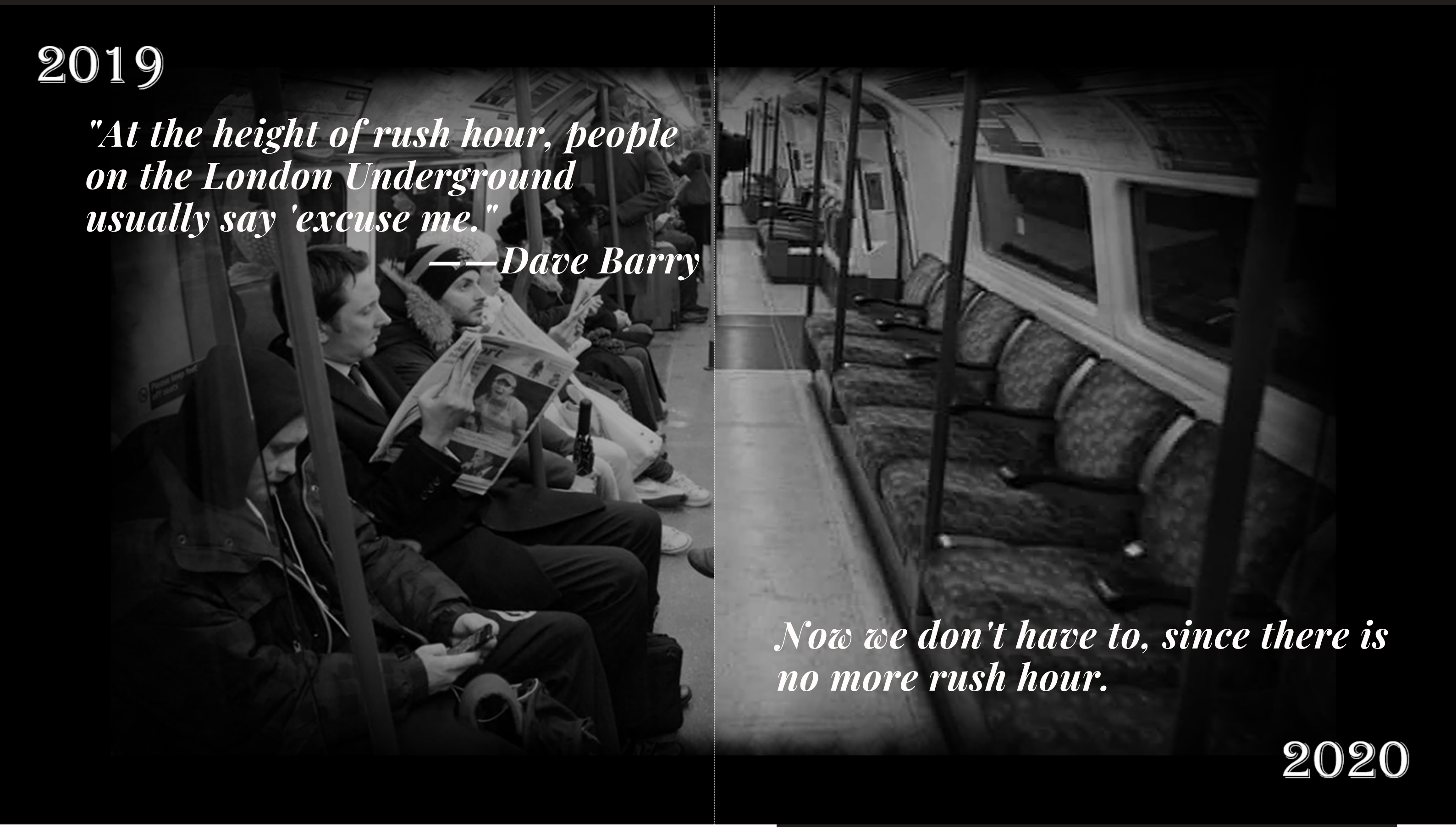
CASA0009 | GROUP REPORT

LONDON TUBE UNDER COVID-19

By Team Bravo



Team members:

Yujue Wang
Haocheng Sun
Zhiyuan Li
Yitong Yue
Yutong Xia

Word Count: 5094

Our Website:
http://dev.spatialdatacapture.org/~ucfnhsu/London_tube_under_COVID19/home.html

Github Repository:
https://github.com/HaochengSun722/SDC_public_html/tree/main

Contents

Introduction

Under the unprecedented shock of the COVID-19, public transportations, which is closely related to life and work have been affected to varying degrees, and the operation of public transportation systems in many countries is also faced with different forms of adjustment. The Underground is one of the most widely used forms of public transportation in London. However, Since the outbreak of COVID-19 in early 2020, remarkable modal shifts away from public transport, especially the Underground, to active and car-dependent transport were reported. Hadjidemetriou et al. (2020) estimated that, in light of restrictions on travel methods and social distancing policies, reduced operation of the London Underground and railway services contributed to reducing daily trips in Mar. 2020. Under COVID-19, the use of the London Underground dropped to a low of 5% of normal levels in early April 2020, gradually recovered to 40% before the second lockdown, and fell below 25% when the epidemic worsened in early November (Vickerman, 2021).

From the perspective of people's health, measures to reduce the movement of people, such as closing underground stations, significantly contribute to public health intervention(Lai et al., 2020). Studies have demonstrated that the boroughs of London with interchange stations have a higher incidence of epidemics(Goscé & Johansson, 2018), which may be the results of more interactive behaviors among people at interchange stations than through direct stations. Therefore, closing such stations may be more effective in slowing down the spread of the virus.

In addition, it is also significant to identify the different impacts on boroughs and stations in terms of social stability. Although closing public transportation during a pandemic can relatively slow the spread of the virus among communities, at the same time, it may also have a non-negligible impact on the social economy because of the reduction of population mobility. Determining the vulnerability of different boroughs and stations to COVID-19 helps achieve this trade-off (Sasidharan, Singh, Torbaghan, & Parlikad, 2020).

Research on the impacts of COVID-19 on transport behavior and the corresponding policy measures has been increasing rapidly (Zhang & Hayashi, 2020). The government has introduced a series of policies to restrict public transport use. For instance, in the pandemic scenario, reducing the frequency of parking at stops attracts large crowds, which is an essential factor in maintaining public transportation (Das et al., 2021). De Vos (2020) points out that even if the number of passengers decreases during the pandemic, public transport operators should not reduce public transport supply to avoid social isolation and financial difficulties. Therefore, accurately understanding the changes brought about by COVID-19 to the travel behavior is an effective shortcut to optimizing the operation of public transport during the pandemic. However, current macro studies are minimal about connections between changes in travel behavior and the spread of COVID-19.

Based on the relevant data and previous research, this study focus on the change in the London Underground traffic flow before and during the COVID-19 pandemic. We first compare the different trends in all modes of public transport and the unique change in London Underground journey pattern.Then we demonstrate journey changes at tube line level and borough level and quantitively analyze the correlation between COVID-19 cases and declines in underground journeys at borough level. Last but not least, we investigate new temporal characteristics in the flow at different metro stations under the pandemic by applying time-series clustering algorithms and regression analysis. Based on the Annual Station Counts dataset (provided by TfL), we calculate the dynamic time warping distance and adopt the hierarchical clustering method to investigate different temporal profiles of travel patterns on the London Underground and the impact of COVID-19 on it.

This study can provide government and related industries with a reference about COVID-19's impact on the London Underground traffic pattern and the new temporal characteristics of the underground journeys to improve related traffic policies and optimize the operation of the London Underground during the pandemic. An interactive website visualizing the dataset and results of the analysis is created, where users can have a better and intuitive understanding of our project.

The remainder of this report is organized as follows: Section 2 describes the data resources and data preprocessing. Section 3 shows our workflow and the main methods adopted in this study. Our results of analysis can be found in Section 4. Section 5 describes how we design the website. Lastly, the conclusion and discussion can be found in Sections 6 and 7.

Data

In this report, we collect **geological data** with spatial attributes of London underground lines, stations, London boroughs for visualization and geological analysis, and **statistical data** about Underground journeys before and during the pandemic, and COVID-19 cases to quantitively analyze COVID-19's impact on London Underground. The **statistical data** includes:

Annual Station Counts for 2019 and 2020, collected from TFL websites[1], which contain the entry / exit counts of every 15 minutes in a typical Monday to Thursday, Friday, Saturday or Sunday at each station, annualized to an annual total. The sample window of the 2020 dataset is during the UK Government's second national lockdown, which benefits our analysis.

Public Transport Journeys by Type of Transport for every month of 2019 and 2020 from the GLA website[2], including Bus, Underground, DLR, Tram, Overground, Emirates Airline, and TFL Rail as a supplement for overall comparison.

Coronavirus Cases and Vaccinations for each borough since the COVID-19 outbreak from the GLA website[3];

The **geological data** includes **London Underground Lines and Stations shapefiles** from Doogal[4] and **London Borough shapefiles** from the GLA website[5].

The statistics have 426 stations named with appendixes, but no information about lines or boroughs, and the geometry has 662 stations named without appendixes, and 28 tube lines, and 33 boroughs with no journey information. And We preprocess the data in 3 ways: unifying the numbers, names, and locations of stations and tube lines with TFL statistics and norms; assigning line and borough information to the corresponding stations, i.e., which lines and borough the station belongs to; setting station information to the corresponding lines and boroughs, i.e., which stations the line or borough contains.

---

1 http://crowding.data.tfl.gov.uk/

2 https://data.london.gov.uk/dataset/public-transport-journeys-type-transport

3 https://data.london.gov.uk/dataset/coronavirus--covid-19--cases

4 https://www.doogal.co.uk/london_stations.php

5 https://data.london.gov.uk/dataset/london-borough-profiles

Methodology

Work Flow

Our analysis is a three-step process based on data about underground journeys and COVID-19 cases. At the macro-level, we demonstrate the general trend in all modes of public transportation and the unique change in underground journey patterns through data visualization. At the medium-level, we use data visualization and linear regression to demonstrate journey changes at tube line level and borough level and quantitively analyze COVID-19's impact on underground journeys at borough level. At the micro-level, we apply Dynamic Time Warping (DTW) Clustering to conclude new temporal characteristics in the flow at different metro stations under the COVID-19. At each stage, we have further findings.

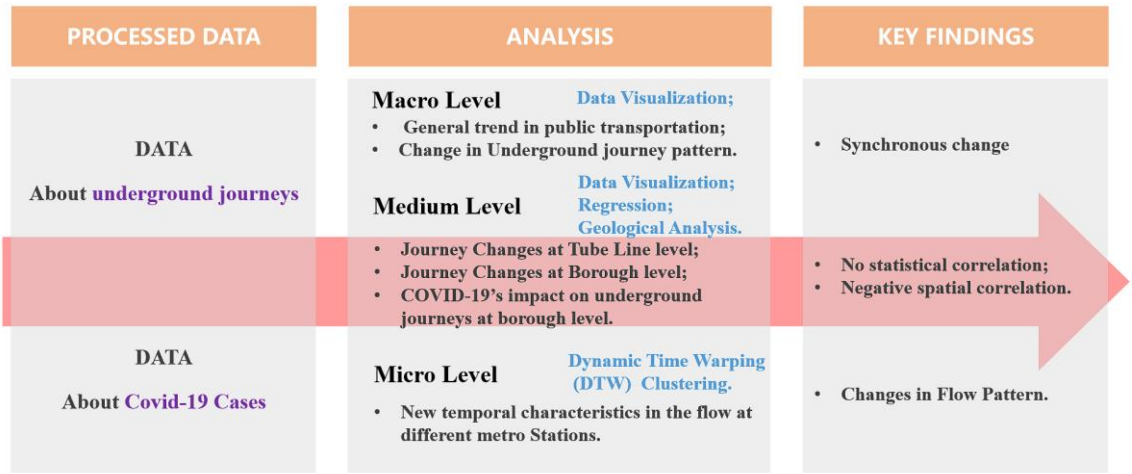


Figure 1 Workflow of our analysis

Clustering Algorithm

Normalisation of time series data

Due to the high dimensionality of time series data, it is often necessary to preprocess it to produce a smaller data set. Each original time series data consists of 96 time points with 15-minute intervals. We reduce the dimensionality by sum processing every hour.

The normalisation of time series data before distance measurement can improve the accuracy of clustering results. L2-normalization method is selected to normalise the data in this study, which is described as follows:

$$x = [x_1, x_2, \dots, x_d]$$

$$y = [y_1, y_2, \dots, y_d]$$

$$y = \frac{x}{\sqrt{\sum_{i=1}^{d} x_i^2}} = \frac{x}{\sqrt{x^T x}}$$

where x is the vector of time series data and y is the result of normalisation.

Distance (Dissimilarity) Measure

The commonly used measures of distance between time series include Euclidean distance, Dynamic Time Warping (DTW) distances and correlation and autocorrelation-based distances. The Euclidean distance measurement is widely used in cluster analysis for it performs well when data have compact or isolated clusters (Jain, Murty, & Flynn, 1999). However, when comparing the profiles of time series data, its sensitivity to distortions on the time axis may lead to less satisfactory results. The DTW approach can make up for this shortcoming of Euclidean distance, which is a shape-based method and allows two time series with similar shapes but locally out of phase to align in a non-linear manner (Ratanamahatana & Keogh, 2004). Thus, we choose the DTW distance measurement in this research.

Clustering Algorithm

After determining the distance measurement method, we will select a clustering algorithm based on the distance matrix. There are four types of clustering algorithms, namely partitional, hierarchical, density-based and grid-based clustering(Jain et al., 1999). The hierarchical cluster method, which is commonly used in previous research, can be divided into the agglomerative cluster analysis (a "bottom-up" approach) and the divisive cluster analysis (a "top-down" approach) (Rokach & Maimon, 2005). The dendrogram generated by the agglomerative hierarchical cluster analysis method can show many neighbor relationships and classification relationships in the data (Murtagh & Contreras, 2012), which is adopted in this study.

Analysis

Macro-Level: London Underground

1. Spatial Scale

According to the London underground annual counts (and the TFL website), in the London Underground, there are 16 underground lines and 426 stations distinguished by name, among which 392 are within the boundaries of the 33 boroughs (Figure 2, left). According to the Coronavirus Case dataset, as of 28 April 2021, there are 719,190 confirmed COVID-19 cases, with Newham, Redbridge, and Ealing reporting the highest numbers of cases among all boroughs (Figure 2, right).

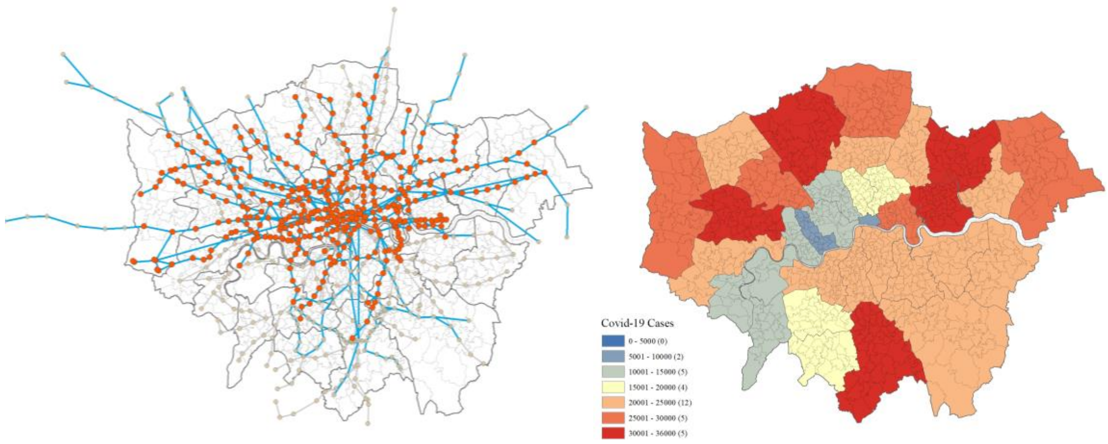


Figure 2 left: London underground stations, lines and boroughs; right: Cumulative cases in different London boroughs

2. Trends in all modes of public transport

Firstly we dip into different kinds of public traffic and check whether they were affected by the pandemic differently. During the pandemic, different modes of public transport journeys were all severely affected. From Feb. 2020 to May 2020 and Oct. 2020 to Feb. 2021, there were 2 cliff-like drops in public traffic journeys, the first of which dropped by 2/3, and the second by 1/2. The two declines coincide with the UK Government's 2 national lockdowns.

Generally, **all modes of public transport kept changing at a synchronized pace during this process, and the proportions of them before and during the pandemic remained stable** (Figure 3).

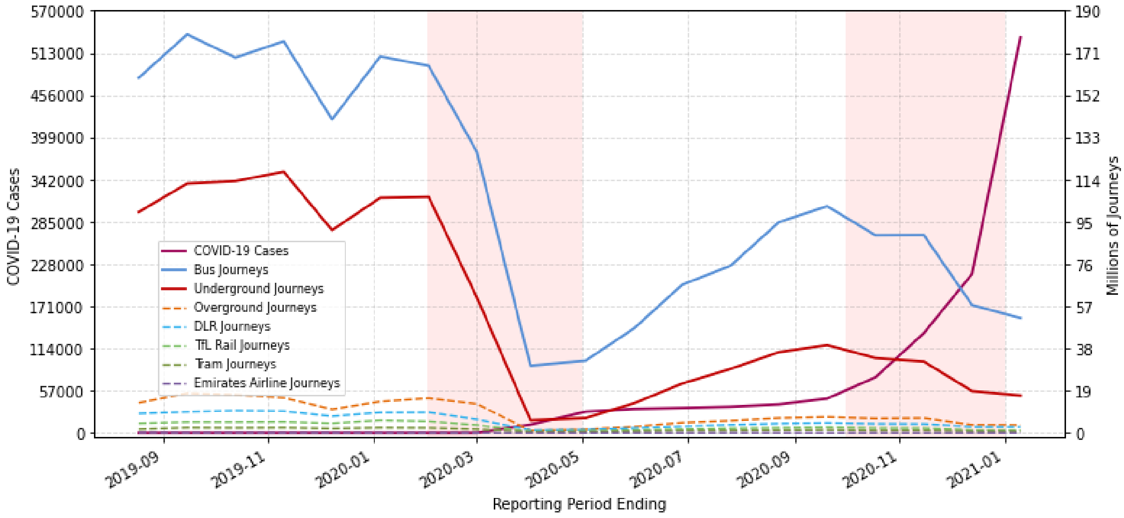


Figure 3 Cumulative cases in London and variation of journeys of London public transport

3. Change in the Underground's hourly/daily Fluctuations

After visualizing the overall trend of the London Underground along with other various modes of public transport, we take a closer look at the specific changes in London Underground journeys under the ongoing pandemic.

Underground journeys during a typical day declined by 2/3 to 3/4. The overall pattern of two peaks in the traffic fluctuations during a day still exist. But before the pandemic, the PM peak outnumbered the AM peak by 1/4 and was followed by a small night peak around 10:00 PM. During the pandemic, the decline in PM peak is much stiffer so that it becomes as small as the AM peak. The small evening peak at around 10:00 PM also disappears, indicating that **there is a more significant decrease in unnecessary evening trips for recreational activities** (Figure 4).

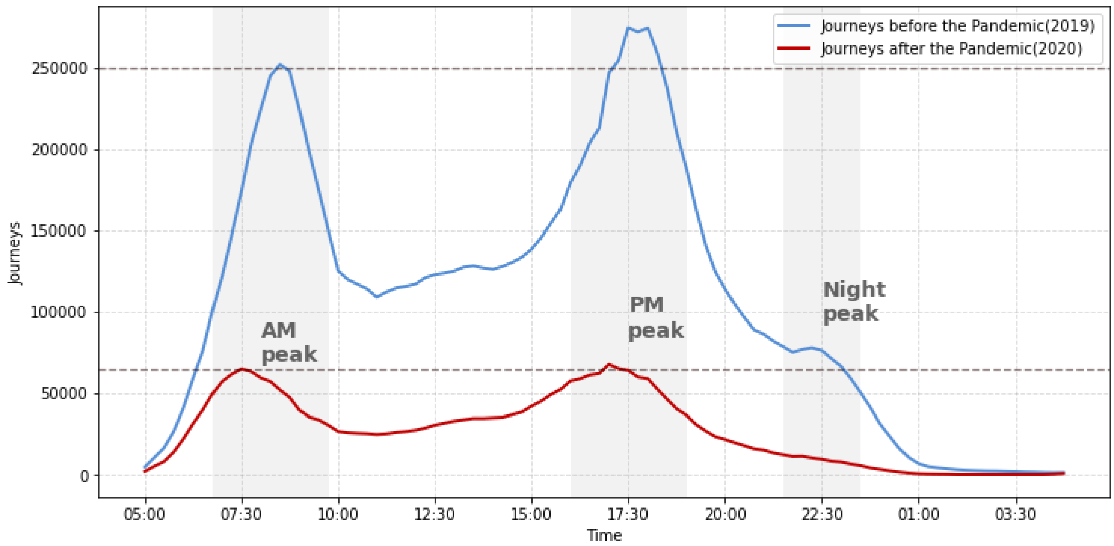


Figure 4 Journey pattern in a day before and during the pandemic

Change in Journey patterns over a typical week before and during the pandemic shows a similar trend (Figure 5). Before the pandemic, weekend journeys were about half of weekday journeys. During the pandemic, overall Underground journeys declined to a quarter of the former state. Meanwhile, **weekend journeys witness a greater decrease in percentage**, accounting for only 1/3 of weekday journeys.

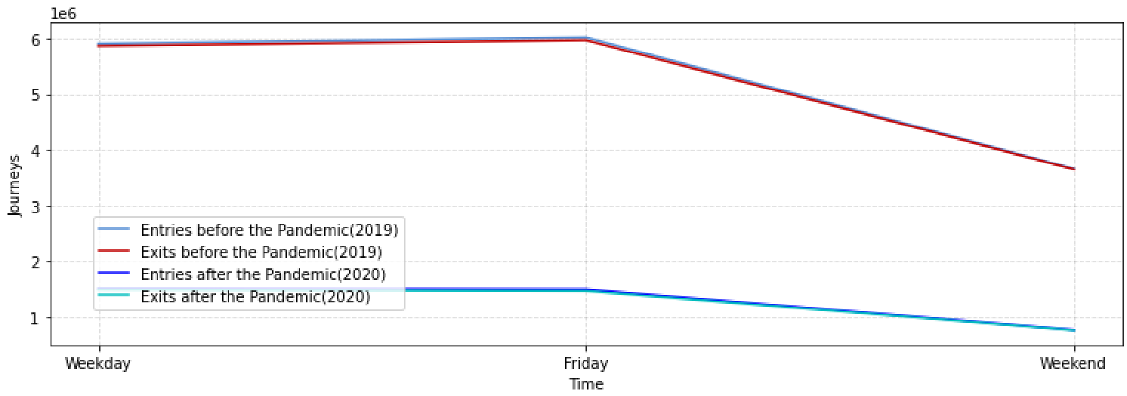


Figure 5 Journey pattern in a week before and during the pandemic

When comparing different journey patterns during a typical weekday or weekend, the left 2 subplots in Figure 6 shows that journeys on weekdays decreased to 1/4 of the original number, while the trips at weekends further dropped to 1/6, which is to say, the decline on weekends is greater. Besides, during the pandemic, the two peaks at weekends are much flatter and almost invisible. The right two subplots in Figure 6 shows during the pandemic, there is no significant difference between changes in entries and exits pattern.

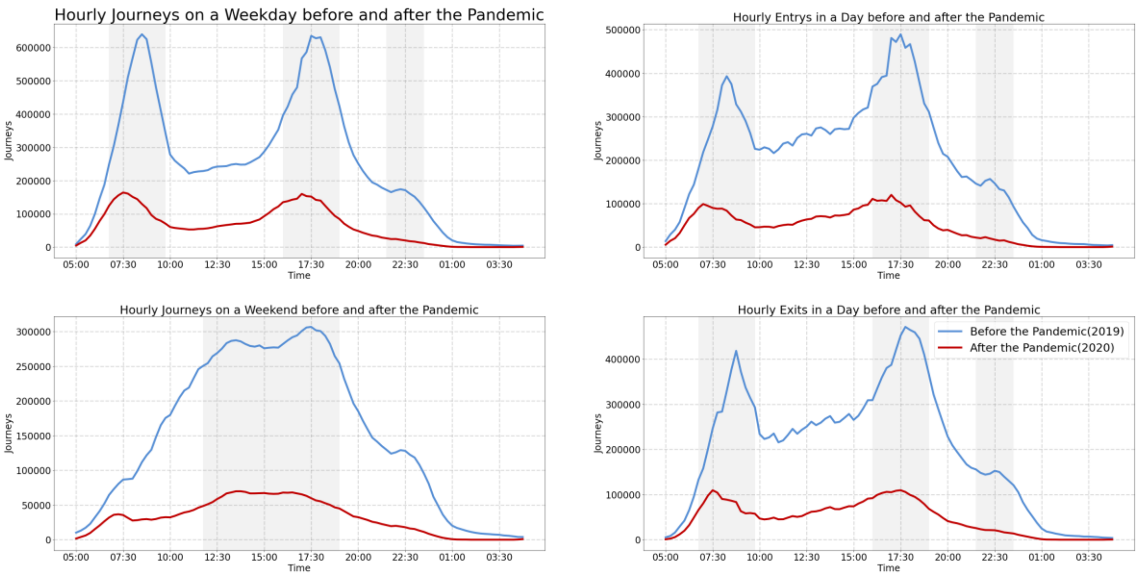


Figure 6 :a. Journey pattern on a weekday before and during the pandemic (upper left); b. Journey pattern on a weekend before and during the pandemic (lower left); c. Entry

pattern in a day before and during the pandemic (upper right); d. Exit pattern in a day before and during the pandemic (lower right)

Medium-Level: lines and boroughs

1. Journey declines in tube lines

Then we go down to the tube line level. Since coronavirus cases can't be summed up by different lines, we can only demonstrate changes in the journey patterns of different tube lines.

As is shown in Figure 7, tube lines witness different degrees of decline, from a smaller drop of 2/3 at DLR to a more significant decrease of 6/7 at Waterloo and City from a proportion perspective.

There are some lines where the 2 peaks are less pronounced, such as Central, and lines where the 2 peaks remain noticeable, such as Overground. We speculate that lines where the 2 peaks are still obvious, are connecting more concentrated residential areas and employment.

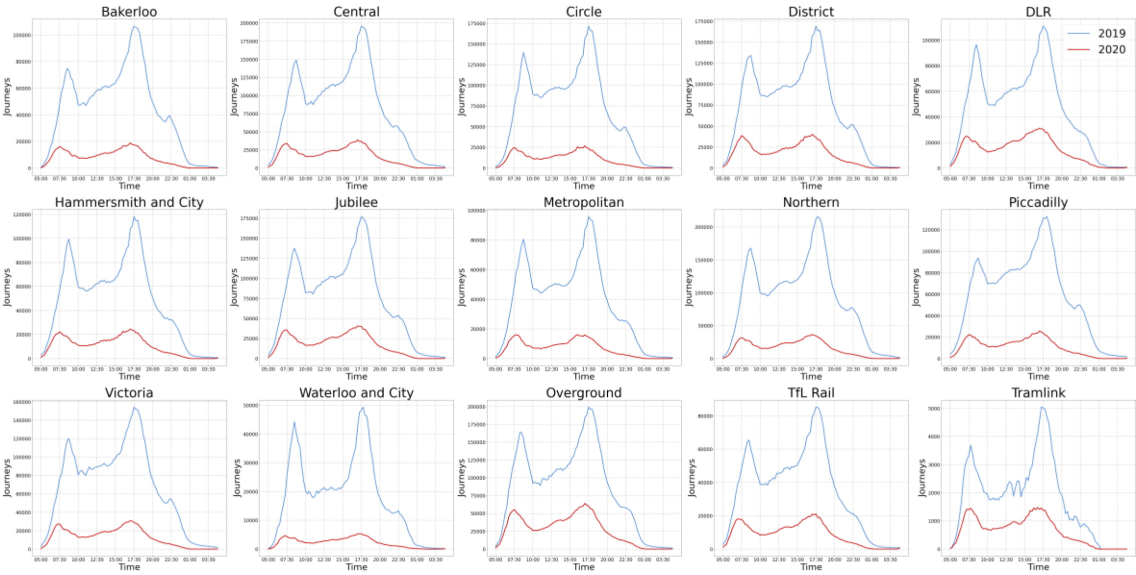


Figure 7 Journey patterns of different lines on a weekday before and during the pandemic (Emirate Air Line excluded)

In terms of absolute declines, Northern has the most significant declines of 7,038,411, followed by Central and Overground at 6,251,568 and 5,633,466 respectively. Tramlink and Waterloo and City have the smallest decreases, which is also related to their small base numbers (Figure 8).

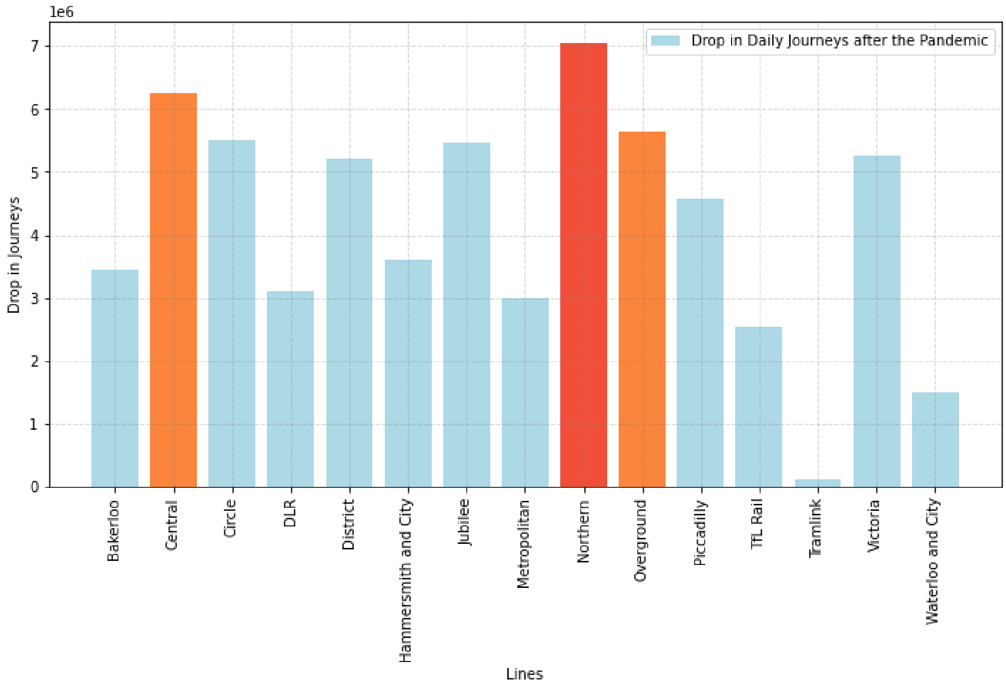


Figure 8 Decrease in daily journeys of different lines after the pandemic (Emirate Air Line excluded)

2. Journey declines in boroughs

Then we look at Journey declines in different boroughs. Only 30 boroughs have tube stations. Boroughs, including Sutton, Bexley, and Kingston upon Thames, have no tube stations.

The boroughs with tube stations have all seen journeys reduced to between 1/7 and 2/3 of the original numbers during the pandemic, with the most significant decreases in Westminster and City of London, and mildest drops in Bromley and Barking and Dagenham (Figure 9).

Like tube lines, some boroughs have less pronounced AM and PM peaks, such as Westminster and City of London, but some boroughs still have more pronounced peaks, such as Barking and Dagenham. This disparity is because boroughs with more obvious peaks have more residential areas and jobs.

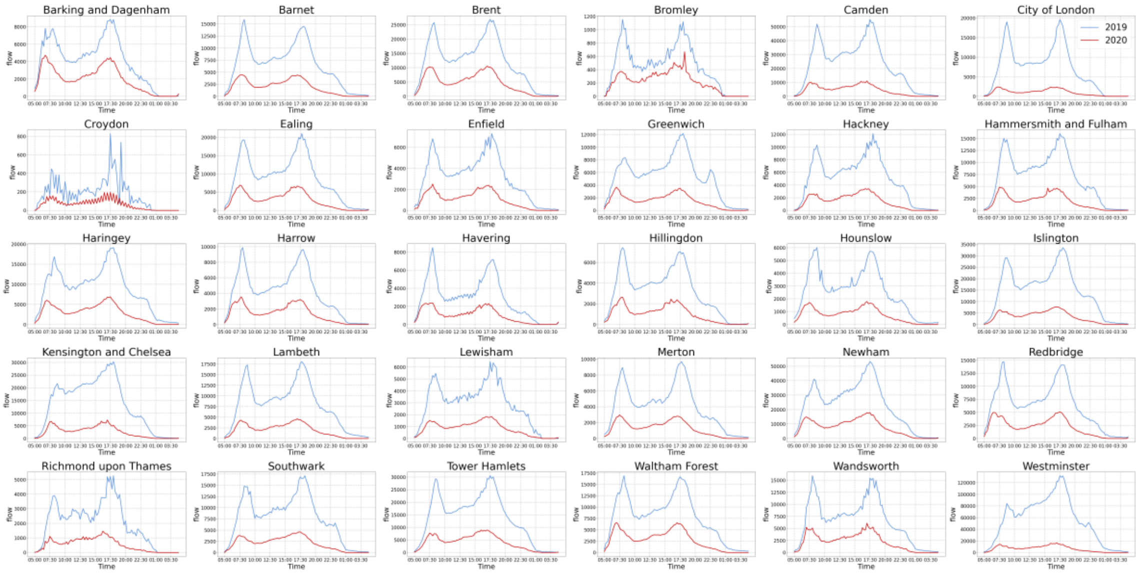


Figure 9 Journey pattern of different lines on a weekday before and during the pandemic (Emirate Air Line excluded)

In terms of absolute declines, Westminster experienced the greatest decline of 4,619,379, followed by Camden with declines of 1,740,022. The mildest decline was seen in Croydon, with a reduction of 10,650 (Figure 10).

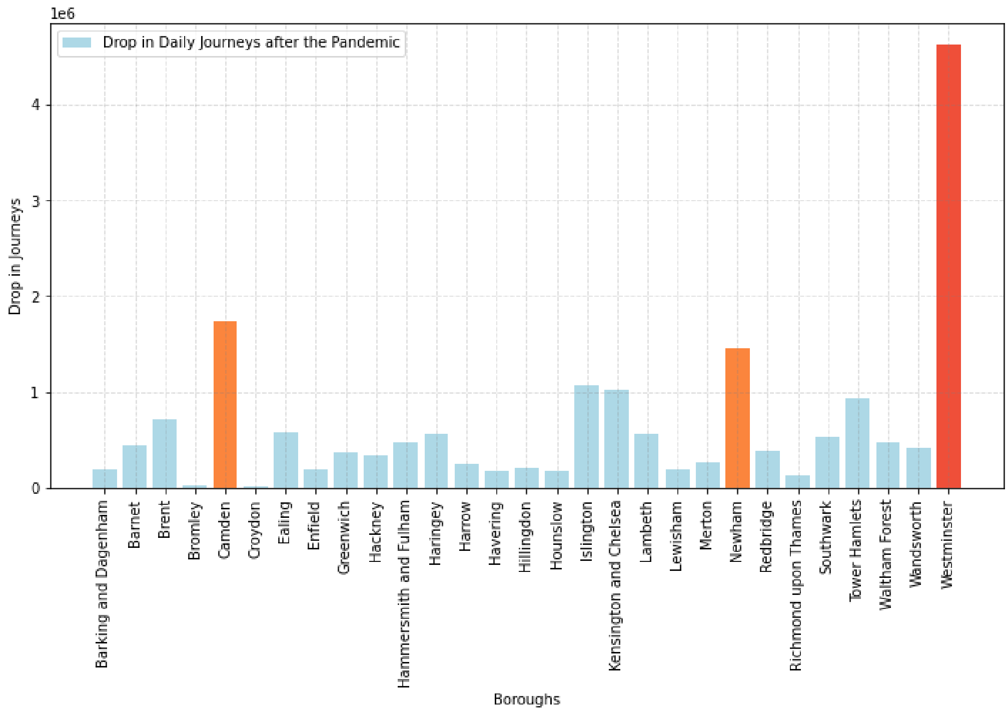


Figure 10 Decrease in daily journeys of different boroughs after the pandemic(only boroughs with stations)

3. COVID-19's impact at borough level

In addition to visualizing the change in Underground journey patterns, we also attempt to analyze COVID-19's impact on Underground journeys quantitively. We make a linear regression of the cumulative number of reported cases and the decline in the London Underground journeys for each borough. As the variables are all positively skewed (Figure 11), we also fit a linear regression after taking the logs for the variables. But as is shown in Figure 12, whether taking logs or not, the relationship between the two variables is not significant, with low R2 values (0.2% and 0.7% after log transformations) and high p-values (0.822 and 0.663 after log transformations）.

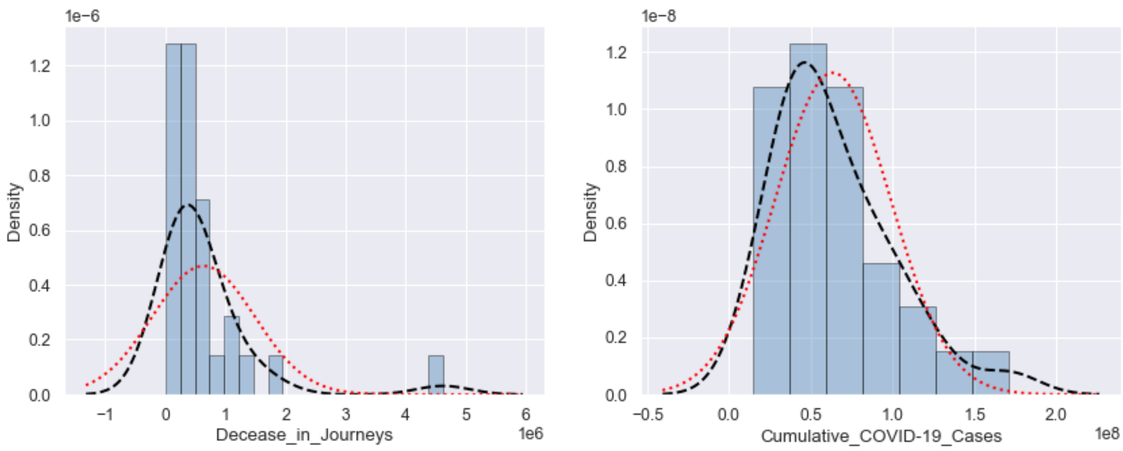


Figure 11 Histograms of cumulative COVID-19 cases and decrease in daily journeys of different boroughs

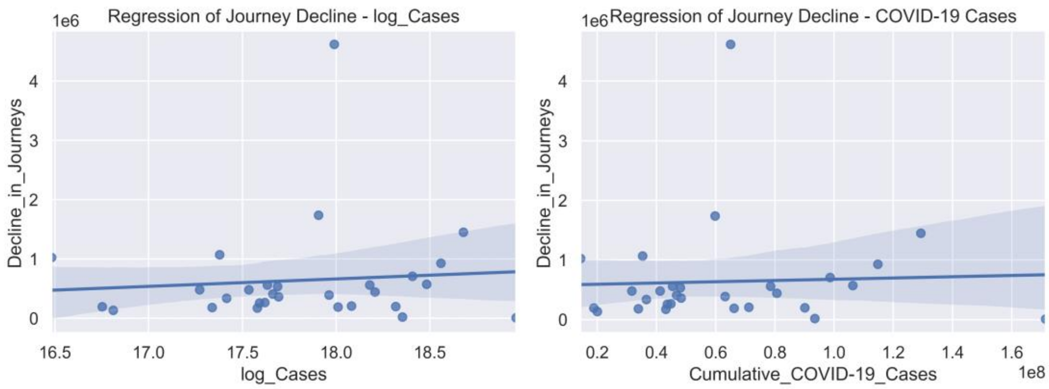


Figure 12 Regression of cumulative COVID-19 cases and decrease in daily journeys of different boroughs (left: before log transformation; right: after log transformation)

However, A divergence between COVID-19 cases and the decrease in underground journeys can be evident when looking at the map (Figure 13). Boroughs closer to central London and the city of London witnessed a greater decline in journeys, which can be explained by the distribution of residence and jobs or services. Jobs and services tend to gather in the city center. When the pandemic hits, trips to and from work, for instance, are the first to be affected. On the contrary, the confirmed cases are reported more at the peripherals. Geological distribution patterns of COVID-19 cases and decline in tube journeys are the opposite of each other.

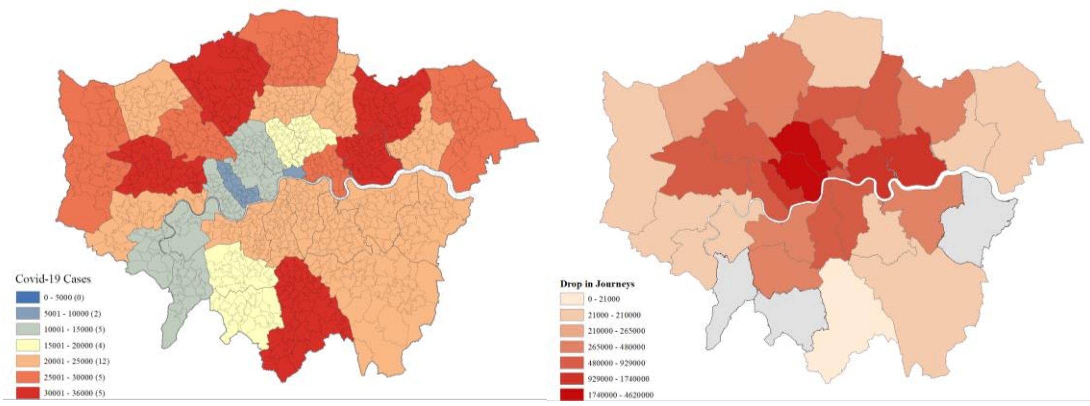


Figure 13 Regression of cumulative COVID-19 cases and decrease in daily journeys of different boroughs (left: before log transformation; right: after log transformation)

Micro-Level: stations

Clustering Results

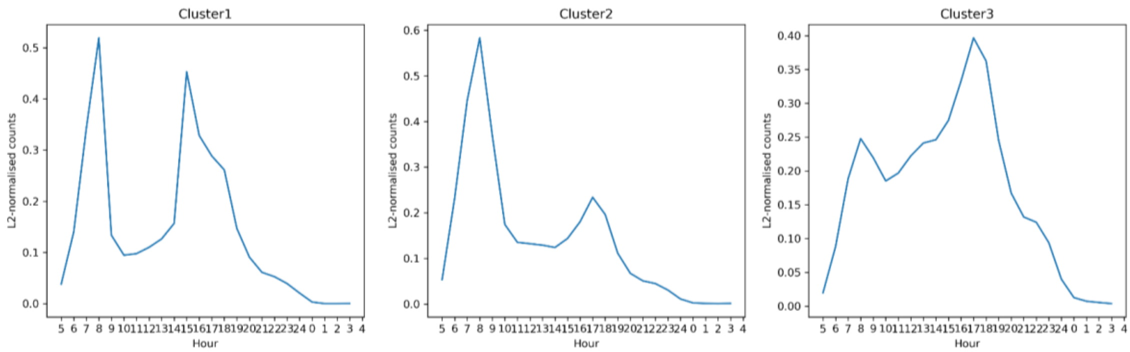


Figure 14 Three clusters of Transport of London Underground counts

After analyzing the pandemic's influence on the journey decline in terms of tube lines and boroughs, we want to study further on the station level with clustering methodology.

Figure 14 and Table 1 show the result of clustering. These three kinds of shapes can be described as **Two-peak pattern, AM-peak pattern**, and **PM-peak pattern.** It is found that the percentage of cluster 1 is low, while that of cluster 3 is relatively high, exceeding 85%. Although the counts of cases are unevenly distributed among the three clusters, it is acceptable because if we reduced the number of clusters to obtain a more even distribution, some unique shapes could not be obtained.

Table 1 The breakdown of cluster cases

Cluster	Description	Cases	Percentage (%)
1 Two-peak	Two peaks are obvious	78	1.16
2 AM-peak	Morning peak is obvious	808	12.00
3 PM-peak	Evening peak is obvious	5850	86.85
	Total	6736	100

According to Table 1, The most common temporal profile in the London underground is cluster 3 with a prominent PM peak, where the maximum appears at around 5:00 PM. Although an AM peak exists in this profile, it is not as evident as the PM peak. After the morning peak, there is a slight drop in passenger flow, and it begins to slowly increase from about 10 in the morning until the PM peak. These stations have a consistently high passenger flow most afternoons and are relatively less busy in the morning. The second most common temporal profile is cluster 2, with an eye-catching AM peak. The peak appears at around 8:00 AM and then drops sharply until 11:00 AM. After a flat period from 11:00 AM to 2:00 PM, it then slowly rises to usher in another smaller peak (5:00 PM). The less common one is the pattern with two peaks (cluster1), where two obvious peaks of about the same height occur at 8:00 AM and 3:00 PM.

Table 2 The composition of clusters

Cluster	2019				2020			
	Cases			Percentage (%)	Cases			Percentage (%)
1	12	MTT	5	6.41	66	MTT	28	35.90
		FRI	4	5.13		FRI	34	43.59
		SAT	1	1.28		SAT	1	1.28
		SUN	2	2.56		SUN	3	3.85
2	400	MTT	219	27.10	408	MTT	189	23.39
		FRI	173	21.41		FRI	173	21.41
		SAT	4	0.50		SAT	32	3.96
		SUN	4	0.50		SUN	14	1.73
3	2956	MTT	618	10.56	2894	MTT	625	10.68
		FRI	665	11.37		FRI	635	10.85
		SAT	837	14.31		SAT	809	13.83
		SUN	836	14.29		SUN	825	14.10

Table 2 shows the composition of each cluster. It can be seen that in 2020 there are significantly more stations belonging to cluster 1 (two-peaks pattern) but fewer stations classified to clusters 3 (PM-peak pattern) than in 2019, indicating that the impact of the COVID-19 on London underground travel patterns is mainly reflected in the increase in two-peak patterns (cluster 1) and the decrease in PM-peak patterns (cluster 3). The reason for this shift may be that the COVID-19 has led to the implementation of remote working and learning policy and 2 major lockdowns, which has reduced the demand for unnecessary commuting, especially for shopping, recreation, and shopping in the afternoon or the evening. Thus, some stations have smaller journey counts during the original PM-peak, transferring their temporal journey patterns from cluster 3 to the other two clusters. Also, the change of numbers of stations in cluster 1 mainly occurs on a weekday (the proportion of MTT and FRI increased from 74% to 93%), which further supports this conclusion.

Case Study – Brentwood

A significant change of travel patterns can be observed in Brentwood station (Figure 15), located on the Great Eastern Main Line in the East of England. We selected the time series data of Brentwood station on a typical Friday and Sunday in 2019 and 2020 to present a case study.

Figure 15 The location of Brentwood station in London

Figure 16 shows the temporal profiles of Brentwood in 2019 and 2020. In the terms of the overall size of the flow, it significantly decreased by around 70%, which is indicated by their y-axis maximums. In terms of the temporal profiles, we can see that the travel patterns of Brentwood on Friday transformed from cluster 2 and 3 to cluster 1. According to the blue and orange lines in the first figure, there was a significant morning peak at around 7-8 AM (consistent with our description of cluster 2) as well as a significant PM peak at around 5-6 PM (consistent with our description of cluster 3), which indicates that many people go to work in the morning at this station and come back here after getting off work in the evening. While these two temporal profiles changed to two-peak patterns after COVID-19, indicating the most common travel behaviour is no longer going out in the morning and back in the evening, which can be explained by implementing the remote working policy. While travel patterns on Sunday did not significantly change, inferring that residents living nearby less go out or less use rail to travel except commuting on weekdays.

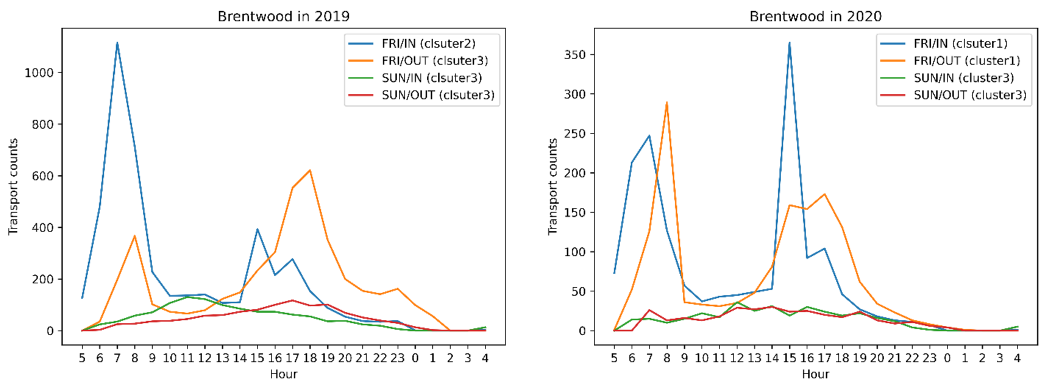


Figure 16 The temporal profiles of Brentwood in 2019 and 2020

Website Design

Design Concept

We hope our website can answer several questions that people have been concerned about: First, has the usage pattern of London Underground really changed after the epidemic regarding relevant polices and measures? On the other hand, since the peak periods of different stations various in different regions and on different workdays, how should travelers identify the traffic of a specific station on a specific day to help them better plan their trips?

Therefore, we would like to build an interactive map that can help people explore the London underground under three different scales: On the time scale, the website can show us the London Underground traffic changes before and during the pandemic; On the city scale, it can show us the distribution of flows and patterns; On the station scale, it can show us the specific flow changes by yearly, daily, and hourly.

Target Users

This website is mainly designed for the users of the London Underground. During the epidemic, it can help travelers understand the usage pattern of different tube stations on different days and time periods to help them better plan their travel time in response to the government's call for avoid unnecessary traveling during rush hours.

Web Framework

This website is designed into four parts: HOMEPAGE, STORY TELLING, INTERACTIVE MAP, and ABOUT.

The HOMEPAGE as the website portal mainly displays the theme of our website, namely the London Underground analysis under COVID-19, which can help users have an idea of the general content of our website at the moment they entered.

The STORY TELLING part mainly introduces the research background, major findings, and case study in the form of text, pictures and plots, so that users can develop a preliminary understanding of the use of the London Underground during the epidemic.

The INTERACTIVE MAP is the key part of our website, which can help users explore the traffic pattern of London Underground by themselves. Users can directly observe the inbound and outbound flow of every stations on the map by selecting different filter conditions, and they can find out even more about the station details when clicking on a specific station.

The ABOUT part mainly introduces the relevant background of the website, including project introduction and designer information.

Backend Implementation

In addition to the user-oriented design of the website, the map interface calls the Google Maps API (including London Underground lines) as the base map. Stations with coordinates and various data tags (such as station name, year, date, time, etc.) are also made into API as the data source of the station appearance, MODE SELECTION section, and STATION DETAILS section.

UI Design and Visualisation

In order to reflect the major impact of the COVID-19 epidemic on people's lives, we set the main color of our website to be heavy black. Dark red is also chosen as the secondary color to deliver a high visual impact. On the one hand, it is the color of blood and dangers that makes people vigilant. On the other hand, it also represents vitality and warmth, in contrast to the cold black, which reflects people's courage in fighting the epidemic that will enhance the perceptual effect of visual design and mobilize the emotional resonance of users. We adopt these colors to create a simple icon to serve as the logo of this website（Figure 17）.

Figure 17 Logo of our website

The navigation bar is set in the upper right corner of each web page to provide users with links to the four sections of the website (Figure 18).

Figure 18 Navigation bar of our website

At the HOMEPAGE, we inserte two comparative pictures of the London Underground before and during the epidemic as the background in attempt to leave users great impressions. Users can observe the huge difference by sliding the slider (Figure 19).

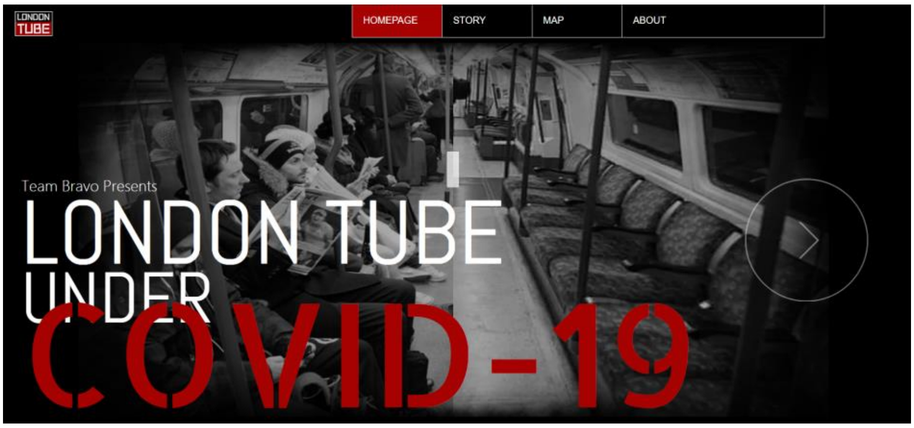


Figure 19 HOMEPAGE of our website

In terms of the STORY page design, we set the page to turn left and right to enhance the user's interactive experience. A progress bar was inserted at the bottom right of the page to display the reading progress at the mean time. Statistical plots are used to represent flow and pattern changes (Figure 20).

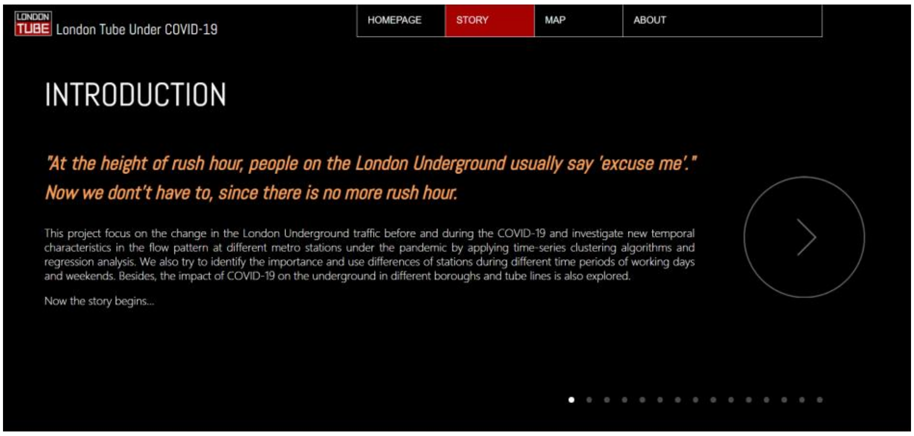


Figure 20 STORY page of our website

At the INTERACTIVE MAP interface, the base map is also made into dark grey and black, and the London Underground lines are made in different shades of red. (Figure 21).

Figure 21 Base map of the INTERACTIVE MAP

A SELECT MODE section will be displayed at the left side of the webpage to choose different conditions. The year, entry or exit flows, and weekday options are selected in the form of a drop-down list, and the time is selected in the form of a sliding bar. Navigations will guide website users through all the steps (Figure 22).

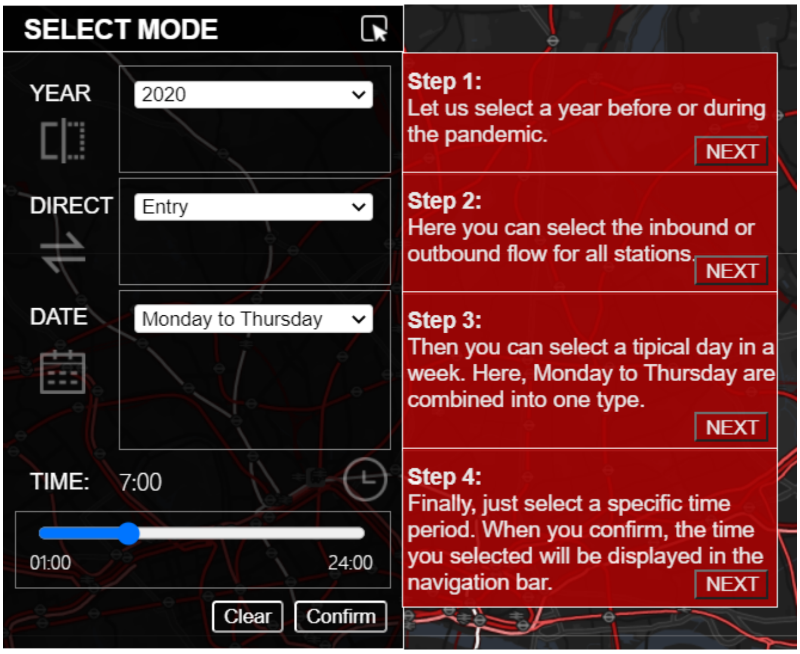


Figure 22 SELECT MODE section in the INTERACTIVE MAP interface

After confirming the choice, the map will display the London Underground traffic under the conditions by using station size to represent traffic flows and the station color to represent flow patterns (Figure 23). A LEGEND section is also displayed at the bottom left of the webpage to provide annotations (Figure 24). Both a SELECT MODE section and LEGEND section can be hidden to facilitate users to observe the map.

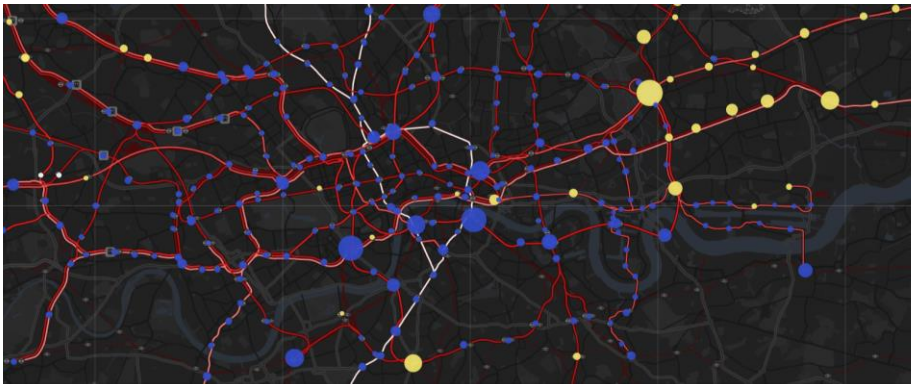


Figure 23 Map displayed after confirming the choice

Figure 24 LEGEND section in the INTERACTIVE MAP interface

When the cursor moves to each station, the corresponding station name and flows will be displayed in the form of a small box. When clicking on a specific station, the STATION DETAILS section will be displayed on the right side of the interface, which includes the station name, the station type (individual TFL managed rail modes), the NLC (National Location Code), the ASC (Alphabetic Station Code), a brief description of the flows and traffic pattern changes of the corresponding station before and during the COVID-19 pandemic, and time series line charts to show the details on flow changes (Figure 25).

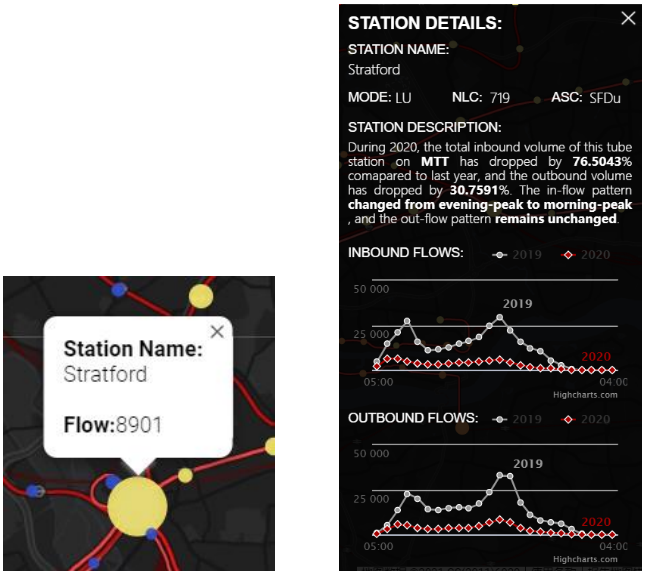


Figure 25 Information box and STATION DETAILS section for selected stations

The ABOUT section is divided into four parts including PROJECT, TEAM, RESOURCES and CONTACT, and page is set to scroll up and down for users to browse quickly. (Figure 26).

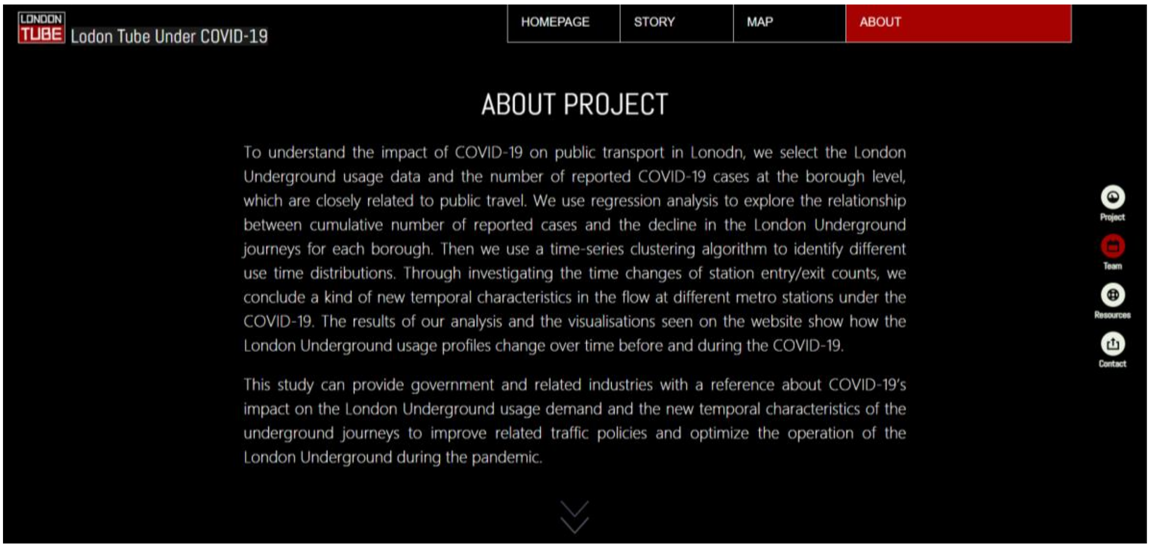


Figure 26 ABOUT section of our website

Conclusion

Based on our analysis and visualization, our key findings are as follows:

At the macro-level, a) journeys of all modes of public transport, including London Underground, keep changing at a synchronized pace during the pandemic and witness 2 cliff-like decreases due to the 2 national lockdowns; b) there is a greater decrease in unnecessary evening trips for recreational activities than journeys to and from work; c) weekend journeys experience a greater decrease in proportion.

At the medium-level, a) among tube lines, Waterloo and City has the greatest decline in terms of proportion, and Northern has the most significant declines of 7,038,411in terms of absolute number; b) among boroughs, Waterloo has the greatest decline in terms of both proportion and absolute number by 2/3 and 4,619,379 respectively; c) there is no significant statistical relationship between the COVID-19 cases and the decrease in the underground journeys at borough level; d) geological distribution patterns of COVID-19 cases and decline in tube journeys are the opposite of each other as boroughs closer to central London witnessed greater declines in journeys and more confirmed cases were reported at the peripherals.

At the micro-level, the impact of the COVID-19 on London underground travel patterns is mainly reflected in the increase in two-peak patterns and the decrease in PM-peak patterns. Moreover, the change in the numbers of stations in each cluster mainly occurs on a weekday. Due to the implementation of remote working and learning policy and 2 major lockdowns, some stations have smaller journey counts during the original PM-peak, transferring their temporal journey patterns from cluster 3 to the other two clusters.

Limitations

There are several limitations to our project.

As for the analysis, firstly restricted by the data we collected, we can only compare the underground journeys based at 2 discrete time windows in 2019 and 2020, so its relationship with the up-to-date COVID-19 cases would be undermined. Secondly, we didn't perform further regression based on geographical correlations between COVID-19 cases and traffic decreases, which can be carried out in future research. Thirdly, our clustering analysis only focuses on the shape of the time series of transport counts, without considering the size of the flow, as the data have been normalized before distance measurement. However, for some stations with few flows throughout the day, their temporal profiles may not be accurate when explained by the above description of each cluster. An extreme example is that if there was a station with only one person going out at 6:00 PM, it would be inaccurate to describe it as the PM-peak pattern.

As for the web design, due to the huge decline in the use of London Underground in 2020, it is difficult for our interactive map to set an appropriate station size when displaying station flows to make stations in both 2019 and 2020 have good visibility. If the stations in 2019 during peak hour are set to an appropriate size, it will be difficult to see any difference in the size of stations of 2020. So we adopted a relatively intermediate station size in reflecting total flows.

We also planned to add more charts reflecting macro changes and use the number of epidemics as a base map in the interactive map interface initially to help users develop a more intuitive understanding. However, due to the limited time and insufficient layout space for the interactive map, we failed to complete this task.

References

Das, S., Boruah, A., Banerjee, A., Raoniar, R., Nama, S., & Maurya, A. K. (2021). Impact of COVID-19: A Radical Modal Shift from Public to Private Transport Mode. *Transport Policy*.

De Vos, J. (2020). The effect of COVID-19 and subsequent social distancing on travel behavior. *Transportation Research Interdisciplinary Perspectives, 5*, 100121.

Goscé, L., & Johansson, A. (2018). Analysing the link between public transport use and airborne transmission: mobility and contagion in the London underground. *Environmental Health, 17*(1), 1-11.

Hadjidemetriou, G. M., Sasidharan, M., Kouyialis, G., & Parlikad, A. K. J. T. r. i. p. (2020). The impact of government measures and human mobility trend on COVID-19 related deaths in the UK. *6*, 100167.

Jain, A. K., Murty, M. N., & Flynn, P. (1999). Data clustering: a review. *ACM computing surveys, 31*(3), 264-323.

Lai, S., Bogoch, I. I., Ruktanonchai, N. W., Watts, A., Lu, X., Yang, W., . . . Tatem, A. J. (2020). Assessing spread risk of Wuhan novel coronavirus within and beyond China, January-April 2020: a travel network-based modelling study. *medRxiv*, 2020.2002.2004.20020479. doi:10.1101/2020.02.04.20020479

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *2*(1), 86-97. doi:https://doi.org/10.1002/widm.53

Ratanamahatana, C. A., & Keogh, E. (2004). *Making time-series classification more accurate using learned constraints.* Paper presented at the Proceedings of the 2004 SIAM international conference on data mining.

Rokach, L., & Maimon, O. (2005). Clustering Methods. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 321-352). Boston, MA: Springer US.

Sasidharan, M., Singh, A., Torbaghan, M. E., & Parlikad, A. K. (2020). A vulnerability-based approach to human-mobility reduction for countering COVID-19 transmission in London while considering local air quality. *Science of The Total Environment, 741*, 140515.

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

Zhang, J., & Hayashi, Y. (2020). Impacts of COVID-19 on the Transport Sector and Measures as Well as Recommendations of Policies and Future Research: Analyses Based on a World-Wide Expert Survey. *Available at SSRN*.