Exploring Public Wi-Fi Log Data to Understand City-Level Travel Demand for Enhancing Microtransit Services in North Carolina Cities

NCDOT Project 2024-33
FHWA/NC/2024-33
June 2025

Lei Zhu, Ph.D.
Rachael (Yuqiu) Yuan
Department of Industrial and Systems Engineering
The University of North Carolina at Charlotte

STATE OF NORTH CAROLINA · DEPARTMENT OF TRANSPORTATION

RESEARCH & DEVELOPMENT

Exploring Public Wi-Fi Log Data to Understand City-Level Travel Demand for Enhancing Microtransit Services in North Carolina Cities

Draft Report RP2024-33

Submitted to

North Carolina Department of Transportation
Research and Development Unit
(Research Project No. RP2024-33)

Prepared By

Lei Zhu, Ph.D.
Rachael (Yuqiu) Yuan

Department of Industrial and Systems Engineering
The University of North Carolina at Charlotte
9201 University City Boulevard
Charlotte, NC 28223-0001

June 30, 2025

Technical Report Documentation Page

1. Report No. FHWA/NC/2024-33	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Exploring Public Wi-Fi Log Data to Understand City-Level Travel Demand for Enhancing Microtransit Services in North Carolina Cities		5. Report Date June 30, 2025
		6. Performing Organization Code
7. Author(s) Lei Zhu and Rachael (Yuqiu) Yuan		8. Performing Organization Report No.
9. Performing Organization Name and Address Lei Zhu, Ph.D., https://orcid.org/0000-0003-0170-1178 Rachael (Yuqiu) Yuan Department of Industrial and Systems Engineering The University of North Carolina at Charlotte 9201 University City Boulevard Charlotte, NC 28223-0001		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address North Carolina Department of Transportation Research and Development Unit 1549 Mail Service Center Raleigh, North Carolina 276699-1549		13. Type of Report and Period Covered Final Report September 1st, 2023 – June 30th, 2025 14. Sponsoring Agency Code RP2024-33
Supplementary Notes:		

16. Abstract

Microtransit services have gained popularity due to their convenience and other benefits; however, the service performance of microtransit, including passenger waiting time, needs improvement. Understanding city-level travel demand and mobility patterns is essential for developing improvement solutions. Passively generated and low-cost Wi-Fi log data provides an opportunity to analyze city-level travel patterns. This project aims to analyze and compare Wi-Fi-based travel patterns and microtransit travel patterns, to identify and validate potential improvement solutions for microtransit services in North Carolina cities. The city of Wilson, North Carolina, collected Wi-Fi log and microtransit travel data. A Wi-Fi log data processing framework was developed to obtain multi-modal travel activities from city-level Wi-Fi log data. The accuracy of Wi-Fi detection was validated by a field test conducted in Wilson. Also, the study analyzed the microtransit passengers' travel information obtained from the microtransit travel data in space and time. A comprehensive spatial-temporal clustering analysis of travel activities derived from the Wi-Fi log and microtransit data was conducted to understand city-level human travel patterns. By investigating the common and unique hotspots, it's possible to identify potential improvement solutions and make recommendations. Additionally, a microtransit simulation platform that integrates background traffic, passenger behavior, and service vehicle operations was developed to quantitatively evaluate the benefits of potential improvement solutions. The results and analysis approach developed in the project will help Wilson and other North Carolina cities with public free Wi-Fi to improve existing microtransit services or deploy new ones by identifying city-level travel demand and mobility patterns.

17. Key Words *Microtransit service, Public Wi-Fi log data, City-level travel demand, Human mobility pattern*	18. Distribution Statement		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 60	22. Price

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

DISCLAIMER

The contents of this report reflect the views of the authors and not necessarily the views of the University of North Carolina at Charlotte (UNC Charlotte) or the North Carolina Department of Transportation (NCDOT). The authors are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of either UNC Charlotte, NCDOT, the Federal Highway Administration, or the Federal Transit Administration (FTA) at the time of publication. This report does not constitute a standard, specification, or regulation.

ACKNOWLEDGMENTS

The authors acknowledge the Integrated Mobility Division (IMD) at NCDOT to provide financial support for this project. Special thanks are extended to Sarah Searcy and Darcy Downs for the excellent support, guidance, and valuable input, which were crucial for the successful completion of this project.

The efforts by Rachael (Yuqiu) Yuan and Mohini Joshi in the field test are appreciated. Additionally, Mr. Rodger Lentz, City Manager of Wilson, NC, kindly supported Wilson's RIDE microtransit data sharing. Mr. Will Aycock and his team (Winston Lee, Donald Richardson, and others) from Greenlight in Wilson, NC, are acknowledged and appreciated for Wi-Fi log data collection and sharing.

EXECUTIVE SUMMARY

Microtransit services have gained popularity for the convenience they offer, including providing first-and-last-mile (FLM) connections to public transit systems. The City of Wilson, North Carolina, launched a microtransit service in 2020, replacing its regular fixed-route bus service. Microtransit can offer a larger service area and cater to a greater volume of travel demand than traditional transit services. However, the service performance of microtransit, such as passenger waiting time, needs improvement. Understanding city-level travel demand and mobility patterns is essential for developing improvement solutions. Passively generated, low-cost Wi-Fi log data, along with city-level coverage, provides an opportunity to analyze city-level travel patterns. This project aims to analyze Wi-Fi-based travel patterns, as well as microtransit travel patterns, to identify potential improvement solutions and validation approaches for microtransit services in North Carolina cities.

The City of Wilson collected and shared Wi-Fi log and microtransit data. This study proposed a Wi-Fi log data processing framework to obtain travel activities from Wi-Fi log data. The study validated the accuracy of Wi-Fi detection by a field test conducted in Wilson. The study also analyzed the microtransit travel information obtained from the microtransit data in space and time. We applied a comprehensive spatial-temporal analysis to the user travel activities derived from the Wi-Fi and microtransit data to understand city-level human travel patterns. For instance, density-based spatial clustering analysis identified the origin and destination (OD) hotspots from the Wi-Fi and microtransit datasets. By investigating the common and unique hotspots, it's possible to identify potential improvement solutions and make recommendations. Additionally, we developed a microtransit simulation platform that integrates background traffic, passenger behavior, and service vehicle operations to quantitatively evaluate the benefits of potential improvement solutions.

Primary Observations and Preliminary Conclusions

- Our analysis of Wi-Fi detected travel activities shows the temporal and spatial patterns. The peak hour for employees is around 7 am, and the peak hours for non-employees are around 8 am and 3 pm. Additionally, we observed an increased travel flow around the city center and between the bus garage and the medical center. The field test indicates that the overall sampling rate in the study area is approximately 75%, which is comparable to the rates reported in the literature. In terms of trajectory tracking, Wi-Fi can accurately detect the sequence of visited locations. The mean absolute percentage errors (MAPE) of detected travel time between visited locations are 20%, 16.7%, and 13.4% for driving, walking, and microtransit modes, respectively.
- The analysis of microtransit data shows that the average passenger waiting time for the ride is 29.4 minutes for weekdays and 21.7 minutes for weekends (Saturdays). The demand for microtransit service is higher in the afternoon than in the morning and usually peaks at 3 pm.

The OD hotspots are generally in the city center, commercial areas, some recreational fields, and residential areas with a relatively high poverty rate.

- The comparison of travel patterns from Wi-Fi and microtransit services indicates that the "common hotspots" are located in the city center and the surrounding low-income residential areas. The "unique hotspots" of microtransit are the commercial areas, while the "unique hotspots" of Wi-Fi are parks and government buildings. 40.7% of microtransit trips can access Wi-Fi when a service request is placed. However, actual Wi-Fi and microtransit common trips are few, making up only 0.63% of the total trips.
- Based on the travel pattern comparison results, the following recommendations can be made.1) Installation of more Wi-Fi APs around the commercial area (grocery stores, shopping centers, and restaurants) to serve more users and travel activities. 2) Resuming the fixed route bus line connecting low-income residential areas and commercial areas to improve mobility efficiency. This study constructed a preliminary microtransit simulation platform to test and validate the feasibility of improvement solutions.

In conclusion, this project analyzed travel patterns from Wi-Fi log data and microtransit services and provided recommendations for service improvement. Furthermore, we developed a simulation platform to test and evaluate potential service improvement solutions. The results and analysis approach developed in the project will help Wilson and other North Carolina cities with public free Wi-Fi to improve existing microtransit services or deploy new ones by identifying city-level travel demand (OD pairs) and mobility patterns.

Table of Contents

LIST OF FIGURES

1. Introduction

Microtransit improves public transit riders' experience by operating a small-scale, on-demand transit shuttle fleet that can offer either fixed-route and fixed-schedule, as well as flexible-route and on-demand scheduled services [1]. An increasing number of transit agencies across the nation are deploying microtransit to serve residents and visitors, such as L.A. Metro in Los Angeles, CA [2], Central Ohio Transit Authority (COTA) [3], and King County Metro in Seattle, WA [4]. Microtransit integrates into existing public transit systems and enhances public transit service where running fixed-route buses is challenging. Other transit agencies directly deploy on-demand and flexible route microtransit services to replace traditional fixed-route and fixed-schedule buses. For example, in 2017, Arlington, Texas, replaced its fixed-route buses with an on-demand microtransit system, providing nearly half its 400,000 residents with transportation access that was not there before [5].

Southeastern cities are embracing microtransit. RIDE, powered by Via, is an on-demand, dynamically routed, mobile-app-powered shuttle service in partnership with the city of Wilson in North Carolina. Since 2020, the service has deployed about 26 gasoline-powered Toyota Sienna minivans in the county and city of Wilson, NC, which entirely replaced a previously existing fixed-route bus system and expanded the transit service coverage by 60% [6]. Microtransit deployments were incentivized, revealing Wilson's specific market dynamics, travel patterns, demographics, and travel preferences. Travel data, such as trip origins and destinations (OD), were collected. A recent survey indicates the success of the program and confirms that the RIDE program provides a critical service to the low-income, transit-dependent population. 52% of RIDE trips are primarily for commuting, while 38% are for essential errands (grocery, healthcare). Demographically, 57% of survey respondents report an annual household income below $25,000, and 86% do not have access to a personal vehicle. The RIDE program survey data indicate that economic-related service factors (e.g., fare, service time, fleet size) significantly sustain a microtransit system that provides an indispensable mobility option to transportation-disadvantaged groups.

Although the preliminary success of the RIDE program indicates a promising future for microtransit in Wilson, North Carolina, the microtransit service can be further improved in several aspects, including reducing riders' waiting time, vehicle re-dispatching, expanding the utility of transit and fleet size, and optimizing management. Understanding city-level travel demand and mobility patterns based on real-world travel data is considered a cornerstone to improve the system. However, collecting detailed city-level travel demand data is not a trivial task. Traditional travel surveys generally have a low sampling rate (about 5% of the population). As a result, such travel survey data and estimates may not be sufficient or accurate for estimating travel demand.

Aside from traditional travel surveys, Wi-Fi log data recording the user communication actions with detailed time and location information can more accurately reflect city-level human travel

activity. The widespread deployment of free public Wi-Fi services and the high penetration rate of smartphones enable the passive collection of Wi-Fi log data over a relatively long period. Wi-Fi log data indicates the number of people and travel activities of individuals who carry and use Wi-Fi-enabled mobile devices. Previously, public Wi-Fi log data collected on university campuses have been studied to understand the daily movement and travel behaviors of students and faculty [7, 8]. Beyond that, public Wi-Fi service deployments have recently gained momentum in a few city communities or neighborhoods, such as New York City, NY [9], and Copenhagen, Denmark [10]. Those studies analyzed Wi-Fi-based human travel in large metropolitan areas and densely populated regions. Still, they did not address Wi-Fi-based human travel patterns in small-scale, transportation-disadvantaged, and underserved cities, especially in North Carolina. Greenlight in the city of Wilson, NC, provides free public Wi-Fi services throughout the city and in key locations, including downtown, the Amtrak train station, Gillette Athletic Complex, and Southern Bank Stadium. Greenlight's raw Wi-Fi log data records city-wide Wi-Fi user daily activity information, reflecting human travel demand and movement, including transit riders' travel activities. Therefore, Wi-Fi-based city-level travel demand data, such as origin-destination (OD) pairs, provides an opportunity to understand city-level travel patterns, thereby improving microtransit service and transit fleet operations.

As a result, this project explored how city-level, Wi-Fi-based human mobility data can be used to enhance microtransit services. By identifying city-level travel demand (e.g., OD pairs) and mobility patterns derived from Wi-Fi data, the project can help Wilson and other North Carolina cities with public free Wi-Fi to improve existing microtransit or deploy new microtransit services. The major achievements of this project include:

1) **Developing a processing and analyzing framework for public Wi-Fi log data** to estimate and validate city-level human travel demand and mobility patterns.
2) **Conducting a field test** in Wilson to validate Wi-Fi data accuracy and Wi-Fi detection performance.
3) **Characterizing microtransit operation and rider behaviors** by analyzing the microtransit data in Wilson.
4) **Identifying potential improvements** for the microtransit system by correlating and comparing travel demand, spatially and temporally.
5) **Evaluating network performance and benefits** of the proposed microtransit service improvements by developing a microtransit simulation platform in Wilson.

2. Literature Review

2.1. Wi-Fi Log Data

2.1.1. Wi-Fi log data processing approaches

Wi-Fi log data contains detailed communication actions between clients (e.g., smartphones, laptops), access points (APs), and servers. Generally, the messages describing the connection, disconnection, and roaming process are more critical since they indicate the client's activity, i.e., when and where the clients arrive, leave, or travel between APs. The related messages in syslog, including association, disassociation, re-association, authentication, and de-authentication [11-13] were considered and processed. Approaches have been proposed to process the Wi-Fi log data based on key information, such as association messages from the university Wi-Fi data [8, 14]. There are several challenges in processing Wi-Fi log data, including the "ping-pong" effect and multi-device issues. Jahromi et al. pointed out that **ping-pong events** occur when clients frequently (in a short period) change their associations to nearby APs, while these changes aren't the result of actual movements [13]. Traunmueller et al. removed entries captured by multiple Wi-Fi APs simultaneously in the procedure of detecting client count and user trajectory [9]. The **multi-device problem** arises when a person carries more than one device to communicate with APs while moving around. An average of 1.3 devices per person on campus [15, 16] was estimated. To solve such a problem, Trivedi et al. used the device-user mapping information from the authentication events in Wi-Fi log files to ensure that only one device of the same user was counted [17]. However, the data provider may anonymize user identity-related information due to privacy concerns, making it more challenging to use device-user mapping. Additionally, data-driven methods have been developed to address multi-device problems. Trivedi et al. classified the type of each device by checking its network behavior and found out that 73.08% of the total devices in the University of Massachusetts campus are always-on devices, and the remaining 26.91% are hibernating devices [16, 18]. Zhang et al. identified three types of noisy devices that cannot reflect human mobility well by setting empirical rules on their network connection behavior, which took up to 63% of the total devices in Wi-Fi raw data obtained from the University of Macau [16].

2.1.2. Wi-Fi-based Human Activity Analysis and Applications

Wi-Fi log data has been widely utilized to support the analysis of human travel and mobility. Several analysis methods have been conducted to reveal the travel activity patterns from Wi-Fi log data, including spatial and temporal analysis, indoor hotspot analysis, and trajectory analysis. **Spatial and temporal analysis** have been applied to reveal mobility patterns in both temporal and spatial dimensions. Zakaria et al. studied the daily building occupancy trends and occupancy heatmaps in buildings before and during the COVID-19 outbreak to validate the effectiveness of crowd management policies during the pandemic using Wi-Fi data [8]. Soundararaj et al. demonstrated the temporal trend of high-street footfall using the proposed Wi-Fi-based footfall counting methodology, enabling a better conceptualization of urban mobility with a people dimension [19]. Hang et al. examined the Wi-Fi-detected hourly activity preferences at various locations within Purdue University on weekdays and weekends to develop a more accurate point of interest (POI) prediction method [20]. Traunmueller et al used origin-destination (OD) flow

between APs to build a network graph for different times of day to check the variation of mobility patterns in the Lower Manhattan section in NYC [9]. For **indoor hotspot analysis**, Trivedi et al. used Wi-Fi log data to estimate indoor occupancy hotspots to understand indoor human mobility better [21]; Marakkalage et al. detected indoor POI by clustering Wi-Fi fingerprints, addressing the limitations of GPS, which loses service in indoor environments [22]. In terms of **trajectory analysis**, human contact tracing was studied using Wi-Fi log data. Trivedi et al. [17] proposed a Wi-Fi Trace method to generate location reports and proximity reports for Wi-Fi users, supplementing the limitations of Bluetooth sensors that rely on mass adoption. Tu et al. leveraged Wi-Fi log data to obtain both contact duration and social distance between users, thereby improving the quality of contact tracing [23].

2.1.3. Wi-Fi Data Validation

Extensive research has been conducted on the validation of data quality, as data may yield biased results due to limited signal coverage and sensing performance across different locations. Data validation experiments have been conducted in various locations, including university campuses, city downtown and uptown areas, and shopping center complexes, from multiple perspectives, such as the accuracy of people count, location, session duration, and temporal lags in trajectory. For instance, Iresha et al. compared the room occupancy or people count obtained from Wi-Fi sensing for four classrooms on a university campus (UNSW) with the observed room occupancy and got an overall symmetric mean absolute percentage error (SMAPE) of 12.1% [15]. Trivedi et al. collected data from a group of 18 volunteers over 37 days, visiting 19,000 locations on campus, and achieved a precision of 0.93 and a recall of 0.94 for testing the location accuracy of clients by comparing Wi-Fi sensing with observation. Additionally, they achieved 100% accuracy for duration of sessions longer than 3 minutes. Trivedi et al. also compared the temporal lag between computed Wi-Fi trajectories with ground truth user trajectory and found out that temporal lags can be as high as 10 minutes for APs closely located, but are less than 1 minute for most of the other APs [17]. Similarly, Swain et al. inferred a university location of 46 students in 34 lectures over three months and got a precision of 0.89 and a recall of 0.75 in terms of people count [14]. For accuracy tests in cities, Traunmueller et al. compared the official manual daily pedestrian count with Wi-Fi probe requests in several locations of the Lower Manhattan section in New York City, with an offset from -11.37% to 10.36% [9]. At a retail high street in London, Soundararaj et al. compared the Wi-Fi-detected footfall from 5 locations with manual counts and got a mean absolute percentage error (MAPE) from 10% to 50% [19]. For building complexes, Wi-Fi can identify 40% to 60% of the customer count detected by human detectors at a large underground shopping mall in Nagoya [24]. Fukuzaki et al. demonstrated that the maximum error in the number of people detected by Wi-Fi packet sensors in the mall for each hour, accumulated over a day, is less than 7.9% [25].

2.2. Microtransit Service

Microtransit, a common form of demand-responsive transportation (DRT), utilizes vehicles such as vans and relies on app-based platforms for real-time booking and optimized routing [26]. It

complements the public transportation system by providing first-and-last-mile (FLM) connections or offering mobility services for low-density areas where public transportation service may be inefficient [26]. Several benefits have been identified for microtransit services. For example, Mayaud et al. concluded that the high induction rate for microtransit among women, older adults, people with disabilities, and low-income households highlights microtransit's ability to address long-standing inequities in transit planning [26]. Liezenga et al. found that microtransit addressed transportation issues by increasing accessibility, effectively reaching vulnerable rider groups, and creating travel opportunities in less spatially concentrated areas [27]. Lucken et al. found that microtransit vans have the potential to reduce vehicle miles traveled (VMT) more than taxis or transportation network companies (TNCs) by enabling greater vehicle occupancy [28]. Aside from the benefits, potential disadvantages have also been observed. Lucken et al. mentioned that microtransit services may pull riders away from more sustainable modes, such as buses, rail, bicycles, or walking [28]. However, Geržinič concluded that combining subsidized microtransit with higher parking prices is the most effective strategy for achieving a modal shift, primarily from car to microtransit, while minimizing the impact on public transport [29]. Furthermore, Mayaud et al. proposed a framework to redesign the transit system based on microtransit data, aiming to improve overall transportation efficiency [30].

To conclude, with proper planning and management, microtransit services can improve the accessibility and effectiveness of regular fixed-route public transit systems, thereby enhancing the overall efficiency of the transportation system.

2.3. Transportation Simulation

Transportation simulation can be applied to test, evaluate, and optimize traffic operations by replicating real-world traffic scenarios. Generally, there are three types of simulation models: macroscopic, mesoscopic, and microscopic. Macroscopic models are based on the deterministic relationships of flow, speed, and density of the traffic stream [31]. It can build a large-scale traffic simulation on an aggregated level; however, the dynamics of vehicles cannot be simulated [32]. On the contrary, microscopic models can simulate the movement of individual vehicles based on car-following and lane-changing theories [31]. It can model detailed travel behavior and complex traffic phenomena; however, it is relatively hard to calibrate and has a higher computational burden [32]. In between, mesoscopic models combine the properties of both microscopic and macroscopic models[31]. It can capture the aggregated-level fluctuation of traffic and is relatively easy to calibrate with higher computational efficiency; however, it cannot model vehicle behavior and vehicle-pedestrian interaction [32].

To evaluate the performance of microtransit operation, this project applied a microscopic simulation model. Microscopic simulation models have been applied to various types of evaluations involving detailed vehicle behaviors. For example, Zhang et al. constructed a simulator, CAVSim, to evaluate the decision, planning, and control methods for CAVs at different levels [33]. Zatmeh-Kanj et al. used TRANSMODELER to study car-following

behavior in the context of distracted driving [34]. Zhu et al. used Simulation of Urban Mobility (SUMO) to test and evaluate the proposed demand-side cooperative shared automated mobility (DC-SAM) service. It found that the proposed DC-SAM service can significantly reduce the SAV's operating costs in terms of vehicle miles traveled (VMT), vehicle hours traveled (VHT), and vehicle energy consumption (VEC) by up to 53%, 46%, and 51%, respectively. [35]. Meshkani et al. proposed an on-demand transit service for the first-mile connection and used SUMO for performance evaluation. Compared to the existing bus transit service, the proposed on-demand transit service is expected to result in a 36% reduction in total travel time and a 41% reduction in detour time [36].

2.4. Point Data Clustering for Mobility Analysis

Clustering is a technique that groups unlabeled data with little or no supervision into different classes, by dividing objects with similar characteristics into the same class and different objects into other classes [37]. Mobility data typically includes GPS trajectories, origin-destination (OD) locations of individual trips, or OD stations/stops for transit systems, which can comprise a large number of records and thus represent a type of big spatial data. To enhance data interpretability, individual origin and destination locations can be clustered to reveal aggregated mobility patterns, such as identifying urban hotspots.

Urban hotspot refers to regions with frequent human mobility activities, heavy traffic flow, and prosperous economic activities [38]. A comprehensive understanding of the characteristics and patterns of travel activities with urban hotspots is critical for the sustainable development of modern cities [39]. Xia et al. used a network-based spatiotemporal field clustering method to identify urban hotspots from pick-up events extracted from taxi trajectories in downtown areas of Nanjing, China, and used an isoline model to delimitate the hotspot centeredness [38]. Ran et al. proposed a novel K-means clustering method based on a noise algorithm to capture urban hotspots from taxi GPS trajectory points [40]. Density-based clustering methods have also been widely applied to detect urban hotspots. Li et al. proposed the Spatial-Temporal Hierarchical Density-Based Spatial Clustering of Applications with Noise (ST-HDBSCAN) model to discover uneven clusters from spatiotemporal taxi GPS trajectory data and to understand the evolution of urban hotspots [41]. Cesario et al. pointed out that though classic density-based clustering algorithms are suitable for discovering hotspots characterized by homogeneous density, their application on multi-density data can produce inaccurate results. Cesario et al. compared the multi-density-based clustering algorithm called City Hotspot Detector (CHD) with other classic density-based clustering methods, including Density-Based Spatial Clustering of Applications with Noise (DBCSAN), Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), and Ordering Points To Identify the Clustering Structure (OPTICS), and found that CHD outperformed other methods [42].

To summarize, mobility point data can be clustered for hotspot analysis, providing better support for urban and transportation planning and management. Different clustering models have been applied to mobility data, especially density-based clustering methods such as DBSCAN.

3. Data Preparation

3.1. Wi-Fi Log Data

Wi-Fi log data records communication activities between network devices (i.e., clients) and access points (APs), as well as interactions with internet servers. Figure 1 shows an example of raw Wi-Fi log data. Since the log data contains the timestamps of each client connecting to or disconnecting from each AP, the movement sequence of each client can be obtained once the location of each AP is known.

```
2/18/2024 5:13:03 PM    Daemon.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: hostapd: ath0: STA dev_002316206 IEEE 802.11: disassociated
2/18/2024 5:13:03 PM    Daemon.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: hostapd: ath2: STA dev_002316206 IEEE 802.11: disassociated
2/18/2024 5:13:03 PM    Daemon.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: hostapd: ath1: STA dev_002316206 IEEE 802.11: disassociated
2/18/2024 5:13:03 PM    Kernel.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: kernel: [6327960.097703] ieee80211_sta_leave: dev_002316206
2/18/2024 5:13:03 PM    Kernel.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: kernel: [6327960.551759] ieee80211_sta_leave: dev_000497454
2/18/2024 5:13:03 PM    Daemon.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: hostapd: ath2: STA dev_002580556 DRIVER: Sead AUTH addr=dev_000497454 status_code=0
2/18/2024 5:13:03 PM    Kernel.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: kernel: [6327960.563067] ieee80211_sta_leave: dev_000497454
2/18/2024 5:13:03 PM    Daemon.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: hostapd: ath2: STA dev_000497454 IEEE 802.11: disassociated
2/18/2024 5:13:03 PM    Daemon.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: hostapd: ath1: STA dev_002580195 DRIVER: Sead AUTH addr=dev_000497454 status_code=0
2/18/2024 5:13:03 PM    Kernel.Info     172.30.1.198    U7MP,dev_002581641,v4.3.13.11253: kernel: [6327960.565753] ieee80211_sta_leave: dev_000497454
2/18/2024 5:13:03 PM    User.Info       172.30.1.209    U7LT,dev_002581595,v4.3.13.11253: : stahtd[23746]: [STA-TRACKER].stahtd_dump_event():
{"message_type":"STA_ASSOC_TRACKER","mac":"dev_002367110","vap":"ath0","auth_ts":"1563875.312954","event_type":"failure","assoc_status":"32","event_id":"1"} - n2p, flags: 0x20
2/18/2024 5:13:03 PM    User.Info       172.30.1.209    U7LT,dev_002581595,v4.3.13.11253: : stahtd[23746]: [STA-TRACKER].stahtd_dump_event():
{"message_type":"STA_ASSOC_TRACKER","mac":"dev_002367110","vap":"ath1","auth_ts":"1563875.313010","event_type":"failure","assoc_status":"32","event_id":"1"} - n2p, flags: 0x20
2/18/2024 5:13:03 PM    User.Info       172.30.1.209    U7LT,dev_002581595,v4.3.13.11253: : stahtd[23746]: [STA-TRACKER].stahtd_dump_event():
```

Figure 1: Example of Wi-Fi log Data

The City of Wilson provided raw Wi-Fi log data from 02/14/2024 to 02/27/2024, as well as the location of APs. Figure 2 shows the AP locations. The city center has the densest APs, and other APs are located around public institutions, hospitals, schools, parks, some residential areas, and commercial areas. For privacy protection, the city IT department anonymized the Media Access Control (MAC) addresses of the clients (e.g., mobile phones, tablets, and laptops) by replacing them with the corresponding device identification numbers.

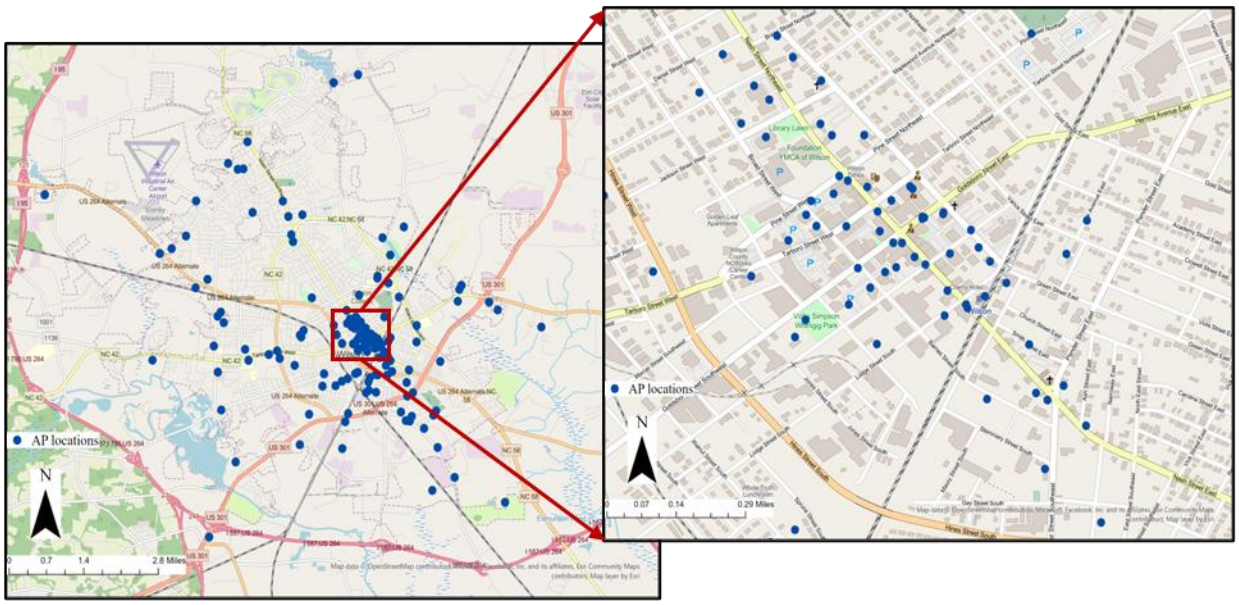


Figure 2: Location of APs in the city of Wilson

3.2. Microtransit Data

We obtained microtransit data from the microtransit system in the city of Wilson, NC. The microtransit service has been using 15 on-demand shuttles to replace the fixed-route transit service since 2020 [43]. Riders can use the mobile application (Via) to request or reserve microtransit service, and a shuttle is matched and sent to service the ride request. The Via app records all trip information, including the origin, destination, pickup time, drop-off time, and the related rider ID, vehicle ID, and driver ID, as shown in Figure 3. With the information we obtained, the origin-destination (OD) flow of the microtransit service in February 2024 is shown in Figure *4*. The microtransit service covers the whole city area. There was a total of 23,018 requests in February 2024. Figure *5* shows the distribution of request status for seven categories, with 70.1% (16,101) of the total requests being completed. Figure *6* shows the daily number of requests. There are more requests on weekdays than on Saturdays. The average daily number of requests on weekdays is 960, and the count is 660 for Saturdays. Note that although there is no microtransit service on Sundays, an average of 50 requests are placed on Sundays.

Request Creation Time	Session ID	Request Status	Rider ID	Driver ID	Vehicle ID	Origin Lat	Origin Lng	Destination Lat	Destination Lng
2/15/2024 5:30	212571412	Completed	3506372	102927	147490	35.69868188	-77.93357447	35.78324071	-77.95032654
2/15/2024 5:30	212571413	Completed	3334360	106866	148994	35.7084461	-77.8975727	35.7645592	-77.9626443
2/15/2024 5:30	212571414	Completed	2922284	102927	147490	35.72067118	-77.8896258	35.73413	-77.94874
2/15/2024 5:31	212571418	Completed	3028772	106866	148994	35.74929593	-77.91695114	35.74369393	-77.96780415
2/15/2024 5:31	212571419	Completed	3623775	106866	148994	35.73185979	-77.91330569	35.77046333	-77.93972075
2/15/2024 5:33	212571454	Completed	2972161	104987	148539	35.69244759	-77.91718148	35.71992072	-77.89856795
2/15/2024 5:33	212571458	Completed	3227265	104987	148539	35.75267539	-77.95858406	35.694331	-77.915124
2/15/2024 5:33	212571459	Completed	3132987	104987	148539	35.71823875	-77.92694103	35.6931944	-77.8904015
2/15/2024 5:35	212571483	Completed	2913332	102927	147490	35.7124	-77.9108	35.7340846	-77.9487005
2/15/2024 5:35	212571484	Completed	3539045	102927	147490	35.71791046	-77.88922817	35.71907	-77.94361
2/15/2024 5:35	212571486	Completed	3104918	102927	147490	35.7159289	-77.9070815	35.715236	-77.9419178

Figure 3: Example of Microtransit Data

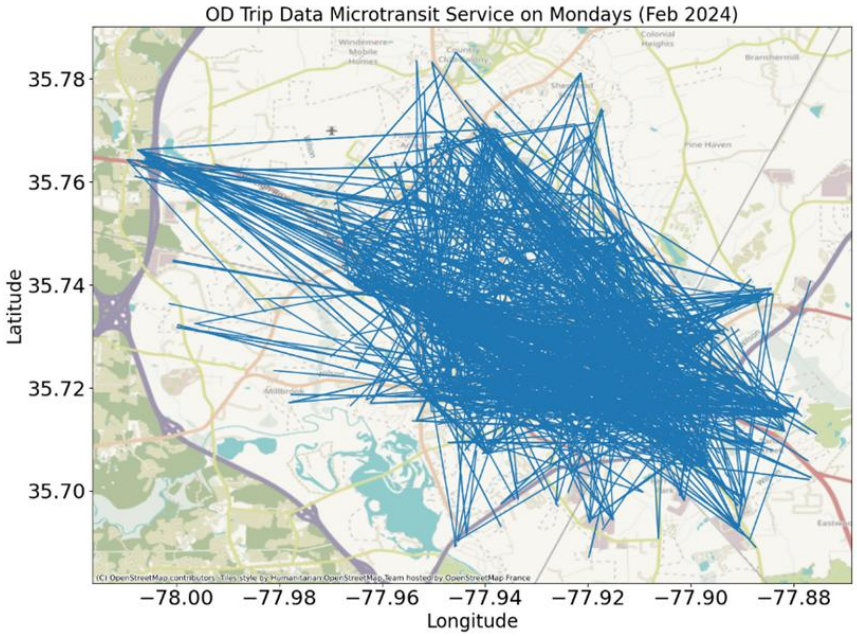


Figure 4: OD Flow from Microtransit Service

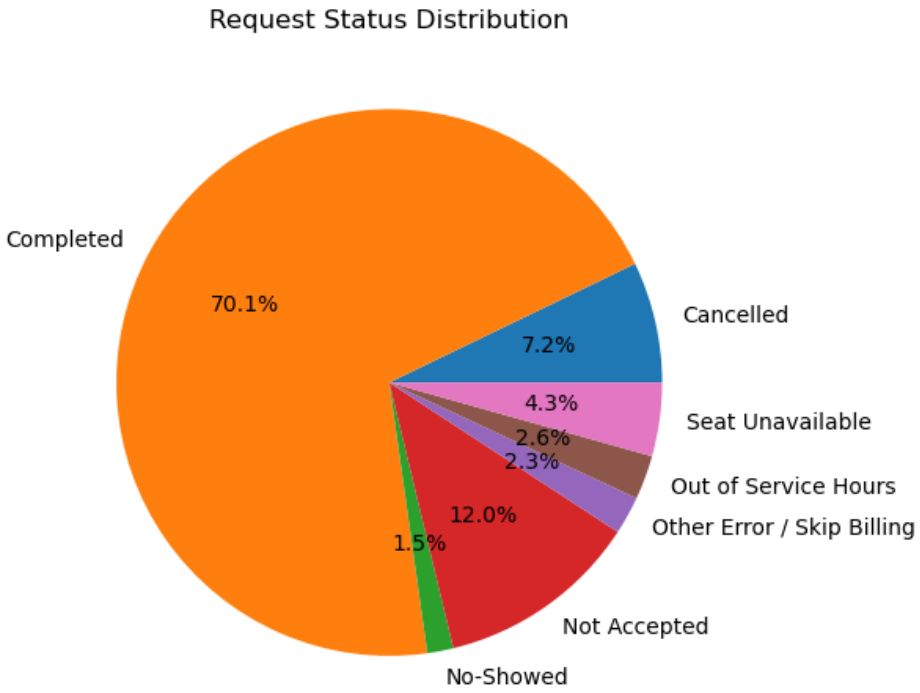


Figure 5: Distribution of Request Status for Microtransit Service in Wilson during Feb 2024

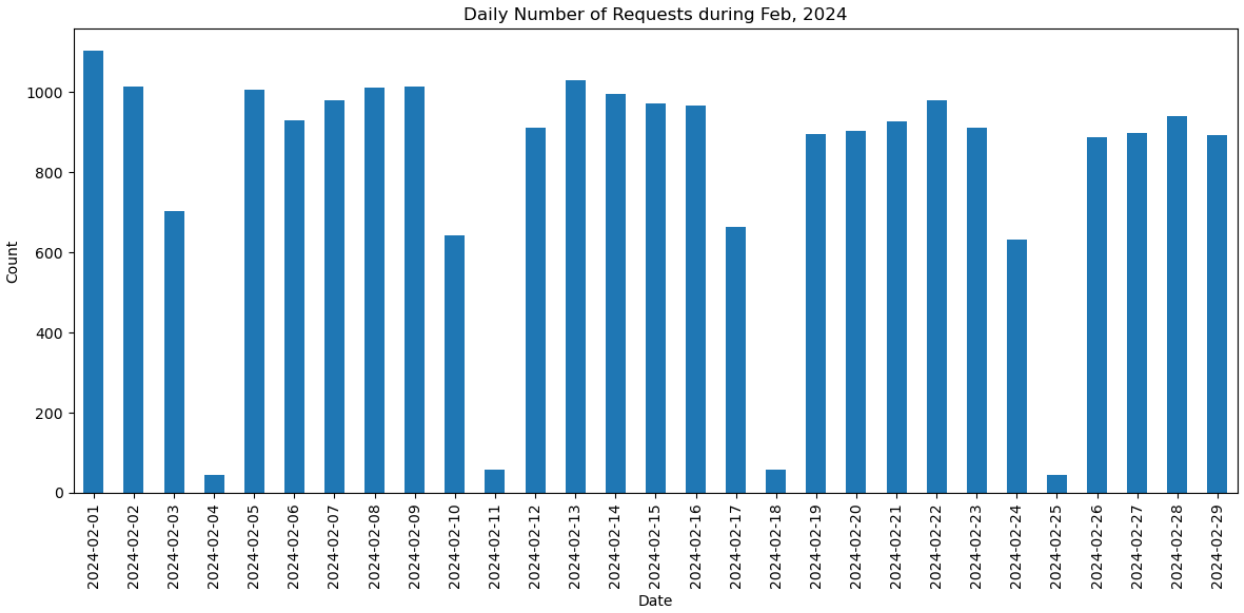


Figure 6: Daily Number of Requests during Feb 2024

3.3. Other Datasets

To understand the impact of demographic features on microtransit and Wi-Fi services, we obtained census block groups and related features from the 2021 American Community Survey

(ACS) [44]. Figure *7* illustrates the polygons of census block groups that cover the city of Wilson, along with the population of each census block group.

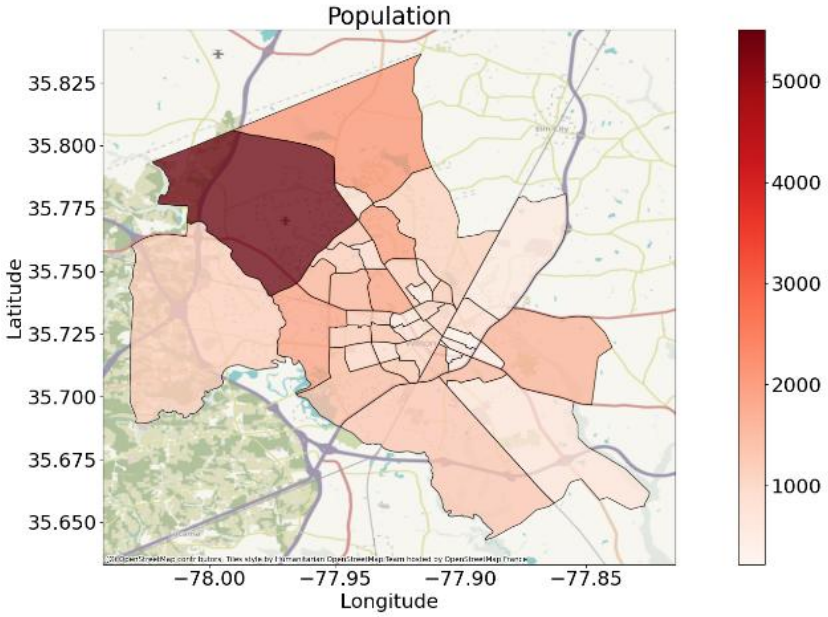


Figure 7: Census Block Groups and Related Population

We collected fatal and serious injury crash locations were or the city of Wilson from the NCDOT website [45]. 63 crashes were recorded from 2014 to 2023 in the city, including 14 types of crashes, as shown in Figure *8*. The top 3 crash types are 1) pedestrian-involved crashes, 2) angle crashes, and 3) run-off-road-right crashes. Figure 9 shows the spatial distribution and density of the crashes. It shows that the high-density area is concentrated around the city center and extends south to the residential area.

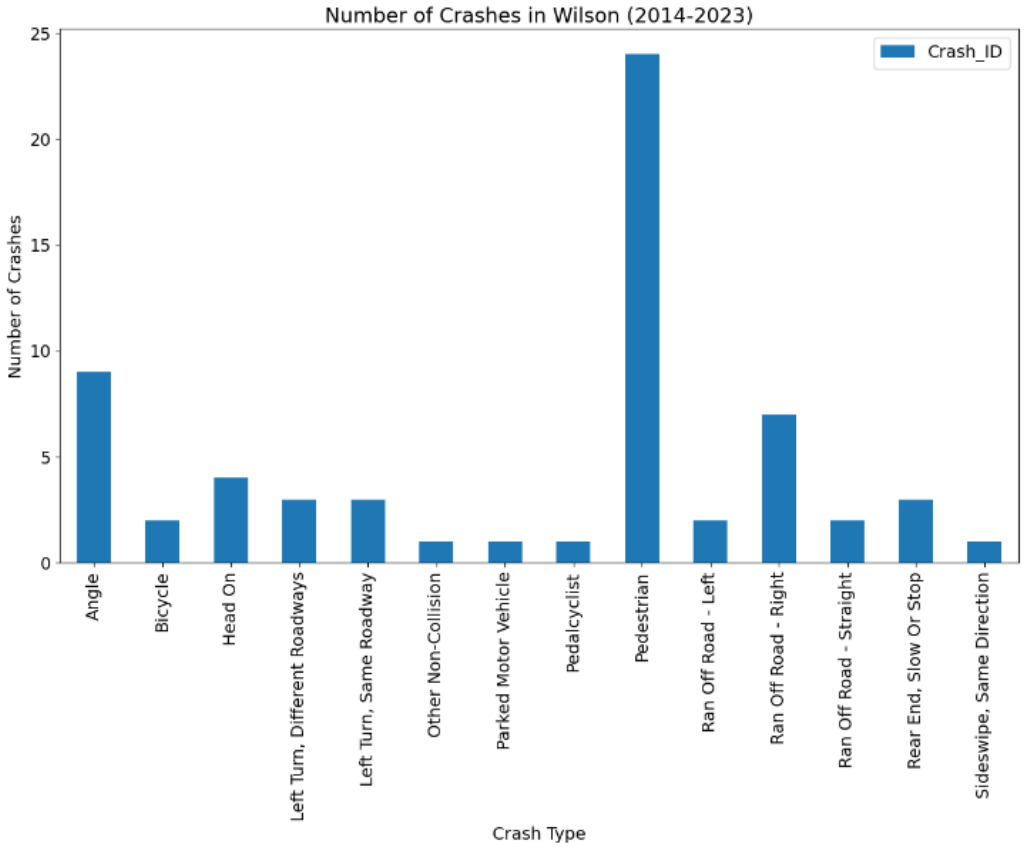


Figure 8: Number of Crashes for Each Crash Type

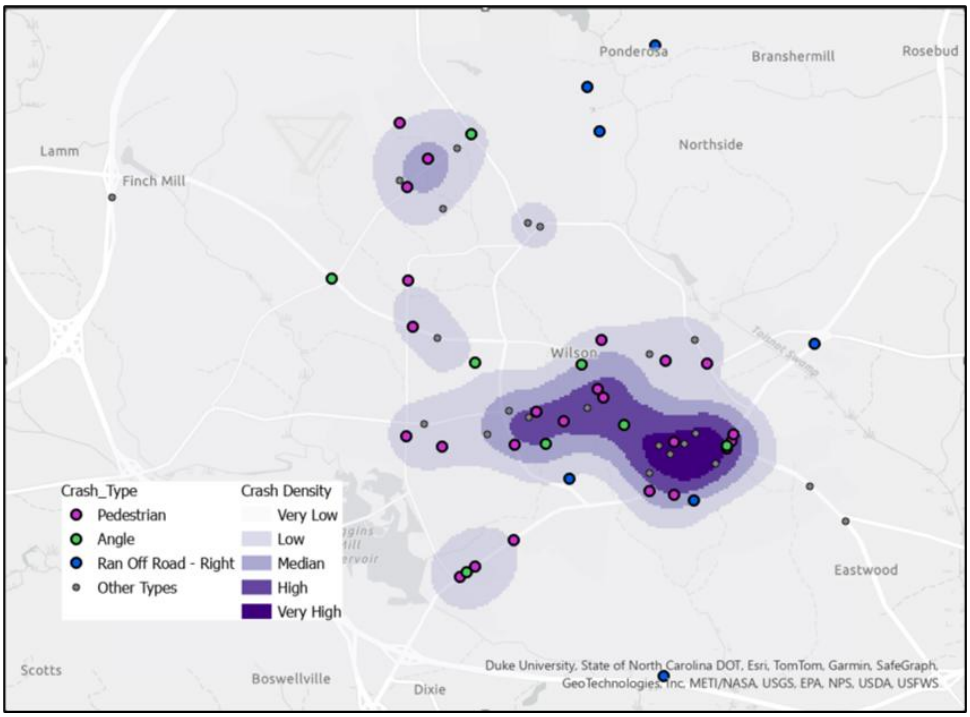


Figure 9: Spatial Distribution of Crashes

4. Methodology

4.1. Wi-Fi Log Data Processing and Validation

A Wi-Fi data processing framework was developed to process Wi-Fi log data and identify client travel activities. Further, field tests were conducted in the City of Wilson to validate the accuracy of Wi-Fi detection.

4.1.1. Wi-Fi Log Data Processing

IEEE 802.11x standards indicate that a Wi-Fi connection includes three major steps [46], including the scanning process, authentication process, and association process. For the disconnection procedure, disassociation or de-authentication is also recorded. According to that, the Wi-Fi log data processing approach has four steps, including 1) raw data pre-processing; 2) AP-level session archiving; 3) location-level session reconstruction; 4) API based user travel activity generation.

- **Raw Data Pre-Processing**

Key log messages in raw Wi-Fi log data, such as authentication requests and successes, as well as association requests and successes, as well as de-authentication and disassociation, were extracted to form Wi-Fi communication actions with the attributes of Time, Type, Client ID, AP ID, and Location ID. Time *(*or t*)* represents the timestamp of a log. Type indicates the type of communication actions, i.e., connection(s) or disconnection(s), according to communication status messages. The Client ID is a unique identifier for each client, such as a device's MAC address. For privacy protection, the device's MAC addresses were substituted and re-coded. APID identifies each AP, while LocationID denotes the location of the AP.

- **AP-Level Session Archiving**

An AP-level session describes the connection and disconnection of an AP for a client, which can be constructed by matching a client's communication action pair, comprised of connection and disconnection for each AP. The attributes of the AP-level session include ClientID*,* APID*,* StartTime*,* and EndTime. StartTime is the connection time, while EndTime is the disconnection time.

It is easy to match the connection and disconnection actions in typical cases, where each connection action has a corresponding disconnection action. However, some exceptional cases, related scenarios, and solutions are described in Table *1*. Those special cases cannot be removed from the data, as they may comprise a significant portion, especially in areas where Wi-Fi coverage is unstable or clients are detected unintentionally. Sometimes, there are only connection messages between a client and an AP, as stated in case 1. This might be because the actual network connection is not established successfully, while the AP does discover the client nearby during the scanning process. The solution is to set the EndTime the same as the StartTime. Case 2 shows that there are only disconnection messages between a client and an AP. Two scenarios

may contribute to this case. One reason is that the connection message may be missing because no actual network connection message has been established. The solution is to set the StartTime the same as the EndTime. Another scenario is that the duplicated disconnection messages are generated after the client leaves the AP's signal range due to system lag. Messages resulting from system lag were removed. The two scenarios can be separated by using a time threshold for the gaps between consecutive disconnection actions from the same client and AP. If the gap is less than a threshold, the message is generated due to system lag, which belongs to scenario 2 and was removed. The threshold is determined by calculating the average gap time between consecutive disconnection actions from the same client and AP, which may differ in study areas.

Another case that needs attention is the Ping-Pong effect, shown in Case 3 in Table *1*, which causes unnecessary fluctuation of communication actions and blurs the client's activity. The ping-pong effect can be detected by three conditions [13]: 1) the duration of the AP-level session is shorter than a threshold T_{TD}; 2) the time gap between two continuous AP-level sessions is shorter than a threshold T_{TT}; 3) the occurrence of AP change for the same client during $T_{TD} + 2T_{TT}$ is more than one time (e.g., more than two AP-level sessions occur within 20 seconds). In this project, T_{TD} is set to 10 seconds, and T_{TT} is set to 5 seconds. Once detected, the AP-level sessions with a ping-pong effect were smoothed and merged with the immediately preceding or following non-ping-pong AP-level sessions of the client, whichever had a longer duration.

After the client leaves the signal range of the AP, multiple disconnect messages may be generated, which is normal due to the log message generation mechanism. The solution is to use the first consecutive disconnect message to avoid lag. Additionally, some disconnect messages may be generated later than the connection message of the subsequent session between the same AP and clients, due to the lag in generating syslogs, especially when APs are located close to each other. To avoid overlapping, the solution is to use the StartTime of the following session as the EndTime of the current session.

Table 1: Special Case, Related Scenarios and Solutions for AP-level Session Construction

Case No.	**Case**	**Scenario**	**Solution**
1	The client has only connection actions for some or all APs	Client detected by APs without an actual connection (scanning)	Set the EndTime to be the same as the StartTime
2	The client has only disconnection actions for some or all APs	Client detected by APs without an actual connection (connection message missed)	Set the StartTime to be the same as the EndTime
		Generated because of system lag after the client leaves	Remove the disconnection action

3	Ping-Pong phenomenon	Clients in the signal overlap area from several APs	Ping-Pong Smoothing
4	The disconnect message is generated multiple times after disconnection	Normal due to the log message generation mechanism	Use the first disconnect message
5	The disconnect message is generated later than the connection message of the following session.	APs are located close to each other.	Use the StartTime of the following session as the EndTime of the current session.

- **Location-Level Session Reconstruction**

On top of AP-level sessions, location-level sessions describe how a client travels across a sequence of visited locations and stays in each location. Each location may have more than one AP, and all AP-level sessions are merged as location-level sessions. Each client (ClientID) needs several records to describe the complete location-level travel activity, including visited locations (StartLocation, EndLocation) and their timestamps (StartTime, EndTime). According to different activity statuses, the location-level client activity contains two types of sessions, Stationary Sessions [47] and Transitional Sessions [48], which describes the status of clients staying in one location or traveling between locations, respectively. It is worth noting that the StartLocation equals the EndLocation for a stationary session. Figure *10* shows an example of client activity that contains three transitional sessions (TS1, TS2, TS3) and three stationary sessions (SS1, SS2, SS3), where TS1 is represented as $\{\mathrm{Client1}, \mathrm{t1}, \mathrm{t2}, \mathrm{l1}, \mathrm{l2}\}$, SS1 is represented as $\{\mathrm{Client1}, \mathrm{t2}, \mathrm{t3}, l2, \mathrm{l2}\}$ and so on. For TS1, the StartLocation and EndLocation are $l1$ and $l2$ respectively and $t1, t2$ are the StartTime and EndTime; for SS1, the StartLocation and EndLocation are both $l2$. $t2$ *and* $t3$ are the StartTime and EndTime.

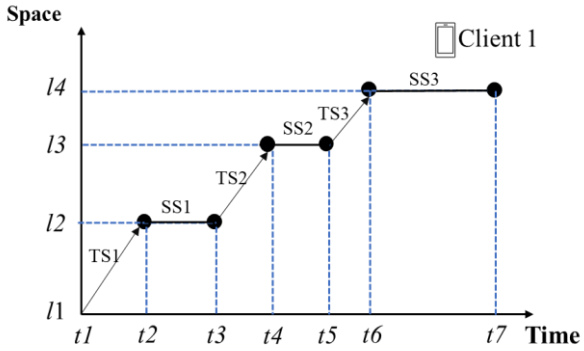


Figure 10: Example of Client Travel Activity

Determination of Start and End Location

The location can be obtained from a single AP or the center of the AP cluster. For AP cluster center cases, such as APs in the same location, the AP-level sessions of all APs in the cluster for

the same client are merged into a combined AP-level session by ordering the AP-level sessions of the cluster by their start times. The first AP-level session's StartTime is the new StartTime, and the last AP-level session's EndTime is the new EndTime. The AP cluster was assigned to a unique identifier, ClusterID. After that, the **processed AP sessions** have the attributes of ClientID, APID/ClusterID, StartTime, and EndTime. As shown in Figure 11, the processed AP sessions are combined from 2 AP-level sessions.

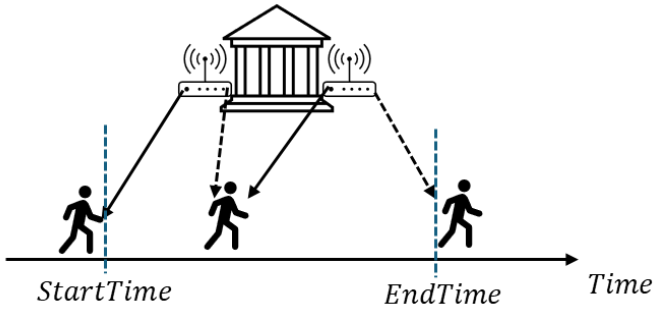


Figure 11: Example of Processed AP Sessions

The duration of processed AP sessions or the time gaps between StartTime and EndTime can be used to determine the client's (stationary or transitional) status around the location. The client can either pass by a location (e.g., duration <= 5 minutes, defined as a waypoint) or stay at a location for a period (e.g., duration > 5 minutes, defined as origin or destination). In other words, the location can be either a waypoint or an origin/destination, depending on the duration of the post-processed AP-level session. Suppose the APID/Cluster ID represents a waypoint. In that case, the StartLocation of such client activity is the location of the current post-processed AP-level session, and the EndLocation is the location of the next post-processed AP-level session, indicating that the client travels from one location to the other. Suppose the APID/Cluster ID represents an origin or destination. In that case, both StartLocation and EndLocation are in the same location as the APID/ClusterID, indicating that the client remains in that location for a period.

Determination of Start and End Time

The StartTime and EndTime of a client activity session indicate when a client passes a waypoint or arrives at or departs from an origin or destination. However, the timestamps in the post-processed AP-level sessions differ, representing when the client connects to or disconnects from an AP or AP cluster. Modifications are required to accurately reflect the pass-by time and the arrival/departure time, enabling a more accurate estimation of client or user travel time between waypoints and origins/destinations, taking into account the Wi-Fi signal range and connection features. If the location is a waypoint, the visit time (VT) or pass-by time is the middle timestamp between the StartTime and EndTime of the post-processed AP-level session, as shown in equation (1). If the location is an origin/destination, the arrival time (AT) is calculated as the StartTime of the post-processed AP-level session plus an approaching time; the departure time

(DT) is the EndTime of the post-processed AP-level session minus a leaving time" as shown in equations (2) and (3). Because the AP signal has a specific and limited range, the client may be connected before arriving and disconnected after departing from the location. Assuming an AP has a range of 100 meters outdoors [49] and the average walking speed is 1.5 meters per second [50], both approaching and leaving times are estimated as 67 seconds.

$$VT_t = \frac{EndTime_{ap_t} - StartTime_{ap_t}}{2} \quad (1)$$

$$AT_t = EndTime_{ap_t} - 67\ sec \quad (2)$$

$$DT_t = StartTime_{ap_t} + 67\ sec \quad (3)$$

Therefore, for stationary sessions (origins/destinations), the StartTime for the session t is the arrival time AT_t, and the EndTime is the departure time DT_t. For transitional sessions (waypoints), the StartTime for the session t is the departure time of a destination DT_t, and the EndTime is the arrival time of the following origin AT_{t+1}. After Wi-Fi data processing, a client activity table of all clients was generated. As shown in Table 2, the client activity contains the following attributes: ClientID, StartTime, EndTime, StartLocation, and EndLocation. The example shows that client 00001 first started the trip from L1 at t1, passed by L2 at t2, and arrived at L3 at t3. After arriving, the client stayed at L3 from t3 to t4 and then left the study area from L4 at t5.

Table 2: Example of Client Activity Table

ClientID	StartTime	EndTime	StartLocation	EndLocation
00001	t1	t2	L1	L2
00001	t2	t3	L2	L3
00001	t3	t4	L3	L3
00001	t4	t5	L3	L4

Note that stationary devices, such as desktops, were removed since they cannot accurately represent users' travel activity. Thus, devices with only stationary sessions or those that have an average daily connection time greater than 16 hours were removed before further processing.

- **API-Based User Travel Activity Generation**

User travel activity, including visited location sequences, travel routes, and travel modes, was generated from location-level sessions. The visited location sequence includes a list of locations (i.e., OD and waypoints) ordered by the time of visit during the day. With origins, destinations, and waypoints, travel modes and routes of segments in a visited location sequence were generated with road network information. This study utilizes the Google Distance Matrix API and Google Directions API to obtain detailed routing information. The Google Distance Matrix API can return the travel distance and duration between an origin and a destination by accepting an HTTPS request containing the origin and destination for a given mode of transport [51]. The

Google Directions API can return detailed directions, broken down into routes, legs, steps, and encoded polyline representations of the entire travel route between the specified origin, destination, and waypoints for a chosen mode of transport [52].

In terms of travel mode, this study only considers walking and driving. Considering that clients may choose multiple travel modes to travel from their origins to destinations, such as first walking to the parking lot and then driving to a destination, the transitional sessions were first divided into sub-sessions by waypoints, and the travel mode of each sub-session was determined. By comparing the durations of sub-sessions with the durations for both travel modes (driving and walking) obtained from the Google Distance Matrix API, the travel mode of each sub-session was determined based on the closeness of the duration. Consecutive sub-sessions with the same travel mode were combined to form a single mode trip. The origin of the single-mode trip is the first location of the consecutive sub-sessions, and the destination is the last location. The waypoints are the locations in between. Then, the Google Directions API can generate a detailed route for a single-mode trip, including a series of turning or shaping points, given the coordinates of the origin, destination, waypoints, and travel mode. For example, as shown in Figure *12*, a client is detected leaving a location for location 4, passing through locations 2 and 3. Thus, there is a transitional session between locations 1 and 4, which can be divided into three sub-sessions, including locations 2 and 3. The travel mode for sub-session 1 was walking, and the travel mode for sub-session 2 and 3 was driving. As a result, a walking route from location 1 to location 2 was generated, and a driving route from location 2 to location 4, passing by location 3, was generated for the client. By combining these trips, the topology of the user's travel route between origins and destinations was obtained.

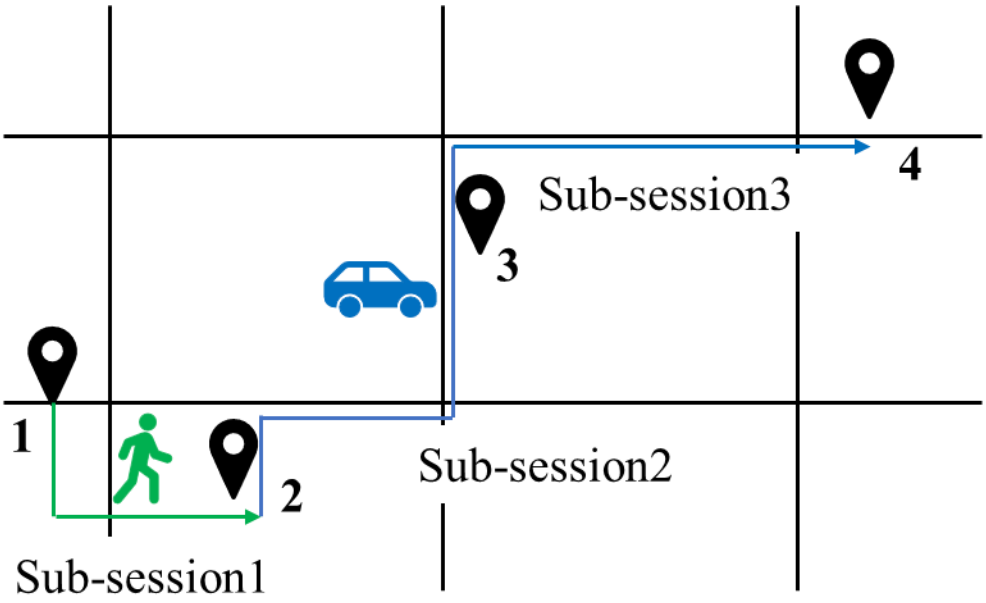


Figure 12: Example of Travel Activity Generation

4.1.2. Wi-Fi Log Data Validation

To validate the quality and accuracy of Wi-Fi log data, this study conducted two types of tests: a sampling rate test and a trajectory tracking test. Sampling rate tests assess Wi-Fi's capability to detect human presence, while trajectory tracking tests evaluate the performance of Wi-Fi user tracking when clients travel among different locations. Volunteers were manually connected to the public Wi-Fi network using their smartphones for the first time. They then enabled an auto-connection function, allowing their devices to connect to Wi-Fi automatically.

- **Sampling Rate Test**

A sampling rate is the ratio of the number of Wi-Fi detected clients to the total number of people visiting the location. Thus, the number of people visiting the locations, serving as the ground truth, was collected. To make the obtained sampling rate more representative, the study used three distinct types of locations in downtown Wilson, including Wilson Utilities Center (WUC), a government institution, Gig East Exchange office (GEEO), a shared working space that sits on the roadside, and the Wilson Public Library (WPL), a location for public activity. Table 3 shows the data collection period for each location. Volunteers were present at the test locations and manually counted the time and number of people moving in (inflow) and out (outflow). The sampling rate of inflow (SR1) and outflow (SR2) was calculated as shown in the equation (4) and (5).

$$\textit{Sampling Rate 1(SR1)} = \textit{Inflow recorded by Wi-Fi/ Inflow collected manually} \quad (4)$$

$$\textit{Sampling Rate 2(SR 2)} = \textit{Outflow recorded by Wi-Fi/ Outflow collected manually} \quad (5)$$

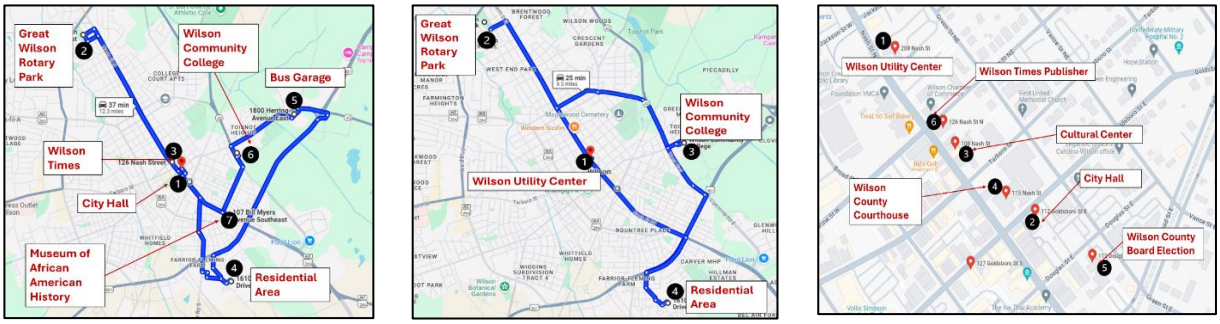
Table 3. Experiment Sessions for Sampling Rate Test

Sessions	Time Period	Location
2/21/2024		
1	13:00-17:00	WUC
2	14:00-17:00	GEEO
2/22/2024		
3	9:00-16:30	WUC
4	9:00-12:00	GEEO
5	14:00-16:30	WPL

- **Trajectory Tracking Test**

Trajectory tracking tests compared Wi-Fi-detected user trajectories with the actual user trajectories, which included walking tests, driving tests, and microtransit tests. In the walking tests, a volunteer walked through the designated locations in downtown Wilson with Wi-Fi signal coverage and recorded the actual arrival and departure times at each location. In the driving tests, one volunteer drove to the designated AP locations in order, and the other volunteers recorded the arrival and departure times. During both walking and driving tests, the volunteers stopped at

the designated location for a few minutes to ensure their devices' Wi-Fi auto-connection was enabled. In the microtransit test, a volunteer traveled between locations with Wi-Fi signal coverage using microtransit shuttles by requesting service via a smartphone when connected to Wi-Fi. Figure 13 shows the designated locations and related routes for walking, driving, and microtransit.



(a) Driving Trajectory (b) Microtransit Trajectory (c) Walking Trajectory

Figure 13: Trajectory tracking test

4.2. Spatial Clustering Method

To explore the spatial clustering features of travel activity from both Wi-Fi log data and microtransit services, the study applied density-based clustering methods to identify origin and destination hotspots.

- **Density-Based Spatial Clustering of Applications with Noise (DBSCAN)**

The goal of density-based clustering algorithms is to group points that are densely packed and label points in low-density regions as noise. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [53] achieves this goal by identifying the **core object**, **border object**, and **noise**. Assuming $X = \{x_1, x_2, \dots, x_n\}$ is a dataset containing n objects and D is a distance matrix with a dimension of $n \times n$, containing pairwise distance $d(x_p, x_q)$ where $x_p, x_q \in X$. As defined by Campello et al. [54], an object x_p is called a core object with respect to ε and m_{pts} if its ε-neighborhood contains at least m_{pts} objects. An object is called a border object if it has fewer than m_{pts} but is within ε of a core object. An object is considered noise if it is neither a core object nor a border point. As shown in Figure 14, point A is a core point since its ε-neighborhood contains 7 points, and given m_{pts} is 6. Point B is a border point since it is in the ε-neighborhood of A, but its ε-neighborhood contains 3 points. Point C is an isolated point, which is a noise point.

The cluster detection process of DBSCAN:

1. For each point or object in the dataset, we identified its ε-neighborhood and the core object. We initialized the clusters with the core object.

2. Starting from one initialized cluster or core object, we checked the objects in the ε-neighborhood and marked them as visited:

- If the neighbour is a core object, we added the neighbour and its neighbour to the cluster.
- If the neighbor is a border object, we add the neighbor to the cluster.

3. Moving on to the next unvisited core object, we continued the process in step 2 until all core objects had been visited. The objects not added to any cluster were noise.

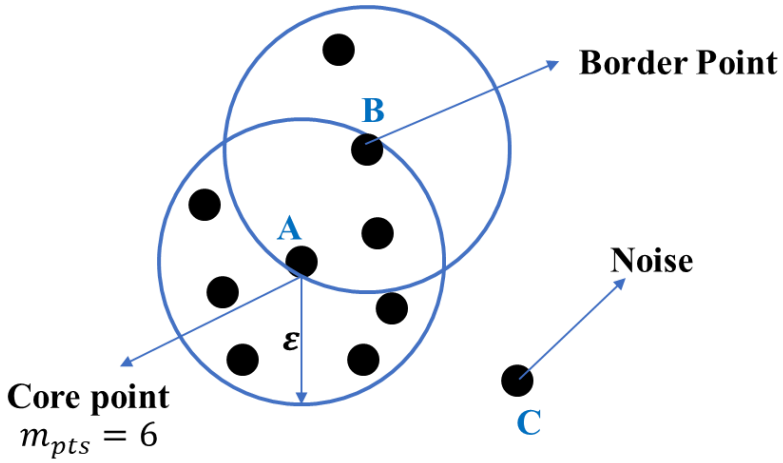


Figure 14: Core Point, Border Point, and Noise of DBSCAN

- **Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)**

DBSCAN detects clusters using uniform density features based on the parameters ε and m_{pts}. However, in the real world, the point distribution and clusters' density may be heterogeneous, which can't be well clustered by using uniform density features. Thus, HDBSCAN [54] was employed by extending the concept of core objects and reachability measures in DBSCAN to convert the space into a reachability graph, and utilizing a hierarchical structure to extract stable clusters. Campello et al. defined the following concepts, including **core distance**, **mutual reachability distance**, and **mutual reachability graph** [54]. The **core distance** of an object x_p ($d_{core}(x_p)$) with respect to m_{pts}, is the distance from x_p to its m_{pts}-nearest neighbour. The **mutual reachability distance** between x_p and x_q is defined as $d_{mreach}(x_p, x_q) = \max\{d_{core}(x_p), d_{core}(x_q), d(x_p, x_q)\}$. For example, given m_{pts} is 4, $d_{core}(x_p)$, $d_{core}(x_q)$, and $d(x_p, x_q)$ are shown in Figure *15*. By comparing the three distances, the mutual reachability distance between x_p and x_q is $d_{core}(x_q)$. The purpose of the mutual reachability distance is to discourage clustering of objects in sparse areas. Then, the **mutual reachability graph** $G_{m_{pts}}$ was

constructed where the objects of X are vertices and the weight of each edge is the mutual reachability distance.

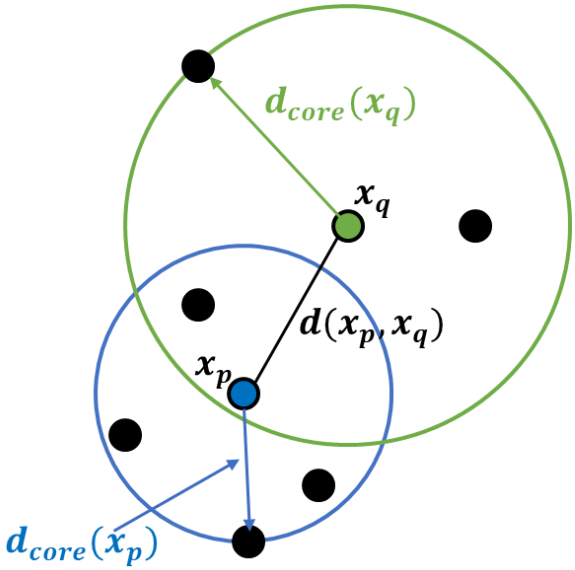


Figure 15: Core Distance and Mutual Reachability Distance in HDBSCAN

With the concepts mentioned above, the cluster detection process of HDBSCAN:

1. We built a minimum spanning tree (MST) from $G_{m_{pts}}$, to identify a hierarchy of clusters by connecting points in a way that respects both distance and density.

2. We identified the cluster hierarchy by removing edges from the MST in increasing order of mutual reachability distance. Cluster hierarchy at different distance levels was obtained.

3. To get stable clusters, we used m_{pts} to filter out small clusters and retain only branches of the hierarchy that persist for a long time. This means that the cluster should exist over a wide range of density thresholds and contain a sufficient number of data points. To achieve the result, a cluster stability measure was calculated as in the equation (6), where $\lambda = \frac{1}{\text{mutual reachability distance}}$. $\lambda_{death}(p)$ is the value of λ when point p joined cluster C; $\lambda_{birth}(p)$ is the value of λ when point p left cluster C.

$$Stability\ of\ a\ cluster\ C = \sum_{p \in C} (\lambda_{death}(p) - \lambda_{birth}(p)) \quad (6)$$

The final clusters were selected by choosing clusters that maximize the total stability.

5. We labeled the objects or points that did not belong to any cluster as noise.

4.3. Simulation Development

4.3.1. Background Traffic Estimation

To better simulate real-world microtransit scenarios, background traffic, including vehicles and pedestrians, is a necessary input to the simulation. In this project, we estimated the background traffic from the Wi-Fi detected travel flow in the city of Wilson by an augmentation approach. First, we calculated the average hourly OD flow for vehicles and pedestrians between AP locations on weekdays based on Wi-Fi-detected travel activity. Then, based on the sampling rate from the field tests in different location types, we calculated the adjusted OD flow and input it into the simulation as background traffic, as shown in the following equations. Equation (7) aggregates the OD flow from the origin i to all destinations to get the total outflow of origin i. Then the OD ratio from the origin i to destination j was calculated as in the equation (8). By applying the sampling rate from field tests, the adjusted outflow at each origin is shown in the equation (9). Finally, the adjusted OD flow between i and j was obtained by multiplying the OD ratio (assuming unchanged from Wi-Fi-based travel data to overall travel data) and the adjusted outflow, as shown in the equation (10).

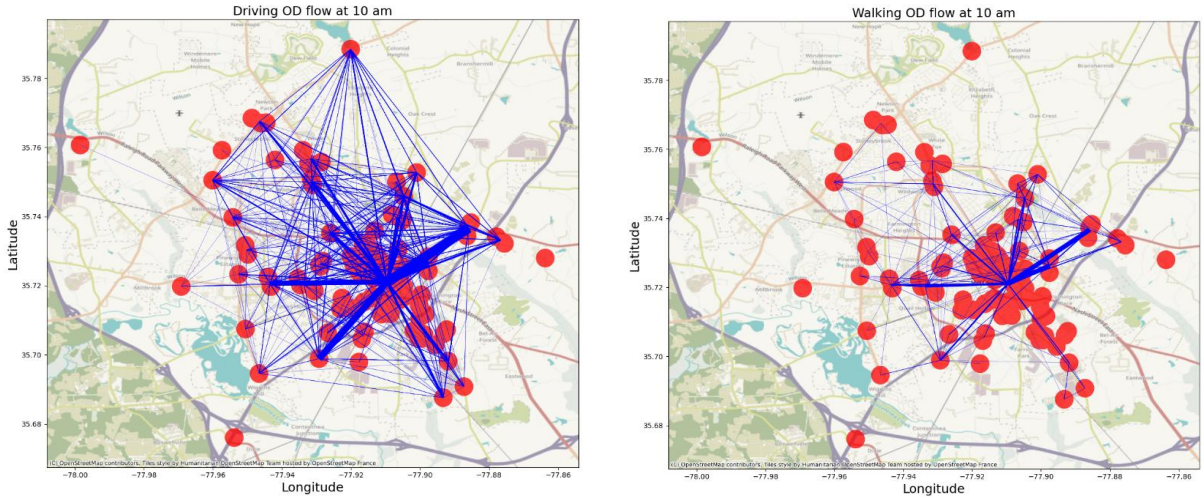
$$Outflow_i = \sum_j OD_{ij} \tag{7}$$

$$ratio_{ij} = \frac{OD_{ij}}{Outflow_i} \tag{8}$$

$$adj_{outflow_i} = \frac{Outflow_i}{Sampling\ Rate} \tag{9}$$

$$adj_{OD_{ij}} = adj_{outflow_i} \times ratio_{ij} \tag{10}$$

Access points (APs) are not distributed evenly in Wilson. Areas around government buildings have more APs, which may have better Wi-Fi signal coverage and quality to be detected by more clients. Different types of locations have different sampling rates. According to the field test, the sampling rate for locations around government buildings is 0.62; for locations around service buildings such as the library, the sampling rate is 0.55; and for other locations, the sampling rate is 0.59. Figure 16 shows the average adjusted OD flow results at 10 am for driving (a) and walking trips (b) on weekdays, with 19,721 driving trips and 15,573 walking trips.



(a) (b)

Figure 16: Adjusted OD Flow at 10 AM for (a) Driving and (b) Walking

4.3.2. Simulation Platform Construction

The study developed a microscopic simulation platform that coordinates the road network, passengers, and service vehicles to evaluate the performance of microtransit services and potential improvement solutions. The simulation platform was built on an open-source microscopic traffic simulation software package - SUMO (Simulation of Urban Mobility)[55]. By applying the built-in Traffic Control Interface (TraCI) with a Python script, the behavior of vehicles and passengers in the simulation was set and monitored. Thus, different types of microtransit operation strategies were tested and evaluated in a simulation environment. As shown in Figure 17, the simulation platform needs three components as input: 1) road network, 2) simulation settings, and 3) passenger and vehicle routing plan. The **road network** of the study area was obtained from OpenStreetMap and automatically converted into a simulation road network using SUMO tools. **Simulation settings,** such as the color, size, and type of the vehicle, were configured using a separate XML (eXtensible Markup Language) file. The **passenger and vehicle routing plan** was generated using the developed Matching and Routing Module (MRM), which took travel activity data as input. This data included microtransit requests with details such as origin, destination, ride start time, and service vehicles. The output of the simulation platform is **evaluation measurements** in terms of traffic efficiency and service performance. Traffic efficiency encompasses the total vehicle travel time and travel distance, while service performance includes passenger travel time, travel distance, and waiting time.

The working flow of the simulation platform is shown in Figure *18*. First, the service vehicle location, as well as the simulation settings, were initialized in the simulation network. After the simulation started, ride requests and available service vehicles were scanned at 5-minute intervals. It assumes that the vehicles continue to serve passengers whenever they're available. The scanned ride requests and available service vehicles were input into the MRM module to generate routing plans for passengers and vehicles using various strategies, ranging from simple (e.g., first-request-first-serve) to complex (e.g., vehicle routing optimization), which will be

explained in detail in the following paragraph. The routing plans were input into SUMO with a configured road network, and the output evaluation measurement was performed after the requests were served. The process was repeated at regular intervals (i.e., every 5 minutes) until all requests were served.

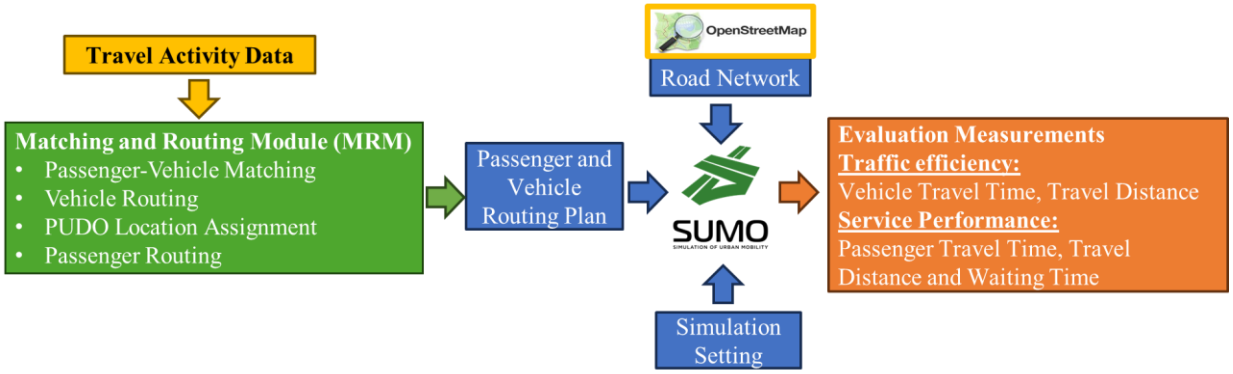


Figure 17: Component of the Microscopic Simulation Platform

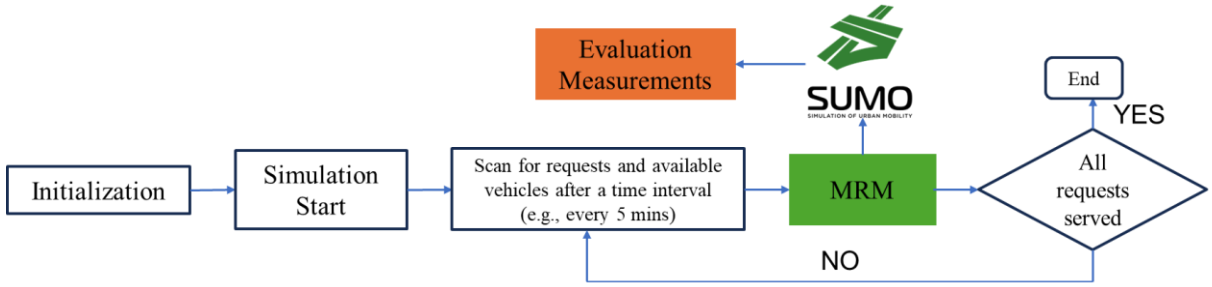


Figure 18: Working Flow of the Microscopic Simulation Platform

The purpose of MRM is to generate routing plans for passengers and vehicles in simulation, given a list of requests and the availability of service vehicles. As shown in Figure *19*, the MRM contains four parts, including 1) pickup and drop-off (PUDO) location assignment, 2) passenger-vehicle matching, 3) vehicle routing, and 4) passenger routing. Each part coordinates with different logic or algorithms for testing. Assuming that there is a list of requests $[r_1, r_2, r_3, r_4]$ waiting for service, and a list of vehicles available $[v_1, v_2, v_3]$. The passenger will first walk to a designated pick-up location. After arriving at the drop-off location, the passenger walks to the final destination. The setting of PUDO locations is within walking distance of both origins and destinations, reducing service vehicle detours [35]. The first step of MRM is assigning PUDO locations. Suppose that each origin or destination has three potential pick-up and drop-off locations. Each request can be denoted as $r_i = ([o_{i1}, o_{i2}, o_{i3}], [d_{i1}, d_{i2}, d_{i3}])$, where o stands for pick-up locations, d stand for drop-off location, and r_i is the ith request. A straightforward assignment logic is to pick locations on major or main roads that are nearest to the origins or destinations as potential PUDO locations. The second step is to match requests with the service vehicle. One simple method is to match requests to the closest vehicle. After matching, the third step is to identify appropriate PUDO locations and visiting sequences for each vehicle to serve

assigned requests efficiently. Different vehicle routing algorithms can be applied, such as the deterministic optimization method for exact optimal solutions or the heuristic methods for near-optimal solutions. Note that the previous second and current third steps can also be combined into a single optimization problem to solve. After identifying the PUDO locations for each request in the third step, the fourth step is to set the path for each passenger to walk from the origin to the pickup location and then to the destination from the drop-off location. One simple method is to use the shortest path algorithm. We generated the routing plan for passengers and vehicles in four steps.

Note that it is flexible to skip the first and fourth steps if alternative pick-up and drop-off locations are not preferred. Though the approaches in the MRM are relatively simple, due to their flexibility, they can be extended to adopt more sophisticated methods in the future.

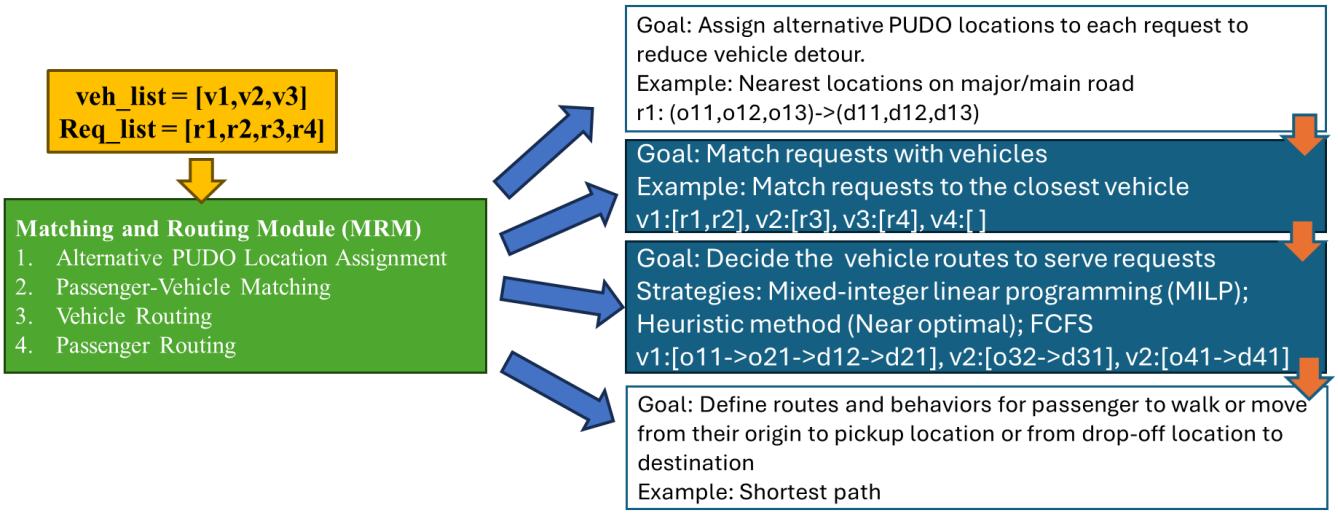


Figure 19: Matching and Routing Module

5. Experiments and Results

5.1. Wilson Wi-Fi Processing and Field Test Results

This section presents the Wi-Fi processing results and field test results in the city of Wilson, including sampling rate tests and trajectory tracking tests.

5.1.1. Wilson Wi-Fi Processing Results

Wi-Fi log data from 02/14/2024 to 02/27/2024, covered 292 APs and 21,877 clients. Among the detected clients, 46.52% had only stationary sessions or an average daily connection time greater than 16 hours. These data points were excluded. 18.62% of the clients had no stationary sessions, which may be due to clients passing by the Wi-Fi-covered area. Among the regular clients with both stationary and transitional sessions, 5.2% appear at least two days each week, with an average daily stationary duration of between 4 and 10 hours. These clients may be employees, and the remaining clients are considered residents, visitors, or non-employees. Visitors (28.81%) are clients who are detected less than twice a week or have an average daily stationary duration of less than 4 hours. Residents (0.84%) are clients detected more than twice a week and have an average daily stationary duration of more than 10 hours. Figure 20 shows the composition of the detected clients in the city of Wilson.

From the clients, excluding stationary devices and random connections, the OD flow for both driving and walking modes between locations was obtained as in Figure 21 for a typical weekday. The size of the red dot indicates the number of trips related to the location, and the width of the blue links illustrates the number of trips between the locations. It can be observed that the **travel flow radiates from the city center to suburban areas**. Strong connections can be observed around the city center, where numerous government buildings, including City Hall and the Police Department, are situated. The strongest connections are observed between the parking lot and the City Hall. Additionally, places located outside the city center, such as the bus garage and the medical center, have high travel volumes.

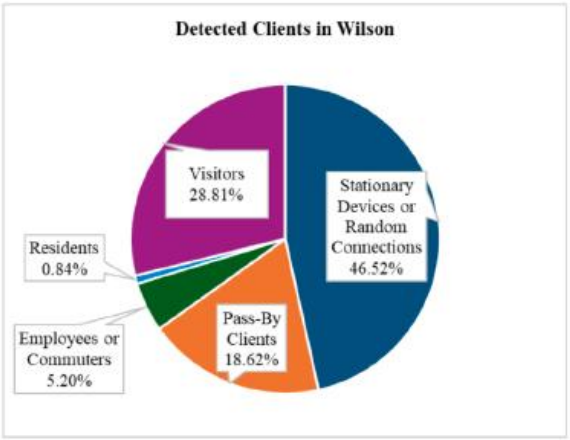


Figure 20: Detected Clients in the City of Wilson

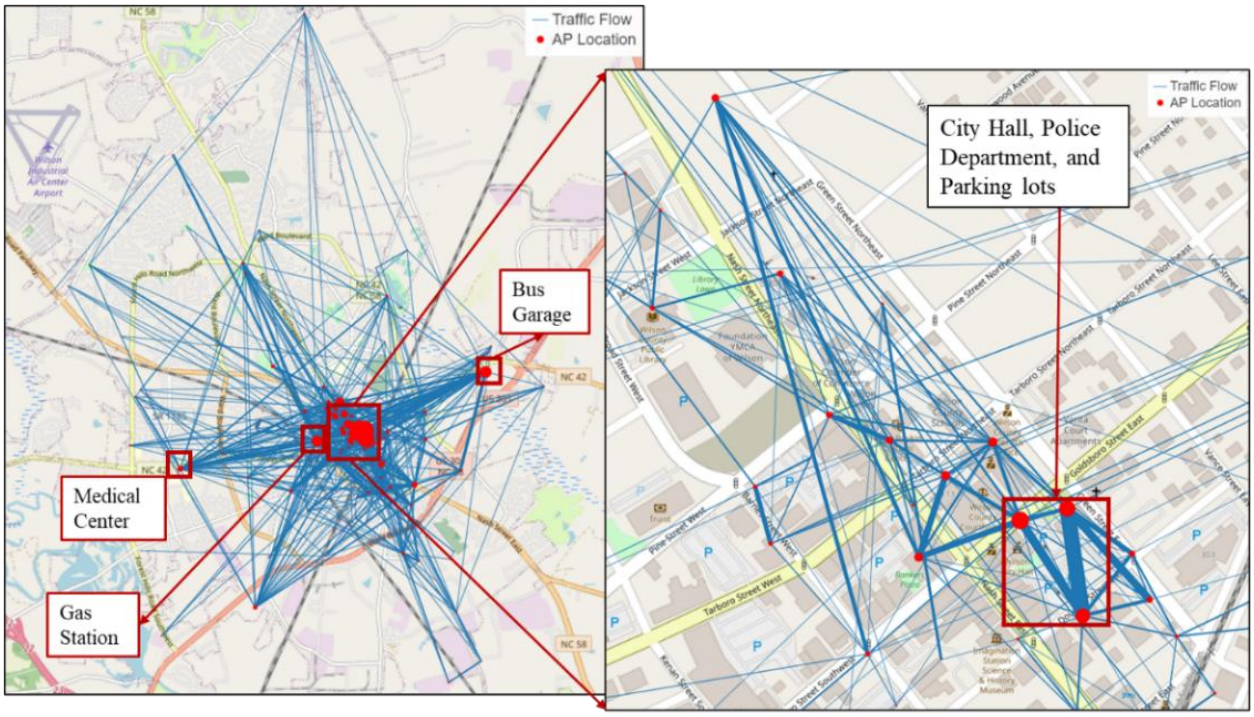
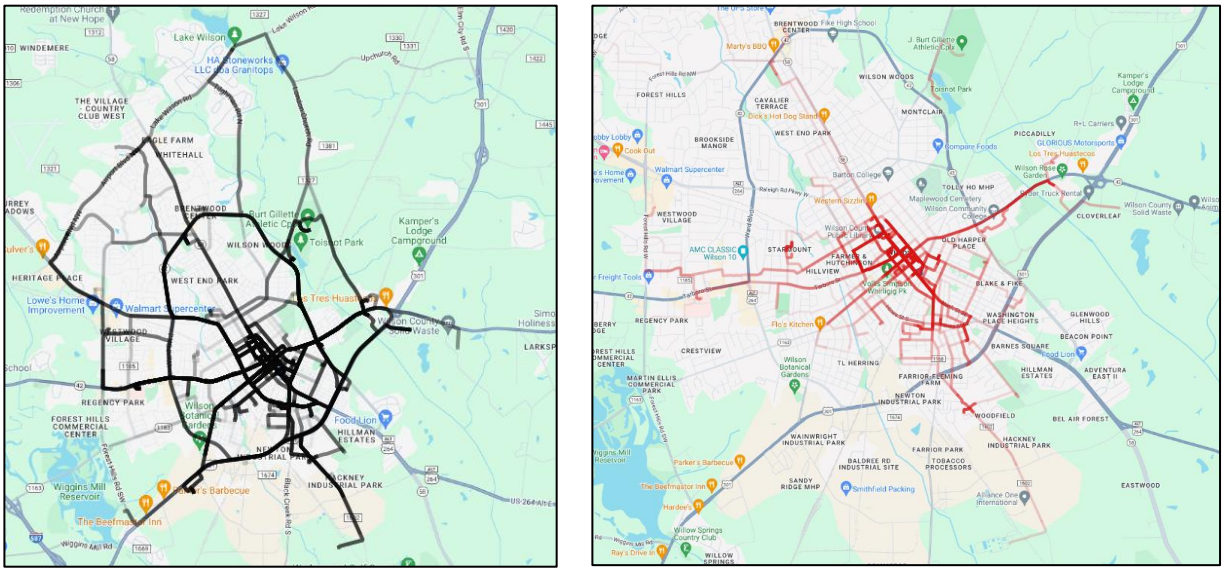


Figure 21: OD Travel Flow between AP Locations

Besides the OD flow, the detailed travel routes and related travel modes were obtained using navigation APIs (Google Directions API). Figure 22 illustrates the driving routes and walking routes on a typical weekday in the city of Wilson. In total, 2,786 driving routes and 1,400 walking routes were detected. Driving routes were mainly located around the city center and peripheral highways, while walking routes were more concentrated in the city center, as expected.



(a) (b)

Figure 22: Travel Route on a Typical Weekday for (a) Driving and (b) Walking Mode

5.1.2. Wilson Field Test Results

- **Sampling Rate Test**

The manually collected people count, and Wi-Fi-detected inflow and outflow people count, as well as the sampling rate, are shown in Table 4. It shows that the highest detection sampling rates exist at Wilson Utilities Center (WUC), WUC in session 1, which can detect around 80% of the ground truth. The sampling rates vary across different periods, dropping to around 60% in Session 3. Additionally, Wi-Fi may overestimate the number of people, as seen in session 2 at the Gig East Exchange office (GEEO). This might be because people wandering around GEEO on the roadside but not entering the building are counted twice. In the morning period (session 4), there is no overestimation, and the sampling rates range from 0.75 to 0.67. The Wilson Public Library (WPL) consistently has the lowest Wi-Fi detection sampling rate, at around 55% of the ground truth. Wilson's **overall sampling rate** (the total flow recorded of Wi-Fi divided by the total flow calculated manually) of all locations **was 0.74**. It is comparable to most literature.

Table 4. Sampling Rate Test Results

Sessions	Time Period	Location	Inflow-Wi-Fi	Outflow-Wi-Fi	Inflow-Manual	Outflow-Manual	SR1	SR2
1	13:00-17:00	WUC	70	79	89	101	0.79	0.78
2	14:00-17:00	GEEO	70	74	40	47	1.75	1.57
3	9:00-16:30	WUC	90	89	161	146	0.56	0.61
4	9:00-12:00	GEEO	41	37	55	55	0.75	0.67
5	14:00-16:30	WPL	51	54	94	97	0.54	0.56

- **Trajectory Tracking Test**

Driving Mode

The driving test results are shown in Table 5, including Wi-Fi connection and disconnection times, as well as the actual arrival and departure times for each location. The stop times (durations) of Wi-Fi and actual cases are calculated by the difference between Wi-Fi connection and disconnection time and the actual arrival and departure time, respectively. Since the actual stay durations in the test are short, the estimated visiting time detected by Wi-Fi is the average of the connection time and disconnection time. The actual visiting time, on the other hand, is the average of the actual arrival and departure times. The average difference between the estimated and actual visiting time is 1.2 min for all locations. With the estimated visiting time, the estimated travel time between the locations can be calculated and compared with the actual travel time. The Mean Absolute Percentage Error (MAPE) is 20%. The MAPE between Wi-Fi detected stop time (duration) and actual stop time (duration) is 13.4%. It shows that the Wi-Fi-detected stop time at some locations is very close to the actual stop time. However, Wi-Fi detected a much longer stop time at locations such as the Bus Garage and the Museum of African American History.

Table 5. Results of Driving Trajectory Test

Locations	Wi-Fi Log Data			Ground Truth Data		
	Connect	Disconnect	Stop Time (min)	Arrival	Departure	Stop Time (min)
City Hall	5:06:21 PM	5:07:17 PM	0.93	5:06 PM	5:07 PM	1
Greater Wilson Rotary Park	5:13:29 PM	5:15:27 PM	1.97	5:15 PM	5:17 PM	2
Wilson Times	5:23:29 PM	5:25:01 PM	1.53	5:23 PM	5:24 PM	1
Residential Area	5:31:16 PM	5:32:53 PM	1.61	5:30 PM	5:32 PM	2
Bus Garage	5:43:18 PM	5:49:53 PM	6.58	5:43 PM	5:47 PM	4
Wilson Community College	5:49:51 PM	5:53:26 PM	3.6	5:49 PM	5:52 PM	3
Museum of African American History	5:55:31 PM	6:01:25 PM	6.25	5:55 PM	5:56 PM	1
City Hall	6:01:26 PM	6:11:12 PM	9.77	6:01 PM	6:11 PM	10

Walking Mode

The walking test results are shown in Table 6. It can be observed that the average difference between the estimated and actual visiting time is 1.06 min. However, the MAPE between the Wi-Fi-detected and actual stop times is around 132%, which is significantly greater than the driving mode (although the walking mode's absolute time difference is only 1.06 minutes). One possible reason is that in the downtown area where APs were densely located, the interference between those APs and auto-connect features (e.g., blurring the disconnection time) may worsen the Wi-Fi detection accuracy. In addition, the MAPE of estimated and actual travel time between the locations is 16.7%.

Table 6. Results of the Walking Trajectory Test

Locations	Wi-Fi Log Data			Ground Truth Data		
	Connect	Disconnect	Stop Time (min)	Arrival	Departure	Stop Time (min)
Wilson Utility Center	12:51:32 PM	12:53:47 PM	2.25	12:52 PM	12:53 PM	1
City Hall	1:01:44 PM	1:04:40 PM	2.93	1:01 PM	1:02 PM	1
Cultural Center	1:06:50 PM	1:10:02 PM	3.2	1:06 PM	1:08 PM	2
Wilson County Courthouse	1:11:09 PM	1:18:19 PM	7.16	1:11 PM	1:13 PM	2
Wilson County Board Election	1:19:04 PM	1:21:38 PM	2.57	1:19 PM	1:21 PM	2
Wilson Times	1:28:22 PM	1:30:38 PM	2.26	1:29 PM	1:30 PM	1

Microtransit Mode

Table 7 shows the results of the microtransit mode tests. Unlike the driving mode test, where the tester was instructed to remain at the test location only briefly after establishing a Wi-Fi connection, the microtransit mode test required the tester to wait at the location for the service vehicle to arrive, which typically took over 10 minutes. Thus, instead of checking the difference between Wi-Fi detected visiting time and actual visiting time, the difference between the detected and actual arrival and departure times was checked for the microtransit mode test. The average difference between the Wi-Fi detected arrival and departure times and the actual arrival and departure times is approximately 42 seconds. The MAPE of the detected and actual travel time between locations is 13.4%, and the MAPE for stop time is 5.9%.

The test results are summarized in Table 8, showing the average time difference between the actual and detected visiting time for driving and walking mode tests, and the arrival/leaving time for the microtransit mode test. The MAPEs for actual and detected travel time and stop time for the three tests are also shown in Table 8, indicating that the microtransit mode test yields the best results. The reason might be that for the microtransit mode, the volunteer needed to wait at the location for the service vehicle to come, giving the smartphone enough time to establish a stable connection with APs.

Table 7. Results of the Microtransit Trajectory Test

Locations	**Wi-Fi Log Data**			**Ground Truth Data**		
	Connect	**Disconnect**	**Stop Time (min)**	**Arrival**	**Departure**	**Stop Time (min)**
Wilson Utility Center	2:31:49 PM	2:46:00 PM	14.18	2:32 PM	2:46 PM	14
Greater Wilson Rotary Park	2:53:11 PM	3:09:59 PM	16.8	2:53 PM	3:12 PM	19
Wilson Community College	3:19:41 PM	3:29:32 PM	9.85	3:20 PM	3:30 PM	10
Residential Area	3:36:14 PM	3:59:06 PM	22.87	3:38 PM	3:59 PM	21

Table 8: Summary of the Trajectory Tests

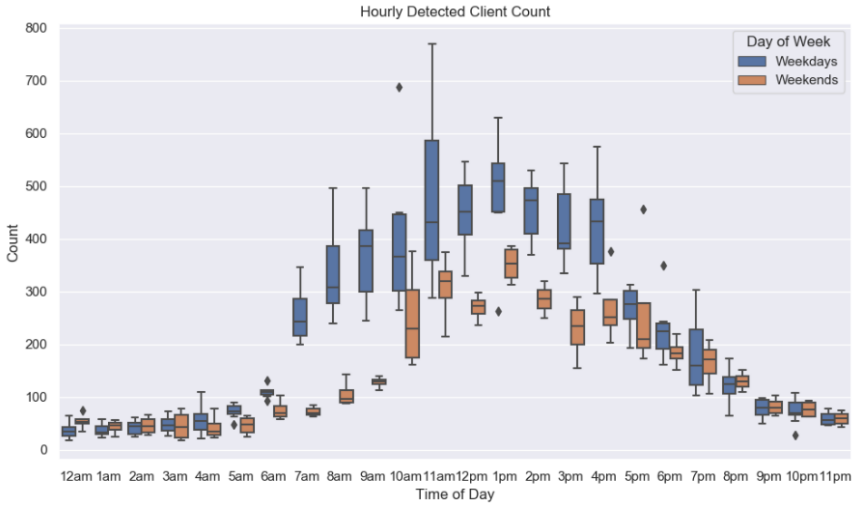
Tests	Difference in Visiting/Arrival/Leaving Time	MAPE for Travel Time	MAPE for Stop Time
Driving Mode Test	1.2 min	20%	13.40%
Walking Mode Test	1.06 min	17%	132%
Microtransit Mode Test	0.7 min	13.40%	5.90%

5.2. Wi-Fi and Microtransit Human Travel Patterns Spatial-Temporal Analysis

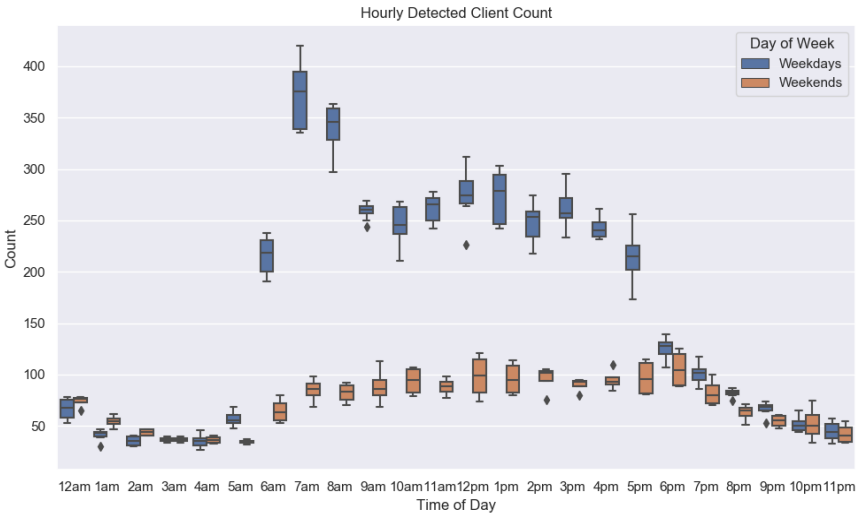
This section analyzes the travel pattern from Wi-Fi log data and microtransit data spatially and temporally.

5.2.1. Wi-Fi Detected Travel Pattern

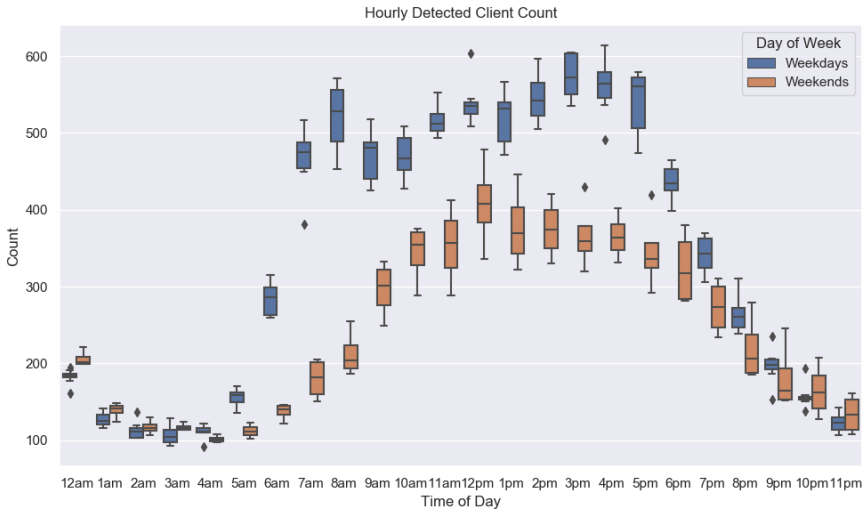
Based on the Wi-Fi connection behavior presented in Section 5.1.1, the detected clients can be classified into three categories: pass-by clients, employees, and non-employees, which include residents and visitors. **Error! Reference source not found.** Figure *23* shows the **hourly clients** detected on **weekdays and weekends** for the three types of clients. It can be observed that for pass-by clients, the hourly count exhibits more variation and does not have obvious peak hours. For employees, there is an obvious morning peak hour at around 7 am, and the count drops quickly after 5 pm. For non-employees, a morning peak is observed around 8 am, and an afternoon peak occurs around 3 pm.



(a)



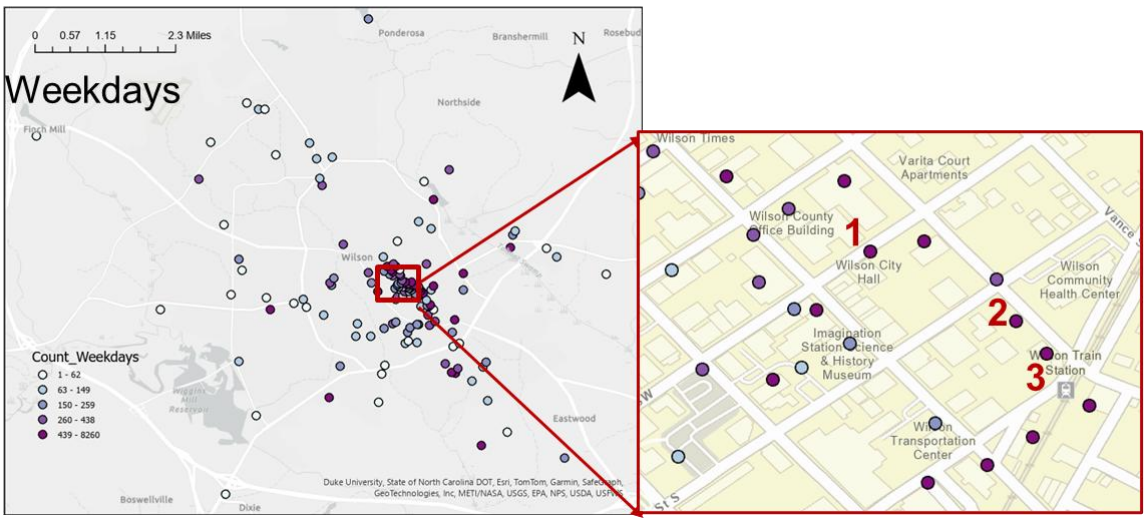
(b)



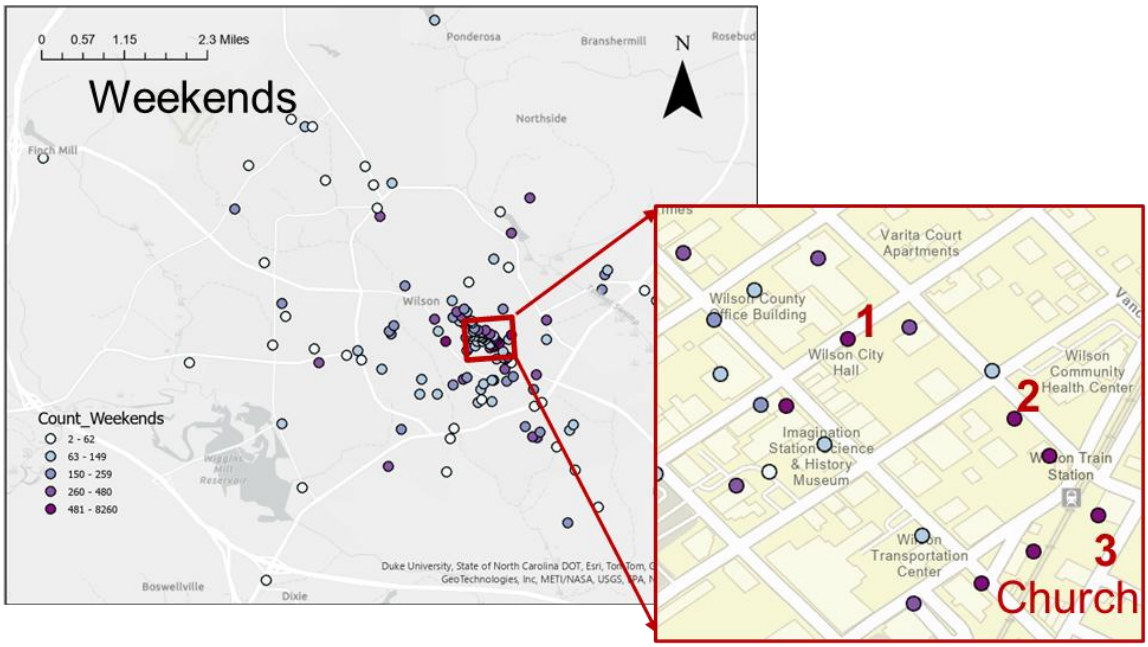
(c)

Figure 23: Hourly Number of Clients Detected on Weekdays and Weekends for (a) Pass-By Clients, (b) Employees, and (c) Non-Employees

Spatially, Figure 24 illustrates the **average daily client count** detected on weekdays and weekends. Generally, there are more clients detected on weekdays than on weekends. The highest number of clients is detected in the city center. The busiest location on both weekdays and weekends is **Wilson City Hall**, where 8,260 clients are detected on weekdays and 3,370 clients are detected on weekends. The second busiest location, on both weekends and weekdays, is the **Wilson Community Health Center**. A church located near the city center is identified as the third busiest location on weekends, with 1,101 clients recorded.



(a)



(b)

Figure 24: Average Daily Client Count Detected on (a) Weekdays and (b) Weekends

5.2.2. Microtransit Travel Pattern

- Basic summary

Wilson's microtransit data provides information from the following perspectives: demand and supply, service performance, trip-related information, transportation equity, and emerging mobility, as shown in Table *9*. The findings are listed below:

- 42.5% more ride requests were observed on weekdays than on weekends.
- Both the cancellation rate and completion rate are higher on weekdays than on weekends.
- Actual waiting time is higher on weekdays and also higher than the estimated waiting time given by the Via App.
- The rider tends to cancel the request if the estimated waiting time is around 30 minutes on weekdays or 20 minutes on Saturdays.
- Before confirming cancellation, riders usually have already waited for around 18 minutes on weekdays or 12 minutes on Saturdays.
- The ride-sharing rate is higher on weekdays, and there are more requests on weekdays.

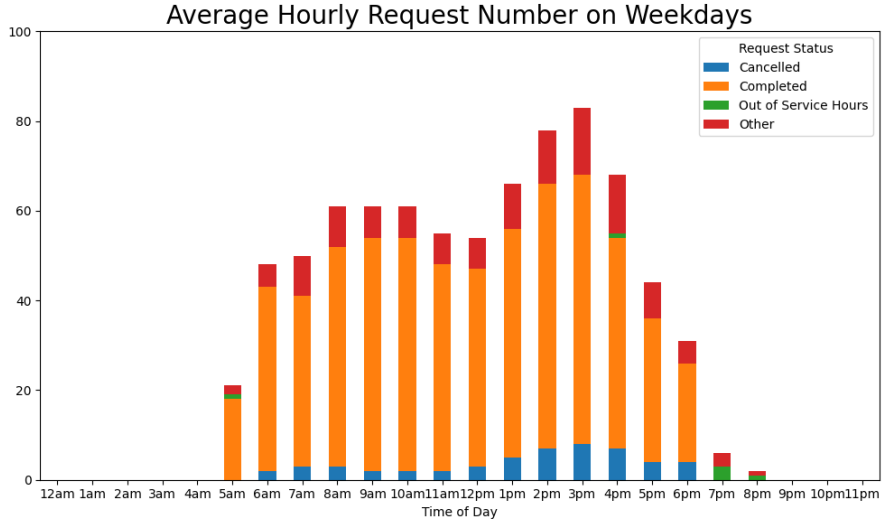
Table 9: Basic Statistic Analysis for Microtransit Data

Category	Attributes	Weekdays	Saturdays
Demand and Supply	Average number of requests per day	938	659

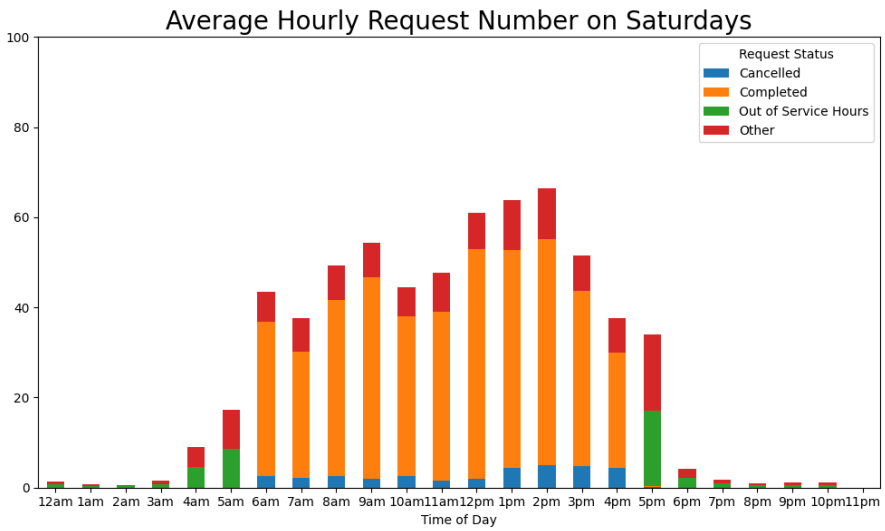
	Number of vehicles in operation per day	17	17
	Cancellation rate	7.00%	5.50%
	Completion rate	71.30%	69.90%
Service Performance	Estimated waiting time	22.2 min	15.5 min
	Actual waiting time	25.1 min	17.4min
	Average estimated waiting time for cancelled requests	29.4min	21.7min
	Average actual waiting time for cancelled requests	17.9min	12.0 min
	Drop-off delay	5.9min	3.7min
Trip Information	Pick up walking distance	87.4 meters	86.2 meters
	Prop off walking distance	89.2 meters	88.3 meters
	Ride distance	3.6 mile	3.2 mile
	Ride duration	10.8 min	9.3 min
Transportation Equity	Wheelchairs pct	5.80%	5.90%
Emerging Mobility	Ride-sharing rate	55.80%	37.80%

- Temporal pattern

From the microtransit data, the **average hourly number of requests** can be obtained for weekdays and Saturdays, as shown in Figure *25***Error! Reference source not found.**. Note that microtransit doesn't operate on Sundays. It shows that demand is high during morning peak hours, from 8:00 am to 10:00 am on weekdays, and peaks around 3:00 pm, marking the afternoon peak hour. On Saturdays, the demand peaks at 9 am and 2 pm. The demand in the afternoon is higher than the demand in the morning. Additionally, the blue bars in the diagram indicate that more requests are cancelled as demand increases.



(a)



(b)

Figure 25: Average Hourly Request Number on (a) Weekdays and (b) Weekends

- Spatial pattern

By applying the HDBSCAN clustering method on the origin and destination (OD) of microtransit trips, the hotspots can be obtained as convex polygons, shown in Figure *26*. Several hotspots are observed around **commercial areas**, **recreational fields**, **public institutions**, and **residential areas** close to the city center.

Additionally, the spatial correlation between hotspots and the poverty ratio is analyzed. The green color showing the **poverty ratio** in each census block group is calculated by the equation (11). IPR represents the income-to-poverty ratio, which can be obtained from the American Community Survey (ACS) [44]. The poverty ratio range is from 0 to 1, where a value closer to 1 indicates a higher proportion of people with incomes below the poverty threshold. Figure *26* shows that residential areas with a high poverty ratio are a hotspot for the microtransit service. Additionally, trips from such residential areas are primarily connected to the downtown and commercial areas.

$$poverty\ ratio = \frac{population\ with\ IPR < 1}{population} \tag{11}$$

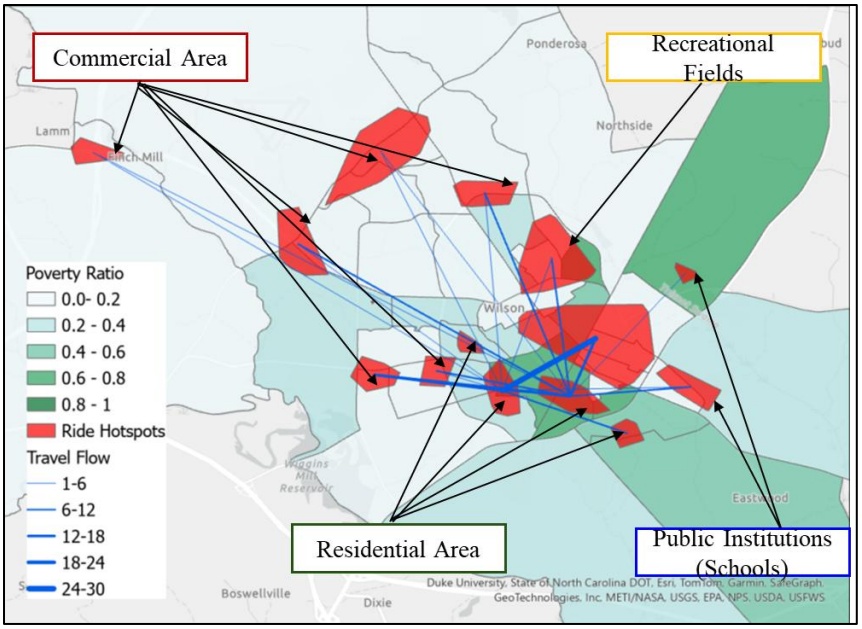


Figure 26: Microtransit Hotspots, Hotspot Related Travel Flow and Poverty Ratio

5.3. Travel Pattern Comparison and Potential Transit Enhancement Solutions

This section compares the hotspots from Wi-Fi detected travel activity and microtransit trips. The potential microtransit riders who use the Wi-Fi service are identified and analyzed to determine potential solutions for transit enhancements.

Hotspots of origins and destinations (ODs) from both datasets were identified using various density-based clustering methods that utilized different features of Wi-Fi-based travel activity and microtransit trips. As shown in Table 10, DBSCAN was applied to Wi-Fi data, and HDBSCAN was applied to microtransit data. This is because APs in Wi-Fi detection as origins or destinations have a fixed signal range, which fits the fixed parameter ε requirement in DBSCAN. On the other hand, microtransit ODs may have various densities in different areas in Wilson, which requires a hierarchical structure in HDBSCAN to capture the variations. With the parameters in Table 10, the OD hotspots from Wi-Fi and microtransit datasets can be detected.

Table 10: Cluster Parameters and Result Number of Clusters

	Wi-Fi Origin	Wi-Fi Destination	Ride Origin	Ride Destination
Clustering method	DBSCAN		HDBSCAN	
MinPts	18	19	13	13
ε	300 meters		None	
n_clusters	17	16	18	17

Figure 27 shows the common and unique OD hotspots from Wi-Fi and microtransit on typical weekdays. Figure 27 (a) and (b) shows that the major common OD hotspots (represented by Wi-Fi AP clusters) are the city center and surrounding residential areas, as well as the Wilson Community College and Wilson Medical Center. Figure 27 (c) and (d) shows the OD hotspots that are unique for the microtransit service (red) and Wi-Fi (green). It can be observed that unique OD hotspots for the microtransit service are located around residential areas and commercial areas, such as grocery stores, shopping centers, and restaurants. The unique OD hotspots for Wi-Fi service are located around parks and some government buildings, such as the Wilson operation center. The origin and destination hotspots don't have a significant difference.

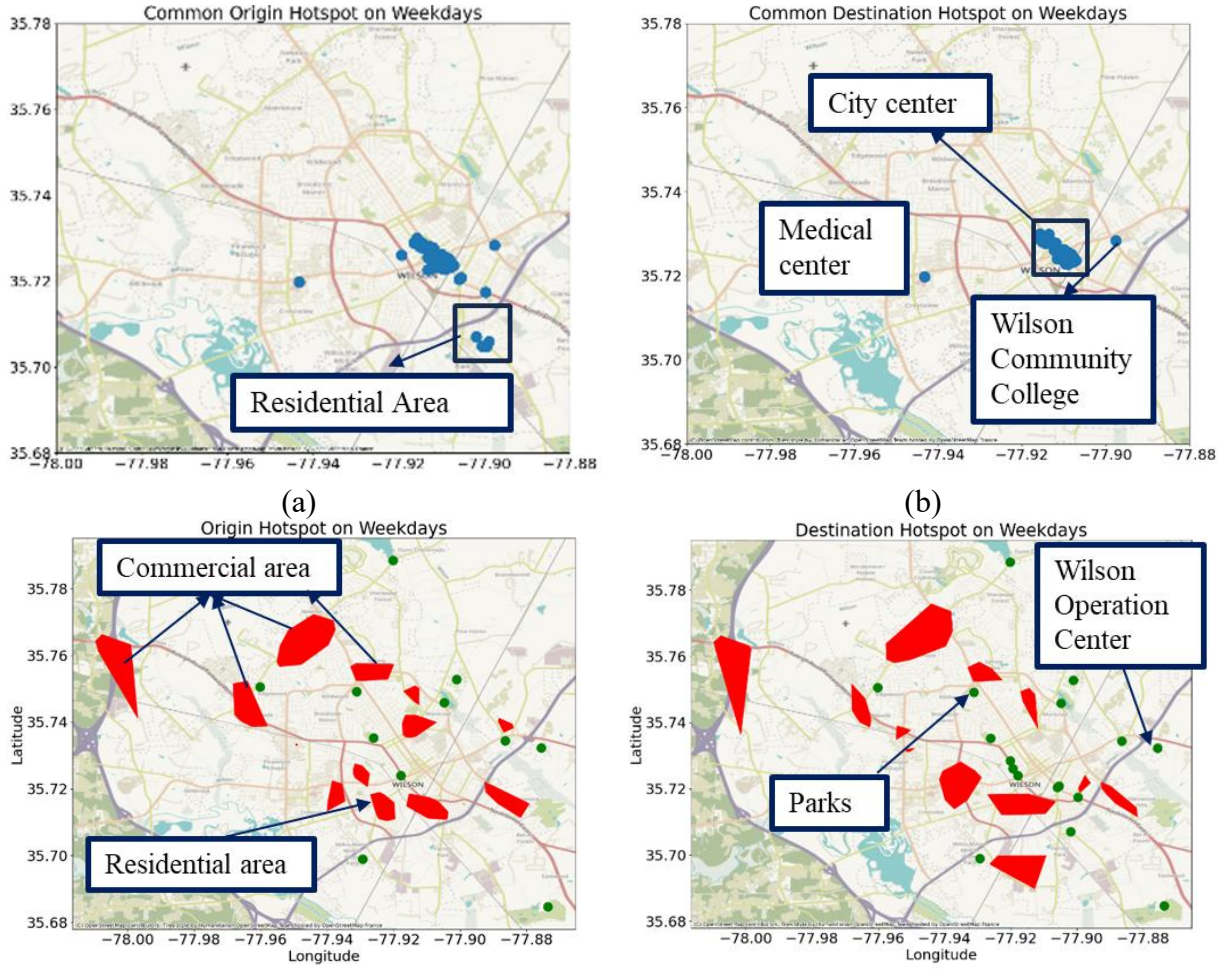
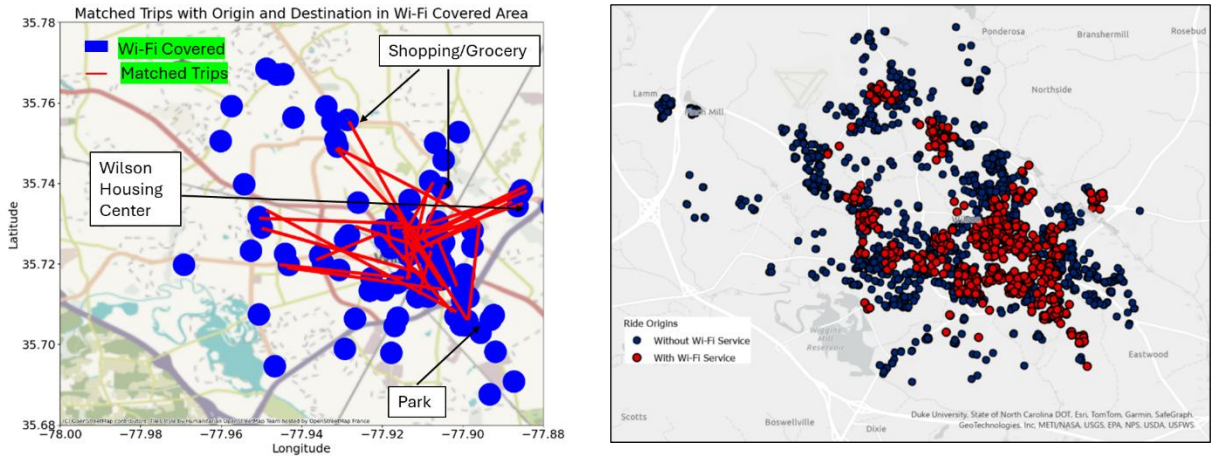


Figure 27: Comparison of Wi-Fi and Microtransit Hotspots: (a) Common Origin Hotspot on Weekdays; (b) Common Destination Hotspot on Weekdays; (c) Unique Origin Hotspot on Weekdays; (d) Unique Destination Hotspot on Weekdays

Aside from the common and unique hotspots for Wi-Fi and microtransit services, **potential microtransit trips that utilize Wi-Fi services were also examined**. Figure *28*(a) shows the matched trips, which are the microtransit trips also detected by Wi-Fi. The matched trip is identified if it satisfies two conditions: 1) The microtransit trip starts and ends within the Wi-Fi

covered area, assuming that the AP signal range is 300 meters; 2) The time difference between the start time of the microtransit trip and Wi-Fi detected trip should be within 30 minutes, as well as the end time. 32 trips are matched out of the 5056 microtransit trips from February 15, 2024, to February 6, 2024, which represents 0.63%. Figure *28*(b) illustrates the microtransit trips originating within and outside the Wi-Fi service area, assuming an AP signal range of 300 meters. The red dots represent the origins within the Wi-Fi service area, while the dark blue dots denote the origins outside the Wi-Fi service area. 2,060 microtransit trips start within the Wi-Fi service area, meaning that 40.7% of the users have the potential to use the public Wi-Fi service to place the ride requests. The significant difference between the existing matched trips (0.63%) and the potential Wi-Fi serviced trips (40.7%) suggests a substantial opportunity for shifting microtransit users to utilize free Wi-Fi service rather than their phone data plan, which could have a socioeconomic impact.



(a) (b)

Figure 28: Potential Microtransit Trips Using Wi-Fi service: (a) Matched Trips; (b) Microtransit Trips with Origins with and without Potential Wi-Fi Service

5.4. Simulation Analysis and Results

This section presents the testing results of the proposed simulation platform, which aims to reproduce real-world scenarios and evaluate different vehicle operation logics.

5.4.1. Real-world Scenario Reproduction

Real-world scenarios are recreated in the simulation by utilizing actual microtransit data, including origin, destination, and request time (O-D-T), as well as the walking distance associated with each request. Figure *29* illustrates a simulation with a road network, background traffic, service vehicles, and passengers. Wilson's road network was obtained from OpenStreetMap. The background vehicle and pedestrians in yellow were generated according to the Wi-Fi detected OD flow in driving and walking modes. Passengers in blue and service vehicles in red, along with their locations and routes, were generated from the microtransit data.

Figure 30 shows an example of the passengers' pickup and drop-off sequence of a service vehicle. The step is the timestamp in the simulation, starting from 0, and was converted from the actual request time in the Time column. The edge represents the origin or destination road link in the simulation network for each passenger, which was converted from the actual origin and destination coordinates. Routes between the origin and destination edges in the network can be generated by the shortest path algorithms. To mimic the walking behavior of passengers, pickup and drop-off locations were selected as the one-step downstream or upstream edges of the real origin or destination edges along the shortest path according to real walking distance, as shown in Figure 31. Other settings of the simulation are described below:

1. Passengers appear at their original edges when making the service requests.
2. Passengers walk to the pickup edges and wait for the vehicle.
3. Vehicles stop at the pickup edges for a fixed time interval (90 seconds after calibration) to pick up the passengers.
4. After delivery, vehicles wait at the drop-off locations and stop until receiving the subsequent request.

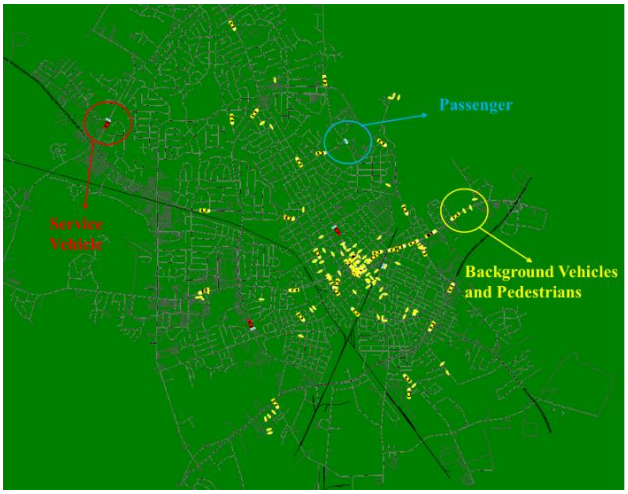


Figure 29: Road Network, Background Traffic (yellow dots - people, yellow vehicles - cars), Service Vehicle (red vehicles), and Passengers (blue dots)

Vehicle ID	Rider ID	Time	step	edge	action
148538	3146723	2024-01-02 09:19:11	909.00000	18930613#0	pu
148538	3146723	2024-01-02 09:23:56	1194.00000	761403822#3	do
148538	2925476	2024-01-02 09:41:42	2260.00000	18927607#4	pu
148538	3588244	2024-01-02 09:44:14	2412.00000	-18931024#0	pu
148538	2925476	2024-01-02 09:47:18	2596.00000	666827321#0	do
148538	3353756	2024-01-02 09:49:09	2707.00000	-18927479#7	pu
148538	3588244	2024-01-02 09:51:32	2850.00000	-18928156#0	do
148538	3520946	2024-01-02 09:56:47	3165.00000	-18924810#0	pu
148538	3353756	2024-01-02 09:59:31	3329.00000	-18929555#1	do
148538	3520946	2024-01-02 10:05:17	3675.00000	-18927603	do
148538	3112795	2024-01-02 10:08:08	3846.00000	18924943#0	pu
148538	3112795	2024-01-02 10:17:07	4385.00000	-18927052#3	do

Figure 30: Service Vehicle PUDO Information from Microtransit Data

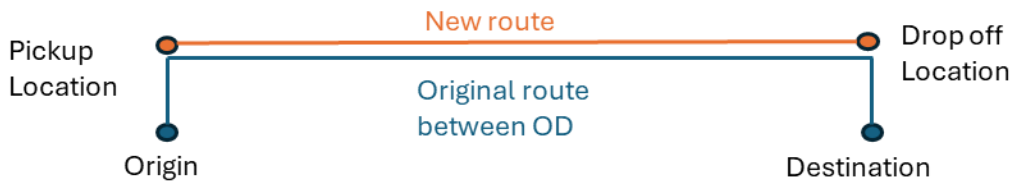


Figure 31: Selection of PUDO Locations

Using those settings, a simulation scenario was constructed according to the microtransit data from 9:00 a.m. to 10:00 a.m. on January 2, 2024, which included 15 service vehicles and 62 requests. Table 11 shows the comparison of results between the simulation scenario and the ground truth (learned from the data) in terms of passenger waiting time, ride time, ride distance, pickup time, and drop-off time. Though the mean absolute percentage error (MAPE) for passenger waiting time and ride time is 15.6% and 20.7% respectively, which is calculated by comparing the simulated waiting time and ride time of each request with the corresponding real-world waiting time and ride time. The mean error is around 0.16 minutes, which is relatively low. The MAPE of ride distance is 25.7%, which may be improved by a more advanced vehicle routing method and a more up-to-date road network. The error between the simulated and real-world pick-up and drop-off times is small, within 10 seconds. Generally, the simulation can efficiently and accurately capture a real-world scenario.

Table 11: Accuracy of Simulation Scenario Compared with Ground Truth

	Passenger Waiting Time	Ride Time	Ride Distance	Pick-up Time	Drop-off Time
Ground truth mean	13.3 min	9.32 min	3.33 mi	-	-
Simulated mean	**13.15 min**	**9.41 min**	**3.59 mi**	-	-
MAPE	**15.60%**	**20.70%**	**25.70%**	-	-
Mean Error (ME)	**-0.16 min**	**0.16 min**	**0.25 miles**	**-0.14min**	**-0.03 min**

5.4.2. Testing and Evaluation of MRM Logics

Given the 15 vehicles and 62 requests, two Matching and Routing Module (MRM) logics were evaluated. Note that the two logics serve as examples to validate the simulation platform's ability for testing and evaluation, rather than aiming to yield optimized results. Table 12 provides further details on the logic. Both logics assign four alternative pick-up and drop-off locations for each origin and destination, respectively, which are the four closest edges located near the origin or destination. The shortest path is used to generate the passenger walking routes from the origin to the pick-up location and from the drop-off location to the destination. In **MRM logic 1**, the passenger-vehicle matching and vehicle routing problems were combined and solved using mixed-integer linear programming (MILP) to minimize the total vehicle miles traveled (VMT). The MILP was constrained by vehicle capacity (set to 5, as it is the capacity of the microtransit minivan). Additionally, the service order is set to first pick up all matched requests before

dropping off the passengers (i.e., pick up all first and drop off all later), to simplify the optimization problem for testing purposes. The matching result and the order of service (both pickup and drop-off) are obtained by solving the MILP. **MRM logic 2** matched the ride requests with the spatially closest service vehicle and used a heuristic method, considering vehicle capacity to generate a near-optimal vehicle routing solution. It may not provide the minimum total vehicle miles traveled, and there is no additional constraint for pick-up and drop-off orders.

Table 12: Details of MRM Logics

<table>
<tr><th></th><th>MRM Logic1</th><th>MRM Logic 2</th></tr>
<tr><td>PUDO Location Assignment</td><td colspan="2">4 nearest locations downstream of the origin; 4 nearest locations upstream of the destination</td></tr>
<tr><td>Passenger-Vehicle Matching</td><td rowspan="2">Optimization approaches: MILP to minimize total VMT; Constraints: Vehicle Capacity, Pick-up First</td><td>Match requests with the closest vehicle</td></tr>
<tr><td>Vehicle Routing</td><td>Heuristic vehicle routing method considering vehicle capacity</td></tr>
<tr><td>Passenger Routing</td><td colspan="2">Shortest paths</td></tr>
</table>

Two MRM logics were then implemented, and the resulting simulation performance was evaluated, as shown in Table 13. It can be observed that **Logic 1** performs better in terms of total VMT, as this is the objective of the optimization problem. **Logic 2** yields better service performance, with shorter walking distances, riding distances, and riding times, as well as reduced passenger waiting times, since the vehicle can pick up and drop off passengers simultaneously to minimize waiting time.

Table 13: Evaluation of MRM Logics

	Average Passenger Walking Distance	Average Passenger Riding Distance	Average Passenger Riding Time	Average Passenger Waiting Time	Total Vehicle Miles Traveled
Logic1	671 meters	3.52 miles	5.5 mins	9.5 mins	**121** miles
Logic2	**600 meters**	**3.22 miles**	**5 mins**	**7.2 mins**	133 miles

6. Conclusions and Future Work

This project aims to identify potential system performance improvement solutions for microtransit services by analyzing comprehensive datasets, including Wi-Fi and microtransit data, in the city of Wilson, North Carolina. Wi-Fi log data was provided by the city of Wilson and processed to help understand the city-level human travel pattern. A Wi-Fi log data processing framework was proposed to obtain human travel activity, including trip origin, destination, travel routes, and travel modes, from Wi-Fi log data. The microtransit service was launched in Wilson in 2020 and is operated by the mobile application Via, which can record the details of requests and trip information.

After processing these datasets, a spatial-temporal analysis was applied to explore travel patterns derived from Wi-Fi log data and microtransit data. By comparing the travel patterns from these two datasets, particularly the origin and destination hotspots, potential service improvement solutions can be identified. To further evaluate the potential solutions, a microscopic microtransit simulation platform was developed, integrating background traffic, pedestrian behavior, and service vehicle operation. The key findings of the study are described below:

- **Wi-Fi log data processing and analysis**
  1) A navigation API-based Wi-Fi log data processing framework was developed to generate human travel activity.

  2) From February 14, 2024, to February 27, 2024, a total of 21,877 unique clients were detected, with 46.52% of the clients identified as stationary devices or random connections. The remaining clients, including those involved in travel activities, can be classified into three categories: **visitors, employees,** and **non-employees**. As expected, the peak hours for employees are 7:00 a.m., and the peak hours for non-employees are 8:00 a.m. and 3:00 p.m. More travel flows are observed around the city center and between specific locations such as the bus garage and the medical center.

  3) A series of field tests was conducted in the city of Wilson to evaluate the accuracy of Wi-Fi detection, including sampling rate tests and trajectory tracking tests. The tests indicated that the overall **sampling rate** in the study area is 75%. For the **trajectory tracking tests**, Wi-Fi can accurately detect the visited location sequence. For instance, the MAPE for detected travel time between visited locations is 20%, 16.7%, and 13.4% for driving, walking, and microtransit modes, respectively.

- **Microtransit travel pattern analysis**

  1) The microtransit data indicates that the average passenger waiting time for a ride is 29.4 minutes on weekdays and 21.7 minutes on weekends (i.e., Saturdays).

  2) The demand for the microtransit service is higher in the afternoon than in the morning and usually peaks at 3 pm. The origin and destination hotspots are generally around the city center, the commercial areas, some recreational fields, and residential areas with a relatively high poverty rate.

- **Potential improvement solution identification**

  1) The common OD hotspot for Wi-Fi travel activity and microtransit trips is the city center and surrounding low-income residential areas, where they have sufficient APs and a quality Wi-Fi signal. The unique hotspots for microtransit trips are the commercial areas, and the unique hotspots for Wi-Fi trips are recreational parks and government buildings.

  2) Under current Wi-Fi AP settings, 40.7% of microtransit trips can potentially access Wi-Fi when a service request is placed. However, actual Wi-Fi and microtransit common trips are

few, accounting for only 0.63% of the total trips, indicating a vast potential for using public Wi-Fi to request microtransit services.

3) Based on the comparison of Wi-Fi and microtransit travel patterns, the following recommendations can be made:

- More Wi-Fi APs should be installed around the commercial area (grocery centers, shopping centers, and restaurants) to serve more users and travel activity.
- A fixed-route bus line may be reinstated to connect low-income residential areas and commercial areas, serving the concentrated travel flow between them to improve the efficiency of the mobility system.

- **Network performance and benefit evaluation**

  1) A microscopic simulation platform was built on SUMO, utilizing Wi-Fi-based human travel activities as background traffic. The simulation platform aims to replicate real-world scenarios, including pedestrian behavior and microtransit services, based on microtransit data. The mean error between the simulated and real passenger waiting times and ride times is -0.16 minutes and 0.16 minutes, respectively.

  2) An MRM module interface was developed for the simulation platform, which can integrate different passenger-vehicle matching, vehicle routing, PUDO location assignment, and passenger routing logic. Two MRM logics, an optimization method for minimizing total VMT and a heuristic method based on rider and vehicle distance, were tested. The performance can be evaluated in terms of traffic efficiency and service performance. Traffic efficiency encompasses both vehicle travel time and travel distance. Service performance includes passenger travel distance, travel time, and waiting time.

Future work could include the further development of the microtransit simulation platform as a compatible tool for real-world applications, by integrating more road networks and MRM logics to test different types of scenarios, services, and mobility improvement solutions. Additionally, the proposed methods, including Wi-Fi log data processing, analysis, and comparison of Wi-Fi and microtransit travel patterns, as well as microtransit simulation methods, can be extended to other cities to provide meaningful improvements and evaluation results, thereby enhancing the efficiency of urban transportation systems.

7. Implementation and Technology Transfer Plan

The research products of the project include: 1) a public Wi-Fi log data processing and analysis approach to process raw Wi-Fi log data and model city-level human travel demand and mobility patterns; 2) analyzing on-demand microtransit trip data (i.e., OD pairs) and safety data in Wilson, and comparing Wi-Fi human travel demand, microtransit ride, and safety factors to identify

potential microtransit service improvements; 3) developing a microtransit simulation tool for evaluating microtransit service improvement solutions.

NCDOT's Integrated Mobility Division (IMD) and the city of Wilson, NC, will benefit from this research outcome. Other NCDOT units, such as the Transportation Planning Division and cities with public Wi-Fi service, like Holly Springs, will also benefit from this research for their existing or future microtransit deployments.

The project explored public Wi-Fi log data to understand city-level travel demand for enhancing mobility services, such as microtransit in North Carolina cities. This is the first-of-this-kind analysis based on public Wi-Fi log data and microtransit operation data to identify microtransit service improvement opportunities in a real city - Wilson, NC. The approaches and analysis results serve as a potential and efficient solution and a benchmark for supporting the cities that provide public Wi-Fi service to improve existing microtransit or deploy new and advanced transit services through understanding the city-level travel demand and safety patterns. This project will greatly help the NCDOT Integrated Mobility Division (IMD)'s mission and city transit agencies with respect to the increase in operational efficiency, time savings, and safety, as well as the development and improvement of specifications or guidelines for microtransit and ultimately advanced transit services. Also, the proposed modeling and analysis framework creates considerable collaboration opportunities for NCDOT. On the other hand, the project required a variety of data. Greenlight provided free public Wi-Fi service and Wi-Fi log data. Also, microtransit data was offered by the city of Wilson's RIDE program. Other data, including road network, demographics, and population (from Census) can be obtained from NCDOT and the city of Wilson, or downloaded from the Internet, such as Open Street Map (OSM). Therefore, the cost of processing and management of data is minimal compared with the significant benefits of the project.

This technical report can be used in a webinar to introduce the major outcomes and products of the project to train users who are interested in public Wi-Fi human mobility and microtransit services. The major findings of technology and knowledge have been presented in multiple professional conferences, such as the Transportation Research Board (TRB), ASCE ICTD, and NCSITE annual meetings. A journal paper, "A hierarchical Wi-Fi log data processing framework for human mobility analysis in multiple real-world communities," has been published in Travel Behaviour and Society. Further research outcomes and findings from the project are planned for publication in journals and conference proceedings.

Reference

1. USDOT. *Shared Micromobility & Microtransit*. 2025; Available from: https://www.transportation.gov/sites/dot.gov/files/2025-01/Shared%20Micromobility%20&%20Microtransit.pdf.
2. Metro. *Metro | Bus, Rail, Subway, Bike & Micro in Los Angeles*. Available from: https://www.metro.net/.
3. COTA. *COTA: Central Ohio Transit Authority*. Available from: https://cota.com/.
4. Metro. Available from: https://kingcounty.gov/en/dept/metro.
5. Transportation, V. *What is microtransit*. 2020; Available from: https://ridewithvia.com/resources/what-is-microtransit.
6. Transportation, N.C.D.o., *Transforming Public Transit with a Rural On-Demand Microtransit Project* 2023.
7. Yuan, Y., L. Zhu, and M. Joshi, *A hierarchical Wi-Fi log data processing framework for human mobility analysis in multiple real-world communities.* Travel Behaviour and Society, 2025. **39**: p. 100985.
8. Zakaria, C., et al., *Analyzing the impact of Covid-19 control policies on campus occupancy and mobility via wifi sensing.* ACM Transactions on Spatial Algorithms and Systems (TSAS), 2022. **8**(3): p. 1-26.
9. Traunmueller, M.W., et al., *Digital footprints: Using WiFi probe and locational data to analyze human mobility trajectories in cities.* Computers, Environment and Urban Systems, 2018. **72**: p. 4-12.
10. Sapiezynski, P., et al., *Tracking human mobility using wifi signals.* PloS one, 2015. **10**(7): p. e0130824.
11. Huang, H., et al., *Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions.* Computers, Environment and Urban Systems, 2021. **90**: p. 101712.
12. Hidayat, A., S. Terabe, and H. Yaginuma, *WiFi scanner technologies for obtaining travel data about circulator bus passengers: Case study in Obuse, Nagano prefecture, Japan.* Transportation Research Record, 2018. **2672**(45): p. 45-54.
13. Jahromi, K.K., F. Meneses, and A. Moreira. *Impact of ping-pong events on connectivity properties of node encounters*. in *2014 7th IFIP Wireless and Mobile Networking Conference (WMNC)*. 2014. IEEE.
14. Swain, V.D., et al., *Leveraging WiFi network logs to infer social interactions: A case study of academic performance and student behavior.* arXiv preprint arXiv:2005.11228, 2020.
15. Mohottige, I.P., et al. *Role of campus WiFi infrastructure for occupancy monitoring in a large university*. in *2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*. 2018. IEEE.
16. Zhang, S., B. Deng, and D. Yang, *CrowdTelescope: Wi-Fi-positioning-based multi-grained spatiotemporal crowd flow prediction for smart campus.* CCF Transactions on Pervasive Computing and Interaction, 2023. **5**(1): p. 31-44.
17. Trivedi, A., et al., *Wifitrace: Network-based contact tracing for infectious diseases using passive wifi sensing.* Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2021. **5**(1): p. 1-26.
18. Trivedi, A., *Human Mobility Monitoring using WiFi: Analysis, Modeling, and Applications.* 2021.

19. Soundararaj, B., J. Cheshire, and P. Longley, *Estimating real-time high-street footfall from Wi-Fi probe requests.* International Journal of Geographical Information Science, 2020. **34**(2): p. 325-343.
20. Hang, M., I. Pytlarz, and J. Neville. *Exploring student check-in behavior for improved point-of-interest prediction*. in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018.
21. Trivedi, A., et al. *WiFiMod: Transformer-based Indoor Human Mobility Modeling using Passive Sensing*. in *ACM SIGCAS Conference on Computing and Sustainable Societies*. 2021.
22. Marakkalage, S.H., et al., *WiFi fingerprint clustering for urban mobility analysis.* IEEE Access, 2021. **9**: p. 69527-69538.
23. Tu, P., et al., *Epidemic contact tracing with campus WiFi network and smartphone-based pedestrian dead reckoning.* IEEE Sensors Journal, 2021. **21**(17): p. 19255-19267.
24. Kawaguchi, N., et al. *Wi-fi human behavior analysis and ble tag localization: A case study at an underground shopping mall*. in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 2016.
25. Fukuzaki, Y., et al. *Statistical analysis of actual number of pedestrians for Wi-Fi packet-based pedestrian flow sensing*. in *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 2015.
26. Mayaud, J., *On the role of microtransit in shaping new mobility patterns.* Travel Behaviour and Society, 2025. **41**: p. 101065.
27. Liezenga, A., et al., *The first mile towards access equity: Is on-demand microtransit a valuable addition to the transportation mix in suburban communities?* Transportation research interdisciplinary perspectives, 2024. **24**: p. 101071.
28. Lucken, E., K.T. Frick, and S.A. Shaheen, *"Three Ps in a MOD:" Role for mobility on demand (MOD) public-private partnerships in public transit provision.* Research in Transportation Business & Management, 2019. **32**: p. 100433.
29. Geržinič, N., M. Guček, and O. Cats, *The potential of microtransit for regional commuting.* European Transport Research Review, 2025. **17**(1): p. 14.
30. Mayaud, J., F. Ward, and J. Andrews, *Microtransit has the potential to flip transit on its head.* Transportation Research Record, 2021. **2675**(11): p. 77-88.
31. FHWA. *Types of Traffic Analysis Tools*. 2024; Available from: https://ops.fhwa.dot.gov/trafficanalysistools/type_tools.htm#:~:text=Mesoscopic%20Simulation%20Models.%20Mesoscopic%20models%20combine%20the,superior%20to%20the%20typical%20planning%20analysis%20techniques.
32. Varga, B. and T. Tettamanti, *Jam propagation analysis with mesoscopic traffic simulation.* IEEE Transactions on Intelligent Transportation Systems, 2023. **24**(12): p. 14162-14173.
33. Zhang, J., et al., *CAVSim: A microscopic traffic simulator for evaluation of connected and automated vehicles.* IEEE Transactions on Intelligent Transportation Systems, 2023. **24**(9): p. 10038-10054.
34. Zatmeh-Kanj, S. and T. Toledo, *Car following and microscopic traffic simulation under distracted driving.* Transportation research record, 2021. **2675**(8): p. 643-656.

35. Zhu, L., Z. Zhao, and G. Wu, *Shared automated mobility with demand-side cooperation: A proof-of-concept microsimulation study.* Sustainability, 2021. **13**(5): p. 2483.
36. Meshkani, S.M., et al., *Innovative On-Demand Transit for First-Mile Trips: A Cutting-Edge Approach.* Transportation Research Record, 2024. **2678**(11): p. 122-136.
37. Oyewole, G.J. and G.A. Thopil, *Data clustering: application and trends.* Artificial intelligence review, 2023. **56**(7): p. 6439-6475.
38. Xia, Z., et al., *Identify and delimitate urban hotspot areas using a network-based spatiotemporal field clustering method.* ISPRS International Journal of Geo-Information, 2019. **8**(8): p. 344.
39. Yi, D., et al., *Identifying urban traveling hotspots using an interaction-based spatio-temporal data field and trajectory data: A case study within the sixth ring road of Beijing.* Sustainability, 2020. **12**(22): p. 9662.
40. Ran, X., et al., *A novel k-means clustering algorithm with a noise algorithm for capturing urban hotspots.* Applied Sciences, 2021. **11**(23): p. 11202.
41. Li, F., W. Shi, and H. Zhang, *A two-phase clustering approach for urban hotspot detection with spatiotemporal and network constraints.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021. **14**: p. 3695-3705.
42. Cesario, E., P. Lindia, and A. Vinci, *Detecting multi-density urban hotspots in a smart city: Approaches, challenges and applications.* Big Data and Cognitive Computing, 2023. **7**(1): p. 29.
43. Wilson, C.o.; Available from: https://www.wilsonnc.org/residents/all-departments/public-works/wilson-transit-ride-wilson-industrial-air-center/ride.
44. Census.gov. *American Community Survey (ACS)*. 2025; Available from: https://www.census.gov/programs-surveys/acs.html.
45. NCDOT. *Fatal and Serious Injury Crash Locations*. 2024; Available from: https://ncdot.maps.arcgis.com/home/webmap/viewer.html?webmap=9a25021dbe91427a92f2eca57bd71ee2.
46. *[802.11] Wi-Fi Connection/Disconnection process*. 2020 [cited 2022; Available from: https://community.nxp.com/t5/Wireless-Connectivity-Knowledge/802-11-Wi-Fi-Connection-Disconnection-process/ta-p/1121148.
47. Grantz, K.H., et al., *The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology.* Nature communications, 2020. **11**(1): p. 1-8.
48. Ichifuji, Y., et al. *A study for understanding of tourist person trip pattern based on log data of Wi-Fi access points*. in *2016 IEEE International Conference on Big Data (Big Data)*. 2016. IEEE.
49. Cisco. *Cisco Meraki Documentation*. 2020 [cited 2024 Jan 16]; Available from: https://documentation.meraki.com/MR/Wi-Fi_Basics_and_Best_Practices/Access_point_range_and_signal_strength_maximization#:~:text=Barring%20physical%20obstruction%20and%20radio,power%20in%20a%20single%20direction.
50. Pearson, R. *What is the average walking speed? - runner's world*. 2023 [cited 2024 Jan 16]; Available from: https://www.runnersworld.com/uk/health/a45709070/average-walking-speed/.
51. Google. *Distance Matrix API overview*. 2025; Available from: https://developers.google.com/maps/documentation/distance-matrix/overview.

52. Google. *Directions API overview*. 2025; Available from: https://developers.google.com/maps/documentation/directions/overview.
53. Ester, M., et al. *A density-based algorithm for discovering clusters in large spatial databases with noise*. in *kdd*. 1996.
54. Campello, R.J., D. Moulavi, and J. Sander. *Density-based clustering based on hierarchical density estimates*. in *Pacific-Asia conference on knowledge discovery and data mining*. 2013. Springer.
55. Behrisch, M., et al. *SUMO–simulation of urban mobility: an overview*. in *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. 2011. ThinkMind.