American University of Armenia

Zaven and Sonia Akian College of Science and Engineering

Geo-Spatial Analysis of Parking Behavior Within Defined Areas with Raw GPS Data

Thesis submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science in Computer Science and Engineering

By **Sargis Hovsepyan**

Under the supervision of **Mesrop Andriasyan**

Spring 2024

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to my research supervisor, *Dr. Mesrop Andriasyan*, for invaluable guidance throughout the process. His enthusiasm, insight, and support have been truly inspiring. He has taught me how to approach my work methodically and present it clearly. It has been a privilege to work and learn under his mentorship.

I would also like to thank Perigon AI for giving me the opportunity to conduct this research and for their guidance, as well as for providing the data, workspace, and general support essential for this research. Furthermore, I would like to extend special thanks to *Dr. Gevorg Yeghikyan* and *Dr. Nerses Srapoinyan* for their significant contributions, which have been instrumental in the success of this work.

I am greatly indebted to my honorable teachers of the Department of Computer Science and Engineering at the American University of Armenia who taught me during the course of my study. Without any doubt, their teaching and guidance have completely transformed me to a person that I am today. Finally, I would like to thank all the people who have supported me to complete the project work directly or indirectly.

Sargis Hovsepyan

American University of Armenia Yerevan,

Date: May 20, 2024

ABSTRACT

This study investigates the parking dynamics within Yerevan, Armenia, using GPS data from Perigon AI spanning 2019 to 2022. We processed and analyzed over 3.7 million GPS points, focusing on those within Yerevan's borders. Initial steps included filtering the data to include only relevant points and eliminating duplicates to ensure data quality. We continued with temporal pattern analysis, revealing a decrease in red-line parking in 2019 and possible impacts of the COVID-19 pandemic on mobility and parking behaviors in 2020. Weekly and monthly trends emphasized subtle differences in parking behaviors between weekdays and weekends.

Furthermore, spatial analysis, including heatmap visualization and unique user assessments, identified high-density activity areas, particularly in central Yerevan. The density analysis showed that central polygons had higher mobility, while unique user analysis revealed an increase in unique visitors in 2020, indicating dynamic parking behaviors. Trajectory analysis provided insights into user mobility patterns, analyzing durations and distances between consecutive GPS points. We reduced the data to about 530,000 points with 10,021 unique devices, facilitating the creation of coherent trajectories and identification of key activity sites and staypoints.

Although we did not fully implement a stop-detection algorithm, we proposed a methodology to identify stops within red line zones, considering duration and distance thresholds. Our findings suggest that trajectory creation enhances the accuracy of analyzing parking behaviors. The results offer valuable insights for urban planners and policymakers, emphasizing the importance of understanding parking dynamics and urban mobility. Future research should integrate additional data sources, apply advanced machine learning techniques, and refine methodologies to enhance urban mobility and parking management strategies. This study provides a solid foundation

for further research, aiming to foster sustainable urban mobility and efficient transportation systems.

Keywords: *Parking dynamics, Urban mobility, GPS data analysis, Temporal patterns, Spatial analysis, Trajectory analysis, Stop detection, Urban planning, Transportation management, Data-driven insights*

Table of Contents

1. Introduction

Due to the use of widespread GPS-enabled devices and the advancement of positioning technologies, a vast amount of spatial data can be collected. The publication and analysis of GPS data allow people to understand the distribution of movable objects in society at a deeper level, especially in terms of the spatiotemporal phenomenon of people. Thanks to the complete processing of GPS data, the research area that used to be confined by human eyes has been significantly expanded; that is, we, aside from those little ants walking on the ground, also acquired a toolset to analyze the movement of objects that glide on the road or even fly in the air. This massive data collection allows researchers to dig into almost everything inside the world of motion.

1.2 The Objective of The Research

This research aims to understand parking behavior within defined parking areas using only raw GPS data collected over time. Our objective is to discover parking behavior cues and irregularities that policymakers and urban planners might use to adjust and optimize the management of parking strategies. We are interested in determining the time of the day that parking is most used, how long parking events are on average, and how outside factors, such as the day of the week, the time of day, or even special occasions, affect parking behavior. Therefore, the main goal of this research is not to locate parking spots but to understand why, when, and how people are parking in specified areas.

The analysis is performed based on the raw GPS mobile data collected from 2019 to 2022, covering the area of Yerevan, Armenia. We are interested in the red line parking lots in the Center and Arabkir regions of Yerevan. It is remarkable to note that data distribution may differ yearly due to changes in user behavior, device exposure, and data-gathering methods. Consequently, this provides both opportunities and challenges when processing the data to understand dynamics over time. However, our data is substantial enough to draw some conclusions about the dynamics around the red lines. Nevertheless, before drawing conclusions, the raw GPS data has to undergo preprocessing stages and must be processed largely by applying mathematical modeling to become suitable for analysis. It is also noteworthy that the following research is going to be used by Yerevan's Municipality for identification of the influence of the policy change during 2023 to 2024.

1.3 Overview of The Research

The upcoming sections of this paper will cover the processing of the collected GPS data, providing a more detailed overview of the research. Furthermore, it will present a comprehensive overview of the methodologies applied and the research findings, focusing on the red lines of the specified regions of Yerevan. In particular, we will cover the steps taken to clean and preprocess the raw data, leaving only the points that matter. Additionally, we will zoom in on how these GPS points are spread across different time frames – weeks and months – and compare weekdays to weekends. Indeed, we cannot miss out on the statistical analysis, which will enable us to make appropriate choices for our parameters in the later stages of analysis. The paper also introduces more in-depth analysis, which involves studying human movement analysis through movement trajectory creation and mobility detection. Therefore, we will present data segmentation for a closer look, tracing vehicle trajectories and analyzing how people move around parking areas. Finally, the results acquired wrap up the paper, offering insights that could shape urban planning and policies.

2. Literature Review

It is already remarkably noted that GPS data analysis helps people create models and extract useful information from them, accelerating the process of urban planning and enhancing the strategies created by policymakers. Research regarding GPS data processing is a vast field that is developing continuously. In order to understand the choices made for GPS data processing and methodology in this research of parking behavior, firstly, let us understand the research carried out by other seminal papers on GPS data preprocessing, activity and trip identification, vehicle movement patterns, and their implications for urban planning.

2.1 GPS Data Pre-processing

One of the critical steps used during the pre-processing of raw data collected through GPS is filtering for accuracy and reliability prior to the subsequent analyses. Stopher, Fitzgerald, and Zhang (2008) place so much focus on the vital aspect of data cleaning and filtering GPS data to avoid undesired noises and inaccuracies. They develop tool segmentation to detect trips and stops, more specifically, and improve the analysis of modality creation. Similarly, the "GPS Data Pre-processing" study sheds more light on improving raw data sets collected from GPS operations for fine-grained analysis and modeling (Stopher et al., 2008). It is also remarkable to note that cleaning is performed based on the aim of the study and the implementation.

2.2 Processing Raw GPS Data

Schuessler and Axhausen's (2009) article, an enlightening work by Siu and Kilian (2009), is about GPS data analytics without any external accessories and describes in detail the possible ways of analyzing the data. The cleaning operation is carried out to eradicate mistakes and anomalies. A smoothing phase follows this operation, and then multi-modal travels are segmented. The choice of mobility, which is used as the element of analysis for such a model, shows its revolutionary impact on the other factors that were integrated (Schuessler & Axhausen, 2009). Indeed, the research presented in this paper also includes data cleaning and segmentation for human trajectory creation.

2.3 Micro-segmenting and Distance Calculation

The GPS ground path calculation precision is the primary factor for trajectory accuracy. The study "Micro-segmenting: An Approach for Precise Distance Calculation for GPS Points" herein refers to micro-segmenting as a method of distance accuracy improvement. This means dividing GPS tracks into shorter pieces, which allows us to accurately monitor distance data, especially in more complicated places, such as urban areas (Smith et al., 2016). This approach is a key factor in trajectory creation, which helps us understand people's movement patterns. The latter can help us understand the mode of the movement of the people more accurately in order to understand their behavior around the defined parking areas (red lines in our case).

2.4 Travel Time and Mode Detection

Determining travel duration and travel modes using GPS logs is crucial for all urban mobility analysis. For instance, Chen et al.'s (2017) article "Road Network Travel Time Calculation Using GPS" and Zhao et al.'s (2015) article " Identifying Transportation Modes from Raw GPS Data" might indicate that these algorithms have been developed explicitly for this purpose. These approaches help to estimate the exact deviation and identify transport modes. As a result, they build a comprehensive urban mobility model (Chen et al., 2017; Zhao et al., 2015). The urban mobility model is crucial not only for parking behavior analysis but also for extended and complete analysis, which might help to extract much more than just the movement of a person.

2.5 Mobility Analysis through GPS Data

The article "Trackintel: An Open-source Python Library for Human Mobility" presents a way to analyze people's movement patterns, focusing on the applications of open sources to study urban walks (Griffin et al., 2020). Trackintel is a valuable tool for analyzing and manipulating geospatial data. Trackintel creates structure and lays the foundation for processing, helping to organize, store, and prepare data for further processing. The approaches developed by this paper are based on the data format provided by trackintel. In order to be more familiar with the inner workings of trackintel it is advisable to read the article.

2.6 Trajectory Data Mining and User Similarity

Zheng (2015) covers the data mining trajectory in extensive detail, emphasizing deriving important patterns from movement data. This project draws attention to the potency of innovative data-mining technology that can subsequently find movement behavior and patterns from GPS raw data. Furthermore, Li et al. (2008) shed more light on group movements by analyzing user similarities using their location history. This is an essential area for analysis since it provides a clear picture of collective movement patterns and clusters of common routes and destinations. The studies thoroughly enhance the ways that researchers may analyze parking behavior through the existence of tools for interpreting the complex trajectory data (Zheng, 2015; Li et al., 2008).). These papers were crucial for staypoint (activity) generation for each person, which helped to find potential activity sites and clean data from uninformative points.

2.7 Comprehensive Methodological Reviews

Indeed, mode detection is one of the crucial parts of the mobility analysis to understand the means of human mobility. The insight we can gain from the information extracted from movement modes is valuable from our research. Systematic reviews such as "Travel Mode Detection Based on GPS Raw Data Collected by Smartphones: A Systematic Review of the Existing Methodologies" by Jahangiri and Rakha (2015) summarize the tools for travel mode detection and traffic analysis employed by professionals today. These reviews provide criteria for assessing methodological strengths and weaknesses and giving directions on future studies and their applications within the field (Mohammed et al., 2015). There are various mode detection strategies, and the choice again depends on the nature of the study. In our study, mode detection is done based on speed, time, and distance thresholding, as we are interested in modes that separate walking from movement by vehicle. Later, we analyze the behavior of vehicles around red lines (parking lots).

2.8 Implications for Urban Planning

The GPS data analysis results show the far-reaching influences that can be used to enhance urban planning and policy development. Through comprehension of the space-time patterns of parking behaviors, urban planners can really maximize the management of parking spots, decrease congestion, and boost the effectiveness of the city transportation system. Therefore, the methodologies used are trying to understand human mobility patterns based on specific data and areas of interest to present the dynamics of parking within specific areas.

3. Methodology

This study intends to analyze the dynamics of the red line parking in Yerevan, Armenia by merging GPS data for the 2019, 2020, 2021, and 2022 periods. The main areas of interest are the red line parking spaces in the center and Arabkir regions of Yerevan. The primary objective is to understand users' spatial distribution and trajectories within the specified regions for exploratory research. The raw data on which the research is based is provided by Perigon AI, a company that provides advanced data analytics and artificial intelligence solutions leveraging large datasets, including geospatial data. The data provided entails minimal yet very elaborate details as it aids in user movement analysis, which is used to measure user density within red-line regions and trajectory explorations.

The data provided by Perigon AI contained identifiers, which are unique per device, identifier type, timestamp, device latitude, and device longitude (Figure 1.1).

1	Unnamed:	identifier	identifier_type	timestamp	device_lat	device_lon
2	1	ccfbeb52-5ada-41a6-8913-542d8f0bb79c	gaid	2020-10-25 09:37:33 UTC	40.3532107	44.0755004
3	2	5055e48d-6b02-4d68-8e93-1e2aa83db2b7	gaid	2020-12-01 18:28:18 UTC	40.3840497	44.3823699
4	3	9b6517fa-2a11-4f13-86a0-359787a8af40	gaid	2020-10-19 03:48:49 UTC	40.3527296	44.3808168
5	9	9b6517fa-2a11-4f13-86a0-359787a8af40	gaid	2020-12-08 05:17:23 UTC	40.3527764	44.3810696
6	10	03e3d35c-6b3c-42e3-93bf-173b9bbeb1ad	gaid	2021-01-02 16:21:47 UTC	40.270725	44.2876964
7	11	1737ff7b-631d-4c30-a35b-91f077ae1b0c	gaid	2020-11-25 00:31:34 UTC	40.2998202	44.3628756
8	12	74746fdf-cd55-4dbd-925c-df71ddc2a492	gaid	2020-11-07 09:31:09 UTC	40.3643767	44.0157781
9	15	585ebf17-eb5a-471e-97af-e3c008add1ea	gaid	2020-11-05 23:34:33 UTC	40.2895224	44.34404
10	17	2bb1c652-9921-462d-b432-10212d168613	gaid	2020-11-01 18:00:07 UTC	40.2579646	44.3453292
11	18	74746fdf-cd55-4dbd-925c-df71ddc2a492	gaid	2020-11-08 20:21:55 UTC	40.3642668	44.0156262
12	19	bd1b0692-6b1f-4ebc-9e15-c13c9d211ec2	gaid	2021-01-24 23:23:30 UTC	40.2793194	44.3583056

Figure 1: The raw GPS data provided by the company Perigon AI

Therefore, each row represents a GPS point that lies within Yerevan. Since the data included points close to Yerevan but not inside the borders, we focused and focused on points lying in Yerevan only. The following libraries were utilized for data handling and analysis: numpy for numerical operations, pandas for data manipulation, geopandas for spatial operations, and shapely for geometric objects. We conformed to pyproj for cartographic projections, osmnx for the street network data, networkx for network analysis, matplotlib for plotting, and trackintel for gathering, storing and analyzing movement data including structures like position fixes and staypoints. These powerful libraries give me enough capacity to read, store, clean, modify, and analyze the GPS data.

3.1 Data Retrieval and Storage

The initial step for our analysis starts with the retrieval and storage of raw GPS points. The whole structure of data manipulation was supported by the library Trackintel.

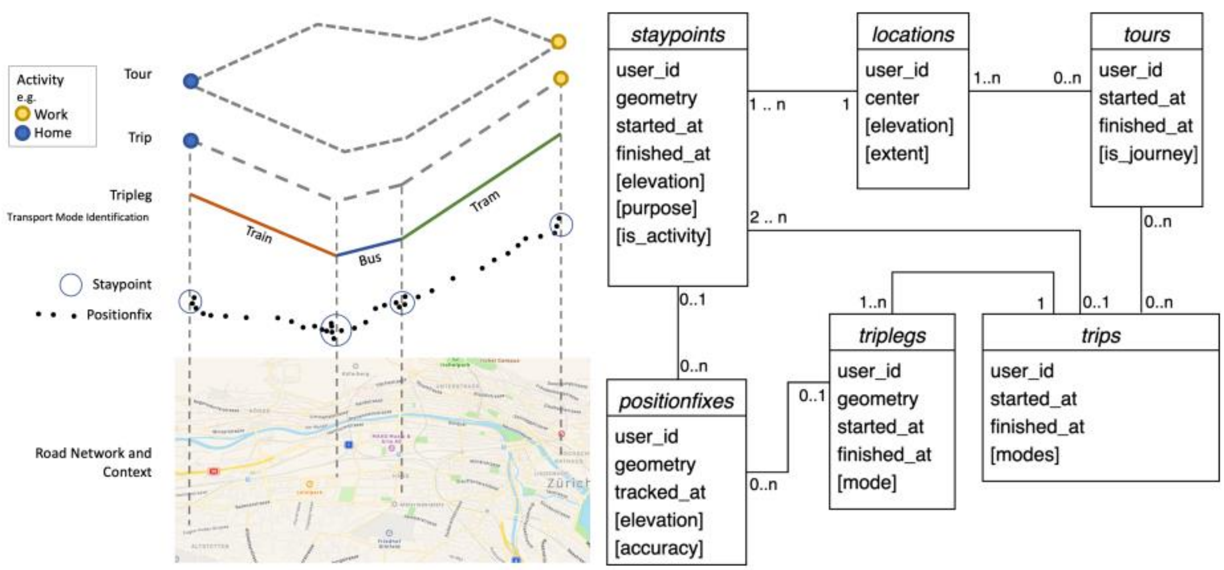

Figure 2.1: Semantic visualization of the Trackintel data models.

Trackintel implements six classes to represent movement data in this hierarchical model: positionfix, staypoint, tripleg, trip, tour, and location (Figure 2.1), where all Trackintel classes are implemented as Pandas Dataframes or Geopandas Geodataframes (Griffin, 2020, p. 5). We start by reading the points from the CSV file into a Trackintel class called positionfixes, which holds the information about GPS data, as shown in Figure 2.2:

	user_id	tracked_at	geom
0	dae4a487-0ce1-4bd9-a90e-955a5ca06dc1	2020-03-24 07:22:46+00:00	POINT (44.42081 40.19434)
1	18ced82b-da82-42a0-967b-03fa2b6aba82	2020-04-05 23:37:37+00:00	POINT (44.42077 40.19358)
2	d155dde3-112a-46a0-8516-e083f1d5daac	2020-01-06 15:02:22+00:00	POINT (44.41815 40.19038)
3	d155dde3-112a-46a0-8516-e083f1d5daac	2020-01-17 10:00:56+00:00	POINT (44.41781 40.19027)
4	d155dde3-112a-46a0-8516-e083f1d5daac	2020-01-14 16:52:39+00:00	POINT (44.41806 40.18994)
...	...	...	...
3768442	b719ad32-f26c-412b-84aa-50064d3758ce	2021-03-02 23:17:41+00:00	POINT (44.48316 40.19336)
3768443	0ff341dd-e8a7-4f0d-9d63-39c49c3fcbf3	2021-02-10 02:08:22+00:00	POINT (44.47050 40.19485)
3768444	8f6bdc17-275e-49ba-892c-6496abfc5a35	2020-09-05 05:29:00+00:00	POINT (44.47579 40.19014)
3768445	8f6bdc17-275e-49ba-892c-6496abfc5a35	2020-09-24 22:22:53+00:00	POINT (44.47554 40.18990)
3768446	8f6bdc17-275e-49ba-892c-6496abfc5a35	2020-08-26 23:32:15+00:00	POINT (44.47580 40.19007)

3768447 rows × 3 columns

Figure 2.2: Storage of GPS data as positionfixes

After the reading, we filtered to include only those inside the boundaries of Yerevan and all the duplicates were eliminated; so that data quality and integrity would be satisfied. The filtered points were written into a CSV file for faster retrieval, which would be used for the next step of the analysis process. This conversion enabled agility in the comprehension of Trackintel functionalities and further data processing and analysis.

3.2 Density Analysis

Before using complex tools to explore people's movement trajectories, we need to explore the properties of the data provided. Therefore, let us understand the behavior of positionfixes inside the red lines. In order to work with the red lines as geospatial entities, we identified the regions of red lines on the map manually and created polygons representing the parking places we were interested in (Figure 3). Furthermore, after identifying red line polygons, we separated points inside the red line polygons, initially

Figure 2.3: About 3.7 million GPS points on Yerevan's map

applying a 0.5-meter buffer to the polygons. The buffer is applied to increase the accuracy of the calculated data by enabling us to capture people who parked their cars but are outside of the vehicle. Consequently, this approach also included points from cars passing by and pedestrians walking near the red lines. To address this, we implemented techniques to clean the data and remove points unrelated to parking. Ultimately, we obtained two sets of data: approximately 3.7 million points within Yerevan and about 22,000 points within the red line areas.

Figure 3.1: Red line parking spaces of Center and Arabkir, Yerevan

3.2.1 Studying the Percentage of Positionfixes in Red Lines

As we have the points inside the red lines, we can analyze their behavior over time regarding the overall points inside Yerevan. First, let us understand the percentage of points inside the red lines in relation to all the data we have on a monthly basis. We presented monthly data for the given years and depicted spatial trends for parking around the red lines throughout the plot (Figure 3.1). It helps to know how the frequency of points appearing inside the red lines changes with time and if these fluctuations are predictable. Moreover, it also helps us understand the data integrity and completeness to understand the reliability of our conclusions.

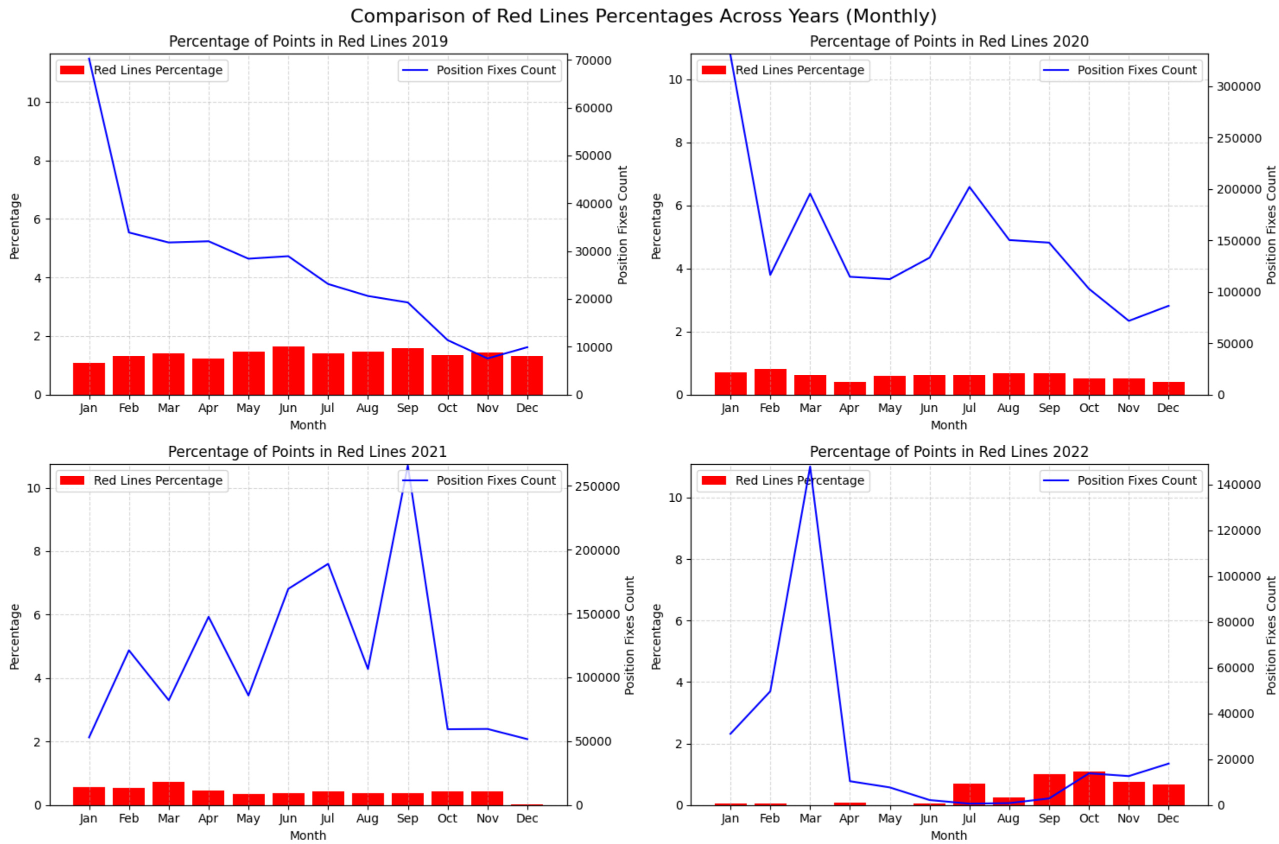

Figure 3.2: Comparison of monthly percentages of points inside red lines across years

The plots demonstrate the overall number of positionfixes by the red line and the percentage of positionfixes inside the red lines by the bars. It can be observed that in 2019, there was a noticeable decrease in the percentage of points within the red lines, which might be an indicator of a potential behavioral change in parking practices.

Furthermore, we can see that the overall number of points in 2020 peaked in January and showed a significant decline till December, which also affected the percentages of the points residing in the red line polygons. This may reflect the impact of external factors, such as the COVID-19 pandemic, on mobility and parking behaviors.

On the other hand, the count of points in 2021 exhibited significant fluctuations, peaking in September. Interestingly, although the overall mobility inside Yerevan, mainly concentrated in the center, might change and fluctuate drastically, the points inside the red lines remain relatively steady.

Lastly, the plot of 2022 shows oddly behaving data, which might indicate particular environmental or social dynamics influencing parking. However, as the ratios are not kept relatively similar and the decline is too drastic, we can claim that the data we have for 2022 is not representative.

In order to understand the dynamics in greater detail, we can create the same plot on a weekly basis. This can be crucial in better identifying fluctuations and understanding their influences on the parking dynamics. Figure 3.3 shows the comparison of the percentages taken weekly, and compared to the monthly graphs, it is evident that the fluctuations and smoothening are more emphasized here.

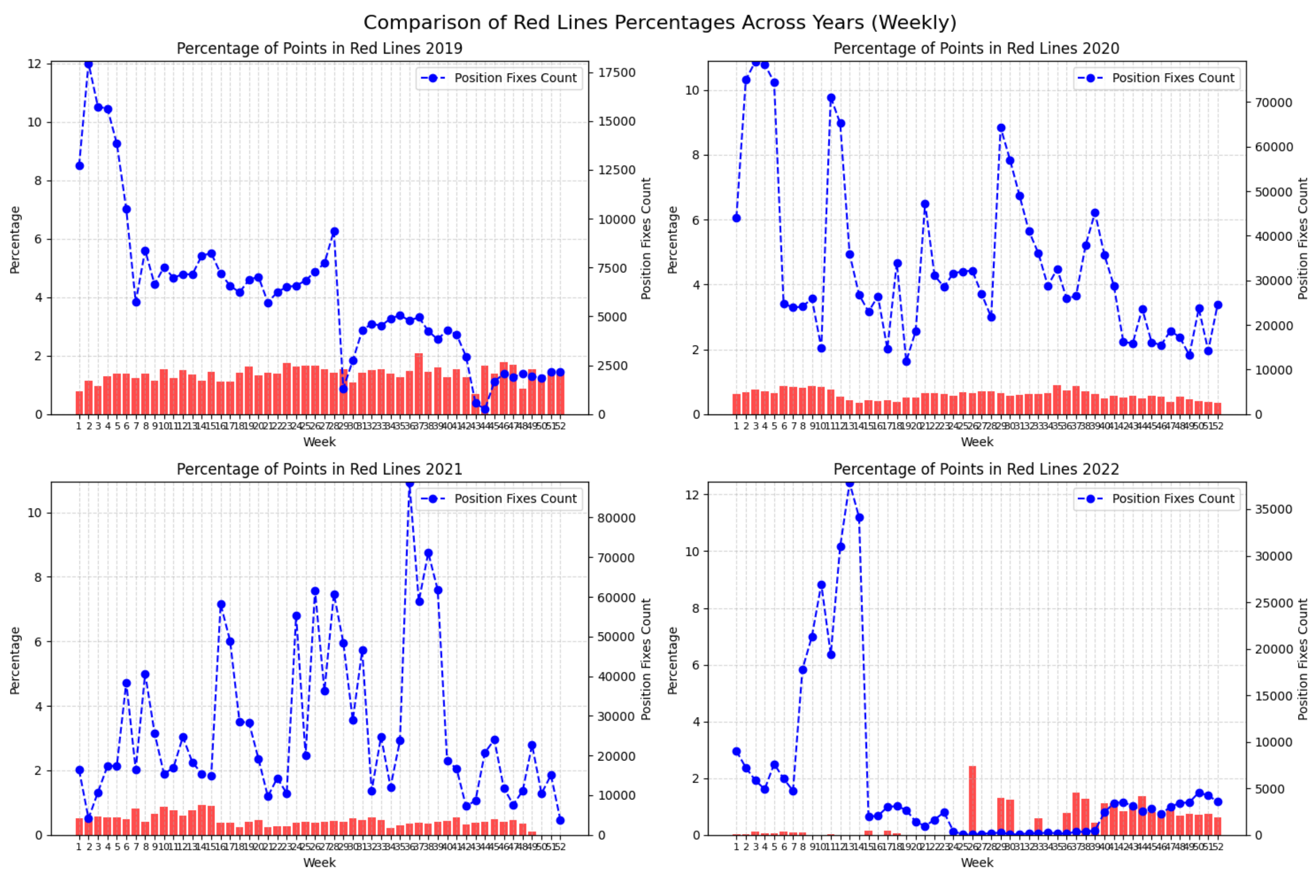

Figure 3.3: Comparison of weekly percentages of points inside red lines across years

3.2.2 Studying the Percentages of Points in Red Lines by Weekdays and Weekends

The analysis can be extended by separately focusing on weekday and weekend percentages. Weekdays typically experience higher activity levels than weekends, suggesting potential movement dynamics and parking behavior differences. However, this does not necessarily mean that there is a significant difference between car parking between weekends and weekdays.

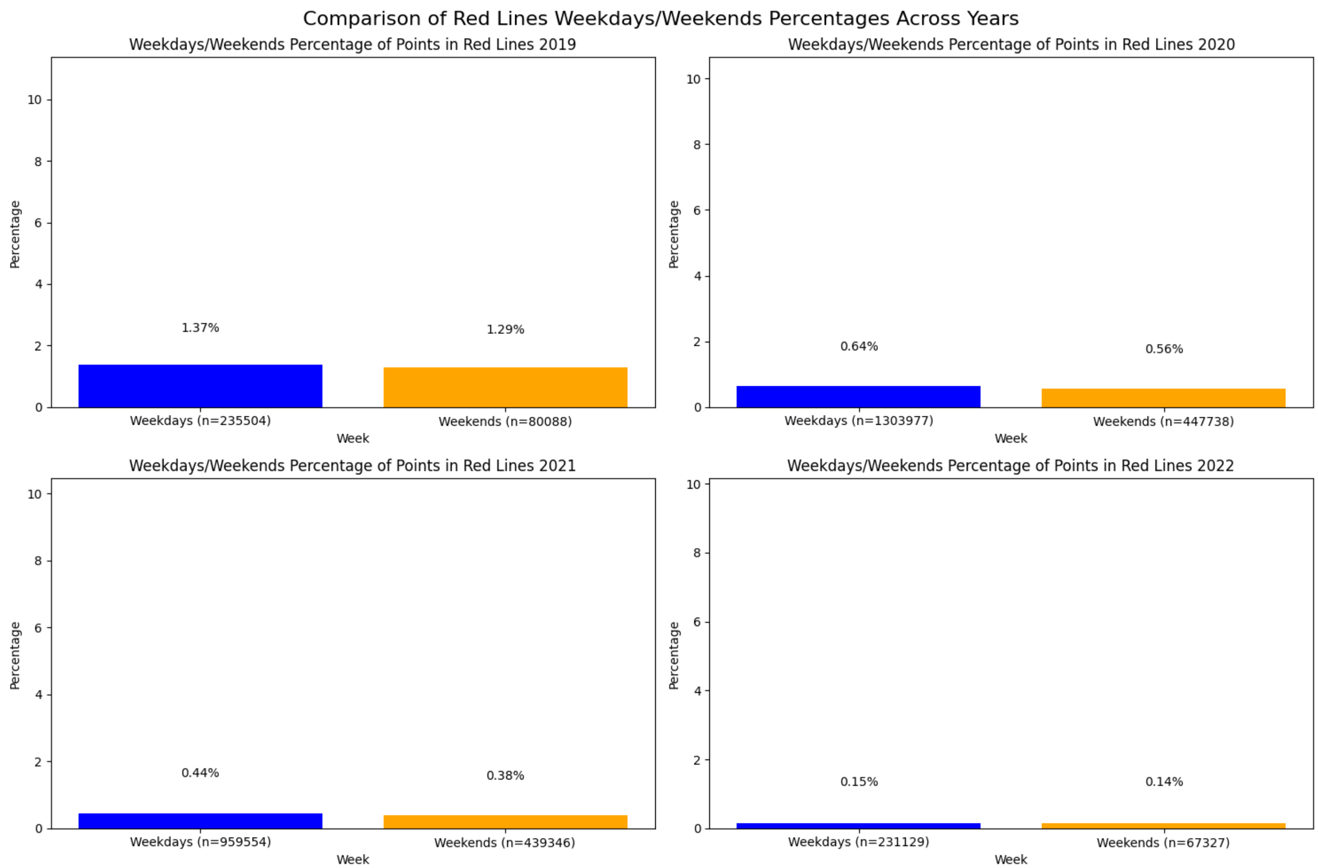

Figure 3.4: Comparison of week(days/ends) percentages of points inside red lines across years

The division of the data between weekdays and weekends might help the research give a comprehensive image of the parking patterns and reveal the possible variations of intensity, duration, and location across different days. Here, we can see that the difference between the number of percentages of points inside red lines during weekdays and weekends is insignificant. At the same time, we can observe the consistency of the plots with the weekly and monthly analyses discussed above. Once again, let us note that gathering this data is of great importance to urban planners and officials for effectively developing parking management strategies that suit the constraints of specific situations.

Additionally, before jumping to the identification of people inside red lines, we can understand the density of points inside the red lines to target the polygons that contain more visits per point.

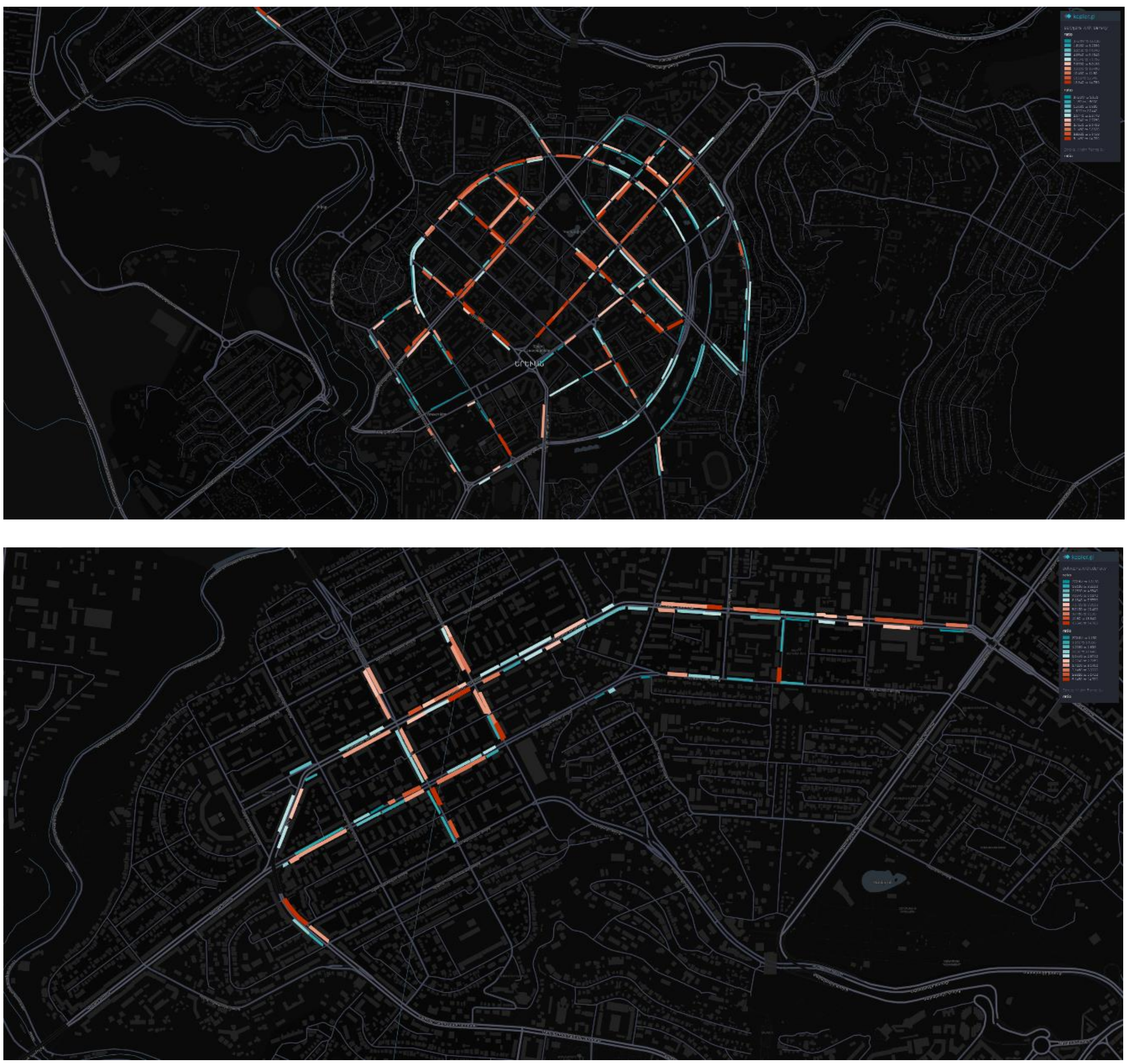

Figure 3.5: Heatmap of the density of points inside red lines in the Center and Arabkir.

The heatmap identifies the dynamics of the regions by showcasing the density. It just claims that certain areas have higher mobility. Interestingly, the polygons in the center have a higher density as more people create more movement. However, the density of the polygons cannot be a factor in claiming that people park there. To delve deeper into the density analysis, let us understand the distribution of unique people inside red lines.

3.2.3 Unique People Dynamics Inside Red Lines

To continue with, let us understand the unique user presence in red line zones. We extracted individual user information by grouping the positionfixes inside the red lines by the user identifier. The aim is to calculate the number of unique people per polygon. The presence of unique users gives us a more informative perspective and helps us further our research. The bar graphs created interprets the concentration of unique users in different red-line parking areas.

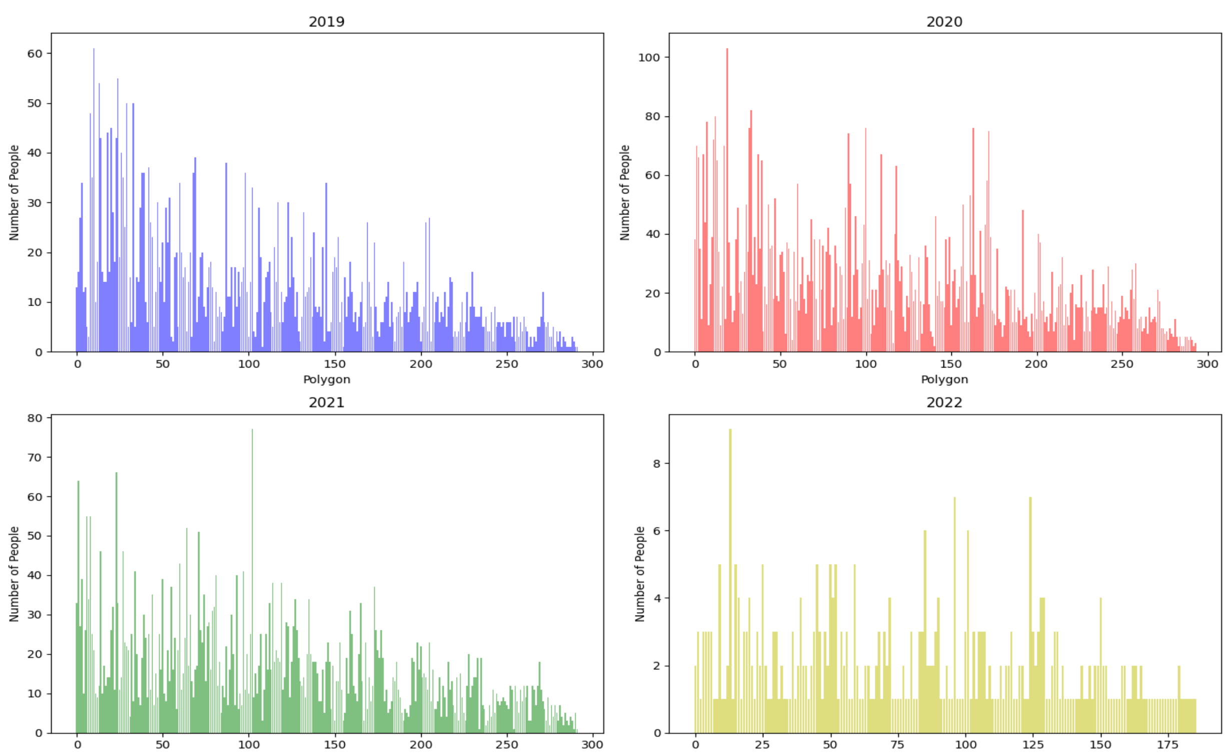

Figure 3.6: The number of unique people per red line polygon across 2019-2022

The plots show how many people were found in each polygon. We have about 300 polygons denoting red-line parking zones, and now we know how many people were found inside them throughout the time. Additionally, if we calculate the mean and median number of unique people inside polygons, we can tell which year showed more visits.

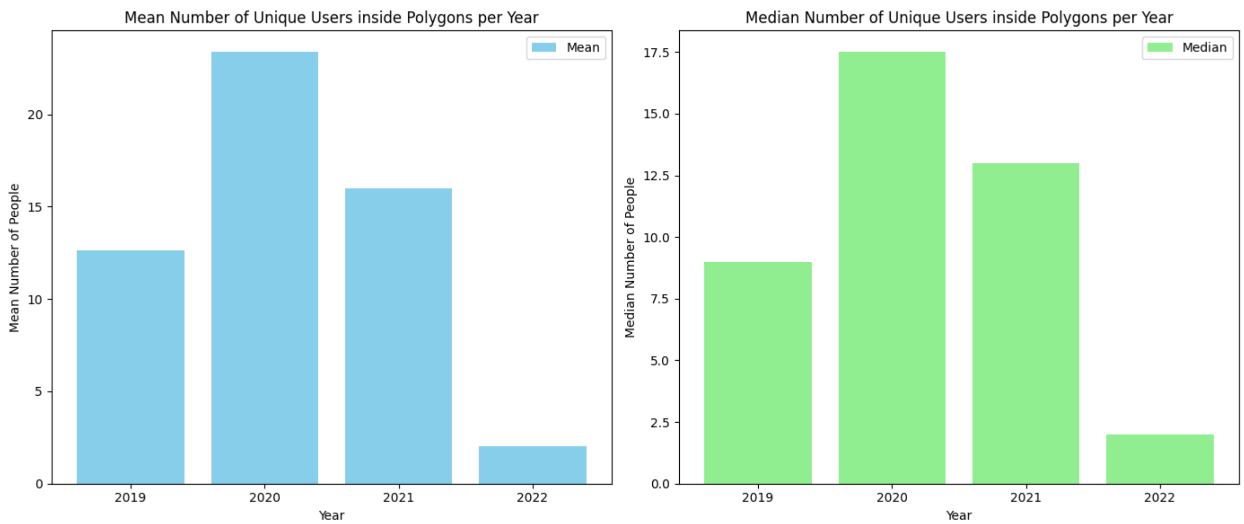

Figure 3.7: The mean and median number of people inside polygons during 2019-2022

In this particular case, it is evident that 2020 shows more polygon visits than any other year. The claim is supported by both mean and median, which means we have a few outliers, which do not interfere with our analysis. Furthermore, we can also state that the dynamics around the parking zones increased from 2019 to 2020 and then decreased. However, the dynamics around red lines do not involve only the people who parked their vehicles. Therefore, we need to continue our analysis to understand the mobility patterns of vehicle owners and explore their behavior around red lines.

3.2.4 Unique People Duration Distributions in Red Lines

Time is a critical factor in our density analysis. In order to understand which people are parking inside the red lines, we need to understand how much time they spend in the red lines. People who pass by in a car or on foot will only have very few points and even might have one point. Therefore, we must target and remove them from the batch of people falling into the red lines. To understand how much time people spend inside the red line, we first need to group them by date and polygon; after extracting the points that belong to a specific person and are in a specified polygon, we just take the difference between the times of the first occurrence inside the polygon and the last one.

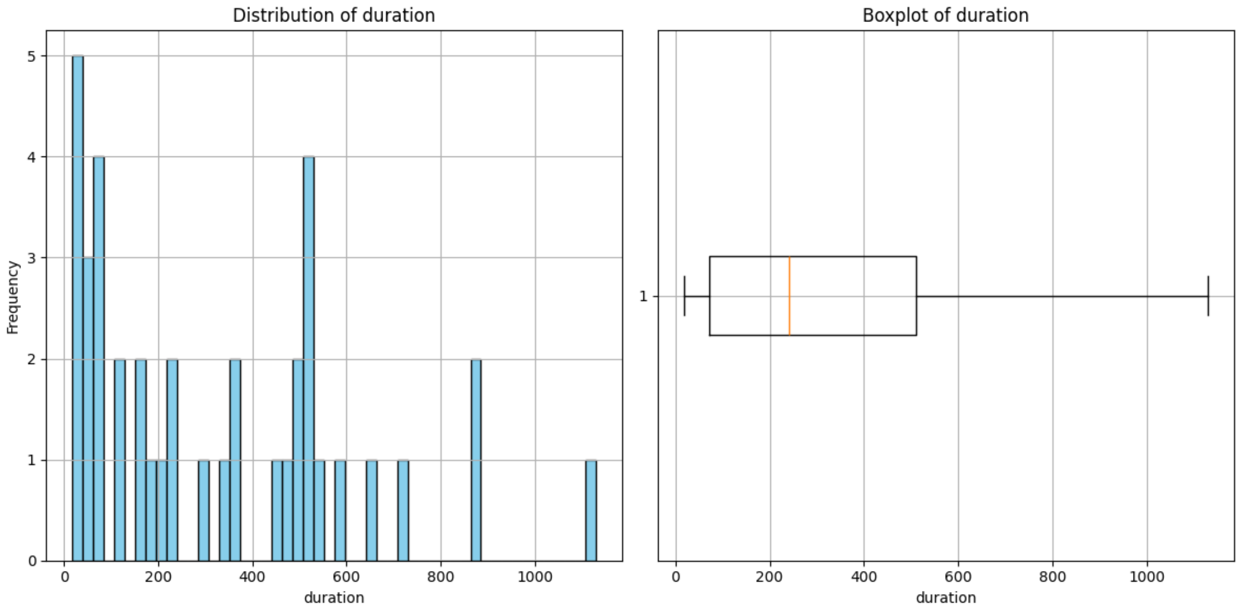

Figure 3.8: The distribution of times spent inside red lines in 2019

This graph shows a rather discrete distribution of time spent inside the red lines during 2019. It is interesting to analyze this, considering the previously acquired graphs. Although it shows a good distribution for unique visits to polygons when we start cleaning people based on a time threshold of 15 minutes, it reduces the number of people being observed. Nonetheless, the results are reasonable, as shown by the following statistics; Mean duration: 325 min., Median duration: 240 min., Mode duration: 18 min., 10% Quartile: 38.813 min., 90% Quartile: 671.007 min.

It is also remarkable to note that we have almost no outliers, which increases the credibility.

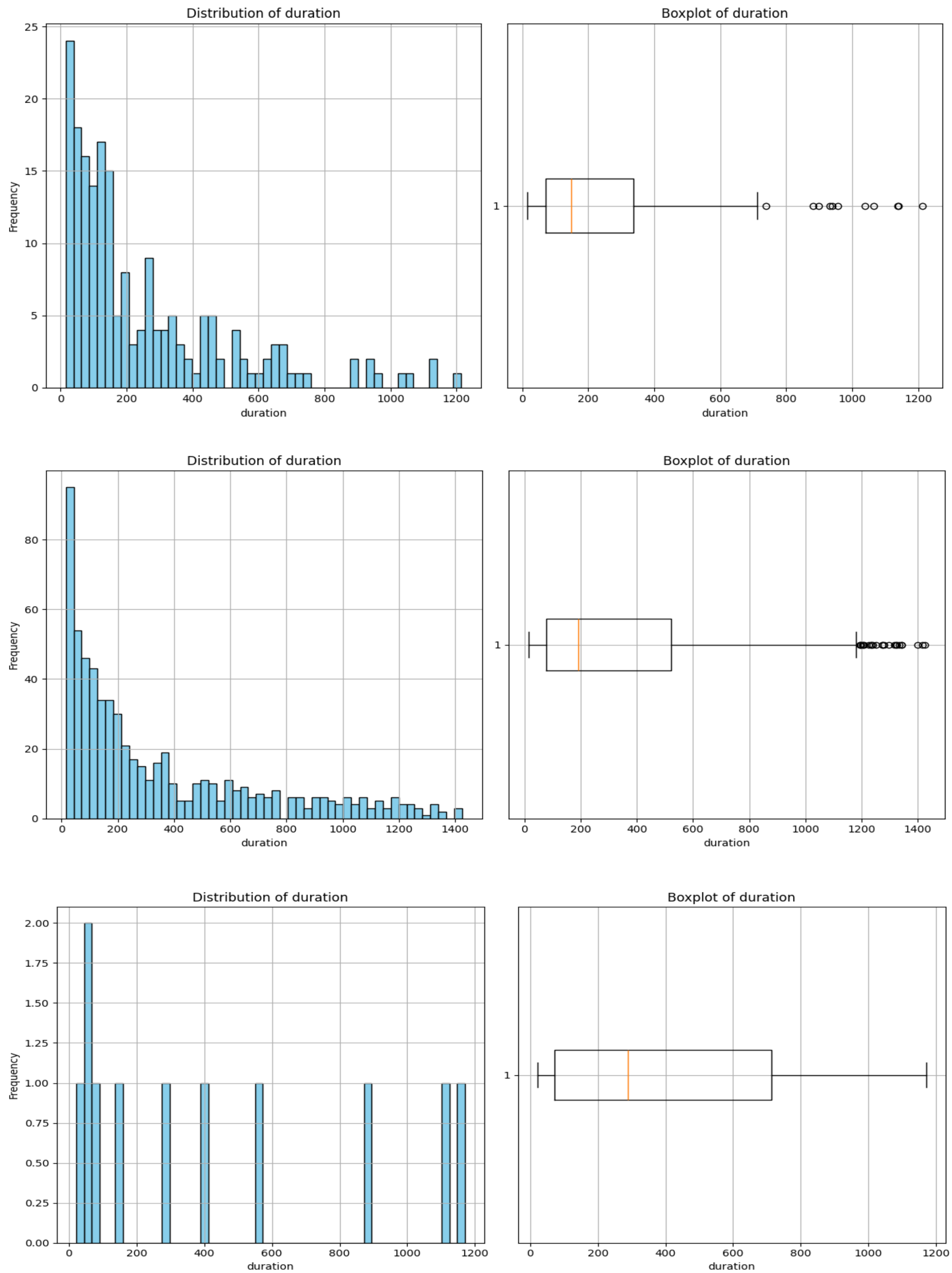

Figure 3.9: The distributions of times spent inside red lines in 2020, 2021, 2022 accordingly

It is also remarkable to note that we have almost no outliers, which increases the credibility of statistics. The distributions of 2020 and 2022 show a better perspective and reasonable timeframes for parking times. Although there might be some outliers, the distributions' statistics can be used to identify the parking events inside red lines. The final step might be taking people in these distributions and considering them potential people who parked their vehicles inside the red lines.

3.3 Trajectory Analysis

The density analysis carried out during this research can be extended to acquire more detailed information about the dynamics of the red line polygons. However, we must be careful with density analysis as it might only reveal accurate information if we have nicely distributed data. Furthermore, we might miss valuable information about the user, only concentrating on the red line polygons. Trajectory analysis suggests identifying user movement throughout the day to understand the movement patterns. After identification of mobility patterns, we can project those patterns around the red lines and understand the user's behavior. The main idea is to identify the movement mode, such as walking, running, traveling by car, bus, or anything else. Consequently, we can use this information to predict whether users parked in the specified parking area.

3.3.1 Data Cleaning Through Staypoint Generation

The trajectory creation is more data-dependent as it involves understanding people's movement patterns. In other words, the data needs to be dense regarding unique users and the number of unique GPS locations a user gives. The first step in trajectory creation will be cleaning data from uninformative data points, such as users having a GPS point daily or a few over four years. We could have removed those points by setting a GPS point count threshold; however, we decided to get a read of those points by staypoint generation. Staypoint is a

Trackintel library class that holds information about potential activity sites. Aside from activities, simple stop points can also be considered as staypoints. Now, in order to generate staypoints on a given frame of positionfixes, we need to have enough points and enough movement performed by the user. The staypoint generation algorithm used by Trackintel is adopted from the methodologies discussed in Zheng's comprehensive overview of trajectory data mining, which highlights various techniques for analyzing spatiotemporal data (Zheng, 2015).

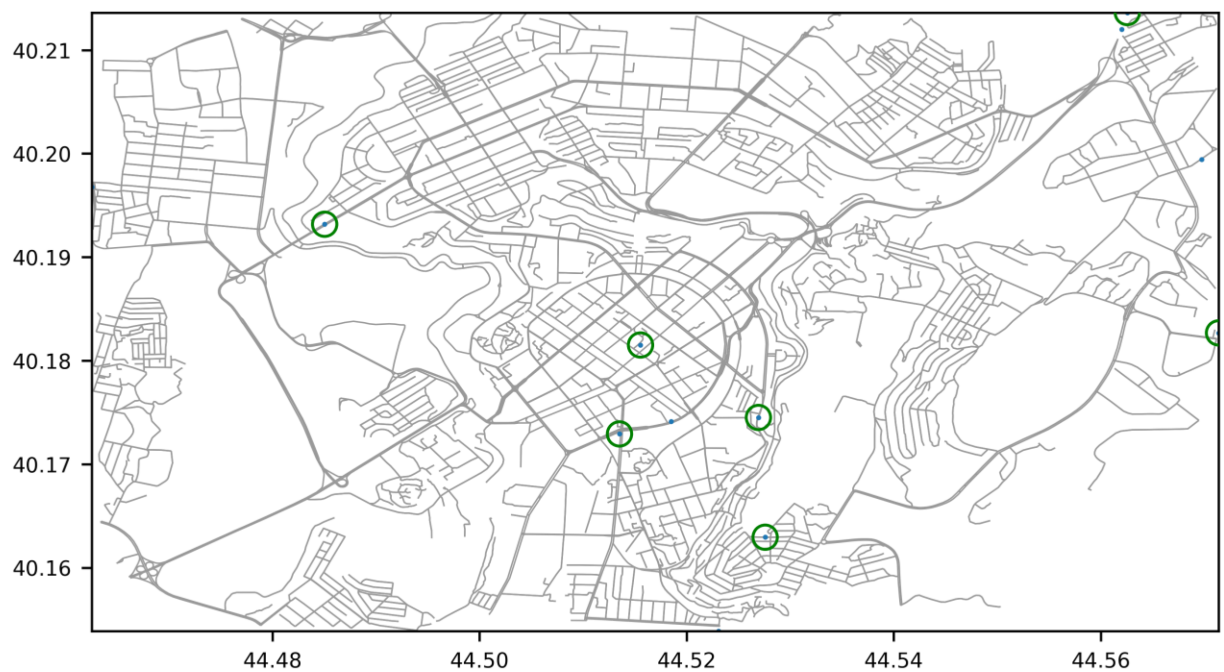

Figure 4.1: The GPS points of a person in a day containing staypoints with green circles

Firstly, positionfixes are used to extract people from them. Each person shows a unique device that has positionfixes over the four years. After user extraction, each user's points are grouped by date, and only then is the staypoint generation applied. If the data per day is not good enough, the staypoints generation fails, and the points of that day are removed. This process cleans data extensively, and after the generation of staypoints, 3.7 million points are cut down to about 530,000 points with 10021 unique devices. Therefore, after this process, we removed a significant amount of data, but we could have gotten no information from those points. Therefore, staypoint generation gave us more representative data for trajectory creation.

3.3.2 Micro-segmentation of Positionfixes

We start the trajectory creation process as long as we have clean and representative data. The process starts with segment creation which is adopted from the works of Smith et al. (2016). A segment represents a path between two consecutive points containing detailed information such as the user identifier, start and end times of the path, the distance, average speed, and the LineString holding the path between two points by the map. The process of this data creation is called segmentation; before that, we need to group points by each user and apply the segmentation to each day separately. This massive data collection lets us explore a device's mobility in greater detail. Therefore, before proceeding with the refinement of segments, let us explore the distribution of duration, distance, and average speed between two consecutive points within a particular day and belonging to a specific user.

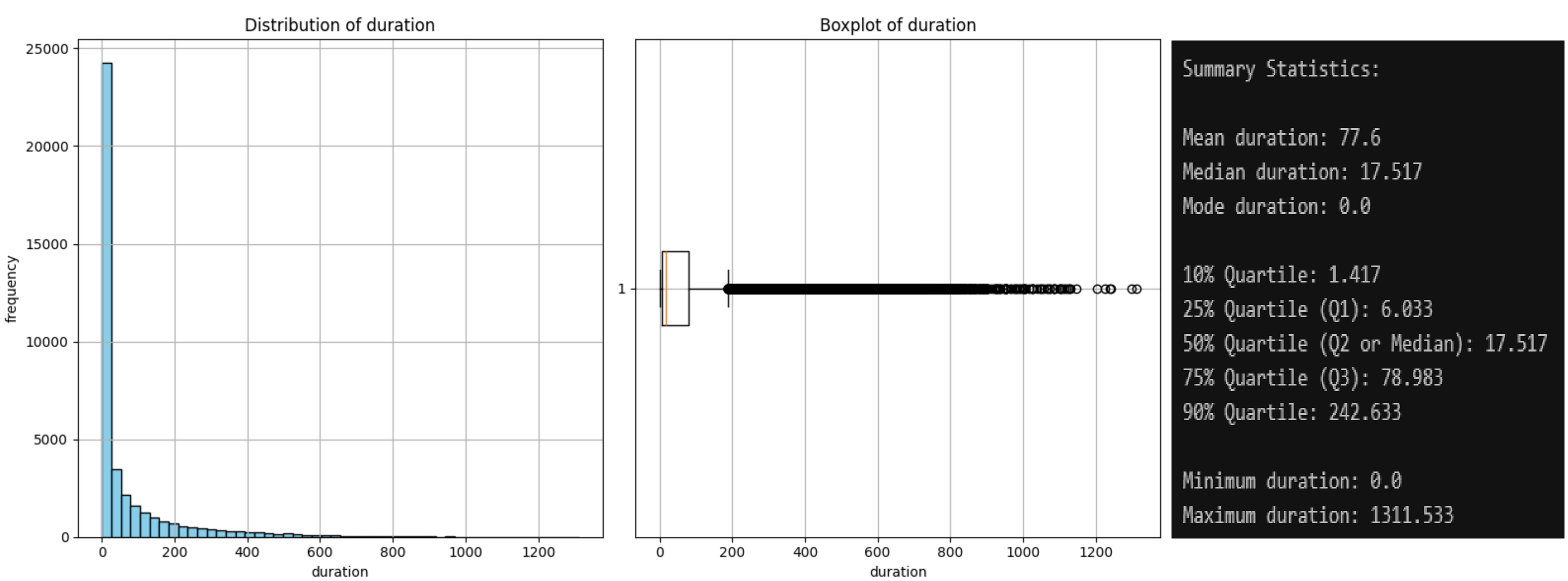

Figure 4.2: The dist. of duration of segments during four years. Statistics in minutes.

We observe a right-skewed graph suggesting that while most of the durations are relatively short, we have a few very long durations that increase the mean. Nevertheless, we can observe from the statistics that we have good conditions for mobility analysis. Mode being 0 and a low 10% quartile (1.417) imply that many recorded durations between consecutive GPS points are extremely short, possibly due to the device being stationary for many short periods or moving frequently with very short intervals between recorded points. Furthermore, the interquartile range (IQR), spanning from 6.033 to 78.983 minutes, indicates that typical durations fall within this range, with a significant portion being on the lower end. In addition, the median time between two consecutive points is 17.5 minutes, which indicates that we could have many moving segments if we have a significant enough distance between two consecutive points. This analysis could indicate days with short-length movements and bursts of motion followed by frequent stops. To better understand the dynamics, let us continue the analysis with the distribution of distances.

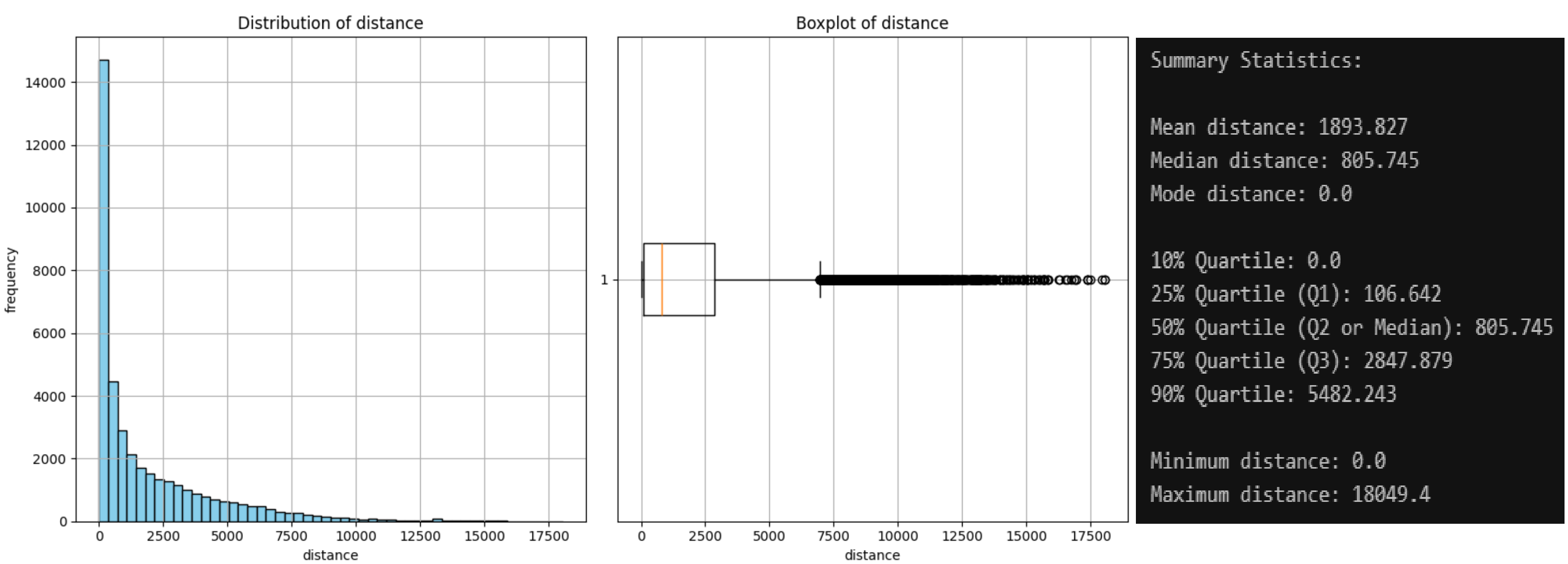

Figure 4.3: The dist. of distances of segments during four years. Statistics in meters.

Here again, we have a skewed distribution of distances. The mode and 10% quartile of 0 suggest that many segments have a distance of zero meters, implying that there are numerous cases where the GPS points are either very close to each other or possibly recorded at the same location due to device accuracy or the nature of the movements being tracked. Exploring the IQR suggests that typical segment distances vary widely, from around 107 meters to approximately 2848 meters. The presence of outliers claims that we have some long-distance consecutive points, which either can be caused by a lack of points for that day or actually be a part of a long-distance movement.

The overall mobility pattern so far shows a diverse combination of durations and distances between consecutive GPS locations in minutes and meters, respectively. The modal values for short durations and distances are frequently observed at zero, which implies many occurrences of the device being static or engaged in very short travels, possibly pointing to stops. The fact that the maximum duration and distances in both datasets were around 22 hours and 18 kilometers, respectively, means there are occasional longer journeys or periods of extended recording. The median values (17. 5 minutes for duration and 805. 7 for distance) indicate that ordinary moves are relatively short, but still, IQRs and high variability emphasize a combination of short daily moves and less frequent longer trips. This is consistent with realistic travel behavior and consists of short daily commuter trips with occasional longer trips and stops, representing various travel types in the dataset. Nevertheless, we should remember that all this information is just about two consecutive GPS points of a user within a day and still needs further processing.

3.3.3 Merging and Adjusting Segments

As we already mentioned, the micro-segments could be stationary or moving. We need to merge consecutive segments of the same modality for further processing. The algorithm for merging and adjusting micro-segments is again adopted from the work "Micro-segmenting: An Approach for Precise Distance Calculation for GPS Points" by Smith et al. The modality split is based on speed threshold of 0.5 m/s. The processed segments now represent data containing moving and stationary segments, possibly more points between the starting and ending GPS points. Therefore, we might have greater distances and duration when describing each segment. Moreover, we have points denoting the segments that have 0 distance. This segmentation process organizes the data but requires further processing to understand users' mobility behaviors clearly.

Figure 4.4: The micro-segments of all users showing movement in Yerevan, Armenia.

Merging and adjusting segments change the dynamics of our data. It can be proved by examining the duration, distance, and average speed distributions after merging and adjusting.

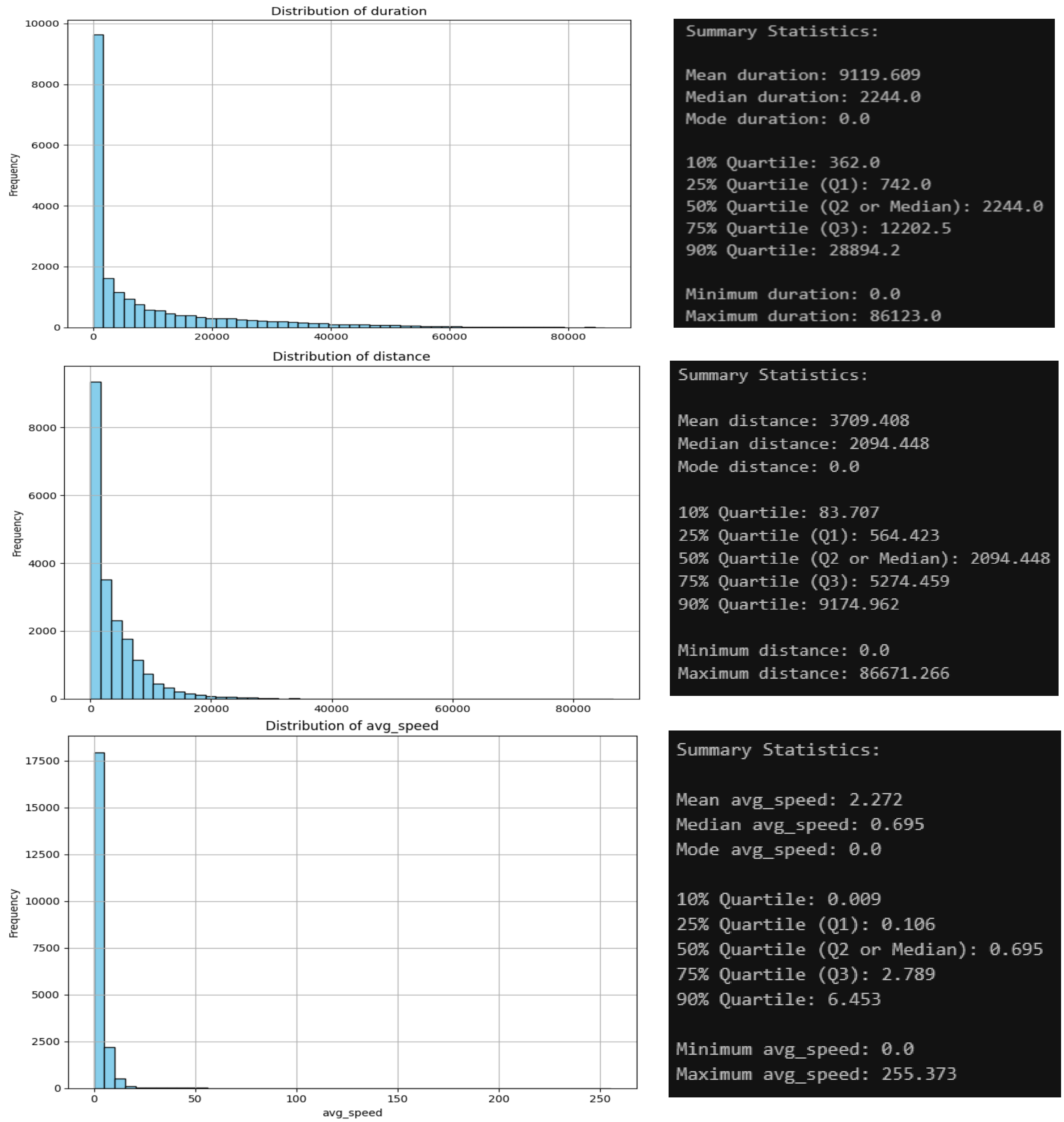

Figure 4.5: Dist. of duration, distance and average speed after merging and adjusting

However, we should remember that before having reasonable data. We must understand how to further process stationary and moving segments to create a reasonable dataset. Therefore, the created segments are still raw in terms of predicting true moving and stationary segments.

3.3.4 Trip and Activity Generation

We need further analysis to understand the user activity sites based on the segments acquired. Activity locations are where users spend a reasonable amount of time remaining within a particular site. According to the duration distributions acquired during segmentation, we have data representative enough to generate activities based on time and distance thresholds. First, previously generated staypoints are not entirely activities, but we can convert them into activities. Based on the duration distributions, we decided to take stationary segments within no less than 5 minutes and about 100 meters distance a place where a user performed an activity. Therefore, we acquired the activity sites from staypoints and merged segments, having points to use for trip generation. Usual activity locations are the home, work, and side activity locations (such as gyms, shopping malls, etc.), which help us better understand the mobility between those sites. However, the usually visited activity sides are more important places classified as locations by Trackintel library. Locations can be useful for further and more detailed analysis, revealing more than parking dynamics.

Having activity sites, we can now connect them through the street map to acquire triplets. Triplegs are small trips created between two consecutive activity locations. They contain all the necessary information to explore the path and the movement. Consequently, we can use that information to predict the mode of movement the user took to get from one activity to another. The mode detection algorithm can be as complicated as creating convolutional neural networks to predict it; however, we are stuck with simple speed, distance, and time thresholding for mobility detection. We are interested in high mobility triplegs to predict vehicle behavior around red line polygons. This research has not explored the specific identification of transportation types.

3.3.5 Stop Detection Around Red Line Zones

Stops made around the red line zones are vital when identifying vehicular behavior and mobility trends in specific areas of interest. We can develop a multi-step methodology to identify stops in those zones based on activity and trip generation data provided earlier.

First, we must separate people with activity sites or staypoints near the red line zones. After that, we devise criteria for detecting stops within these zones. A stop is a stationary segment where the duration exceeds a minimum threshold (e.g., 5 minutes) and the distance covered is less than a minimal threshold (e.g., 100 meters), ensuring that short pauses, like those at traffic signals, are not misclassified as stops.

Furthermore, by analyzing the stop data, we can infer behavioral patterns such as frequent stopping points, duration of stays, and possible reasons for stops (e.g., deliveries, pickups, or regulatory checks) and even try detecting anomalies; stops unusually long or frequent in unexpected locations signaling potential issues, such as unauthorized activities or inefficiencies in traffic management.

Therefore, trajectory creation gives us greater details and increases the accuracy of the analysis. Unfortunately, the research has not implemented the stop-detection part. However, we have enough data processed to continue the research and extract meaningful data for conclusions.

4. Results

The exploration of GPS data encompassing the years from 2019 to 2022, generously provided by Perigon AI, has yielded profound insights into the intricate dynamics of red line parking in the heart of Yerevan, particularly in the central and Arabkir regions. It is important to note that this investigation is a vital part of the research that attempted to identify and define the information about the distribution and mobility of users within the specified regions to understand urban mobility principles and parking practices.

4.1 Temporal Patterns and Trends

The monthly trends analysis revealed intriguing fluctuations in parking behaviors over the four years. Notably, observations include a decrease in the percentage of points within the red lines in 2019, suggesting a potential shift in parking practices. In addition, the year 2020 witnessed peaks in January, followed by a decline, indicative of the seismic impact of the COVID-19 pandemic on mobility patterns and parking activities. Conversely, 2021 exhibited significant fluctuations, peaking in September, possibly reflecting the complex interplay of various socioeconomic factors. Finally, unfortunately, the data for 2022 presented anomalies and disintegration rendering it less representative for conclusive analysis.

A granular examination of weekly trends confirmed the findings from the monthly analysis, emphasizing the subtle distinctions in parking behaviors between weekdays and weekends. Despite minor variations, the overall consistency in parking dynamics underscored the robustness of the observed patterns. The data collected and processed in this stage could be further used for further analysis separately and applied to weekdays and weekends. A good example of the latter is the analysis of the separately distributed durations on weekdays and weekends.

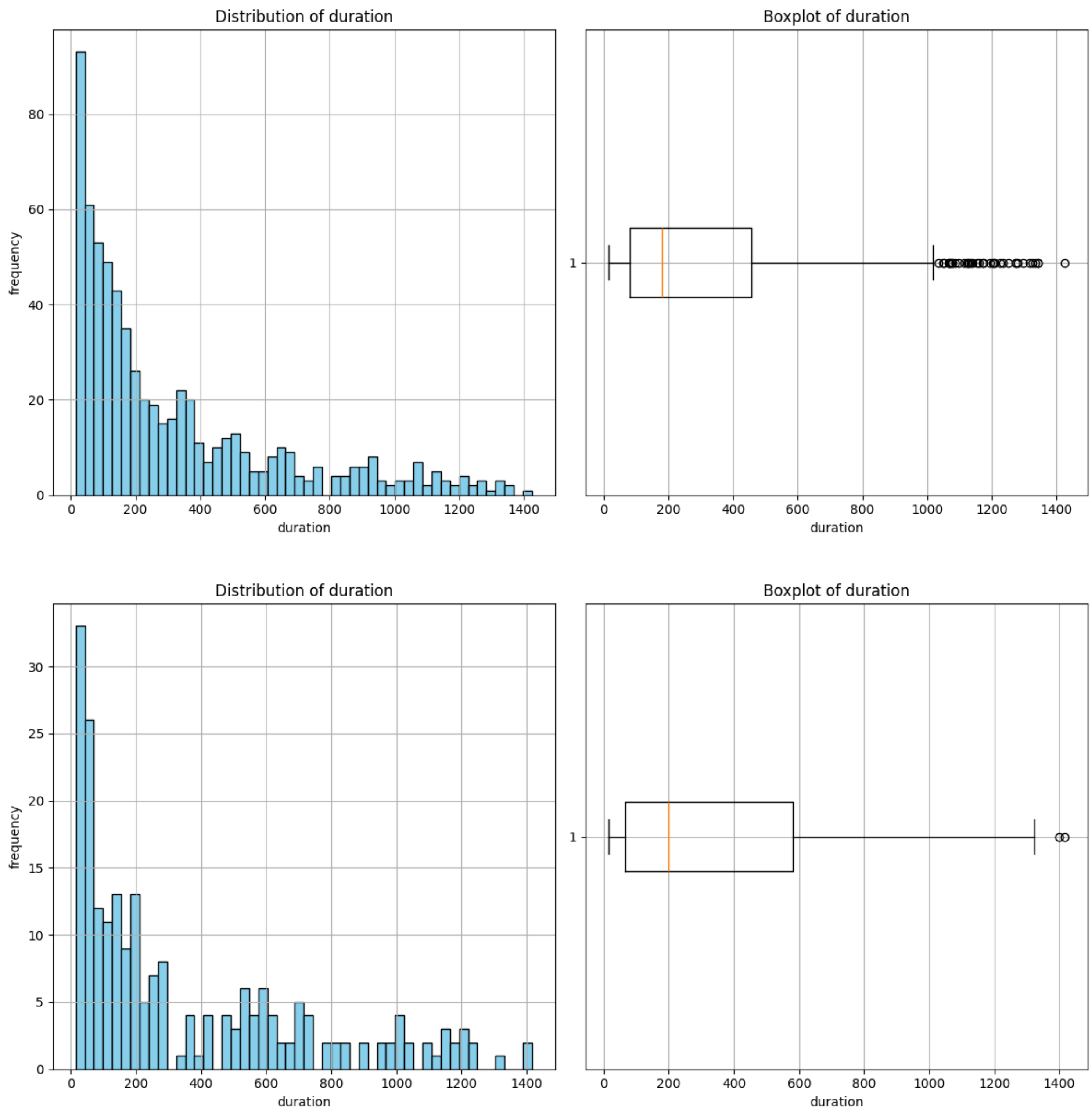

Figure 5.1: Dist. of potential parking durations inside red lines, separately by weekdays and weekends accordingly

4.2 Spatial Distribution and Density

Remarkably, spatial analyses, including heatmap visualization and assessment of unique user presence, offered valuable insights into the density and distribution of parking activities. The heatmap analysis outlined the spatial hotspots of mobility, with central polygons exhibiting heightened density, indicating elevated activity levels in the center of Yerevan. However, it is appropriate to note that density alone does not denote parking activities, as it also encompasses transient movements.

The assessment of unique user presence within red line zones uncovered interesting findings, particularly regarding the influx of visitors in 2020. The overflow in unique user counts during this period suggests heightened activity and underscores the dynamic nature of parking behaviors. With the help of the data extracted by unique user analysis, we can create graphs showing the number of potential barking events during each year presented for each day of the week as shown in Figure 5.1.

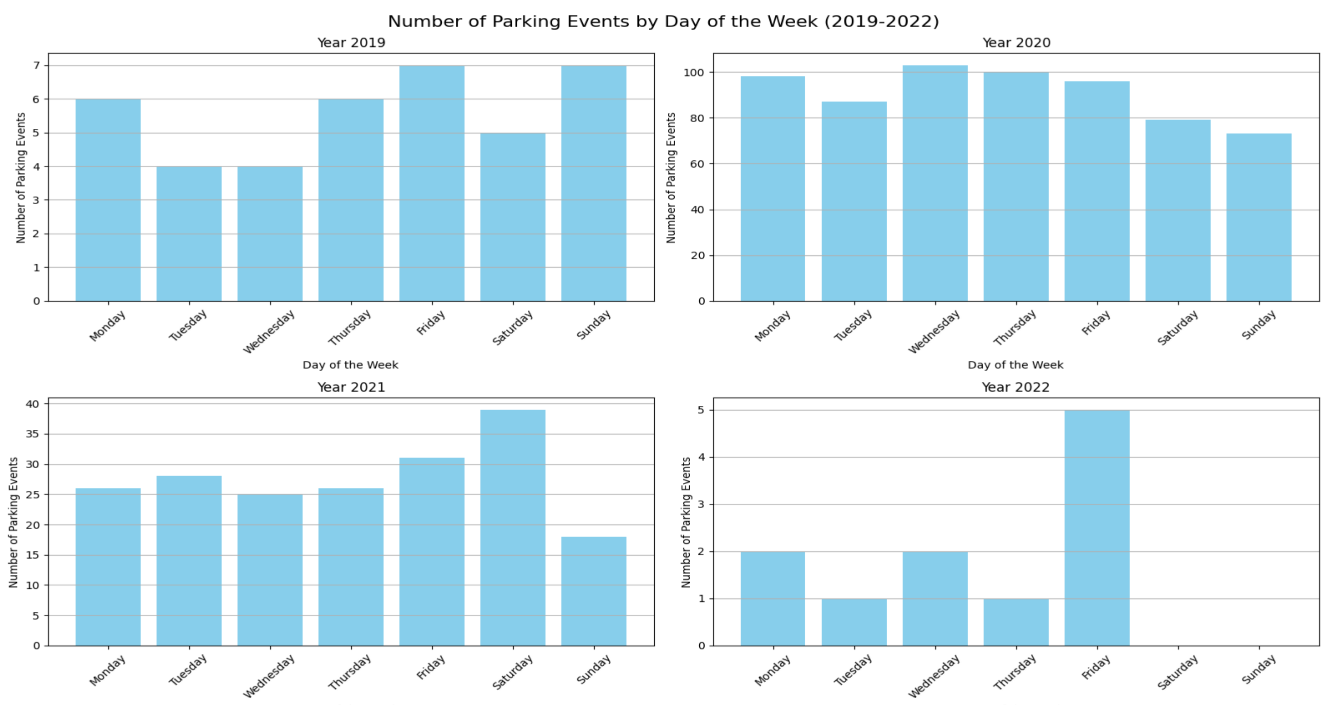

Figure 5.2: Number of parking events by day of the week over the period 2019-2022

4.3 Trajectory Analysis & Stope Detection

The trajectory analysis framework provided a nuanced understanding of user mobility patterns, encompassing diverse durations and distances between consecutive GPS points. The merging and adjustment of micro-segments facilitated the creation of coherent trajectories, enhancing the fidelity of mobility behavior analysis. Moreover, the identification of activity sites offered a comprehensive view of user mobility between key locations, including home, work, and recreational sites. The segments of a user's movement in a day can be observed in Figure 5.2, where each line string describes a merged and adjusted path. Indeed, these segments are not yet triplegs and need to be processed further to identify of activities, and only then tripleg creation.

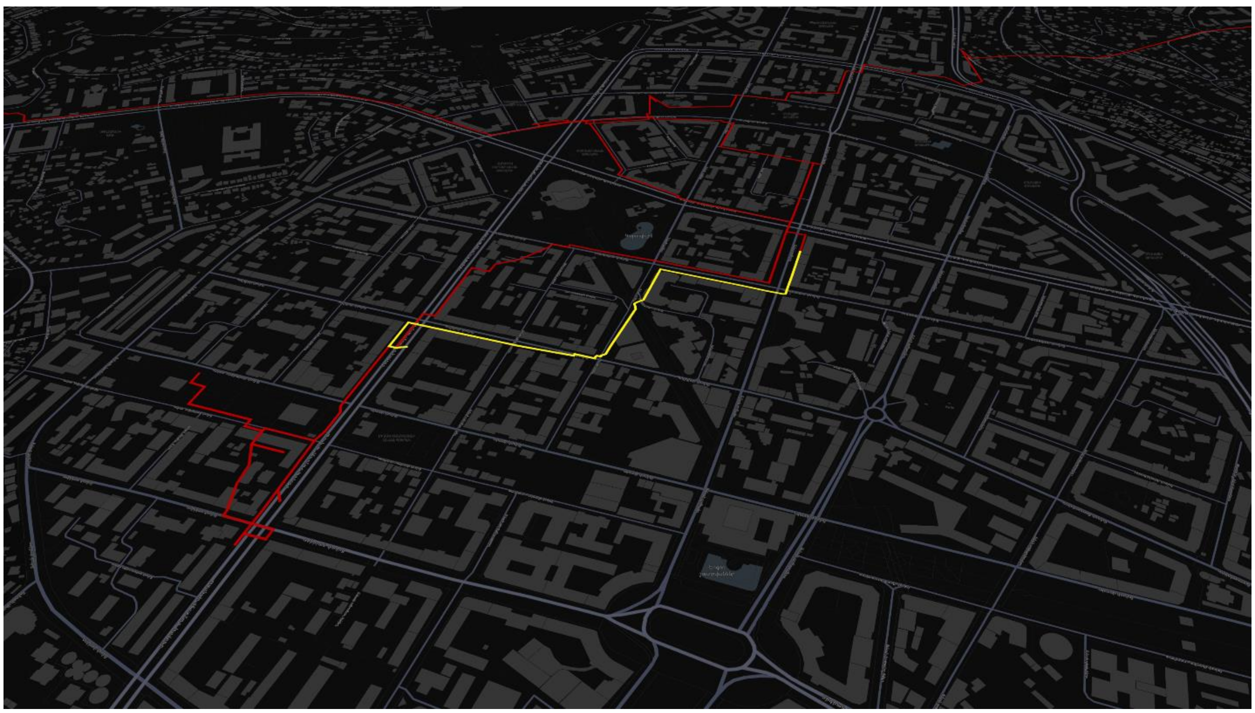

Figure 5.3: Adjusted and merged segments of a user within a day. Each discrete line shows moving or stationary segments.

The stop-detection algorithm is meticulously crafted to determine vehicular behaviors and mobility trends around red-line zones. The segmentation of stationary segments exceeding predefined thresholds facilitated the identification of frequent stops, explaining the underlying behavioral patterns such as deliveries, pickups, and regulatory checks. Therefore, stop detection has a higher accuracy in the prediction of potential parking due to detailed analysis based on user mobility. Furthermore, analyzing time spent within red lines provided insights into parking durations, with notable variations observed across different years. Unfortunately, stop detection is not implemented due to time constraints. However, the research provides a great fundament for the algorithm.

5. Future Directions

We carried out two approaches to explore the red line parking dynamics over four years. We managed to extract valuable insights that could be used for urban planning and strategy development. Nevertheless, future research endeavors could focus on refining our approaches to get a better, more detailed, and more accurate analysis.

The density analysis can be significantly enhanced for future research through several refinements. Incorporating temporal weighting can reveal hidden patterns by assigning different weights to GPS data points based on time, thus highlighting peak hour dynamics. Moreover, integrating additional data sources like traffic cameras, sensor data, and parking meter transactions can validate and enrich GPS data, offering representative data and a great view of parking dynamics. Considering contextual factors like land use, population density, and socioeconomic variables can explain observed density patterns and enhance urban planning. Indeed, machine learning techniques for pattern recognition and anomaly detection can be used to identify recurring patterns and deviations. Finally, assessing the impact of policy changes and interventions on density patterns can offer valuable feedback for urban planning and policy-making. These refinements will enable a more detailed and accurate understanding of parking dynamics, supporting effective urban planning and policy decisions.

To continue with, trajectory creation and stop detection give more room for interpretation and refinement:

- The activity generation can occur for each user not only within a certain day but also considering the sites visited over the years. This will reveal some patterns in the movement of users and might even identify different parking sites in red lines as staypoints, increasing the accuracy of the analysis.
- Tripletgs could be used for trip generation, creating usual and unusual routes people take within a day or over the years.
- More sophisticated algorithms can be used for mode detection, increasing the accuracy of the transportation mode prediction. The latter can be achieved with unsupervised or semi-supervised machine learning algorithms.
- Extending the detailed analysis, we can focus more on each person, trying to categorize them and understand whether they own a car or not.
- Combining this with the usual locations a person visits, we can be more certain that they are parking their car while detecting a stop near red lines.
- Lastly, the density analysis and trajectory creation can be mixed to support each other's choices statistically. The trajectories could be created only for people who appeared near the red line polygons, or the accuracy can be refined for those who are proven to visit the polygons more often.

In a word, we can take this further by taking advantage of every single piece of information we can extract from just raw GPS data.

Future research endeavors can refine the research to perfection, creating a model that extracts far more than we did. They can use our work as a solid foundation and a kickstart for later analysis. Such advancements are imperative for fostering sustainable urban mobility and ensuring the seamless functioning of transportation systems in modern urban landscapes.

6. Conclusion

The analysis conducted on the GPS data provided by Perigon AI over the 2019-to-2022-time horizon provided key information regarding parking factors in red line zones in Yerevan's center and Arabkir region. Utilizing a series of powerful libraries for data processing and spatial analysis, the gained data has been filtered and cleaned to highlight only certain regions of interest in the city where the analysis needs to be performed.

Some notable trends were observed through the temporal analysis, such as the reduction of red line parking in 2019, the peaks and troughs associated with 2020, possibly due to the COVID-19 pandemic, and the variations in 2021. Data anomalies in 2022 may also indicate external factors affecting parking activities, but that also seems to be a less representative year.

The heatmap and unique user analysis helped identify areas of mobility and activity in Yerevan. Central polygons had higher densities, signifying higher activity levels, but density does not necessarily have conclusive relationship with parking behavior. Density analysis helped us understand distributions of durations inside red lines and extract potential parking events.

Trajectory analysis, on the other hand, gave a clear picture of the user's movement and discovered patterns from GPS points by identifying separated trajectories. It was also interesting to see that there is a variety of different time spans and distances between consecutive points, as it happens. It helped us extract more details about the user and lay a foundation for understanding the dynamics near red-line parking areas. However, there was not enough time to implement the stop-detection algorithm, which would have helped define the border of each parking event.

The study's findings underscore the dynamic nature of parking and mobility within Yerevan. Temporal and spatial trends offer valuable insights for urban planners and policymakers. These insights can aid in developing more effective parking management strategies and understanding the broader impacts of socioeconomic factors on urban mobility.

Future research can build on this work by incorporating more sophisticated machine learning techniques for mode detection, integrating additional data sources, and refining algorithms for greater trajectory and stop detection accuracy. These advancements will foster sustainable urban mobility and improve transportation systems in modern urban environments.

7. References

Chen, X., Ma, Z., & Wu, Y. (2017). Road Network Travel Time Calculation Using GPS. Journal of Transport Geography, 60, 65-75.

Griffin, R., Winstanley, C., & Paternò, F. (2020). Trackintel: An Open-source Python Library for Human Mobility. SoftwareX, 12, 100619.

Hovsepyan S., (2024). Yerevan Parking Dynamics Analysis [Source Code]. GitHub. https://github.com/Sargis-Hovsepyan/yerevan_parking_dynamics_analysis

Jahangiri, A., & Rakha, H. A. (2015). Travel Mode Detection Based on GPS Raw Data Collected by Smartphones: A Systematic Review of the Existing Methodologies. Transportation Research Part C: Emerging Technologies, 60, 1-18.

Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., & Ma, W. Y. (2008). Mining User Similarity Based on Location History. In Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (p. 34). ACM.

Schuessler, N., & Axhausen, K. W. (2009). Processing GPS Raw Data Without Additional Information. Transportation Research Record: Journal of the Transportation Research Board, 2105(1), 28-36.

Smith, J., Brown, T., & Wilson, R. (2016). Micro-segmenting: An Approach for Precise Distance Calculation for GPS Points. International Journal of Geographical Information Science, 30(4), 687-703.

Stopher, P. R., Fitzgerald, C., & Zhang, J. (2008). Search for a Global Positioning System Device to Measure Person Travel. Transportation Research Part C: Emerging Technologies, 16(3), 350-369.

Wang, J., Yang, X., & Zhao, Z. (2019). Searching for Parking in GPS Data. Computers, Environment and Urban Systems, 77, 101341.

Zheng, Y. (2015). Trajectory Data Mining: An Overview. ACM Transactions on Intelligent Systems and Technology (TIST), 6(3), 29.

Zhao, F., Tian, X., & Ma, W. (2015). Identifying Transportation Modes from Raw GPS Data. Journal of Transportation Technologies, 5(2), 113-122.