CS 506 - Bus Transit Performance - Deliverable 2

Problem Statement

Public transportation is a pivotal element in shaping the daily experiences and quality of life for residents across Massachusetts and the Greater Boston area. Yet, the assurance of equal and fair access to quality service across all reachable areas remains a question that demands investigation. Given its profound impact on the day-to-day lives of residents, it becomes imperative to quantify the equity and fairness embedded in Boston's public transportation system and to discern the varying perceptions of service quality among different neighborhoods.

Our analysis will leverage a combination of data science methodologies, encompassing the extraction and examination of public transportation data, and demographic information. We explored key questions such as: How does the quality of public transportation services vary across different neighborhoods? Are there discernible disparities in service frequency, reliability, and accessibility? How do factors such as income levels, population density, and geographic location correlate with the perceived quality of public transportation?

Our analysis provides a comprehensive overview of the existing state of public transportation equity. The goal is to pinpoint areas where improvements can be made, ensuring that the benefits of an efficient and reliable transportation network are shared equitably among diverse communities.

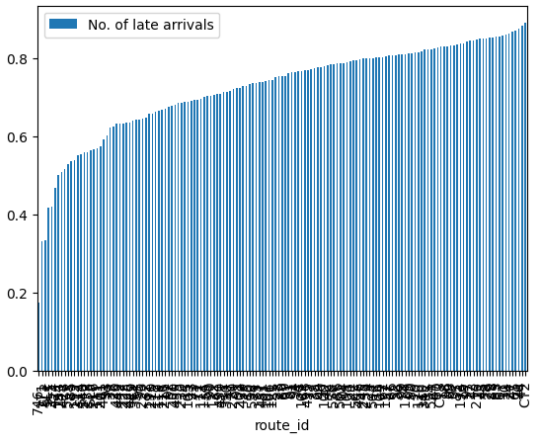
Data Collection and Cleaning

Collected Data from the https://mbta-massdot.opendata.arcgis.com/ website. The following 5 Datasets were used for analysis:
1. PATI Bus Stops
2. MBTA Bus Arrival Departure Times
3. Bus Reliability
4. Bus Ridership by Time Period, Season, Route Line, and Stop
5. Wheelchair/Accessibility

In addition, data from the https://api-v3.mbta.com/ api was utilized to obtain bus level information such as "vehicle_number", "route_id" and the bus stops each particular bus would visit along with their coordinates.

Exploratory Data Analysis

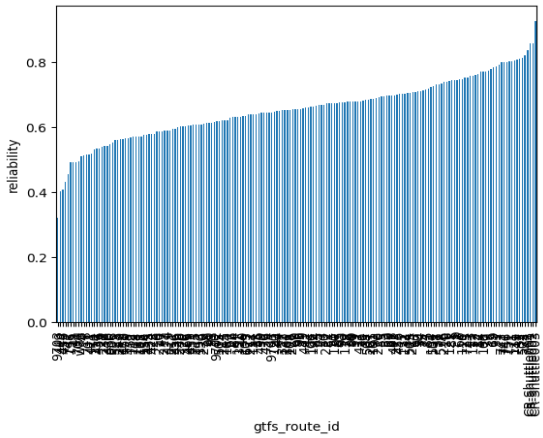
Q: What is the proportion of time the buses are late on a certain route?



Route_id v.s. Number of late arrivals

- *We observed that the frequency of late arrivals varies significantly across different bus routes. The proportion of late arrivals ranged from as low as 0.4 to upwards of 0.8, indicating a substantial variation in punctuality and reliability among routes.*
- *This variation suggests that certain routes constantly struggle with timeliness issues, while others maintain a higher level of on-time performance. Routes with higher proportions of late arrivals (closer to or exceeding 0.8) could be experiencing factors such as traffic congestion, longer route lengths, operational challenges, or scheduling inefficiencies. Conversely, routes with lower late arrival proportions (around 0.4) might be benefitting from better infrastructure, optimal scheduling, and less congested paths.*

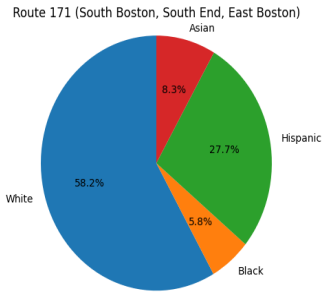
Q: How reliable are the buses on a certain route?

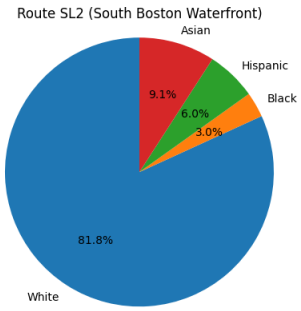


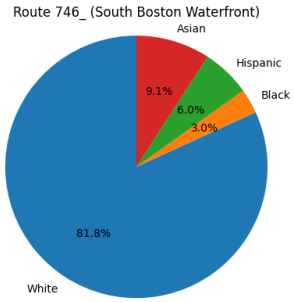
Route_id v.s. Reliability

- The reliability scores ranged from 0.4 to over 0.8. This metric provides a clear indication of how consistently buses adhere to their schedules across different routes.
- Routes with higher reliability scores (approaching or exceeding 0.8) demonstrate a strong adherence to scheduled times, suggesting efficient management and fewer disruptions. These routes likely benefit from effective route planning, less traffic congestion, and possibly fewer operational challenges.
- On the other hand, routes with lower reliability scores (around 0.4) indicate a higher frequency of deviations from the schedule. This could be due to a variety of factors, including traffic issues, longer route lengths, or logistical challenges.

Top 3 Least Late Routes







Visualizations, methodology for finding underlying patterns, and insights for key base questions

Q: How long does it take to travel end-to-end on different routes?

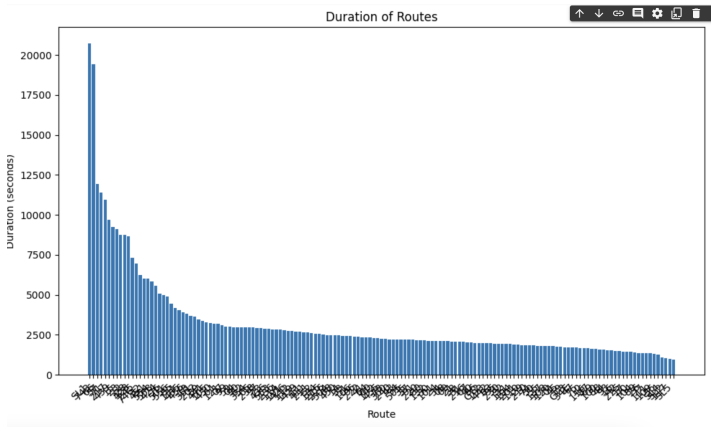
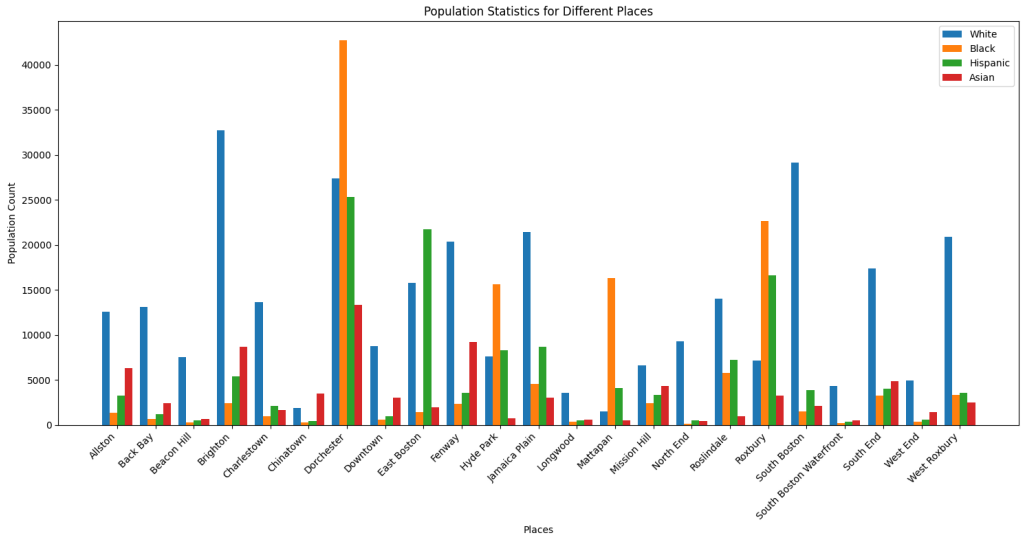


Figure. Route v.s Duration

- The bar chart displays a wide variation in the duration of bus routes, indicating that end-to-end travel times differ significantly across the network. The longest route takes just over 20,000 seconds, while the shortest is only a few hundred seconds.
- We notice that a majority of the lines have end to end travel times of < 5000 seconds.
- Most routes fall between these extremes, with a steep decline in duration observed for the majority. This information suggests that while some routes are quite lengthy, potentially due to extended geographical coverage or slower speeds, many are considerably shorter, highlighting the diversity in the operational characteristics of the bus network.

Q: What are the population sizes and characteristics of the communities serviced by different bus routes (e.g. race, ethnicity, age, people with disabilities/ vulnerabilities)?



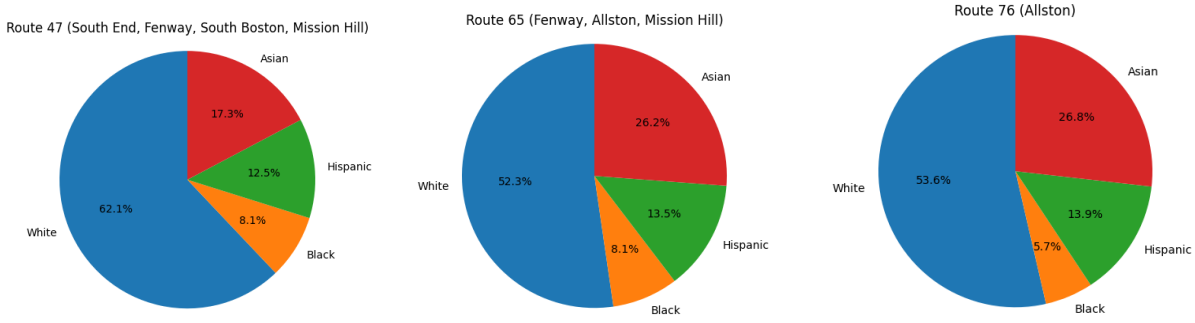
Neighborhoods sorted based on wheelchair accessibility:

Allston, East Boston, Mattapan, West Roxbury, West End, Beacon Hill, Back Bay, South Boston, Waterfront, North End, Fenway, South Boston, South End, Jamaica Plain, Mission Hill, Roslindale

- There are marked differences in population sizes across neighborhoods. Some areas, like Allston and East Boston, have significantly larger populations, potentially indicating a greater demand for accessible transportation.
- The bar chart shows the distribution of White, Black, Hispanic, and Asian populations within each neighborhood. Diverse neighborhoods can be identified by the balance of colors in their respective bars. For instance, Allston and East Boston show large populations, while Mattapan is predominantly Black, and Back Bay is predominantly White. Disparities in wheelchair accessibility across these neighborhoods could signal service gaps, particularly in racially diverse or densely populated areas.
- Assuming that neighborhoods higher on the list have better wheelchair accessibility, it's possible to speculate that neighborhoods with larger populations, especially with higher proportions of vulnerable groups (like the elderly or disabled), may have better access to wheelchair-accessible buses.

Q: Are there disparities in the service levels of different routes? (which lines are late more often than others)

Top 3 Most late Routes



- Route 47 dominantly serves the White population (62.1%), followed by the Asian (17.3%), Hispanic (12.5%), and Black (8.1%) communities. The high lateness frequency on this route could significantly impact the majority White population in these areas.
- Route 65 shows a more diverse demographic with a White majority of 52.3%, a substantial Asian community at 26.2%, followed by Hispanic (13.5%) and Black (8.1%) populations. The lateness on this route could disproportionately affect the Asian and Hispanic populations due to their significant representation.
- Route 76 serves an area with a demographic composition similar to Route 65, with a slight increase in the Asian population (26.8%) and a decrease in the White population (53.6%). The Hispanic (13.9%) and Black (5.7%) populations are also notable. The Asian community's higher representation suggests that lateness on this route could particularly impact this demographic group.

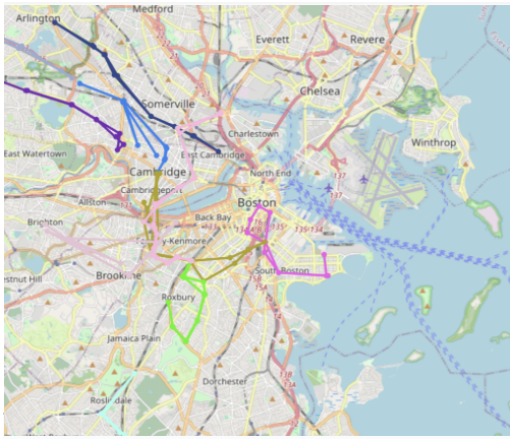


Figure. Top 10 most late routes plotted on the map of boston

- The late routes appear to be distributed across various neighborhoods in Boston, with some concentration in the central and southern parts of the city.
- Areas such as Back Bay, South Boston, and the vicinity of Fenway Park show overlap of late routes, suggesting traffic congestion or other systemic delays that could be affecting these particular corridors more frequently.
- The geographic spread of these routes indicates that the lateness issue impacts a diverse cross-section of Boston's population, including various residential and commercial districts.

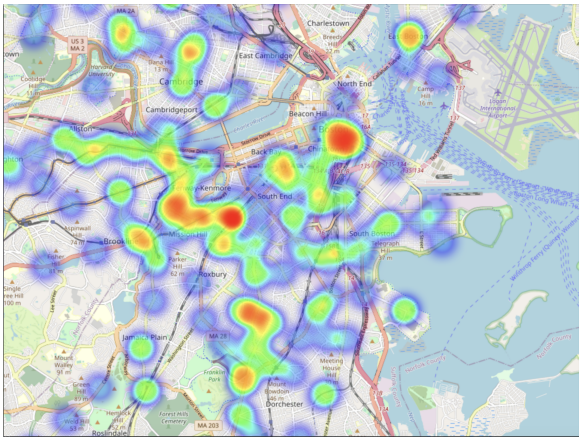
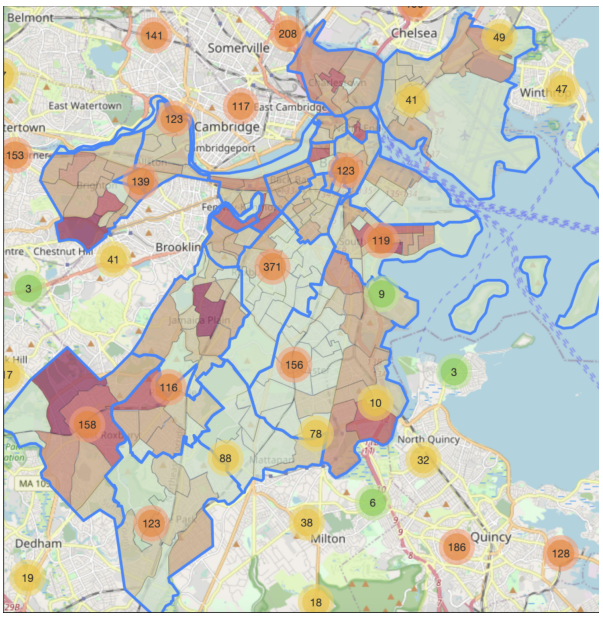


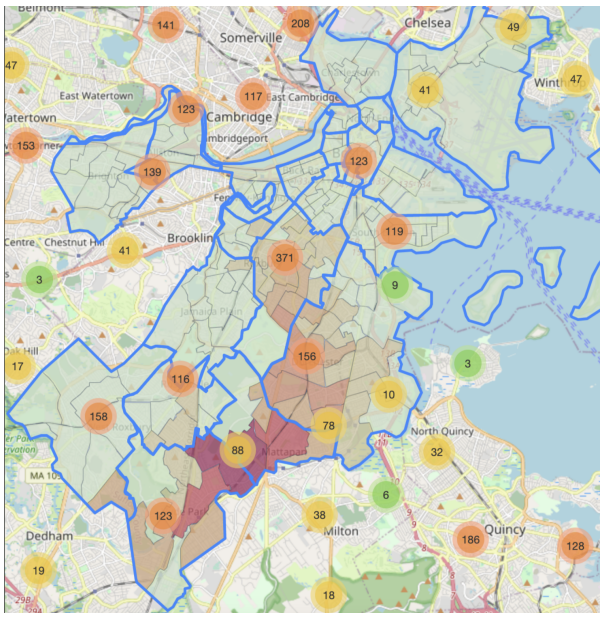
Figure. Late Route Stops Heatmap: East Boston, Chinatown, Mission Hill/Northeastern, Commonwealth Avenue, Roxbury, Cambridge

- A heat map of late bus stops in Boston provides a visual representation of spatial and temporal patterns of delays, revealing clusters of high-delay areas, potential correlations with traffic congestion and infrastructure issues, and insights into route-specific challenges.

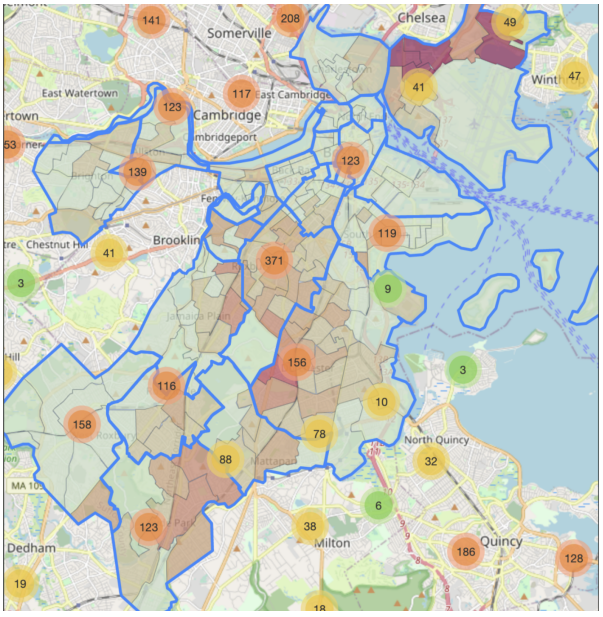
Q: How are different races distributed across Boston?



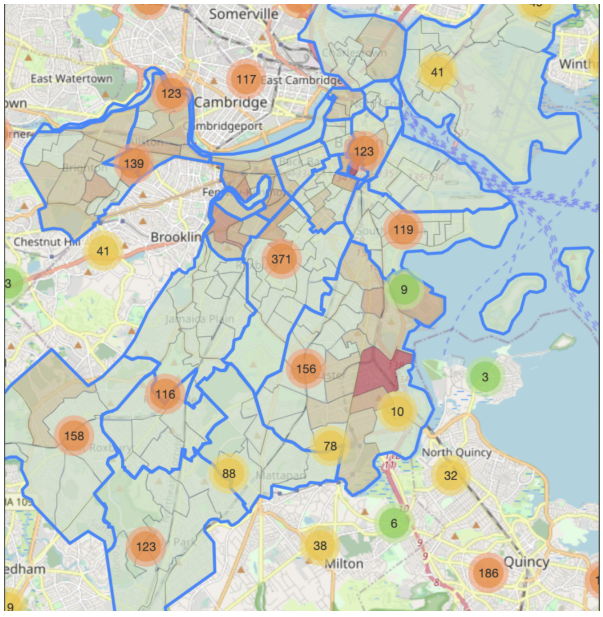
WHITE POPULATION: Roxbury, with 60,705 residents, has a diverse population with a median age of 34, primarily relying on buses for transportation, while Jamaica Plain, home to 41,112 residents, boasts a higher median income of $106,153, with a majority holding college degrees and utilizing various modes of transportation including cars, walking, biking, and buses.



BLACK POPULATION: Hyde Park, home to 38,402 residents, features a diverse population with a median age of 37 and a preference for car transportation, while Mattapan, with 35,997 residents, has a median age of 36, and residents primarily rely on cars followed by buses, with an average household income of $81,033.



HISPANIC POPULATION: Dorchester, with 85,854 residents and a median age of 35.1, exhibits a diverse population with 54% holding college degrees, an average household income of $93,069, and a transportation preference for cars followed by buses. East Boston, home to 45,501 residents with a median age of 33.8, has a mixed demographic, and residents mainly rely on cars before using buses, with an average household income of $98,782



ASIAN POPULATION: Chinatown, with a population of 6,546, features a diverse demographic, with 50% holding some college degree and an average income of $50,652. Meanwhile, Dorchester, has 85,854 residents, a median age of 35.1, 54% college degree holders, and an average household income of $93,069, with a transportation preference for cars followed by buses.

Extension Proposal

Extension Pitch	***Exploring the Intersection of Disability Access and Bus Ridership*** *This project aims to dissect the intricate relationship between disability access features on buses and patterns of ridership. Our focus will be on understanding how these elements intertwine to shape the landscape of public transportation inclusivity. This endeavor is not merely an analytical exercise; it's an opportunity to influence policy and enhance service delivery, ensuring that public transit systems cater to the needs of all, including those with disabilities.*
Rationale	*We hypothesize that socioeconomic factors such as income, employment rates, and educational attainment play a significant role in the use of disability-accessible buses. This exploration is critical for two reasons: firstly, it addresses an often-overlooked aspect of public transportation—equity and inclusivity. Secondly, it provides actionable insights that could reshape how public transit systems approach disability access, thereby influencing a wide range of policy and operational decisions.*
Questions for Analysis	***Socioeconomic Disparities in Ridership:*** *How do socioeconomic factors such as income and education levels impact the utilization of buses with disability access features?* ***Identification of Accessibility Deserts in Low-Income Area****s:* *Are there regions with a high prevalence of disabilities but lower accessibility usage due to socioeconomic factors?* ***Correlation Between Employment Hubs and Disability Access Usage****:* *Does the presence of employment hubs in certain areas correspond to higher usage of disability-accessible buses, possibly indicating a need for accessible commuting options?*
Data Sets & Sources	***Bus Performance Data****:* *The original dataset on bus performance, including ridership metrics and information on disability-accessible buses.* ***Census Data****:* *Socioeconomic indicators such as income, education levels, and employment rates at a granular level, mapped to census tracts corresponding to bus stops.* ***Disability Demographics:*** *Data on the prevalence of disabilities at the census tract level, providing insights into the distribution of potential users of disability-accessible features.*

Data Visualizations	**Heatmap of Disability-Accessible Bus Utilization**: *A heatmap overlaying disability-accessible bus utilization on socioeconomic indicators, highlighting areas with high usage and potential disparities.* *More Details list in next part of "Visualization and insights for extension proposal",*
Additional Information	*Understanding the socioeconomic dimensions of disability access and ridership is crucial for designing targeted interventions and ensuring that public transportation serves all community members equitably. This extension recognizes the importance of socioeconomic factors in shaping transit choices and aims to contribute insights that foster a more inclusive and accessible public transportation system.*

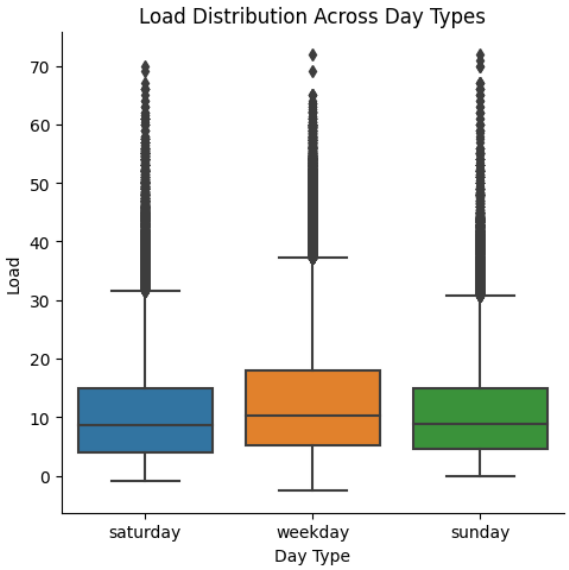
Visualization and insights for extension proposal

- We carried out visualizations and analysis on the ridership data available.

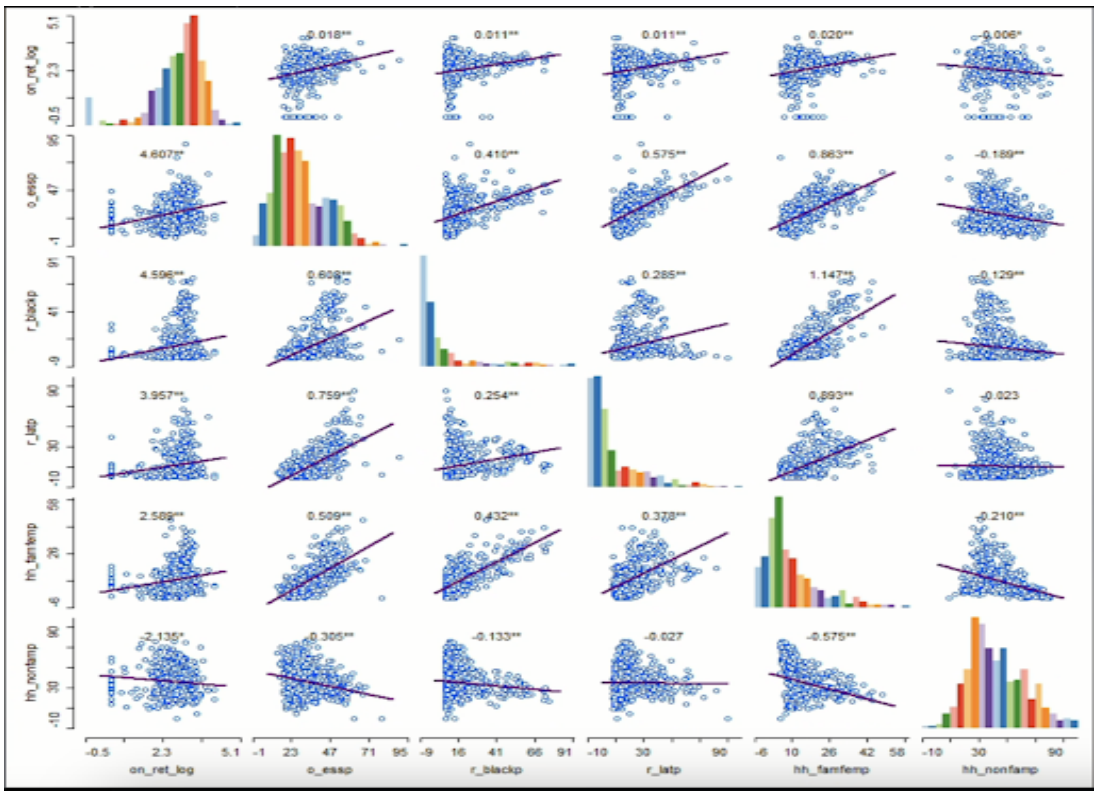
- *Disability-accessibility utilization heatmap*
- This is a heatmap overlaying disability-accessible bus utilization on socioeconomic indicators, highlighting areas with high usage and potential disparities.
- From this visualization, we observe notable differences in the utilization of disability-accessible buses across different areas. Regions with higher usage potentially indicate a strong dependence on public transportation by residents with disabilities, possibly due to limited alternative transportation options. This high usage might correlate

with areas having higher populations of disabled individuals or lower income levels, where affordable and accessible transportation is crucial.

- Conversely, areas with lower utilization rates raise questions about the adequacy and awareness of available services. These disparities could be attributed to a range of factors, including varying levels of income, differences in the distribution of the disabled population, or the availability of alternative forms of accessible transportation.



- Figure. Load (number of people on the bus) across day types
- The boxplot shows the distribution of passenger load across different days of the week. It seems there is a higher variability on weekdays, indicated by a larger interquartile range (IQR) and the presence of more outliers, suggesting that weekdays experience peaks of high ridership.
- The median load appears to be higher on weekdays compared to weekends, which could reflect the routine work commute.
- Notably, there are outliers on all days, with weekends showing extreme values. This might indicate special events or irregular activities that drive up ridership.



- These plots directly comparing the number of boardings on disability-accessible buses with socioeconomic variables can reveal whether higher income or education levels are associated with increased or decreased use of these services.
- If there's a trend indicating higher boardings in areas with more essential workers, this could imply that disability-accessible transportation is crucial for this workforce.
- In terms of ridership and race, the plots might show that certain racial groups are more reliant on disability-accessible services, indicating a need for targeted outreach or service improvements in racially diverse neighborhoods.
- In terms of income level and education, if higher boardings correlate with lower income or education levels, it may reflect the dependency on public transit among these demographics due to a lack of alternatives.

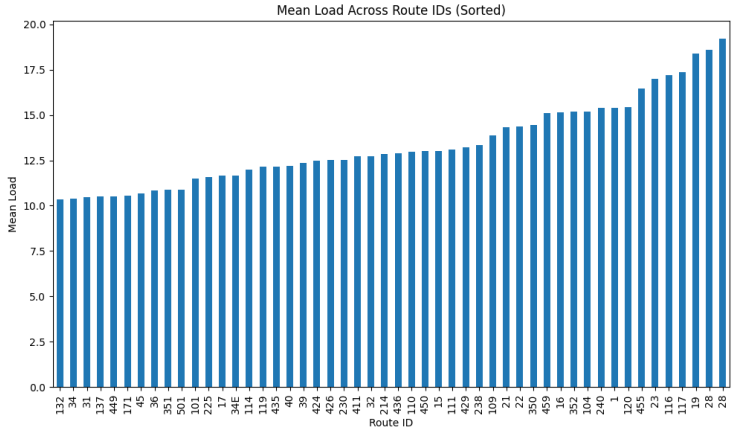


Figure. Mean Load (number of people on the bus) across different route id

- The ascending bar chart illustrates the mean load across different route IDs, sorted by load. It highlights which routes have the highest average number of passengers, potentially reflecting their importance in the network.

- The gradual increase towards certain route IDs suggests a distribution where few routes carry the bulk of passengers, which could be crucial for planning disability-accessible services.

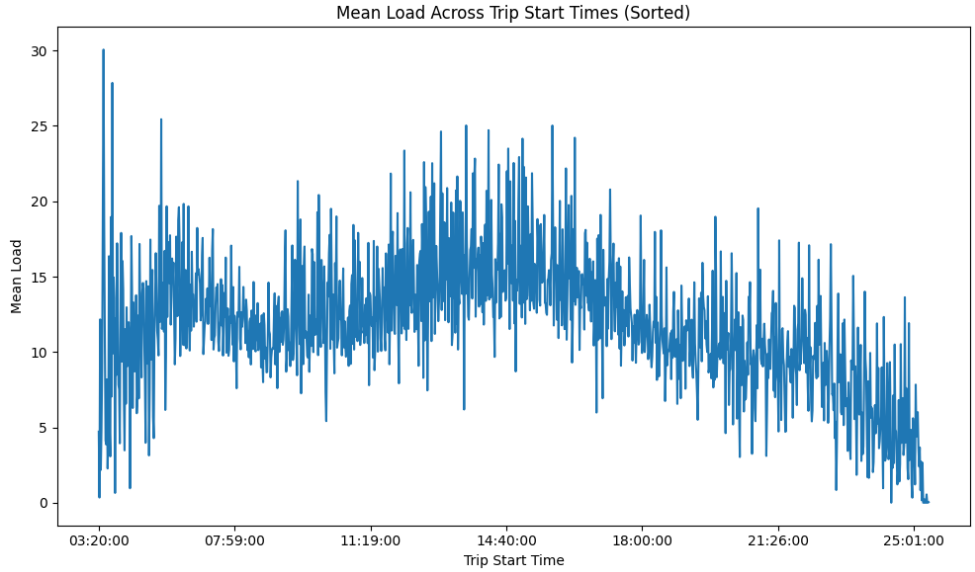


Figure. Mean Load (number of people on the bus) vs Time

- This chart shows fluctuations in mean load throughout the day. There are clear peaks, likely corresponding to rush hours, suggesting higher demand for buses during these times.
- The trend seems to drop towards the midday and then rise again for the evening rush, which is typical for workday commuting patterns.

Individual contribution

- Rishven and Haoxiang - Data preprocessing, base questions 1, 2
- Xavier and Ketan - Data collection, cleaning, preprocessing, base questions 1, 2, 3, 5
- James - Data preprocessing, extension project, presentation and slides