An Analysis of London Underground Traffic Patterns Before and After the Pandemic

Dimitris Nikandrou

Abstract-This project presents a comprehensive analysis of London Underground traffic patterns during and after the COVID-19 pandemic. Utilizing Oyster card data from 2020 and 2022, the study explores changes in commuting behaviour, focusing on entry and exit patterns at different times and locations. The analysis includes visual tools like boxplots, histograms, and maps to identify trends and outliers. It examines shifts in peak usage times, variations in station traffic, and overall resilience of the urban transport network to large-scale disruptions. The findings offer insights into urban mobility trends and future infrastructure planning in response to societal changes.

1. PROBLEM STATEMENT

The onset and aftermath of the COVID-19 pandemic have dramatically reshaped urban mobility patterns, posing unique challenges and opportunities for public transportation systems worldwide. This report focuses on the London Underground, a crucial component of the city’s transport infrastructure, to understand how the pandemic has affected commuting behaviour. The primary aim is to unravel the extent of these changes by comparing passenger traffic patterns from two distinct years: 2020, marked by the height of the pandemic and consequent lockdowns, and 2022, as the city emerged back into normality.

The core questions driving this analysis are: How did the London Underground's usage in terms of entries and exits change between 2020 and 2022? Were there shifts in peak usage times or changes in the most frequented stations? Such questions are crucial for understanding the resilience of urban transport networks and planning future infrastructure adaptations in response to large-scale societal disruptions.

The datasets encompass daily and time-block records of entries and exits from the London Underground’s Oyster card system for 2020 and 2022. These records provide a fine view of commuting patterns, allowing for an in-depth comparative analysis. The choice of these specific years and data types is deliberate; 2020 represents a period of unprecedented disruption in public life and mobility, while 2022 offers a contrast as a period of recovery and adaptation. Thus, this analysis stands to offer significant insights into the impacts of global events on local transit systems, highlighting shifts in urban mobility trends in response to external challenges.

2. STATE OF THE ART

The book "COVID-19 Pandemic: Geospatial Information and Community Resilience - Global Applications and Lessons" is a detailed collection of studies about using geographic and mapping information to understand and respond to the pandemic. It uses different kind of data like satellite imagery, mapping systems (GIS), and health records to study how the pandemic affects places and communities. The methodologies employed include spatial analysis and data visualization, offering insights into pandemic trends and helping in decision making.

For this project, this book's approach presents parallels in data utilization and visualization. Spatial analysis techniques described in the book could be key in understanding travel patterns within the context of the pandemic. However, due to the book's broader focus on global applications, the methods applied might require modifications before being applied to the context of this project.

Someone could obtain valuable knowledge about the integration of spatial data with public health information and the significance of visual analytics in simplifying complex data sets. These insights are invaluable for analysing Oyster card data to understand the pandemic's impact on London's public transport system effectively.

In "Will Covid-19 put the public back in public transport? A UK perspective," Roger Vickerman explores the pandemic's effects on UK public transportation. The paper focuses into how COVID-19 reshaped public transport usage, emphasizing challenges like reduced ridership. It considers long-term changes in commuting habits and public trust in transport systems, suggesting that a return to pre-pandemic transport usage is unlikely. The study primarily uses a theoretical approach, focusing on broader socio-economic changes rather than specific data analysis.

When used in this project, this approach offers a big-picture view that could enhance this data-driven project. His insights into commuter behaviour changes and public perception shifts can guide the analysis of travel pattern changes in the Oyster data. However, the paper is centred around ideas and theories which could be more useful in the interpretation of the results obtained.

The paper "Impact of the COVID-19 Pandemic on Urban Human Mobility - A Multiscale Geospatial Network Analysis Using New York Bike-Sharing Data" presents an in-depth study of urban mobility alterations in New York City during the pandemic using bike-sharing data. It uses a multiscale geospatial network analysis, applying statistical and visualization methods to determine the pandemic's influence on bike-sharing and urban mobility.

This methodology, particularly the multiscale analysis and visualization techniques, could be adapted to this project. The study's emphasis on spatial patterns and network analysis aligns with the objectives of understanding changes in public transport usage in London. However, the specific context of bike-sharing in New York differs from London's Oyster Card data, but with the appropriate modifications to the methodologies, valuable insights could be obtained for applications in this project.

From this paper, lessons in spatial network analysis and the effectiveness of visualization techniques in changing mobility patterns are obtained. These methods could be applied to the analysis of Oyster card data, providing similar insights into London's public transport patterns during the pandemic.

3. PROPERTIES OF THE DATA

The study utilizes four distinct datasets for the years 2020 and 2022 from random weeks of each year, sourced from the Transport for London (TfL) database, specifically focusing on the London Underground system. The first dataset, named AC2020_AnnualisedEntryExit.xlsx, primarily captures the number of entries and exits in each station of the LU system, by daily passenger counts, which are organised in weekdays (Monday-Thursday) and weekends (Friday, Saturday, and Sunday). It comprises fields such as NLC (National Location Code), ASC (Alphabetic Station Code), station name, coverage (method of entry or exit), and entries and exits for each set of days as well as an annualised column for the specific station. The data types in this dataset range from categorical for station names and numerical for counts. The two datasets were nearly identical, with the exception of the 2022 dataset where the weekdays were split on Mondays and Tue-Thu, as well as extra weekly columns.

The other dataset was obtained by the aggregation of ten other datasets which each included time blocks (early entries, am entries etc.,) for entries and exits for a set of days. This resulted in one dataset for each year which included the total yearly counts of entries and exits in each time block as well as net counts. The datasets were then merged by station name, to obtain each one for each year. This complements the first by providing details on the traffic by distinguishing different time blocks, for each specific set of days. This is crucial for understanding how the pandemic affected travel in both the daily and hourly level. As previously mentioned, the datasets, included the NLC and ASC for each station, which were used to create different map styles.

The Oyster card data for the years 2020 and 2022 were analysed for quality using visual charts. Boxplots (Fig. 1) highlighted the regular and irregular entries, identifying outliers that might indicate either errors in data collection or days with atypical transit usage. Histograms (Fig. 2) provided a view of the frequency of travel numbers, aiding in the detection of any unusual patterns. The two charts, can be found below:

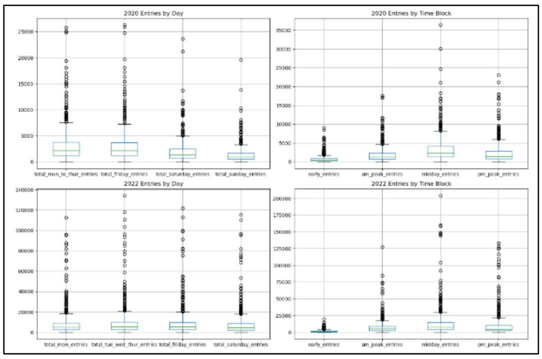

Fig.1: Boxplots for the datasets.

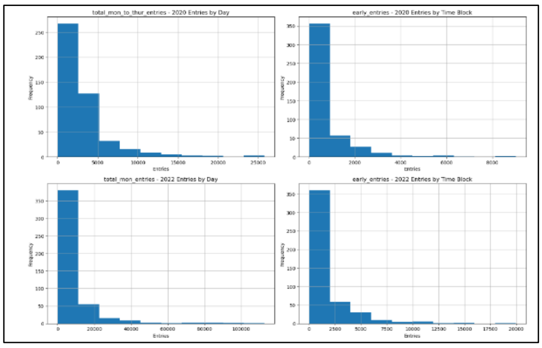

Fig. 2: Histograms for the first column of each dataset.

The investigation also involved searching for missing data, as its absence could skew the results. Consistency in data collection across the two-year span was verified to ensure a uniform dataset for analysis.

The next steps, in data pre-processing included a thorough cleaning of the data. This required a detailed review of the outliers to determine their relevance and whether they should be retained or corrected. Missing information would be addressed either by either imputation of the values or by omitting the incomplete records, depending on the nature and extent of the missing data. These actions are pivotal to refining the dataset, ensuring its reliability for accurately reflecting travel behaviours for further analysis. The data, once processed, is expected to provide insights into post-pandemic transportation trends.

4. ANALYSIS

4.1 Analysis Approach

The analytical workflow process of London Oyster card data starts with a series of methodical steps, beginning with Data Collection. The data is sourced from the Transport for London (TfL) website, where it is ensured that the data is both accurate and, in an appropriate format for analysis, establishing a solid foundation for the workflow.

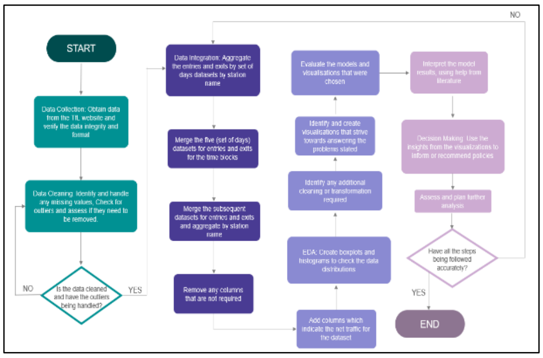

Fig. 3: Analysis Workflow diagram

Upon securing the data, the second step is Data Cleaning. This phase involves a detailed inspection for missing values or outliers that may skew results. The treatment of these outliers is subject to human judgement, determining which data needs to be removed or imputed, considering whether they reflect actual events or simply errors.

Following the cleaning process, Data Integration is undertaken. This step is pivotal as it involves the merging of datasets, aggregating them by station names and other common fields. Such integration ensures that the data is comprehensive, representative, and well-structured to continue with the analysis.

With the dataset cleaned and refined, Exploratory Data Analysis (EDA) then follows. The deployment of visual tools, such as boxplots and histograms, to spot patterns and gain a deeper understanding of the data, is crucial in this stage. This visualization aids in confirming the effectiveness of prior cleaning and transformation efforts.

If through the EDA, further inconsistencies are revealed, further Data Cleaning or Transformation may be required. The datasets are then re-examined and adjusted accordingly, ensuring its readiness for advanced analysis.

Following the EDA, the creation of visualisations or models is the next step, where the datasets are thoroughly examined and manipulated to allow for the effective representation of the results obtained from this process. The visualisations created, should be easily interpreted by the target group of people reading the report and be able to convey information effectively.

In the Evaluation phase, the effectiveness of the analytical models and visualizations is assessed. This evaluation ensures that the methods employed are robust and that the visualizations effectively display the data used.

The Decision Making stage is where the insights gained from the analysis are translated into actionable recommendations or policy suggestions. In this case analysts would synthesize the data findings to inform transport strategy but in this case the findings will be interpreted and explained

An assessment follows, serving as a quality check to validate that each step of the analysis has been thoroughly and correctly executed. This phase is critical to the credibility of the analysis process.

The workflow culminates in the Conclusion phase, where analysts would review their collective findings and reflect on the analysis's success. This final stage confirms that the results align with the analysis goals and that the process has comprehensively addressed the research questions, as well as decide on further future work that would be beneficial to the analysis.

Throughout the workflow, human insight complements computational efficiency. Analysts should always apply their expertise to navigate the analysis, using the computational power to manage the data's volume and complexity.

4.2 Analysis Process

This section explores the practical execution of the analytical approach previously outlined, with special attention to the human element of the process. It showcases the journey of the analysis through the perspective of the analyst, emphasizing the engagement with visual displays, the interpretation of complex data patterns and the derivation of conclusions that answer the research questions. This section will illustrate how the repeated application of both visual and computational methods can refine the outcomes and inform each subsequent step of the analysis. To create these visualisations, Python and R Studio were used.

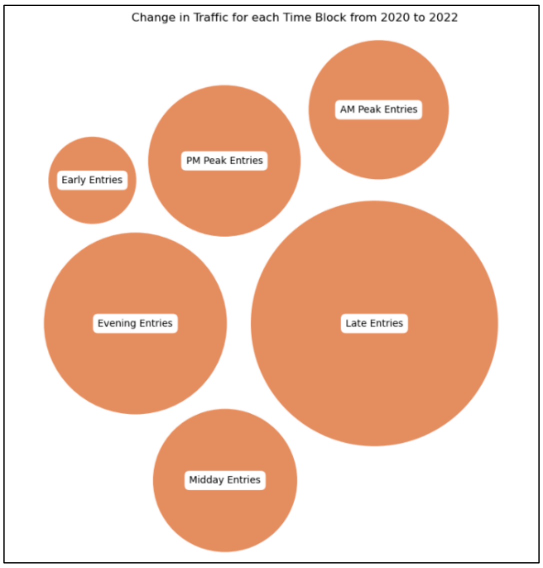

Fig. 4: A circular packing chart indicating the relative changes in traffic for each time block.

The chart visualizes the increase in traffic for each time block between 2020 and 2022 using circles of varying sizes, which are indicative of the magnitude of change. The largest circle represents "Late Entries," suggesting this time block saw the most significant increase in traffic volume over the two years. "AM Peak Entries" and "Evening Entries" also show considerable changes, as denoted by their relatively large circles. "PM Peak Entries" and "Late Entries" exhibit a moderate increase, while the smallest circle for "Early Entries" indicates the least increase in traffic during the early hours.

This type of visual display is instrumental in quickly conveying the relative scale of changes across different parts of the day. It allows the analyst to identify not only the periods of greatest change but also to make deductions about the potential causes, such as shifts in commuter habits or changes in service levels, due to the pandemic. Decisions on where to focus further analysis or policy adjustments can be guided by these visual insights. The chart serves as a foundational tool for discussions about how public transport usage has evolved, which is critical for planning and operational decisions in the post-pandemic period and could inform management teams about the potential changes faced during pandemic-like events. For example, someone could deduce that since the biggest changes have been observed in the Late Entries and Evening Entries, the large change was due to people not commuting back to their homes after work or after a night-out, which all were activities that were considerably limited during the pandemic.

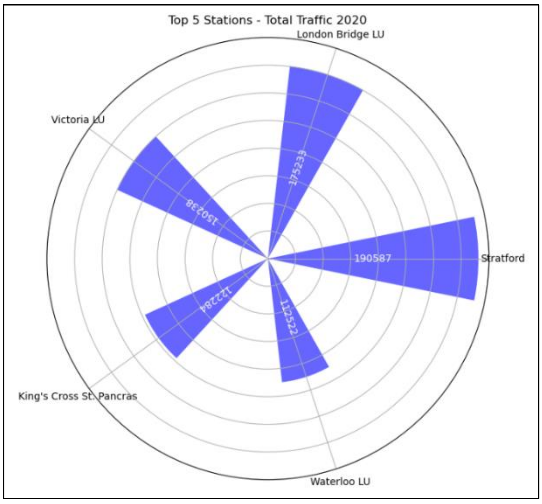

Fig. 5: Radial chart of total traffic in the top 5 stations in 2020.

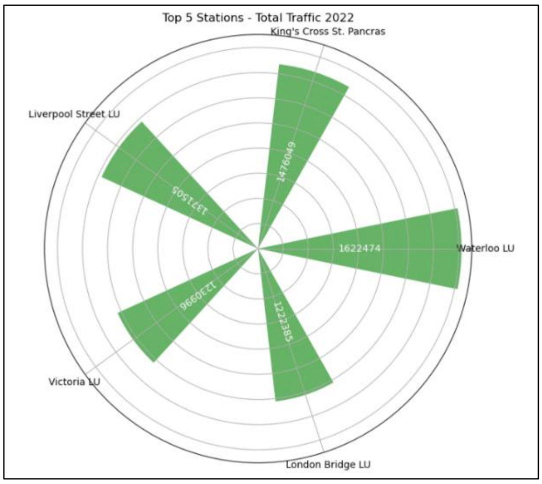

Fig. 6: *Radial chart of total traffic in the top 5 stations in 2022.*

The radial charts for 2020 and 2022 represent total traffic volumes at the top five London Underground stations, offering a direct visual comparison of traffic distribution. In 2020, the chart shows a concentration of traffic at Stratford, signifying its prominence in passenger volume, which is logical since Stratford is considered a very popular residential area, having one of the largest shopping centres in the UK and also being one of the most connected stations in the network. The fact that Stratford is mostly a residential area and experienced the most traffic in 2020, suggests that most people were staying at home when the pandemic hit. The distribution is fairly even among other stations, though London Bridge also commands a significant segment.

Transitioning to 2022, the chart reveals a shift with King's Cross St. Pancras emerging with a substantial increase, possibly indicating a change in commuter flows or an adaptation to new travel habits. Waterloo LU maintains consistent traffic, while Victoria LU and Liverpool Street LU occupy smaller segments, suggesting a possible decrease or redistribution of passenger volume to other stations or modes of transportation. These shifts could be explained by assuming that the reason for the increase in traffic in these particular stations was due to most people returning to their physical work environments (offices), where these stations are very popular destinations since they are mostly located around financial centres and are also well connected with National Rail services (people living outside London are commuting to their offices). This idea is further supported by the fact that Liverpool Street Station, which is located very close to the City of London, was not included in the top 5 stations, during the pandemic, however with the lift of social distancing measures, it experienced an increase in traffic.

The effectiveness of the visualization is evident in its ability to communicate shifts in traffic at a glance, drawing immediate attention to changes over time. It supports a narrative of changing travel behaviours and infrastructure usage without the need for complex data tables. The visual comparison between the two years provides context for these changes, prompting further investigation into underlying causes, such as developments in urban planning, station renovations, or changes in service patterns. These charts are instrumental in forming a basis for strategic planning and operational adjustments within the London Underground network.

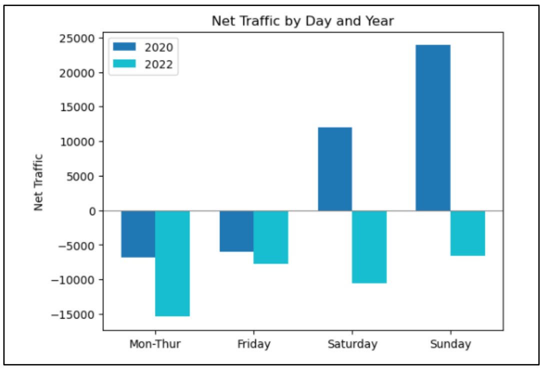

Fig. 7: A bar chart showing the net traffic by day and year.

The bar chart illustrates the net traffic differences by day for the years 2020 and 2022. It shows that Monday through Thursday, the net traffic was consistently lower in 2022 compared to 2020. However, there's a notable increase in net traffic on Saturdays and Sundays in 2022, suggesting a shift in travel patterns, where weekend travel has rebounded or even surpassed pre-pandemic levels.

The visualization showcases the impact of the pandemic on commuter behaviour, with potentially more people traveling on weekends in 2022, reflecting a change in leisure activities or work habits. The decrease in weekday traffic in 2022 could indicate a sustained adoption of remote work or a change in the overall volume of commuters.

This chart is pivotal in understanding how travel preferences have shifted over the course of the pandemic and the subsequent years. It informs transport authorities about potential adjustments needed in services to accommodate the new travel patterns, especially the increased weekend travel demand as observed in 2022.

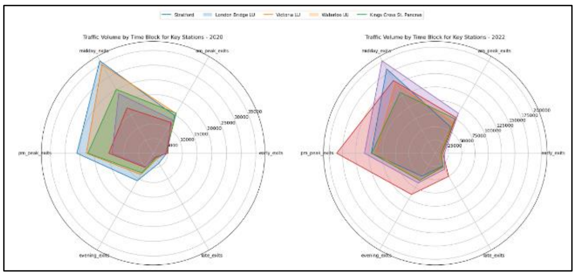

Fig. 8: Radar charts comparing traffic volume in 2020 & 2022.

The radar charts compare the traffic volume by time block for key London Underground stations in 2020 and 2022. The chart for 2020 shows a balanced distribution of traffic across different times of the day, with no single time block dominating. In contrast, the 2022 chart reveals a more noticeable concentration of traffic during the early exits and PM peak times, particularly at King's Cross St. Pancras and Stratford, suggesting a shift in travel behaviour, possibly due to changes in work patterns or social habits after the pandemic.

These visualizations effectively showcase the changes in traffic flow at major stations, illustrating where and when passenger volumes have increased or decreased. The changes observed in the 2022 data may also reflect broader trends in urban mobility and commuter preferences in the wake of the pandemic.

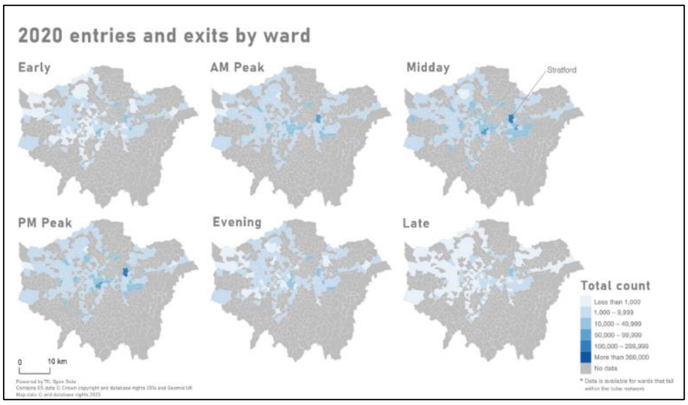

Fig. 9: Maps showing the entries and exits in different wards for 2020.

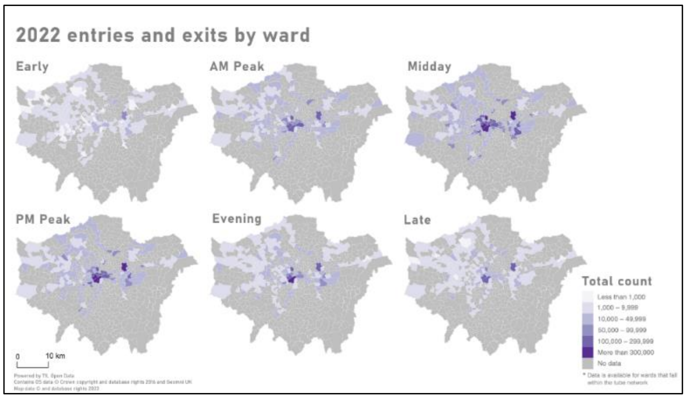

Fig. 10: Maps showing the entries and exits in different wards for 2022.

The 2020 and 2022 maps provide a look into London's Underground transport use across various wards (electoral subdivisions) during different parts of the day. The 2020 map reveals a spread-out pattern of train station use with notable activity in central London during the busy morning and evening rush hours. However, the maps also show that during the middle of the day and later at night, the traffic in most stations was significantly reduced.

Fast forward to 2022, and the maps show significant shifts in travel patterns. More people are using the trains, particularly in the evening. This suggests that as the city started recovering from the pandemic and that habits started changing. These shifts could be attributed to more people going out for leisure activities in the evening or that people started commuting back to their workplaces.

These maps could be quite useful for the Transport for London management, since they uncover patterns and showcase them from a geographical point of view. They highlight which train stations are getting busier and might need more trains or larger facilities to handle the growing number of passengers. They also show which stations aren't being used as much, which might mean train services could be reduced there and trains used in stations where passenger flow is more significant.

By showing where the most people are traveling and when, the maps guide planners in creating train schedules that meet the city's needs. They help ensure that the train system is efficient and ready to serve the residents of London. The analysis came to a natural end when the patterns of train use stopped changing significantly, suggesting that the city has adapted to a new routine in transit use.

The growing evening activity in the 2022 map indicates a city that's becoming lively again after the pandemic, with the train system playing a key role in this revival. Planners can use these insights to continue improving the train network, making it a reliable backbone for the city's nighttime economy and beyond.

4.3 Analysis Results

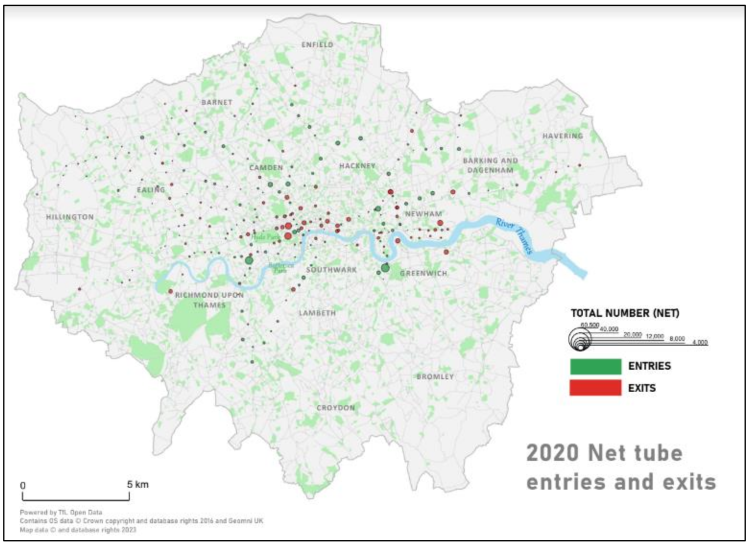

Fig. 11: A map showing the net entries and exits for 2020.

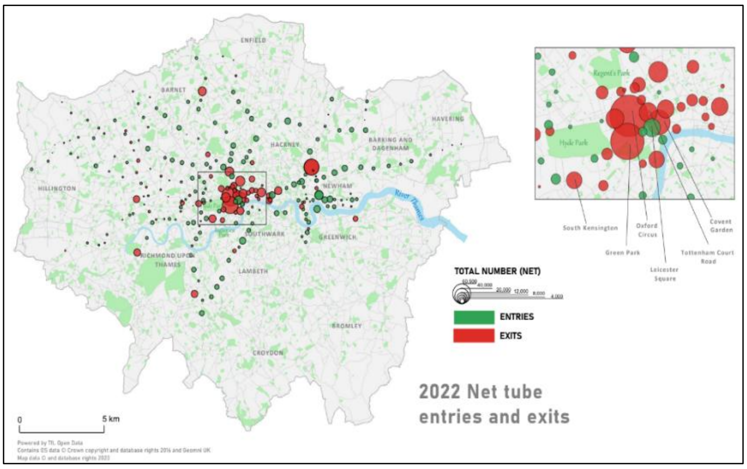

Fig. 12: A map showing the net entries and exits for 2022.

The maps of 2020 and 2022 London tube entries and exits provide a visual depiction of travel patterns across the city's wards. The 2020 map shows a more uniform distribution in entry and exit volumes, suggesting a balanced use of the tube network. In contrast, the 2022 map exhibits a concentration of activity in central London, with significantly larger volumes of exits, especially in areas like Covent Garden and Oxford Circus, indicative of a resurgence in central London's attractiveness post-pandemic.

These findings suggest a shift in the dynamics of urban movement, with 2022 showing a return to or even an increase in central activity compared to 2020. This could imply an economic recovery, or a revival of tourism. The implications for city planning and transport services are substantial, indicating a need to adjust resources to manage the increased central traffic and ensure the network's resilience.

These maps answer the research questions on how the pandemic affected urban mobility and the subsequent recovery pattern. While the full scope of the research questions may not be entirely addressed due to potential data limitations, the maps provide clear evidence of changing patterns in tube use, which is crucial for future transport planning in London.

5. CRITICAL REFLECTIONS

Reflecting on the analytical approach and the process undertaken, it becomes apparent that critical thinking and human judgement was a crucial part of the journey. The choice of visual and computational methods was key, aiming to effectively analyse complex urban transit data into easily understandable visualisations. Key decision points in the analysis were largely guided by the visual representations, which provided immediate and direct understanding of patterns and irregularities in the data.

The visuals served as more than mere illustrations; they were analytical instruments that aided reasoning. The shift in volume and patterns of London Underground usage became clearer through maps and charts, influencing the following steps of the analysis. When the 2020 and 2022 maps were compared, the evident contrast in traffic volumes, particularly in the evening and late hours, uncovered a resurgent city, possibly reflecting a post-pandemic revival.

Despite the robustness of the approach, not all research questions found complete answers. The limitations of the data, such as the lack of details regarding specific user demographics or the reasons behind travel, left gaps in understanding the full context of the observed patterns. Additionally, the techniques employed, while effective, had inherent weaknesses. The reliance on visual patterns and trends may have overshadowed the potential insights that could have been gained from more advanced statistical analyses or machine learning models that account for a wider range of variables and potential correlations.

Considering alternative approaches, software functionalities that allow for real-time data interaction and manipulation could have improved the analytical process's efficiency. The ability to look further into specific data points and test different scenarios interactively would provide a richer, more detailed exploration of the transit system's dynamics.

One key limitation of the approach was that the data assumed that the volume of entries and exits at tube stations would be enough for understanding transit patterns. This assumption may have neglected other influential factors such as surface transport links, pedestrian flows, buses, and bike-sharing. The specific focus on data conforming to the tube network may have also omitted relevant travel behaviours occurring outside its boundaries.

Despite these limitations, the chosen approach offers potential applicability to other datasets and problems, particularly those requiring the analysis of spatial-temporal patterns. The lessons learned here emphasize the importance of integrating diverse analytical methods, and the crucial role of visuals in data exploration.

Analysts working on similar projects are advised to adopt an iterative approach, methodically refining their analysis as fresh insights emerge. A balanced integration of visual data examination alongside statistical analysis is crucial for a well-rounded understanding. Awareness of the data's constraints is essential, and analysts should proactively seek additional information to address any deficiencies. Utilizing interactive and sophisticated tools can greatly enhance the depth of the analysis, providing a more intricate understanding of the complex data patterns.

In summary, this reflection underscores the necessity of a thoughtful, iterative analysis that combines the strengths of visual analysis and quantitative methods. The critical examination of this process offers valuable lessons for future analysis within the urban transit domain and beyond.

6. REFERENCES

[1] Rajabifard, A., Paez, D. and Foliente, G., 2021. *COVID-19 pandemic, geospatial information, and community resilience: Global applications and lessons* (p. 544). Taylor & Francis.

[2] Vickerman, R., 2021. Will Covid-19 put the public back in public transport? A UK perspective. *Transport Policy*, *103*, pp.95-102.

[3] Xin, R., Ai, T., Ding, L., Zhu, R. and Meng, L., 2022. Impact of the COVID-19 pandemic on urban human mobility-A multiscale geospatial network analysis using New York bike-sharing data. *Cities*, *126*, p.103677.

[4] Data Source: http://crowding.data.tfl.gov.uk/

[5] Bansal, P., Kessels, R., Krueger, R. and Graham, D.J., 2022. Preferences for using the London Underground during the COVID-19 pandemic. *Transportation Research Part A: Policy and Practice*, *160*, pp.45-60.

Section	Word Count
Abstract	97
Problem Statement	250
State of the Art	493
Properties of the Data	488
Analysis Approach	490
Analysis Process	1468
Analysis Results	200
Critical Reflections	495