

Data Visualization

Artsvik Avetisyan

Ani Harutyunyan

Gayane Hovsepyan

December 06, 2024

GG Taxi

DS116 / CS343

Final Project

American University of Armenia
Yerevan, Armenia

Abstract

The report analyzes the early operational data of GG Taxi, a ride-hailing service in Yerevan, Armenia. The research focuses on user behavior, operational efficiency, and key performance indicators, highlighting critical aspects such as time-based determinants of taxi performance, geographical demand distribution, user delays, and even user habits. Two global hypotheses are considered during this project. Firstly, the report proves that people mainly request rides when going to work from home and vice-versa. Additionally, it shows whether there are some latency patterns among people. This research provides valuable insights into the early trends, demand distribution, and operational challenges faced by GG Taxi during its formative years, offering actionable guidance for a diverse range of stakeholders. Ride-hailing startups can use these findings to optimize operations, avoid common pitfalls, and better enter the market to meet their needs. Urban planners and transportation authorities can leverage demand distribution data to enhance public transport systems and traffic management. Established ride-hailing companies expanding into similar markets can refine their strategies based on these insights, and technology providers can develop targeted solutions to address operational inefficiencies. Furthermore, readers can use this case study to deepen their understanding of the interplay between mobility, technology, and socio-economic dynamics in developing urban environments.

Introduction

This project report analyzes data from GG Taxi to uncover patterns and trends that could help improve their service. Originally, our dataset contained a large volume of data, but many records had missing or inconsistent entries. We conducted extensive data cleaning and processing to address these issues and ensure the reliability of our analysis. Drawing inspiration from global studies of taxi services, such as those conducted in New York and Chicago, our analysis incorporates methods and insights that have proven effective in other urban contexts. These studies have also helped us shape our approach to examining GG Taxi's operations, particularly through spatial and temporal analyses. We focused on how factors like time of day, weather conditions, and the start and end locations of rides influence service usage. This includes determining whether rides primarily start in residential areas and head towards the city center during morning hours, which could inform strategic fleet management and fare adjustments. This focused analysis not only highlights specific challenges and opportunities for GG Taxi but also contributes to the broader discussion on optimizing urban taxi services using data-driven strategies.

Literature Review

Taxi ride analysis involves examining patterns in ride demand, duration, fares, and geographic distribution, often to optimize operations and improve user experiences. Extensive work has been done in cities like New York and Chicago. However, studies in Armenia still need to be expanded, making it essential to explore insights from international contexts and assess their applicability. NYC taxi data is very famous, and numerous analyses were performed on this dataset after the NYC Taxi and Limousine Commission's (TLC) trip records were released publicly. Studies have explored spatial and temporal demand patterns, identifying high-demand zones and peak travel times. Similarly, in Chicago, taxi ride analysis has leveraged the Open Data Portal, which includes trip records. Research has examined the impact of taxi services like Uber and Lyft on traditional taxi demand. The data frames had similarities with our GG dataset, and these projects inspired some visualization techniques that applied to our data.

Data Cleaning and Preprocessing

The dataset used for this project initially consisted of **2,661,633 observations** and **16 variables**. However, it included a significant number of missing values and inconsistencies, which required extensive preprocessing and cleaning to ensure data quality and relevance for analysis.

Data Cleaning

- Rows where the taxi ride was canceled (**869,582 rows**) were removed, and the cancellation column was dropped.
- Observations with missing destination coordinates were excluded, as these were critical for further analysis and could not be reliably imputed.
- Rows with a travel distance of less than **500 meters** were eliminated. These primarily represented test rides taken by newly joined drivers, which were not considered valid for analysis.

Imputation of Missing Distance Values

- Missing distance values were imputed using the **K-Nearest Neighbors (KNN) method**. For each observation with a missing distance value, we identified observations where the absolute difference in origin and destination coordinates was less than a predefined threshold of **0.003**. The average distance of these neighboring observations was then assigned to the missing values.
- After completing these cleaning and imputation steps, the dataset contained **1,709,400 observations** while retaining the original 16 variables.

Time-Based Feature Engineering

To facilitate temporal analysis, new columns were introduced:

- A **period** column categorized the time of day into four parts:
 - **Night**: 00:00–05:00
 - **Morning**: 05:00–12:00
 - **Afternoon**: 12:00–17:00
 - **Evening**: 17:00–00:00
- A distinct **date** column was derived from the original time variable, which was stored in **POSIXct** format.

Fare-Based Feature Engineering

A **price per kilometer** column was computed to reflect fare structures:

- For rides after April 2016, this value was calculated as **fare/distance**.
- For rides before April 2016, historical fare formulas were used, assigning **100 dram per kilometer** based on the standard pricing model (**600 dram for the first 3 km + 100 dram per km**).

Exclusion of Non-Yerevan Rides

Observations representing rides outside of Yerevan were deemed irrelevant to the analysis and removed. This filtering reduced the dataset to **1,564,761 observations**.

Spatial Feature Engineering

To enhance geographic analysis, origin and destination districts were assigned to each ride:

- A shapefile containing the latitude and longitude boundaries of Yerevan districts was imported.
- The ride origin and destination coordinates were converted into **simple feature (sf)** objects.
- Using spatial operations, the coordinates were mapped to their respective districts within Yerevan.
- The orders were also classified by **origin_district** and **dest_district**.

Time-Based Features

These features facilitate temporal analysis of rides by breaking down time variables into meaningful categories:

- **hour (integer)**: The hour of the day (24-hour format) when the ride was accepted. Useful for studying hourly trends in ride demand.
- **day (factor)**: The day of the week when the ride was accepted (e.g., Monday, Tuesday). Enables weekly pattern analysis.
- **month (numeric)**: Represents the month (1–12) of ride acceptance, useful for seasonal trend analysis.
- **season (character)**: Indicates the season (e.g., Winter, Spring) in which the ride occurred, helping in identifying seasonal variations.

Ride Duration Metrics

Columns like `ride_duration`, `duration_category`, `distance_category`, `is_late`, `response_time`, `speed_kmh` were added to the data to have outlined metrics which contribute to ride overall duration from request time to completed time.

Final Dataset

After completing these preprocessing and feature engineering steps, the dataset was finalized with the following attributes:

- **1,567,461 observations**
- **30 features**, including newly engineered columns such as *time period*, *price per kilometer*, and *origin/destination districts*.

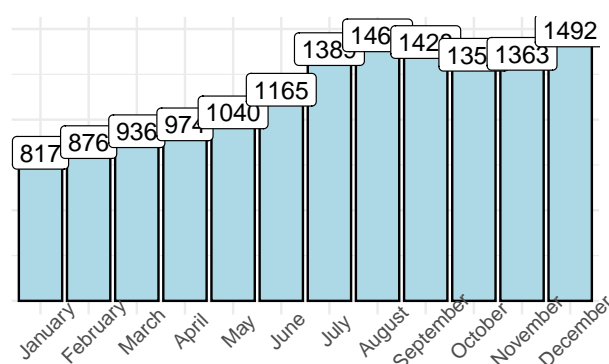
This comprehensive cleaning and transformation ensured the dataset was consistent, relevant, and enriched for a detailed analysis of Yerevan taxi rides.

General Findings

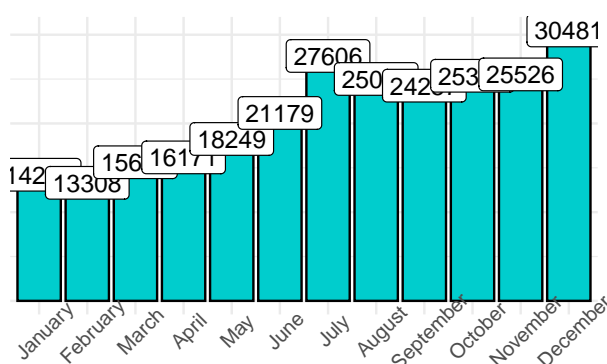
In this section, we delve into the insights uncovered during the analysis of GG taxi rides, focusing on how temporal factors such as the day of the week and the time of day influence taxi performance metrics.

Additionally, a significant shift in GG's pricing and operational policies was observed following the entry of Yandex Taxi into the market. This competitive dynamic appears to have influenced GG's fare structures, particularly after April 2016, when new pricing models were implemented. By analyzing these changes alongside ride performance data, we provide a comprehensive view of how market competition and time-based factors have shaped GG's services over the analyzed period.

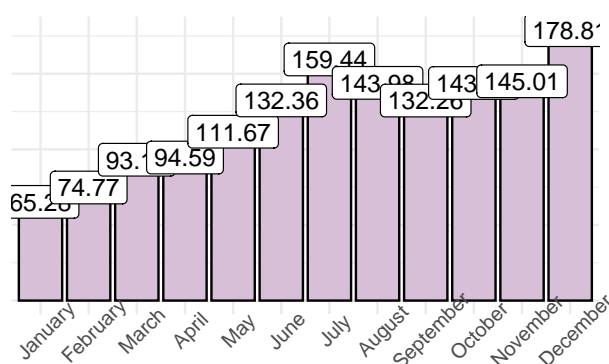
Number of taxi drivers of GG over months in 2016



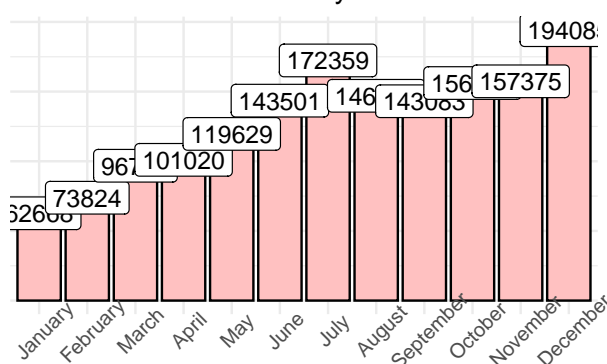
Number of GG users over months in 2016



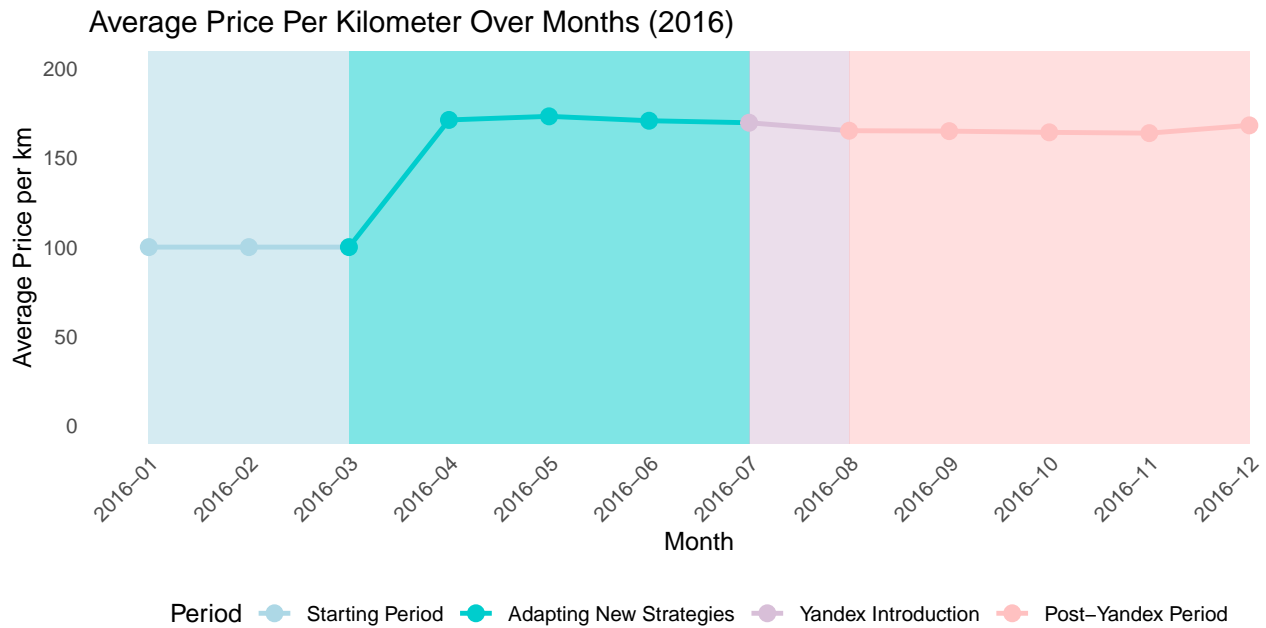
Revenue of GG over months in Mln drams in 2016



Number of orders done by GG over months in 2016

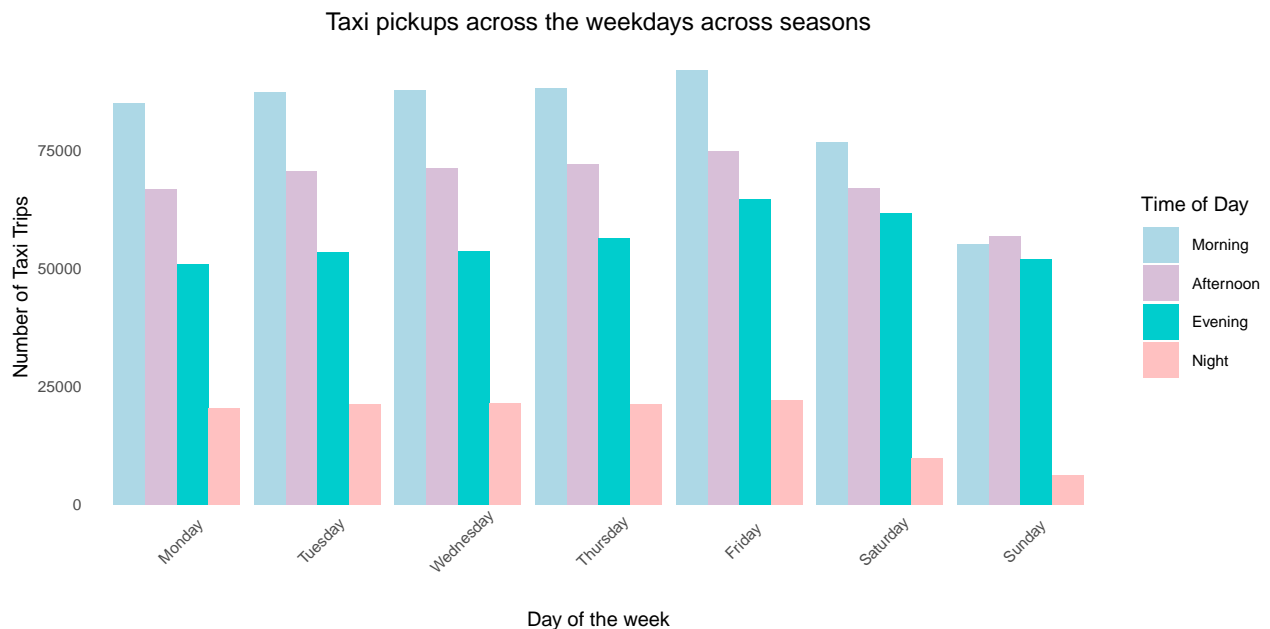


From the graphs, it is evident that all performance metrics for GG Taxi—number of drivers, number of users, revenue, and completed orders—experienced significant peaks in July 2016. This month stands out as a period of rapid growth, which suggests that GG implemented strategic efforts to bolster its position in response to external market pressures, particularly the introduction of the Yandex Taxi platform in Yerevan. The arrival of a major competitor like Yandex Taxi likely disrupted the market dynamics, prompting GG to launch aggressive marketing campaigns to maintain customer loyalty and attract new users. The increased number of drivers in July suggests that GG may have also invested in expanding its fleet, possibly through recruitment campaigns or incentives for new drivers to join the platform.

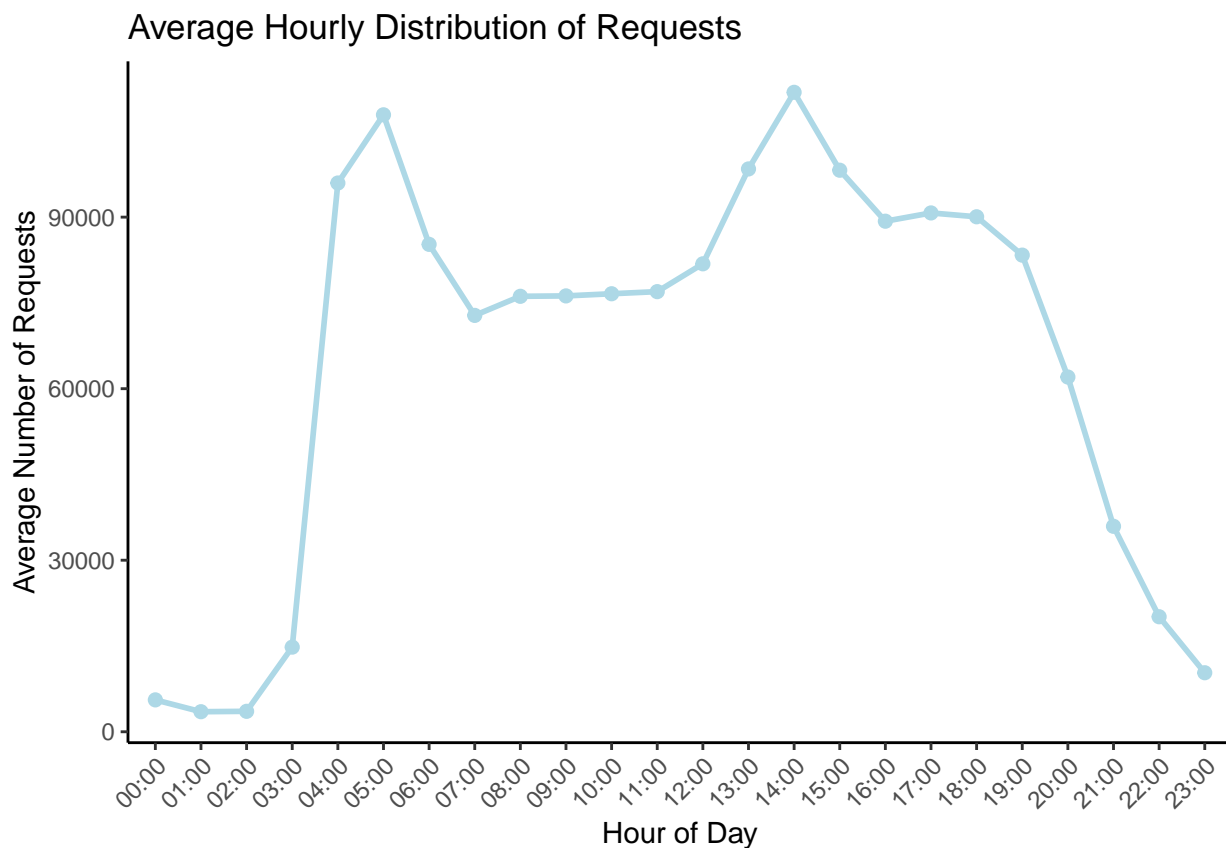


The graph shows the pricing policy of GG Taxi during the initial periods of its operation. As the graph shows, the prices before March were constantly 100AMD, the way it was in old traditional Armenia.

However, after this the prices experience a significant rise, which may be due to some economic factors or based on competitor's strategies. In July The Yandex Taxi entered the Armenian market. In this period there is a reduction in prices which can be due to the fact that GG was really trying to keep its customers and was lowering prices to not lose its place. After this period for 3 months the prices stayed stable, but started to increase in December, which may be connected to different factors, one of which might be holidays.



This graph illustrates the distribution of taxi demand across different times and days of the week. It highlights that the demand patterns remain relatively consistent for weekdays within each period. Specifically, the morning period consistently sees the highest number of taxi trips, likely due to work-related travel, and school. Conversely, night trips are consistently the lowest across all weekdays, reflecting reduced activity during late hours. On weekends, however, a notable shift occurs in the demand distribution. While morning and afternoon periods maintain a steady level of demand, the night period sees a significant drop compared to weekdays. This suggests a decrease in late-night travel activity over weekends, possibly because residents of Yerevan exhibit different mobility habits during the weekend, like staying home, which reduces the need for nighttime taxi services.



The graph illustrates the hourly distribution of taxi requests in Yerevan, revealing distinct patterns in daily demand. The data shows that demand increases sharply in the early hours of the day, reaching a notable peak at 14:00. This mid-afternoon surge in demand likely corresponds to a combination of factors, including lunch-hour activities and school dismissals. The graph also highlights a steady rise in requests starting from the early morning as people go to work or other daily routines. However, after the peak at 14:00, the demand gradually declines into the evening and drops significantly after 20:00, likely reflecting the end of workdays and reduced travel activity during late-night hours.

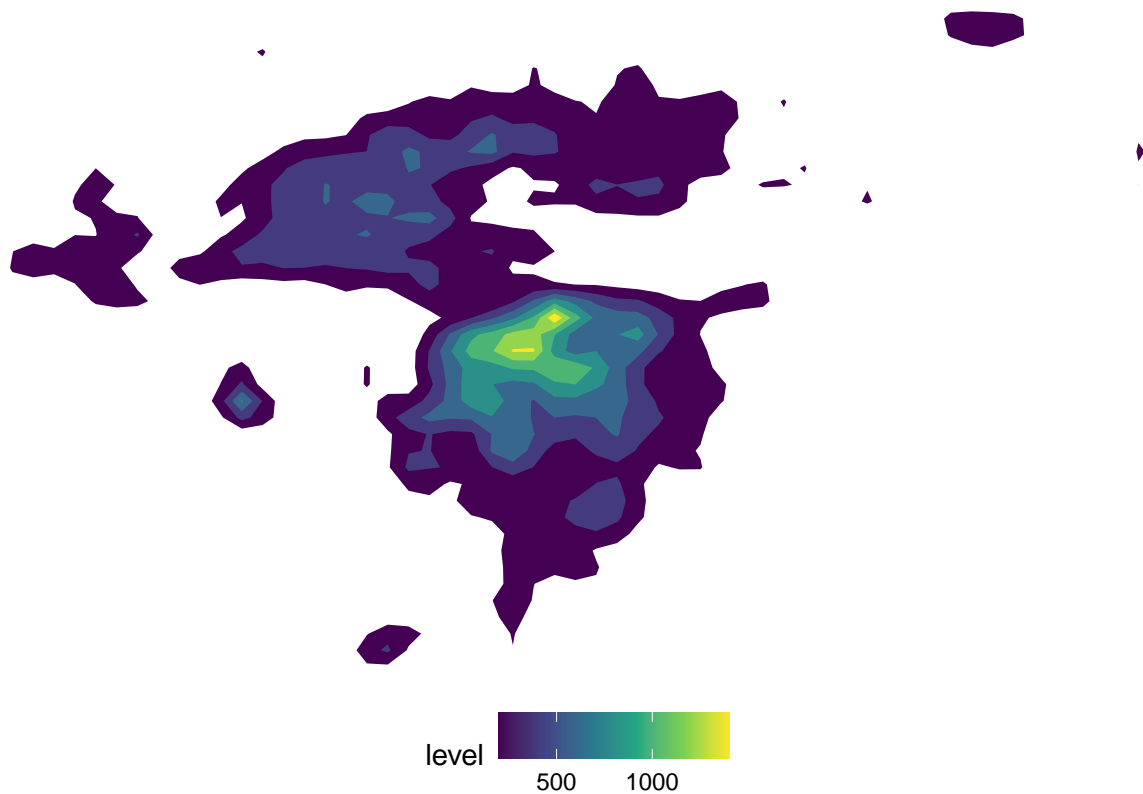
Hypothesis Analysis

Hypothesis

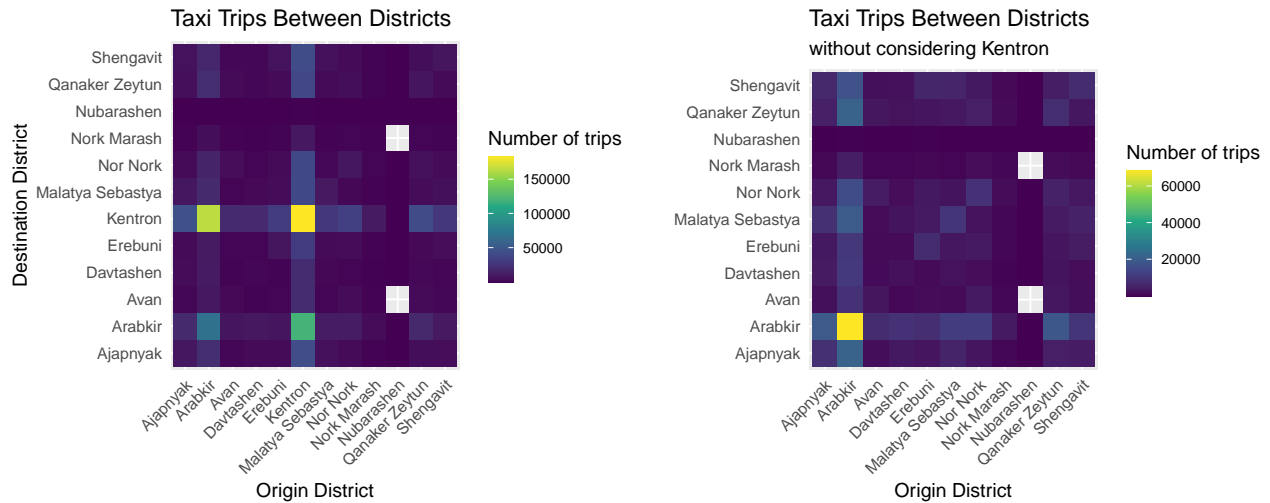
Most taxi rides in Yerevan originate from residential areas, with the city center (Kentron) being the primary destination, especially in the morning. Kentron also serves as a major hub for taxi activity. Similarly, evening rides predominantly return passengers to residential areas, suggesting a pattern of commuting behavior where the most common route is home-to-center and center-to-home.

To assess whether Kentron is indeed the primary hub for taxi rides, we first examined the spatial distribution of taxi origins in Yerevan throughout the year 2016. A heatmap was generated to visualize the concentration of taxi pickups across the city. The intensity of the color on the map, with a yellowish hue indicating higher concentrations, represents areas with more frequent taxi activity.

Heatmap of Trip Origins



The results reveal that Kentron exhibits the highest concentration of taxi pickups, as indicated by its prominent yellow coloration. In contrast, other districts display considerably lower levels of taxi activity, reflected by cooler colors on the map. This suggests that the central district of Kentron is indeed the main area from which taxis are most frequently hailed, supporting the hypothesis that Kentron plays a dominant role in taxi movement within Yerevan.



Origin-Destination Analysis

To further investigate taxi ride patterns, a heatmap matrix is used, where the x-axis represents the origin district and the y-axis represents the destination district. The heatmap highlights that the Kentron-to-Kentron route has the highest demand, represented by a bright yellow cell. Other significant routes include Arabkir-to-Kentron and Kentron-to-Arabkir, shown in greenish hues, followed by the Arabkir-to-Arabkir route with a light blue color.

The influence of Kentron is evident, as its corresponding rows and columns on the heatmap display consistent activity, reinforcing its role as the central hub.

Excluding Kentron from the heatmap reveals that the Arabkir-to-Arabkir route becomes the most prominent, with approximately 60,000 rides. However, this is significantly lower than the Kentron-to-Kentron route, which records 150,000 rides.

This analysis confirms that Yerevan functions as a monocentric city, with Kentron as the focal point of taxi movement. While Arabkir shows moderate activity, it does not rival Kentron's dominance in the city's taxi network.

Insights from High-Frequency Taxi Users

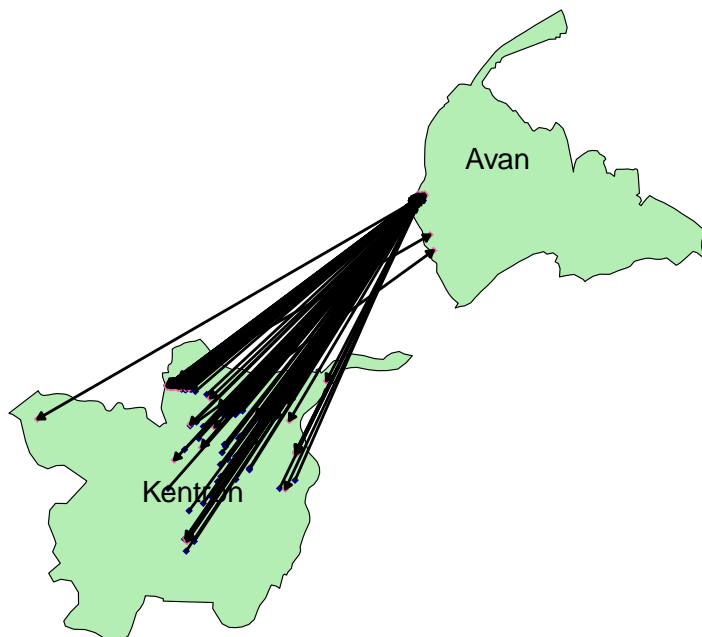
To extract meaningful insights, additional steps were necessary. Given that the dataset covered taxi rides for an entire year, it included many users who had only used the service once. These observations were deemed less relevant for uncovering patterns. To address this, the dataset was filtered to include only users who completed more than 500 rides within the year.

This filtering resulted in a subset of 72 clients, representing the most frequent taxi users. For each client, the data was grouped by the routes they took, and only their two most popular and repeated routes were retained for further analysis.

Visualizations were then created to analyze these users, both collectively and individually, to identify potential patterns. While aggregate visualizations for all users provided some insights, the overlapping of routes reduced the clarity of the patterns. Therefore, individual user cases were examined in detail to uncover distinct trends in their taxi usage.

Home Location as a Key Starting Point One of our hypotheses posits that most taxi orders either originate from users' home locations or have that destination. The visualization shown bellow supports this hypothesis, as it demonstrates that for the users shown in the graph, the majority of taxi rides either originate from a specific location (labeled X) or are destined for it. This location is most likely the users' home, given its centrality in their ride patterns.

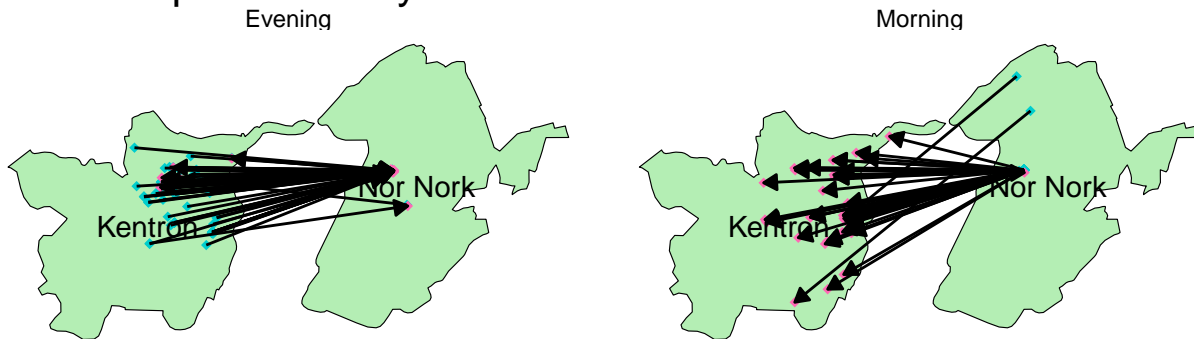
Spatial Distribution for Taxi Rides of user



Temporal Patterns of Taxi Usage

The next step was to examine whether users tended to leave their homes in the morning and return in the evening. To investigate this, we conducted a focused analysis on these users. We began by creating individual spatial visualizations for these users. The results revealed intriguing patterns for some, providing deeper insights into their daily taxi usage behaviors.

Spatial Distribution of Taxi Rides of user 5681 Based on period of day



Case Study: User 5681

In this specific example, we analyze the taxi ride history of User 5681. The visualization reveals a clear pattern: in the morning, the user consistently travels from Nor Nork to Kentron, and in the evening, they leave Kentron and return to Nor Nork. The data shows that all ride points in Nor Nork cluster around the same location, strongly indicating that this is the user's home.

Using facet wrapping, we can observe that most rides taken in the morning originate from the home location and end in Kentron, while in the evening, approximately 100% of rides start in Kentron and return to the home location.

A similar individual example specifically for the user 12496 is presented below to further clarify this recurring pattern and provide additional insights into user behavior.

Spatial Distribution of Taxi Rides of user 12496 in Yerevan



An interesting outlier case appeared when we identified a user who made 1,502 taxi orders over the course of one year, averaging 4 rides per day. This is notably high usage compared to the general trend. What stood out in this case was that there was no recurring route or destination for this user, and none of the destinations overlapped with previous ones. This presented a unique situation, as typically, users follow a pattern of using taxis between consistent locations such as home, work, or other frequently visited places.

Initially, we considered the possibility that multiple family members might be using the same account, but upon further analysis, this explanation did not seem plausible. The variety in destinations—without any returning trips to previous locations—also ruled out the idea that this might be a regular office commute. Instead, our hypothesis shifted towards the idea that this user might be involved in some form of delivery activity. The pattern of ordering taxis multiple times a day to different locations suggested that the user could be using taxis for deliveries, possibly for a business, such as a supermarket or a similar establishment. In this case, the lack of repeated destinations and the high frequency of orders made sense, as the user was likely coordinating deliveries to various locations across the city.

To visualize this further, an animated graph showing the history of rides for this user (ID: 13031) over the course of the year has been created. The animation demonstrates the user's daily ride patterns, revealing the diverse destinations each day without repetition. This visualization reinforces our hypothesis of the user being engaged in deliveries or similar activities, as the movement across different locations aligns with this interpretation.

This extraordinary case highlights the diversity of taxi usage patterns and provides insight into the broader spectrum of possible behaviors within the dataset.

<https://drive.google.com/file/d/1BDWON-3sHC9XX2c3mNpNlt9DeIHRO7lQ/view?usp=sharing>

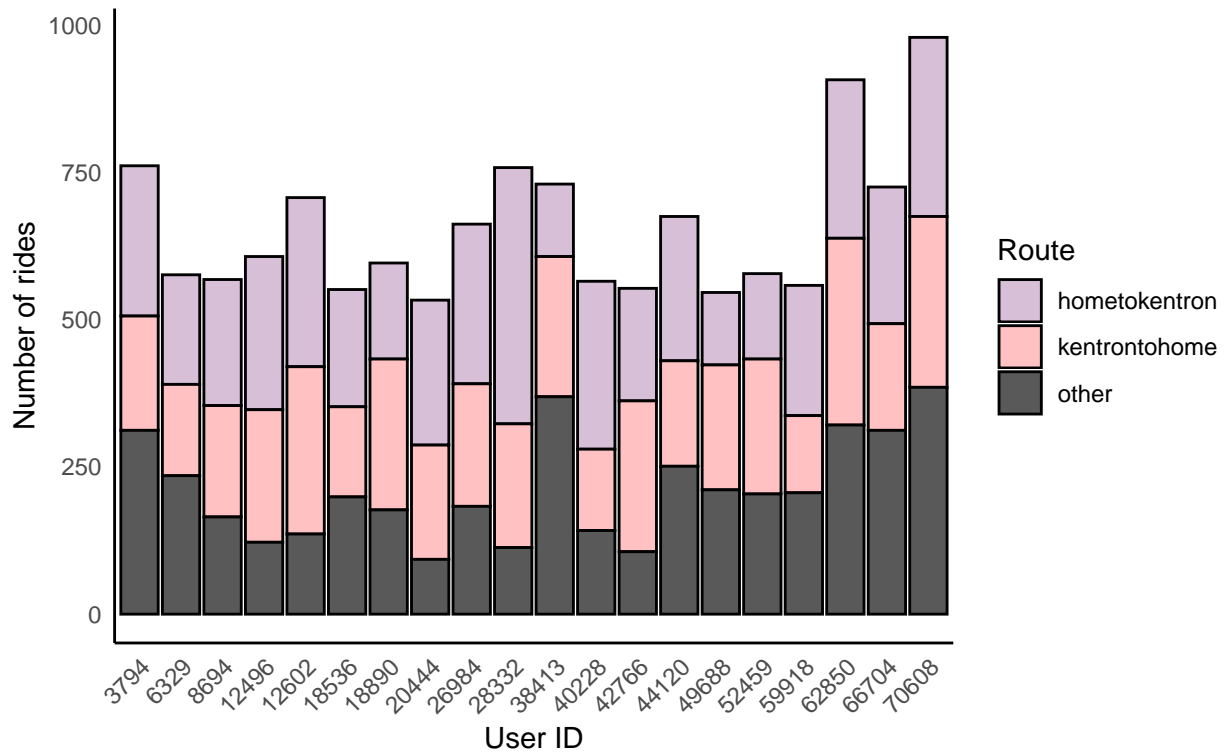
General Analysis of Ride Patterns

With a better understanding of individual user behavior, we proceeded to conduct a more general analysis by examining data for the top 20 clients collectively. For this analysis, we compiled all their rides and added a new column to classify each ride into one of three categories: Home-to-Kentron, Kentron-to-Home, and Other.

This categorization aimed to determine the proportion of these ride types within each user's total rides. To visualize the results, we created two bar charts: one displaying the absolute numeric values of each ride type and the other showing their percentages relative to the total rides.

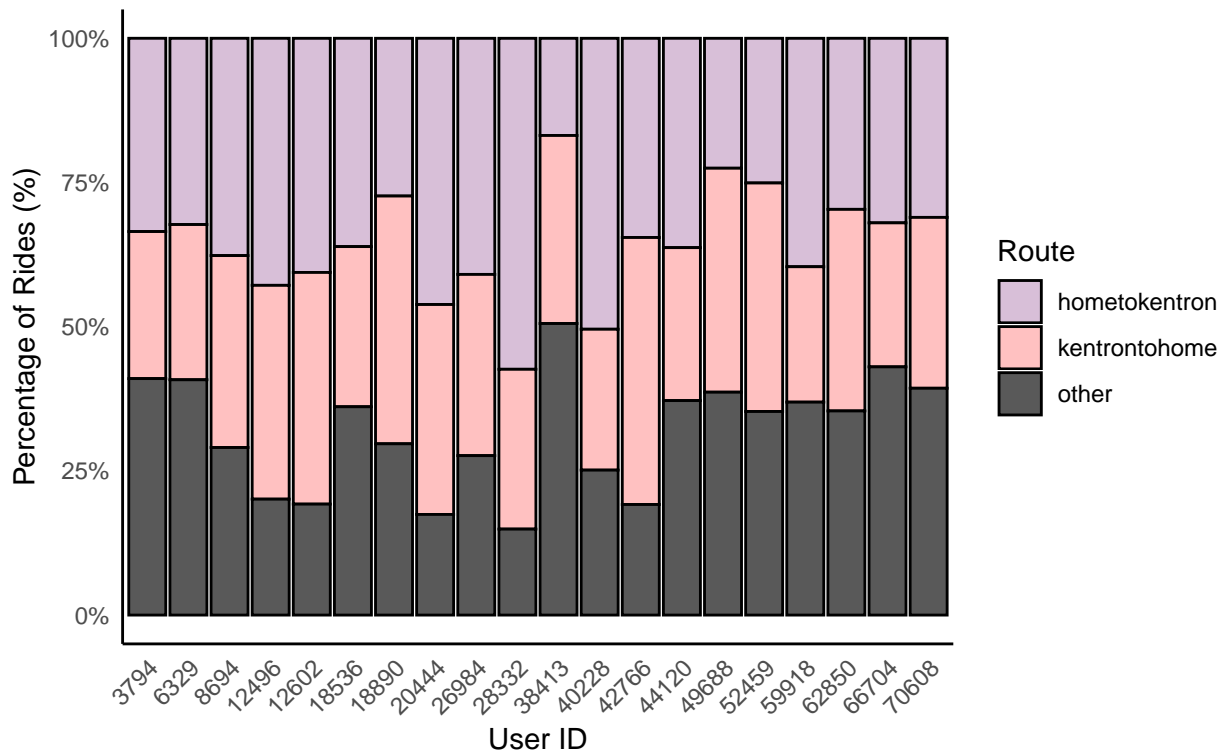
Distribution of Ride Routes for the Top 20 Clients

Home-to-Kentron & Kentron-to-Home vs Other



Distribution of Ride Routes for the Top 20 Clients

Home-to-Kentron & Kentron-to-Home vs Other



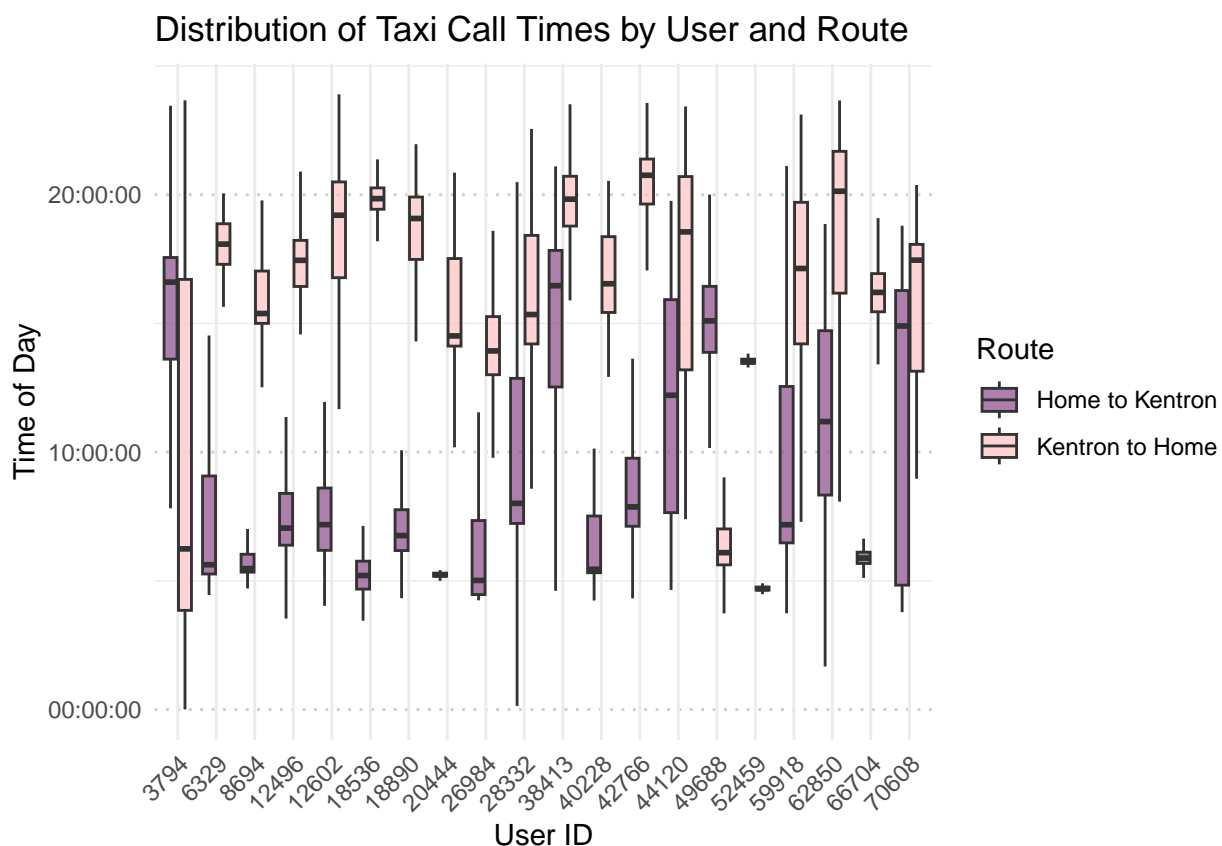
Insights from Bar Charts

The bar charts provide a clear visual representation of the data, supporting our hypothesis that users predominantly use taxis for consistent, repetitive routes. Specifically, the charts confirm that the majority of taxi rides are taken for the same purpose: users travel from their home to Kentron in the morning and return home in the evening. This pattern is evident in both the absolute values and the percentage breakdowns shown in the charts.

Temporal Analysis of Taxi Usage

Having established the location patterns, we now focus on the timing of taxi usage for the most common routes. To do this, we excluded rows categorized as “Other” and retained only the Home-to-Kentron represented with light purple color and Kentron-to-Home represented with pink color.

Next, we created boxplot visualizations to illustrate the time at which each of the 20 users called a taxi for their journey to Kentron and their return trip home. These boxplots provide a clear view of the distribution of ride times for both routes, offering insights into the temporal patterns of taxi usage for these users.



Observations from Boxplot Analysis

The boxplot visualizations reveal a clear temporal pattern in the users' taxi usage. For the morning ride (Home-to-Kentron), the boxplots are narrow, indicating that there is little variance in the time users called taxis for this route. This suggests that most users tend to call a taxi at a consistent time in the morning.

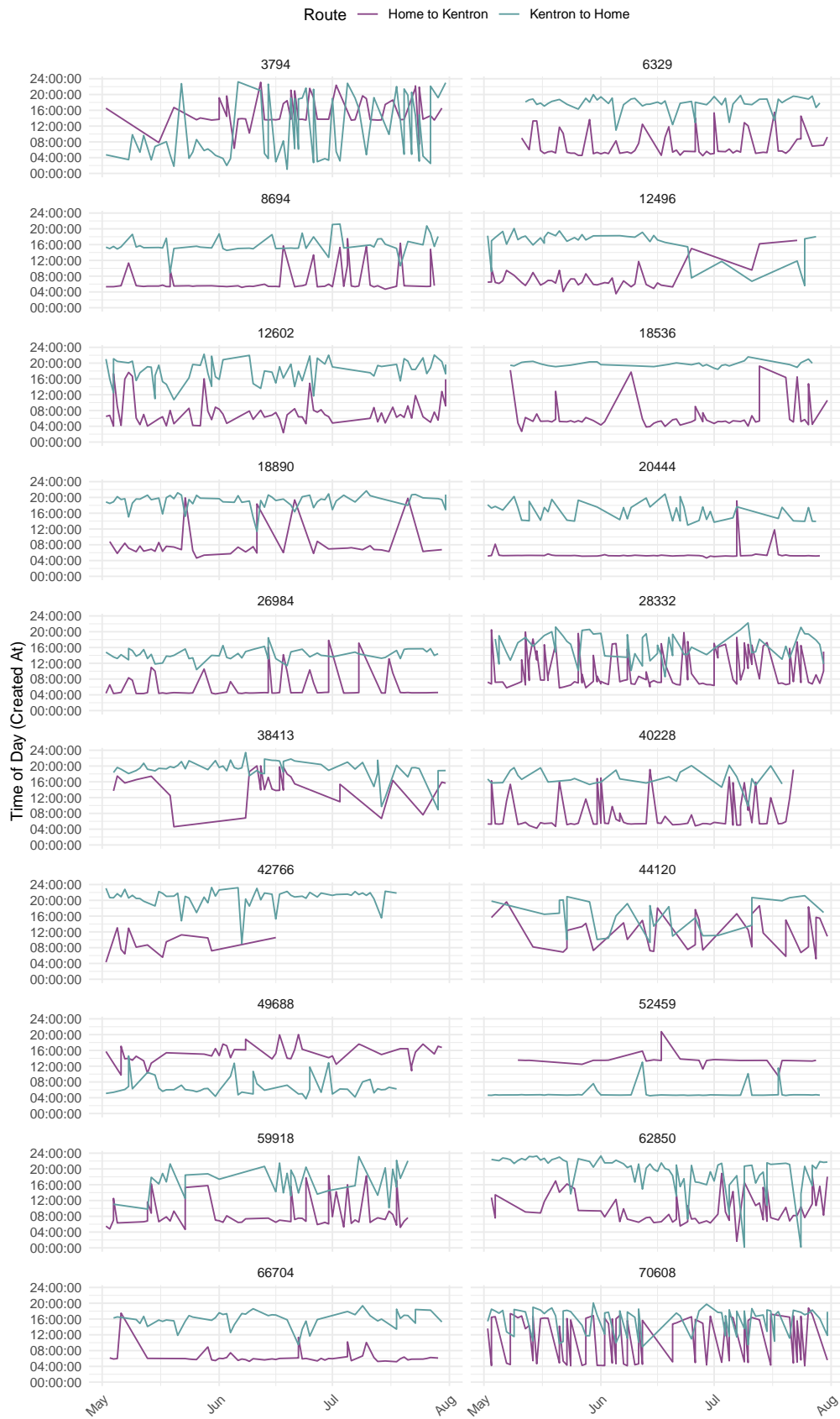
In contrast, the evening ride (Kentron-to-Home) shows wider boxplots, indicating greater variability in the time users called taxis for the return trip. Although there is more fluctuation in the evening, the data still indicates a general trend, with most users calling taxis around 7-8 AM in the morning and between 5-7 PM in the evening. Also we can see that user 49688 and user 3794 had their route times swapped, that is, they left Kentron in the morning and they came back in the evening. This can suggest different things, either their job started and ended at times different from the other users, or they lived in Kentron and they constantly left Kentron for their job location. This is a topic of further investigation, however, since we lacked extra information about the context, we could not tell the exact reason of this happening.

In-Depth Analysis of Taxi Usage Timing

To further investigate the timing patterns, we created individual plots for each of the 20 users. Each line in the graph represented the time at which a user called a taxi for a specific route. For each user we have 2 colors, the purple one shows the time they called taxi to go from their home district to Kentron, while the color blue represents the time they used taxi to go back to their home district from Kentron.

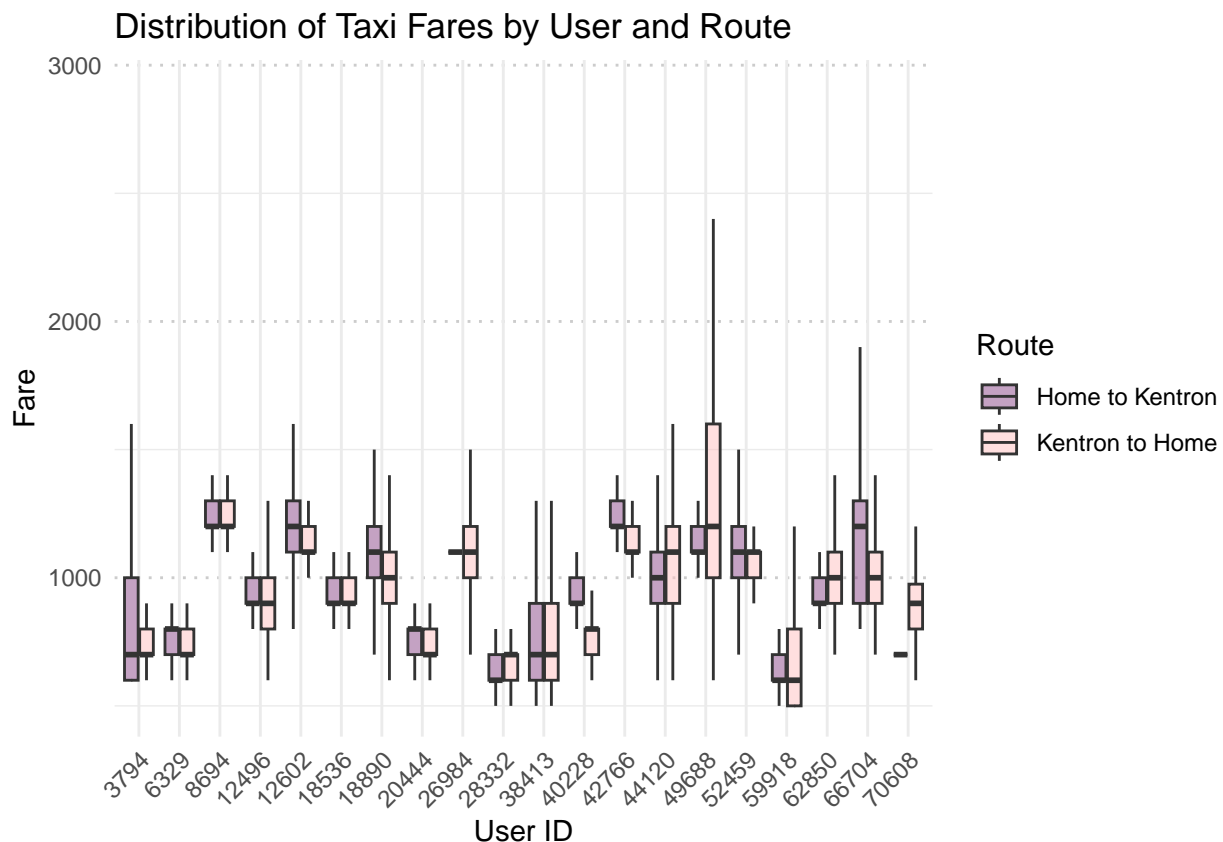
Since displaying the entire year's worth of data in a single plot would result in an overwhelming number of data points, we focused on data from three consecutive months chosen randomly. This allowed for a clearer analysis.

Time of Activity by Date



The findings varied among users. For some, there was a clear, consistent pattern in their taxi usage, with small deviations in timing. These users demonstrated regular, predictable travel times, likely indicating the routes were used for commuting between home and work. However, for other users, the timing patterns were more erratic, showing significant fluctuations without a clear trend. These inconsistencies could indicate different factors influencing their taxi usage, such as varying daily routines or personal schedules.

Unfortunately, the dataset lacked additional contextual information to fully interpret these irregularities. Based on our analysis, we concluded that users with consistent patterns were likely following regular commuting routes, while those with more varied patterns may have had different objectives or lifestyles, affecting their taxi usage timings.



Price Distribution Analysis

Next, we conducted a boxplot analysis to examine the fare distribution for each of the 20 users, categorized by the Home-to-Kentron and Kentron-to-Home routes. Similar to the timing analysis, the boxplots revealed that the prices for these routes were quite consistent, with values clustering closely for both trips. This suggests that users regularly traveled along the same routes with same distances.

This finding provides further evidence that many users followed consistent commuting patterns, traveling from their home location to another destination and then returning home at the end of the day.

Analysis of Taxi Movements by Time of Day

To further highlight the centrality of Kentron District and demonstrate how taxi distribution changes throughout the day, we developed an interactive UI visualization. This visualization displays two maps: one for taxi origins and one for destinations, based on a selected time period and date.

The analysis revealed a clear pattern: throughout the day, Kentron consistently hosts the highest concentration of taxi movements, both in terms of origin and destination. In the morning, the majority of taxi origins were from districts outside of Kentron, with destinations primarily in Kentron. Conversely, in the evening, taxis typically originated from Kentron and had destinations in other districts.

This pattern further reinforces the hypothesis that most riders travel from their home districts to Kentron in the morning and return home in the evening. The UI also includes a table displaying the number of taxi origins and destinations per district, along with the total number of taxis present in the city at each time point. These numbers corroborate the observation that taxis from other districts tend to converge in Kentron during the day and are then redistributed across the city in the evening.

The analysis of taxi rides in Yerevan in 2016 revealed clear patterns that confirm the centrality of Kentron District as the hub for taxi movements in the city. Through various methods, including heatmaps, route visualizations, and temporal analysis, we demonstrated that the majority of taxi rides originated from or were destined for Kentron, with a strong pattern of travel from home districts to Kentron in the morning and a return trip in the evening.

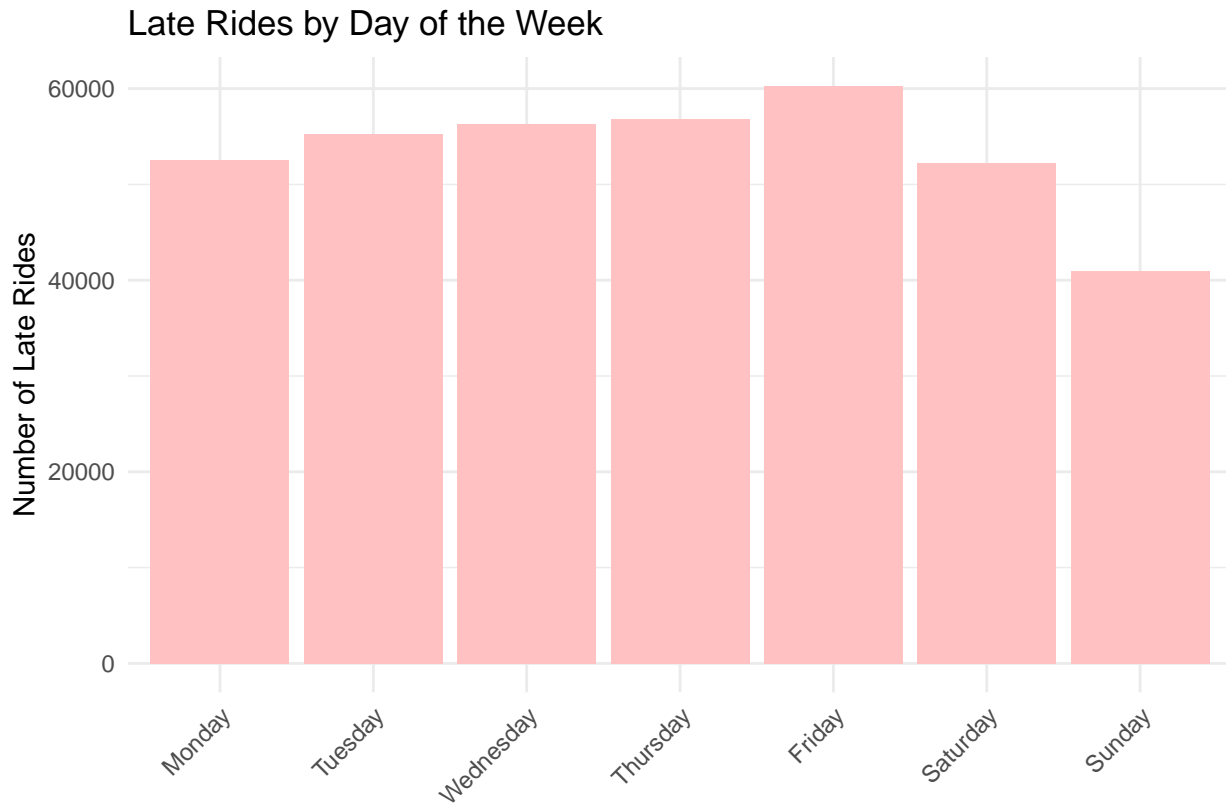
The findings suggest that Kentron plays a crucial role in the daily commutes of Yerevan's residents, reinforcing its position as the city's primary commercial and transportation center. The patterns of usage, such as the consistency in time and route selection for regular users, further highlight the structured and predictable nature of daily taxi movements.

Overall, this analysis provides a comprehensive view of Yerevan's taxi system, illustrating not only the centrality of Kentron but also the regularity and significance of commuter behavior in shaping transportation trends in the city.

https://drive.google.com/file/d/1mK5_nK_6ODaDtwZFOC7MDsjC55jMeHh3/view?usp=sharing

Hypothesis Analysis about Lateness Tendency

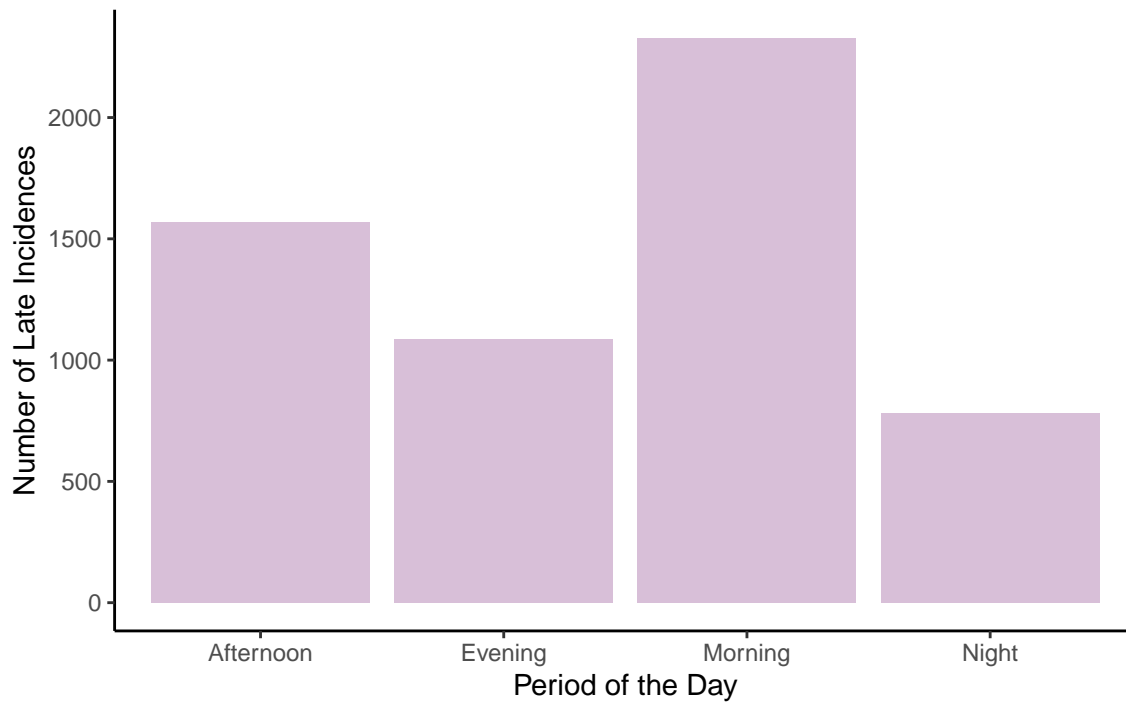
Hypothesis: “Users tend to be later as the week progresses due to cumulative fatigue, with a drop in late rides during the weekend due to fewer commitments.”



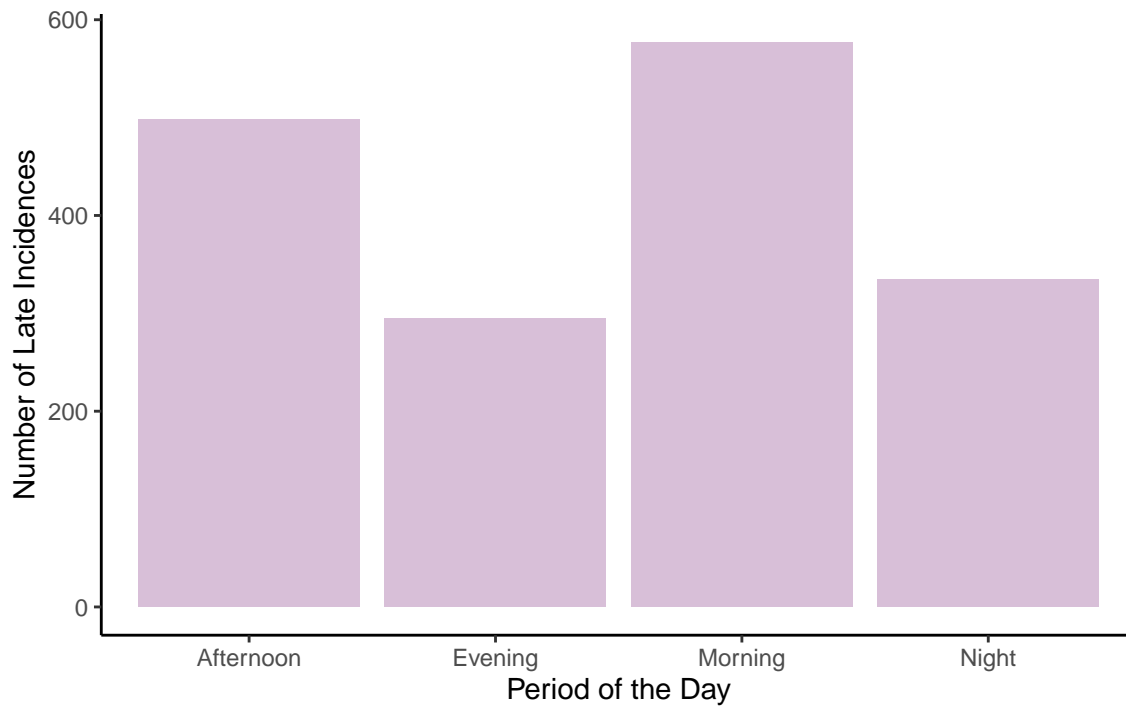
The hypothesis is accepted. The analysis reveals a distinctive trend in the occurrence of late rides throughout the week. Starting on Monday, there is a gradual increase in late rides, peaking on Friday. This pattern could reflect the cumulative burden of professional and personal responsibilities that tend to escalate as the week progresses. Each subsequent day adds a layer of potential delays, whether due to extended work hours, increased traffic, or reduced personal energy, which could impede timely arrivals. Interestingly, the data shows a sharp decline in late rides during the weekends. Saturdays and Sundays exhibit significantly lower instances of tardiness compared to weekdays. This shift might be attributed to fewer occupational obligations and a more relaxed schedule, allowing users greater control over their time and a reduced likelihood of encountering delays. Such insights are crucial for understanding user behavior and optimizing service delivery, suggesting that ride services might need to adjust resource allocation and scheduling strategies based on these weekly patterns to better accommodate user needs and improve punctuality.

Hypothesis: “Lateness varies significantly across different times of the day, especially in the morning, potentially influenced by the morning rush or oversleeping.”

Number of Late Incidences by Period for Top 20 Late Users

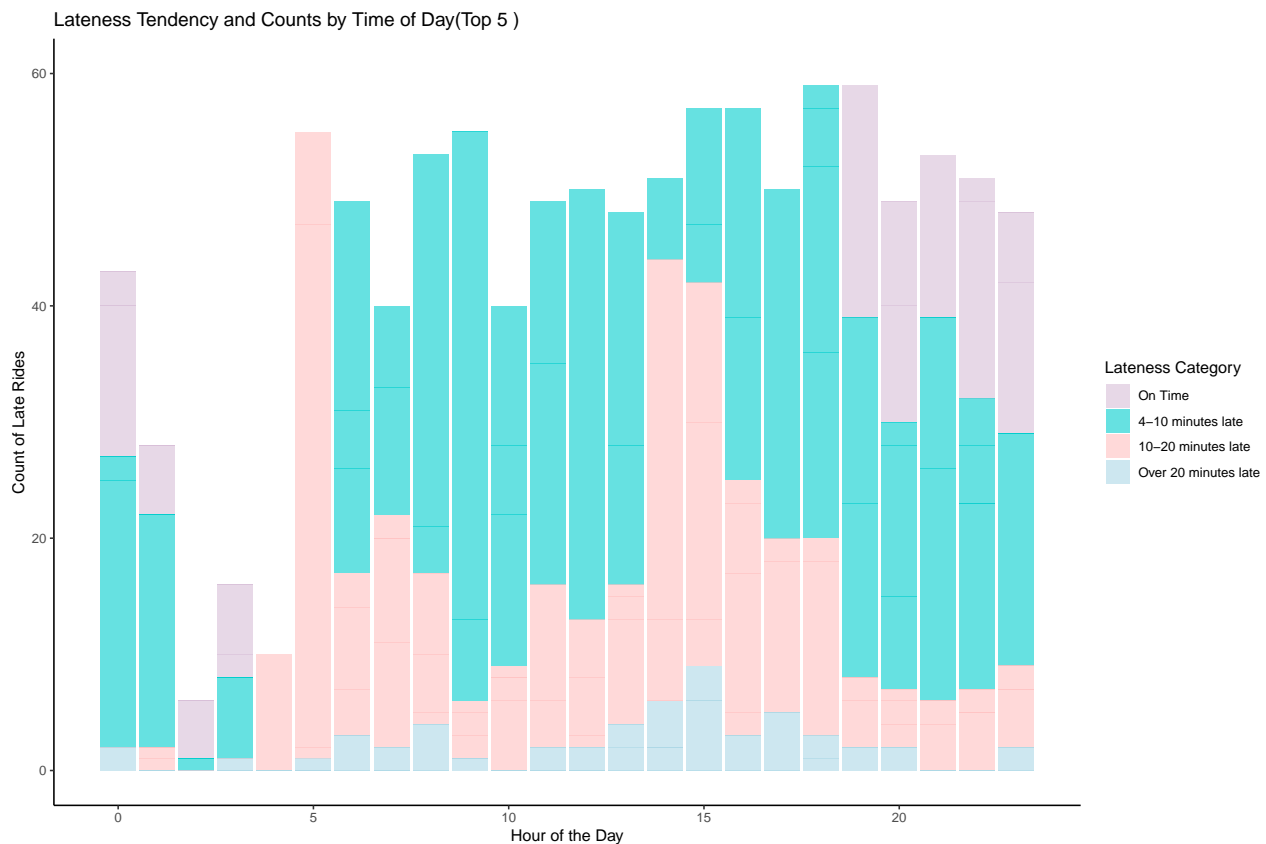


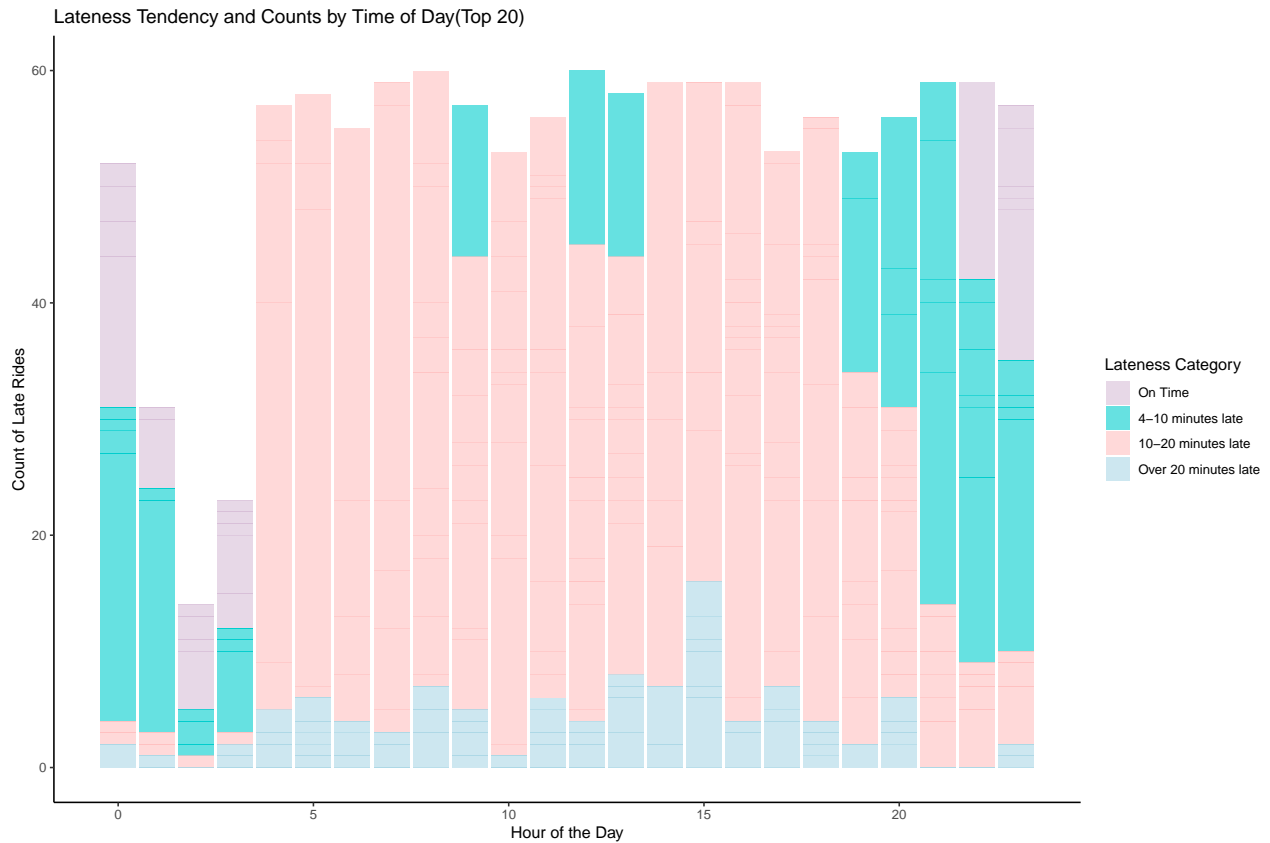
Number of Late Incidences by Period for Top 5 Late Users



The hypothesis is accepted. The analysis of late incidences by period, illustrated through two graphs representing the top 5 and top 20 late users, reveals significant daily timing patterns. The morning period consistently records the highest frequency of late rides across both user groups, reflecting challenges such as the morning rush or oversleeping. Among the top 5 late users, afternoon incidences are slightly lower than in the morning but still significant, while night incidences are higher than those in the evening. For the top 20 late users, the pattern is almost similar, with morning delays dominating, followed by the afternoon, while night and evening incidences change their places compared to the top 20 late users. This trend highlights the morning as a critical period for addressing lateness, driven by predictable factors like traffic congestion and individual preparation routines. Proposed interventions to mitigate lateness during this peak period include real-time traffic updates, flexible work or school start times, and encouraging habits that promote punctuality. Such strategies could address the unique challenges of the morning rush and improve timeliness across user groups.

Hypothesis: “Top 20 late users exhibit more consistent lateness patterns throughout the day compared to the top 5 late users, whose lateness patterns vary significantly depending on the time of day.”

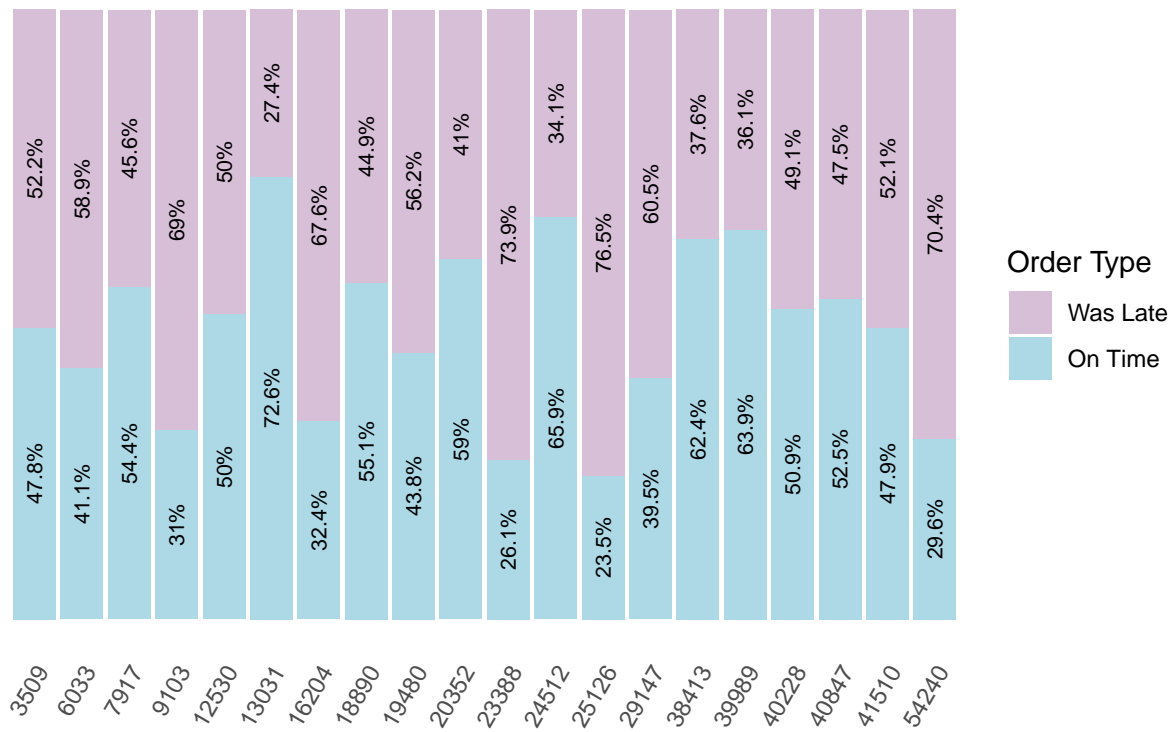




The hypothesis is accepted. The analysis of lateness behavior across the top 5 and top 20 late users reveals distinct patterns that support the hypothesis. For the top 20 late users, the graphs illustrate a relatively consistent trend of lateness across the hours of the day, with the majority of late rides falling within the 10-20 minute range. This consistency suggests that these individuals have habitual lateness tendencies that are less influenced by external factors, maintaining a steady pattern throughout the day. In contrast, the top 5 late users display more variability in their lateness patterns. The lateness counts fluctuate significantly across different hours, with noticeable peaks at certain times of the day. This indicates that the lateness behavior of the broader group is influenced more by external factors such as work schedules, traffic congestion, or personal routines, leading to more unpredictable lateness patterns. The comparison between the two groups supports the hypothesis: the top 20 late users exhibit more consistent lateness behavior throughout the day, while the top 5 late users show greater fluctuations in their lateness across different hours. This contrast affirms the hypothesis, highlighting the role of individual habits versus external influences in shaping lateness tendencies.

Hypothesis: “People who are late more than 200 times have a habitual tendency to be late.”

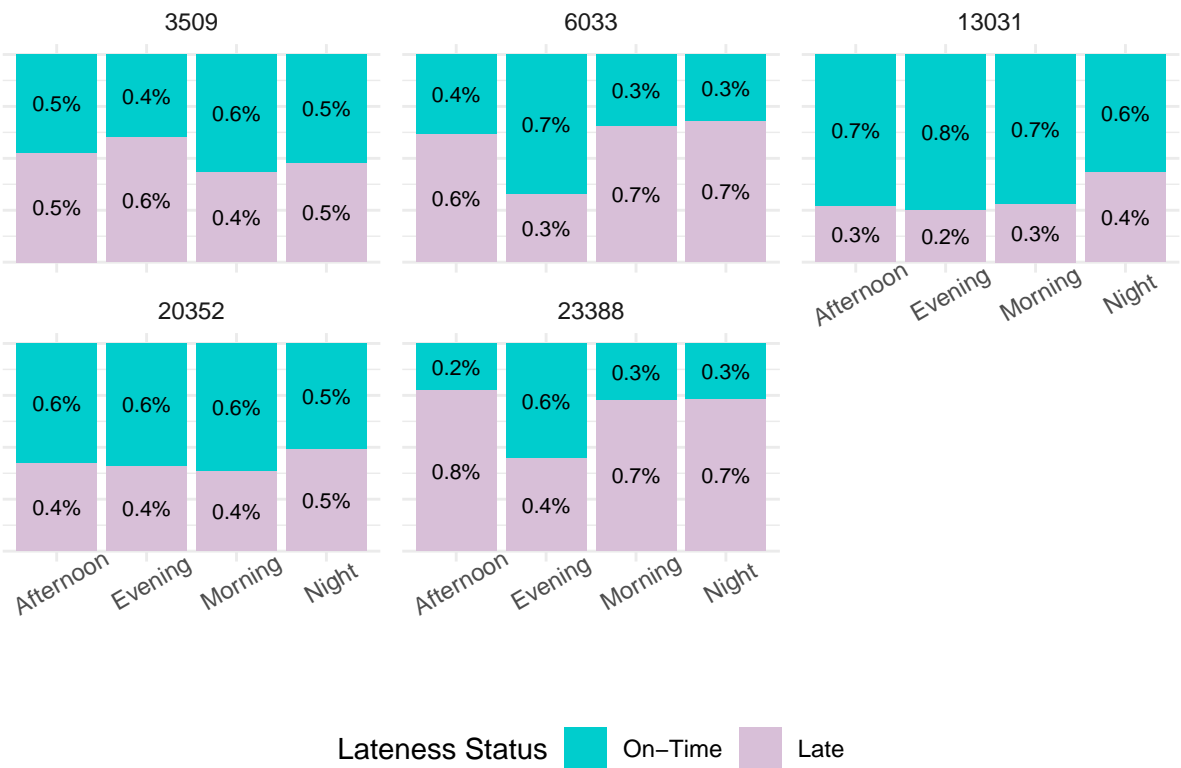
Proportion of Late Rides vs On-Time Rides



The hypothesis that “people who are late more than 200 times have a habit of being late” is supported by the data presented in the graph. Out of the top 20 users analyzed, 12 show a higher proportion of late rides compared to on-time rides. This significant majority indicates a consistent pattern of lateness, rather than occasional delays caused by external factors. The variability in lateness proportions among the users suggests that punctuality may be influenced by individual habits or routines. For example, users with lateness proportions exceeding 70% likely exhibit a habitual tendency to be late, reinforcing the hypothesis. On the other hand, a smaller group of users maintains a balance between late and on-time rides, possibly reflecting sporadic lateness due to unpredictable circumstances. The graph underscores the distinction between habitual and situational lateness, where a majority of frequent latecomers display a clear predisposition toward delayed behavior. This pattern provides strong evidence to accept the hypothesis, emphasizing that habitual lateness is a defining characteristic for a significant portion of the observed users.

Hypothesis: “User lateness is consistent regardless of the period of the day.”

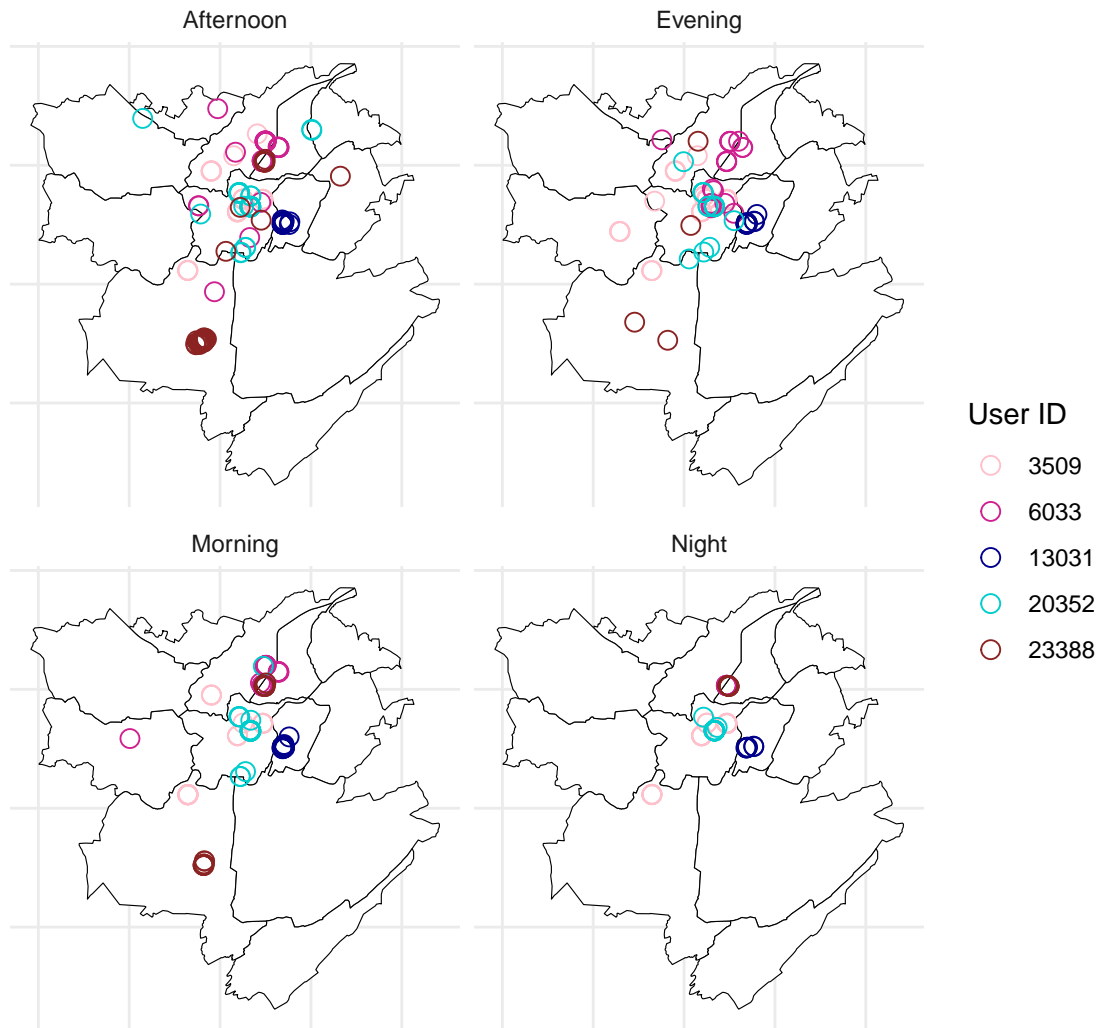
Proportion of Late vs. On-Time Instances by Period for Top 5 Users



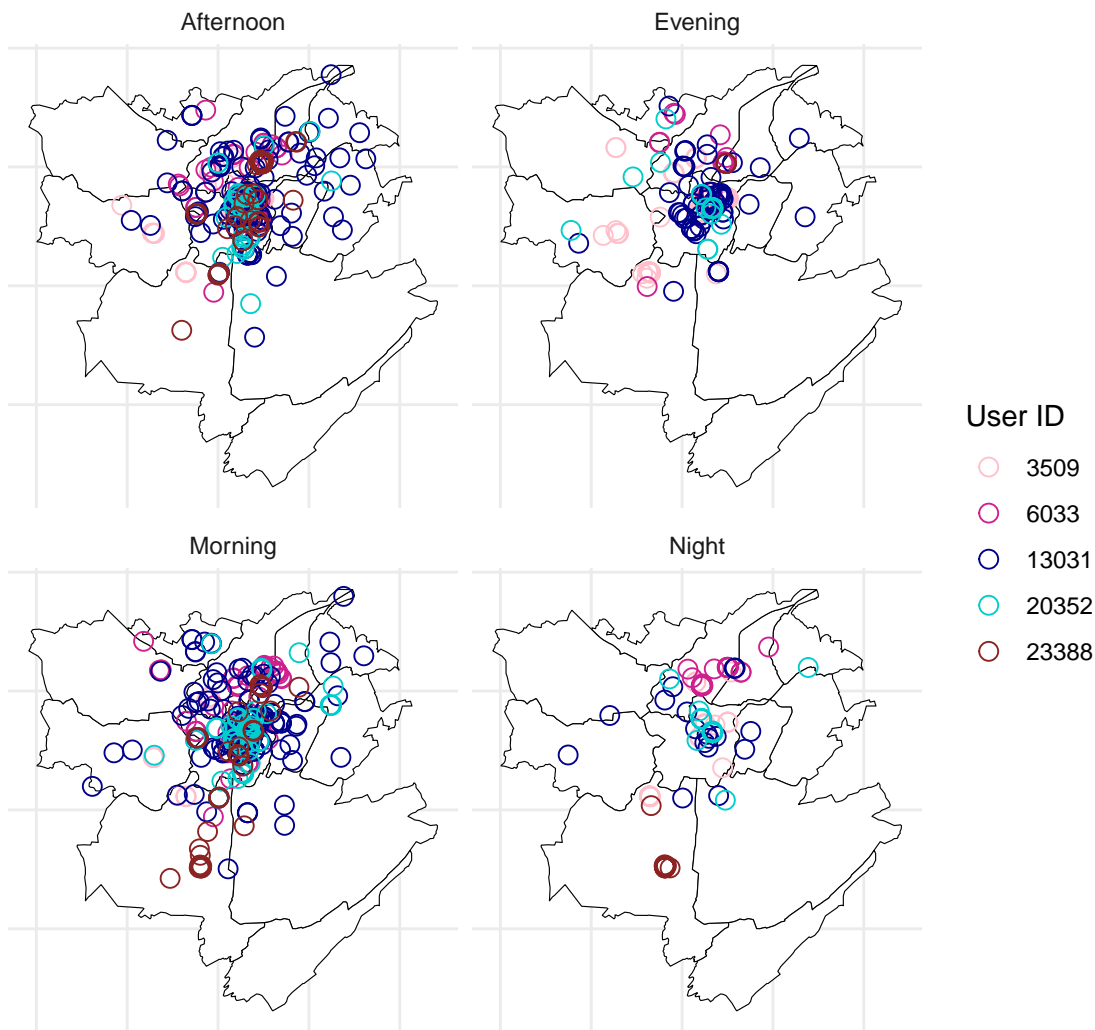
The hypothesis is accepted. The analysis reveals remarkable consistency in lateness rates across different time periods for 4 out of the top 5 late users. These individuals maintain a lateness rate between 0.6 and 0.7, regardless of whether it is morning, afternoon, evening, or night, highlighting their habitual tendency to be late as a defining characteristic. However, one user deviates significantly from this pattern. While their lateness rate aligns with the others during most periods, ranging from 0.6 to 0.7, their evening rate drops drastically to 0.2. This striking exception suggests that this user’s punctuality improves considerably in the evening, possibly due to structured evening plans or reduced external disruptions during that time. This insight underscores the idea that while lateness may often be habitual, it can be contextually influenced by specific periods of the day. The consistency of lateness among the majority of the users emphasizes the strength of habitual patterns, while the deviation of one user reveals that time-specific factors can disrupt these tendencies, introducing variability. The analysis highlights the nuanced interplay between individual habits and contextual factors in shaping lateness behavior.

Hypothesis: “In the morning, individuals consistently start from the same location, while their destinations vary.”

Origins of Late Rides for Top 5 Late Users



Destinations of Late Rides for Top 5 Late Users



The hypothesis is accepted. In the morning, the clustering of origin points suggests that these individuals consistently leave from the same location, most likely their homes. This consistency implies a fixed starting routine, such as preparing to leave for daily commitments. However, the scattered destination points tell a different story. Unlike their predictable origins, their destinations vary significantly, covering a wide geographical range. This dispersion suggests that these users do not commute to a single fixed workplace or educational institution but instead engage in diverse activities, such as errands, or social commitments, which take them to different places each day. This contrast between stable origins and highly variable destinations highlights the potential impact of lifestyle choices on lateness. The absence of a fixed destination or routine could contribute to delays, as planning and timing become more challenging when daily schedules and locations are unpredictable. This analysis supports the idea that lateness among these users is influenced by their dynamic and varied routines rather than external constraints tied to a single, recurring destination.

Conclusion

This project provided an in-depth analysis of user lateness patterns and taxi ride behaviors in Yerevan, revealing critical insights into individual habits, daily routines, and urban transportation dynamics. By examining user lateness and taxi ride data, we identified significant trends that shape both individual behavior and the collective movement of the city.

Key findings include:

User Lateness Patterns:

- **Cumulative Fatigue:** Late rides increase as the week progresses, peaking on Fridays, likely due to the accumulation of professional and personal responsibilities. The sharp decline in weekends reflects the relaxed schedules and reduced commitments.
- **Time-of-Day Trends:** The afternoon period consistently records the highest instances of late rides, influenced by unpredictable lunch breaks, traffic, and end-of-day rush hours.
- **Individual Variability:** Frequent late users exhibit habitual lateness, with top users showing consistent patterns throughout the day, while a broader group displays variability influenced by external factors.

Taxi Ride Behaviors:

- **Kentron's Centrality:** The analysis confirmed Kentron as Yerevan's primary hub for taxi movements, with most rides originating from residential areas in the morning and returning there in the evening.
- **Consistent Commutes:** Regular users exhibit structured routines, with predictable morning departure times and a narrower fare range, reflecting commuting patterns.
- **Dynamic Lifestyles:** Variability in destinations and ride times among some users suggests the influence of diverse, flexible lifestyles on travel behavior.

Implications

The findings emphasize the importance of understanding user behavior for optimizing transportation services. For future studies, integrating additional contextual data, such as road conditions or user demographics, could provide a more nuanced understanding of the factors influencing lateness and taxi usage. By leveraging these insights, urban planners and transportation service providers can enhance the efficiency and reliability of Yerevan's mobility systems. In conclusion, this project sheds light on urban transportation's structured and dynamic aspects, offering a foundation for data-driven decision-making and improved service delivery in Yerevan.