only matched a small number of vehicles. As this study focuses on the feasibility of ride-sharing using vehicular telematics data, it is feasible to show ride-sharing can be matched from these 300 vehicles. It is possible that 300 vehicles matched daily could have been over 20,000 vehicles, assuming around 90% of around 23,000 employees commute using their personal vehicle. Actual matching would be accomplished via new sources of data, including a ride-sharing App that UVA plans to adopt in the near future. The visualizations included in the report also illustrate the clustering results, showing where and when ride-sharing could be most effective. These results provide a foundation for developing a systematic ride-sharing program that could be implemented at UVA.

6 CONCLUSIONS

This research highlighted the potential to match commuters for ride-sharing by leveraging vehicular telematics data. Using the case study of the UVA campus, we demonstrated how ride-share matching could be achieved to reduce traffic and to improve transportation efficiency for individuals commuting to and from UVA. Leveraging both individual preferences in commuting times and their proximity criteria (i.e., home location and parking lot), this study identified opportunities for ride-sharing that could be implemented through a coordinated system, thereby enhancing the overall commuting experience while contributing to sustainability goals. The results suggested that a significant number of trips to and from UVA could be consolidated through ride-sharing, leading to a more efficient and sustainable transportation system on campus.

7 FUTURE WORK

Following the analysis of home location clusters, future work will focus on exploring their implications for optimizing ride-sharing routes to and from UVA. After filtering and clustering trips that start or end at UVA and preparing groups of vehicles, route overlap analysis can be conducted to provide a deeper probability of providing ride- sharing opportunities. Evaluating the overlap of routes for vehicles within the same cluster could suggest potential ride-sharing pairs or groups.

In addition, future work should also focus on evaluating a larger and more recent dataset to enhance the ride-sharing analysis. Incorporating an expanded dataset, researchers would be able to capture more travel patterns to find temporal changes in commuting behavior and identify emerging trends in ride-sharing demand. This would not only improve the robustness and reliability of the findings but also allow for a more comprehensive understanding of current UVA commuting behaviors, and facilitate the design of more effective ride-sharing initiatives and policies.

8 ACKNOWLEDGMENTS

This research was funded by the SMARTER University Transportation Center, and the Virginia Department of Transportation provided data access to the Wejo data. Additionally, this publication is based on research supported by Science Foundation Ire- land through grant 18/CRT/6222 (ADVANCE). The authors recognize Professors Donna Chen, Andrew Mondschein, Afsaneh Doryab, and a graduate student, Hanyi Wu, at the University of Virginia for their earlier efforts and discussions related to this research.

References

[de Palma et al., 2022] de Palma, A., Stokkink, P., and Geroliminis, N. (2022). Influence of dynamic congestion with scheduling preferences on carpooling matching with heterogeneous users. *Trans. Res. Part B: Methodol.*, 155:479–498.

[Hussain et al., 2022] Hussain, I., Knapen, L., Bellemans, T., Janssens, D., and Wets, G. (2022). A matching framework for employees to support carpooling in the context of large companies. *IEEE Trans. Intell. Transp. Syst.*, 23(2):1159–1170.

[Jain and Dubes, 1988] Jain, A. K. and Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc., USA.

[Li et al., 2023] Li, L., Zhang, H., and Gan, Z. (2023). Factors affecting college students' attitudes toward carpooling. *Transportation Safety and Environment*.

[Li et al., 2024] Li, X., Cheng, Y., Peng, H., and Dai, X. (2024). Can ridesharing improve the reserve capacity of transportation network? *IEEE Access*, 12:73196–73209.

[Li and Chung, 2020] Li, Y. and Chung, S. H. (2020). Ride-sharing under travel time uncer- tainty: Robust optimization and clustering approaches. *Computers Industrial Engineering*, 149:106601.

[Masoud and Jayakrishnan, 2017] Masoud, N. and Jayakrishnan, R. (2017). A real-time algorithm to solve the peer-to-peer ride-matching problem in a flexible ridesharing system. *Trans. Res. Part B: Methodol.*, 106:218–236.

[Narman et al., 2021] Narman, H. S., Malik, H., and Yatnalkar, G. (2021). An enhanced ride sharing model based on human characteristics, machine learning recommender system, and user threshold time. *J. Ambient Intell. Humaniz. Comput.*, 12(1):13–26.

[Naumov and Keith, 2023] Naumov, S. and Keith, D. (2023). Optimizing the economic and environmental benefits of ride-hailing and pooling. *Prod. Oper. Manag.*, 32(3):904–929.

[OpenStreetMap contributors, 2017] OpenStreetMap contributors (2017). Planet dump re- trieved from https://planet.osm.org . https://www.openstreetmap.org.

[University of Virginia, 2023] University of Virginia (2023). Updated parkingmastermap. [Online; accessed October 15, 2024].