

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.

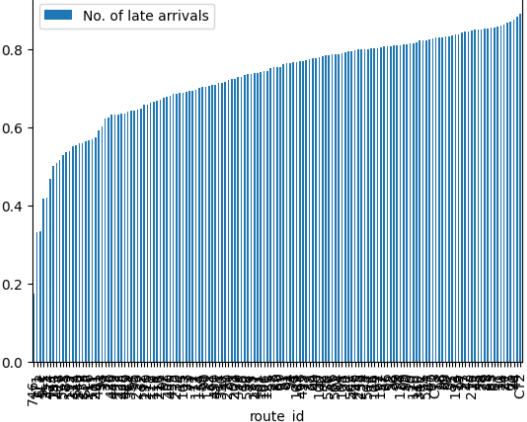
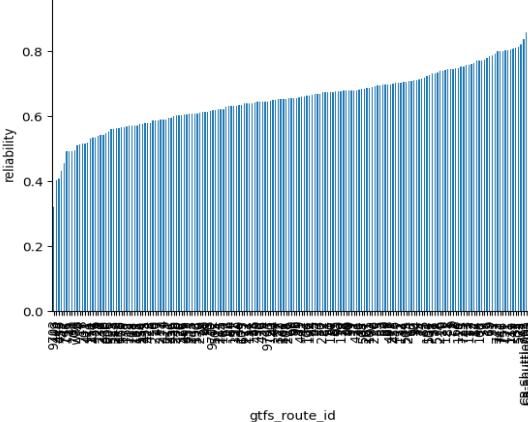
Data Collection and Cleaning

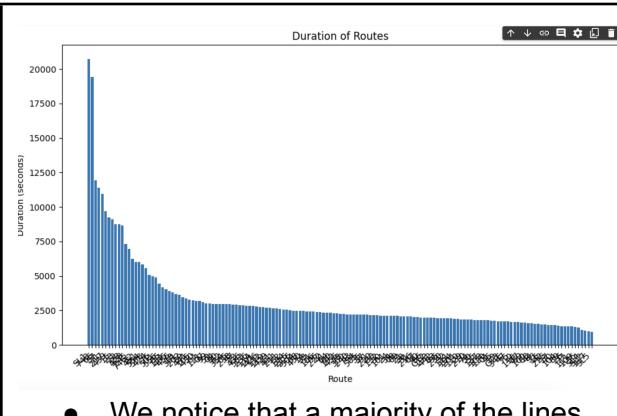
Collected Data from the <https://mbta-massdot.opendata.arcgis.com/> website. The following 5 Datasets were used for analysis:

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

<p>Q: What is the proportion of time the buses are late on a certain route?</p>	<p>Q: How reliable are the buses on a certain route?</p>
 <p>Route_id v.s. Number of late arrivals</p> <ul style="list-style-type: none"> 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. 	 <p>Route_id v.s. Reliability</p> <ul style="list-style-type: none"> 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.
<p>Q: How long does it take to travel end-to-end on different routes?</p>	

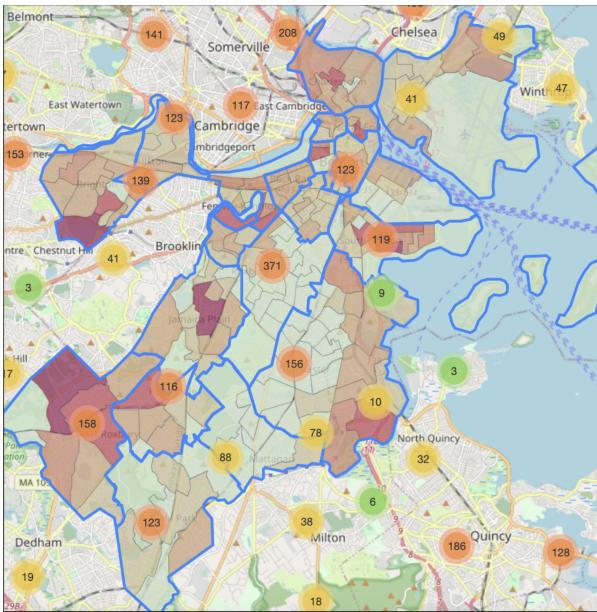


- We notice that a majority of the lines have end to end travel times of < 5000 seconds

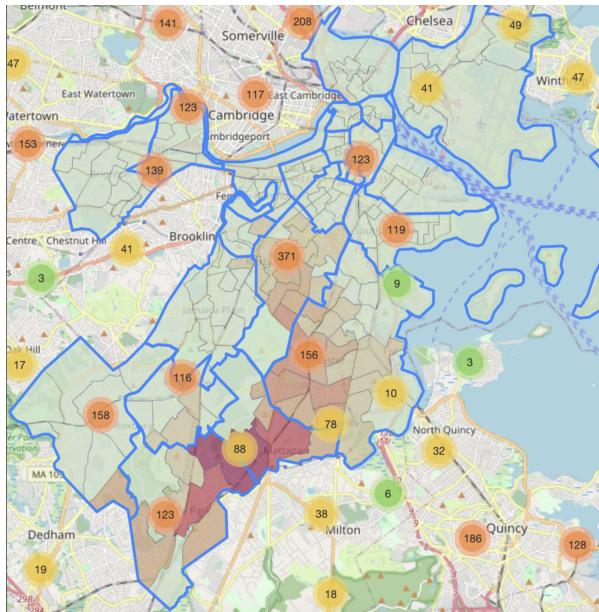
Visualizations, methodology for finding underlying patterns, and insights for key base questions

<p>Q: Are there disparities in the service levels of different routes? (which lines are late more often than others)</p> <ul style="list-style-type: none"> • The plot indicates the Top 10 most late routes plotted on the map. 	<p>Late Route Stops Heatmap: East Boston, Chinatown, Mission Hill/Northeastern, Commonwealth Avenue, Roxbury, Cambridge</p> <ul style="list-style-type: none"> • 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.

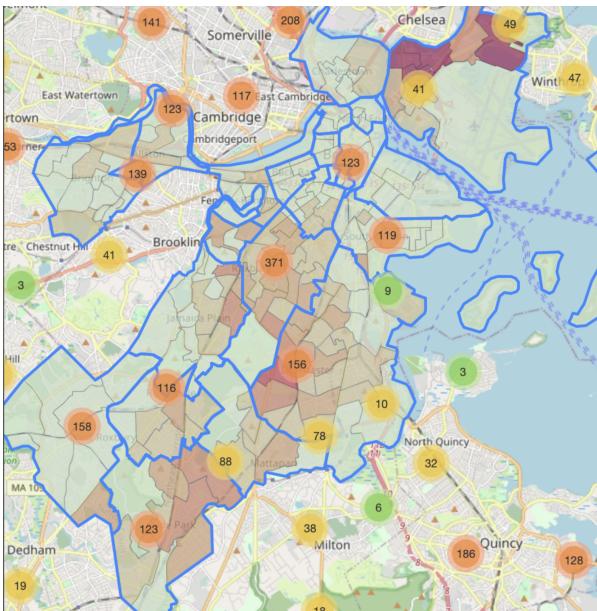
Q: How are different races distributed across Boston?



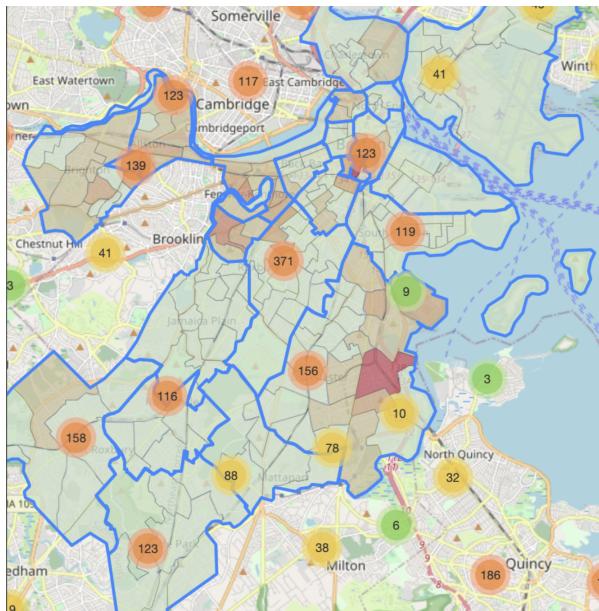
WHITE POPULATION: Roxbury, with 60,705 residents, has a diverse population with a median age of 34, primarily relying on buses for transportation, while Jamaica Plain, home to 41,112 residents, boasts a higher median income of \$106,153, with a majority holding college degrees and utilizing various modes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

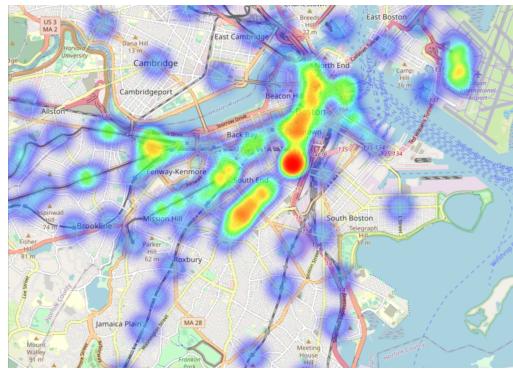
Extension Proposal

Extension Pitch	<p><i>Exploring the Intersection of Disability Access and Bus Ridership</i></p> <p><i>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.</i></p>
Rationale	<p><i>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.</i></p>
Questions for Analysis	<p><i>Socioeconomic Disparities in Ridership:</i> <i>How do socioeconomic factors such as income and education levels impact the utilization of buses with disability access features?</i></p> <p><i>Identification of Accessibility Deserts in Low-Income Areas:</i> <i>Are there regions with a high prevalence of disabilities but lower accessibility usage due to socioeconomic factors?</i></p> <p><i>Correlation Between Employment Hubs and Disability Access Usage:</i> <i>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?</i></p>

Data Sets & Sources	<p>Bus Performance Data: <i>The original dataset on bus performance, including ridership metrics and information on disability-accessible buses.</i></p> <p>Census Data: <i>Socioeconomic indicators such as income, education levels, and employment rates at a granular level, mapped to census tracts corresponding to bus stops.</i></p> <p>Disability Demographics: <i>Data on the prevalence of disabilities at the census tract level, providing insights into the distribution of potential users of disability-accessible features.</i></p>
Data Visualizations	<p>Heatmap of Disability-Accessible Bus Utilization: <i>A heatmap overlaying disability-accessible bus utilization on socioeconomic indicators, highlighting areas with high usage and potential disparities.</i></p> <p><i>More Details list in next part of “Visualization and insights for extension proposal”,</i></p>
Additional Information	<p><i>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.</i></p>

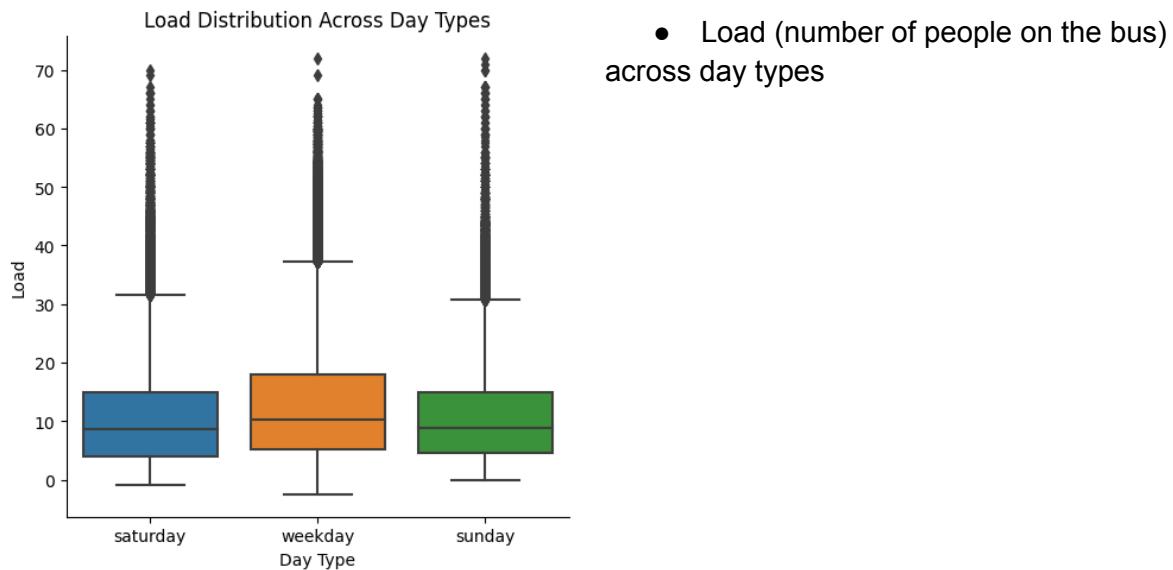
Visualization and insights for extension proposal

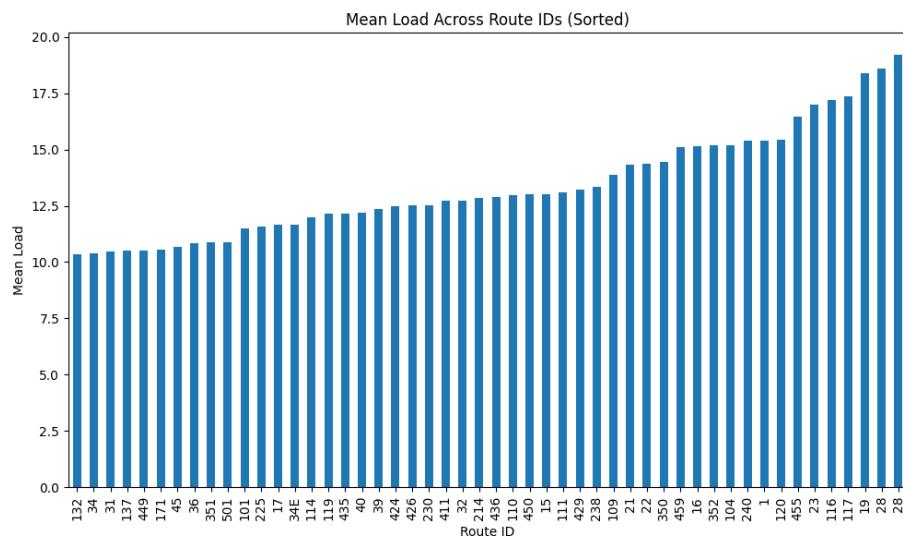
- We carried out visualizations and analysis on the ridership data available.



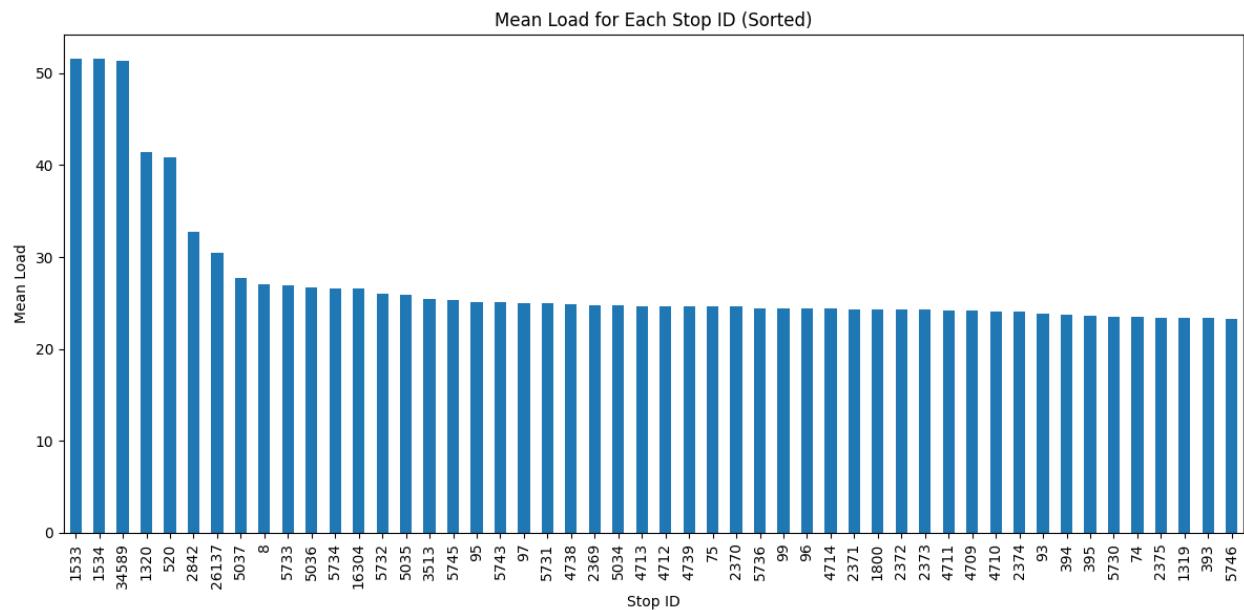
- *Disability-accessibility utilization heatmap*

- 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.

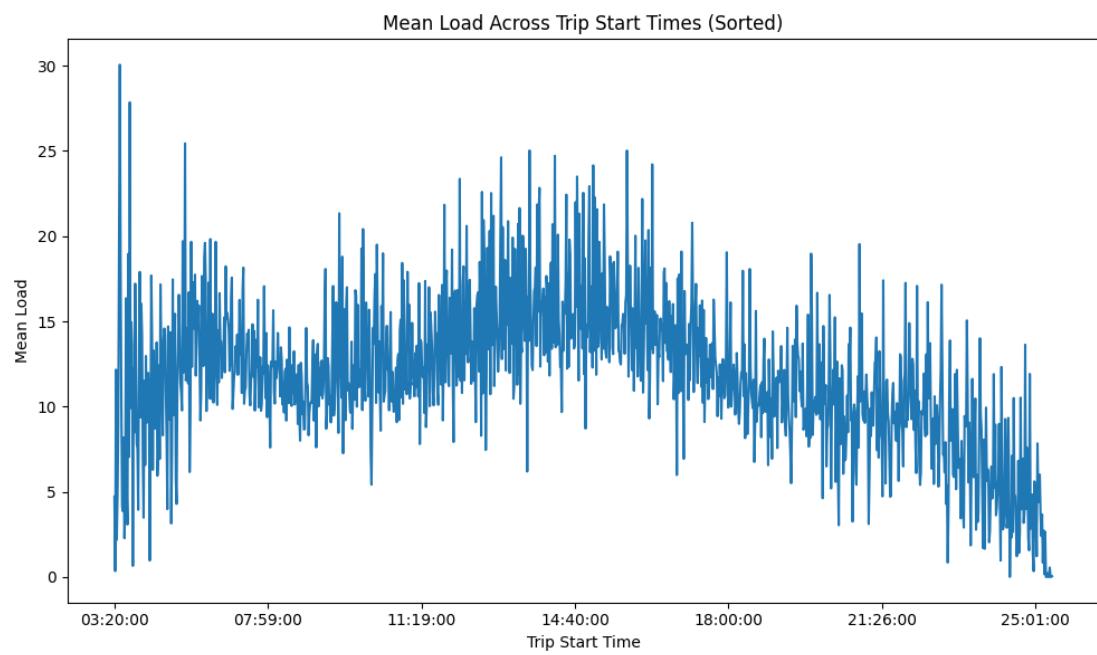




- Mean Load (number of people on the bus) across different route ids



1. Mean Load (number of people on the bus) across different Stop ids



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

Individual contribution

- Rishven and Haoxiang - Data collection, cleaning, preprocessing, base questions 1, 2
- Xavier and Ketan - Data collection, cleaning, preprocessing, base questions 3, 4, 5
- James - Data collection, cleaning, preprocessing, extension project, presentation and slides