





## Article

# Capturing Twitter Negativity Pre- vs. Mid-COVID-19 Pandemic: An LDA Application on London Public Transport System

Ioannis Politis , Georgios Georgiadis , Aristomenis Kopsacheilis , Anastasia Nikolaidou   
and Panagiotis Papaioannou

Transport Engineering Laboratory, Department of Civil Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; ggeorgiadis@civil.auth.gr (G.G.); kopsacheilis@civil.auth.gr (A.K.); nikolaid@civil.auth.gr (A.N.); papa@civil.auth.gr (P.P.)

\* Correspondence: pol@civil.auth.gr

**Abstract:** The coronavirus pandemic has affected everyday life to a significant degree. The transport sector is no exception, with mobility restrictions and social distancing affecting the operation of transport systems. This research attempts to examine the effect of the pandemic on the users of the public transport system of London through analyzing tweets before (2019) and during (2020) the outbreak. For the needs of the research, we initially assess the sentiment expressed by users using the SentiStrength tool. In total, almost 250,000 tweets were collected and analyzed, equally distributed between the two years. Afterward, by examining the word clouds of the tweets expressing negative sentiment and by applying the latent Dirichlet allocation method, we investigate the most prevalent topics in both analysis periods. Results indicate an increase in negative sentiment on dates when stricter restrictions against the pandemic were imposed. Furthermore, topic analysis results highlight that although users focused on the operational conditions of the public transport network during the pre-pandemic period, they tend to refer more to the effect of the pandemic on public transport during the outbreak. Additionally, according to correlations between ridership data and the frequency of pandemic-related terms, we found that during 2020, public transport demand was decreased while tweets with negative sentiment were being increased at the same time.

**Keywords:** COVID-19; Twitter; sentiment analysis; public transport; latent Dirichlet allocation



**Citation:** Politis, I.; Georgiadis, G.; Kopsacheilis, A.; Nikolaidou, A.; Papaioannou, P. Capturing Twitter Negativity Pre- vs. Mid-COVID-19 Pandemic: An LDA Application on London Public Transport System. *Sustainability* **2021**, *13*, 13356. <https://doi.org/10.3390/su132313356>

Academic Editors: Armando Carteni and Ilaria Henke

Received: 7 November 2021

Accepted: 28 November 2021

Published: 2 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In times of emergency, imposed restrictions are mainly aimed at reducing mobility, affecting the economy and social life, as in the case of the coronavirus disease 2019 (COVID-19). COVID-19 is a contagious disease caused by the Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2). In March 2020, the World Health Organization declared COVID-19 a pandemic due to the uncontrolled spread of the virus worldwide [1]. By June 2021, confirmed cases of the disease exceed 180 million, including 4 million deaths worldwide.

Due to the lack of any effective therapeutics or vaccines at the early stages of the pandemic, countries worldwide began taking social distancing measures to control the spread of the disease. To further support their implementation, social distancing measures were accompanied by specifically designed travel restrictions and modifications on transport systems.

Based on results from different countries, it can be argued that the implementation of social distancing measures had a significant effect on personal mobility [2,3]. The daily number of trips was significantly reduced, while the choice of transport mode was also influenced by the spread of the coronavirus, with travelers choosing to avoid public transport in favor of more private means of transport, such as their car or active mobility [4–6]. Additionally, there was a differentiation in mobility patterns in relation to certain characteristics, such as gender, age, or income [7].

Technology advancements and social media create opportunities to keep people safe, informed, and connected. However, the same tools also enable and amplify the current infodemic that continues to undermine the global response and jeopardizes measures to control the pandemic [8]. Social media can be regarded as a useful source for understanding public opinion, especially in relation to numerous controversial issues that people face in their everyday life. As social network sites provide an astonishing amount of information about the number of people who participate in the networks, these sites are potential sampling frames for transport data collection [9]. Social media data can be proven beneficial for transport planners, especially in the era of smart cities, where the concept of the Internet of Things (IoT) provides new prospects for the operation of transport systems through the seamless provision of streams of data and real-time analytics [10].

One of the social media platforms frequently utilized by millions across the globe to share their views with others is Twitter. The microblogging application of Twitter is ideal for conversations and sharing small posts. Twitter data can be processed, and various types of analysis can be performed to understand public opinion [11]. Through social media monitoring, operators and public authorities can get meaningful insights, understand in more depth users' needs and actions, and collect social media content referring to relevant topics (e.g., bike-sharing platforms) [12]. Furthermore, sentiment analysis can be used to compute and measure the sentiment that may be contained in users' posts.

This vast amount of online information available makes social media a reflection of the real world. Motivated by the ongoing global COVID-19 pandemic, we have designed and developed a framework to collect, analyze, calculate sentiment, and identify trending topics of social media posts that are related to public transport use and concern the Greater London Area, one of the most deeply affected areas.

The main objective of this paper is to explore and analyze the textual content of Twitter data in order to obtain transport-related information based on people's sentiment during the pandemic. The following research questions are framed:

- What was the impact of COVID-19 on the sentiment polarity of tweets, what are the main topics discussed, and which are the most important keywords that emerged through tweets with negative polarity?
- Can traditional methods and tools like sentiment analysis techniques and sentiment lexicons operate efficiently during extreme circumstances such as a pandemic, and can they correctly classify tweets based on their sentiment and, in particular, those related to the negative sentiment?

The remainder of this paper is structured as follows. The next section discusses the role of social media and sentiment analysis during the pandemic. In Section 3, the basic steps of the methodology are presented, such as the data collection process and the data analysis methods and tools that were used. The results of the statistical, sentiment, and topic analysis performed are presented in Section 4. Conclusions and recommendations for future research are given at the end of the paper.

## 2. Literature Review

Social media data can help create a dynamic information system that can draw useful conclusions about users' mobility characteristics. Social media enables researchers to assess the sentiment of users towards the provided transport services both by means of transport and for specific routes [9].

One of the social media platforms used largely for sentiment analysis is Twitter because of its nature as short (only 140 characters per tweet) and immediate response to events [11]. Twitter users post transport-related events and exchange information regarding mostly traffic conditions and the provision of transport services [13].

The COVID-19 pandemic has prevailed online social media posts and has been one of the most trending topics on Twitter since January 2020 [14]. Analyzing social media data that were posted about COVID-19 can generate further knowledge about dominant themes, trending topics, sentiments, and changing trends in tweets about the COVID-19

pandemic as well as users' behavior. This type of information can help policy makers and health care organizations assess the needs of people and provide appropriate responses to prevent and control the spread of the pandemic.

There have been several works related to analyzing Twitter data on different topics and themes during the COVID-19 pandemic.

Various studies analyzed users' engagement (posts, likes, comments, etc.) in relation to the evolution of the COVID-19, identified trending topics and created a time series of tweets in an attempt to explore the most important words used in social media posts [15]. Authors applied topic modeling to monitor topics of concern over time and showed how discussions shifted from topics such as the source of the virus, prevention, and number of cases to health and safety protocols as well as vaccination [16]. Others focused on specific topics and keywords such as "mask" and "social distancing" and studied their volume and polarity during the outbreak of the virus [17].

Other works concerned the spatial and temporal distribution of tweets related to the pandemic through the use of interactive visual analytics technologies, such as interactive maps and dashboards [18]. The aim of these visualization tools is to enable policy makers and public authorities to view the increasing social media activity as the contagion spreads [19]. Visual analytics dashboards can also display the most discussed topics, influential users, and alerting incidents alongside a visual overview of confirmed cases and virus spread in examined regions, such as in the USA and UK [20].

A topic of special interest for the research community is the ability to predict the development of a disease outbreak through mining and analysis of social media posts. Authors showed that the maximum number of social media posts were published 10–14 days before the peak of the virus as reported by the official authorities [21]. Other studies collected tweets in real-time in an attempt to monitor the COVID-19 outbreak through social media activity [14]. At the same time, increased attention on the pandemic has led to the viral spread of COVID-19 fake news online [22]. Studies proposed different fake-news detection models and dashboards that can track misinformation on Twitter based on the credibility of news sources shared by users [23].

Social media posts, amongst others, facilitate the measurement of public sentiment. Several studies analyzed topics related to COVID-19 and conducted sentiment analysis on collected Twitter data. It was found that during the first months of the pandemic, the average sentiment of tweets was negative, mainly due to the fact that the majority of the users were misinformed about the new disease. In the next months, users' sentiment became neutral as more news and information were available to a large number of people. Negative sentiment was mainly correlated with the symptoms, the infection rate, and health and safety protocols [24]. Government organizations mostly post tweets with a positive tone, while a lot of mixed sentiments were also recorded [25]. Overall, people's feelings changed over time. Initially, people were in favor of the lockdown and the order to stay home, but their opinions changed later, possibly due to fatigue [26].

A general downward trend was recorded in sentiment in most European countries, with peaks at times when lockdowns were imposed and a slow recovery in the coming weeks. Sentiment was initially very negative and became more positive over time [27]. Another study approach that monitored the COVID-19 discourse found that there were negative sentiments related to the overall outbreak, and positive sentiments related to physical distancing [28].

Social distancing measures had important effects on mobility and activity patterns. A lot of people were temporarily unemployed or worked from home, and most out-of-home activities were canceled. As a result, travel demand decreased, and many countries had witnessed spectacular drops both in car traffic and in public transport ridership [29]. Many people chose not to use public transport as public transport vehicles and facilities can be considered a breeding ground for viruses and places where it might be difficult to avoid contact with other passengers. The constant negative news in the mass media also increased the reluctance to use public transport, as did the disruption of working hours

and places brought about by teleworking. The use of public transport was significantly reduced also due to the limited capacity of buses based on health guidelines and due to the government recommendations [30].

In almost every country, public transport ridership decreased in response to stay-at-home orders and fear of the virus [31,32]. Passenger traffic in some cities' ridership has been reduced by more than 90% [33,34]. This is possibly explained by the limited capacity of buses based on health guidelines, due to the government recommendations for the greatest possible reduction in public transport journeys, and also due to the fear of potential exposure to COVID-19. As a consequence, public transport users switched to more private transport modes and walking [4].

### 3. Methodology

In this section, we describe the basic elements of the methodology that we followed in our analysis. Figure 1 presents the basic steps of the methodology from the data collection process until the extraction of the most predominant topics in our datasets.

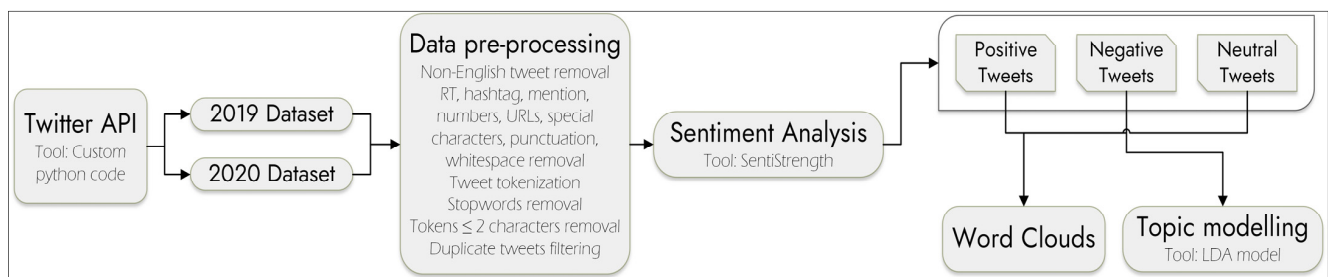


Figure 1. Methodology flowchart.

#### 3.1. Data Collection

The Greater London Area was selected as a case study. The virus reached the UK in late January 2020. The study area was selected based on the specific criteria that are presented below:

- London is the capital and largest city of England and the United Kingdom, as well as one of the world's most important global cities. With a population of nearly 9,000,000 people, London can be considered as an ideal study area for the collection of social media posts from a very large number of people.
- The official language of the country is English, a fact necessary for carrying out specific analytic processes, such as sentiment analysis, as most sentiment lexicon dictionaries are suitable only for the English language.
- London was initially one of the worst affected regions of England. As of 26 June 2021, in the UK, there had been more than 4.7 million confirmed cases and 128,330 deaths among people who had recently tested positive—the world's nineteenth-highest death rate by population and the second-highest death toll in Europe after Russia [35–37].
- Public transport services are overseen by the executive agency for transport in London, Transport for London (TfL) which manages the majority of public transportation in the agglomeration. This is particularly important for the present research as social media data collection for public transport services focuses on a single entity.

The social media platform that we have selected to focus on is Twitter, considering its high popularity (315 million active users by the end of 2020) and the provided free API for collecting real-time and historical tweets [38]. Especially for collecting historical tweets, the present study used the Academic Research product track. This product track enables qualified academic researchers to receive highly elevated access to endpoints that can be used to study conversations on Twitter [39]. Such a method has also been successfully applied for acquiring customer feedback from various transport services and for further understanding travelers' needs and expectations [12].

From 1 January 2019 to 31 December 2020, tweets were collected through Twitter's API. The dates were chosen to include one year before (normal period) and the year during the COVID-19 outbreak (pandemic period). Data collection was performed by searching the Twitter Network for selected keywords. The selection of appropriate keywords was made with the primary aim of collecting data related to the use of public transport in London. Keywords related to public transport terms were used alongside keywords related to the study area, i.e., London. To this end, data collection was performed based on the keyword search: "TfL" OR "Transport London" OR "Public Transport London" OR "Public Transportation London" OR "London Underground".

The collection of data from Twitter was performed by using the Python programming language, version 3.9, and specific code libraries, such as Tweepy and search-tweets-python [40]. Twenty-eight (28) variables were extracted and used during the analysis, such as tweet text, time stamp, author id, user's metrics (followers, following), tweet metrics (likes, retweets, replies), geolocation coordinates (if available), etc.

### 3.2. Data Pre-Processing

The first step towards an accurate sentiment analysis is the pre-processing of the data. During this process, the collected data undergo certain procedures, which include the removal of unnecessary words or characters, the filtering of duplicate tweets, etc. In this paper, all the steps mentioned below for the pre-processing of the tweets were implemented using the Python programming language (version 3.9).

Initially, the collected datasets were filtered for the existence of non-English tweets. The database search returned 17,702 and 16,789 tweets for the 2019 and 2020 datasets, respectively, which were excluded from the datasets. Following this, each tweet in both datasets was further cleaned in order to filter out retweets, mentions, hashtags, numbers, webpage links, whitespaces, punctuation marks, and special characters. This filtering process contributed to the improvement of the relevance of each tweet with the exclusion of characters that cannot be assessed in terms of sentiment. In order to make each tweet more compact, we also filtered out stopwords (e.g., I, no, to, at, etc.) based on the Natural Language Toolkit (NLTK). For the needs of this task, each tweet was previously converted into single tokens using the corresponding 'TweetTokenizer' method of the NLTK suite. Since certain one- or two-word tokens remained in the dataset after the removal of the stopwords, we included one more step in our filtering process, which searched and removed tokens with a length smaller or equal to two (2) characters. The final step in the pre-processing process included the removal of duplicate tweets, which resulted in the removal of 44,283 tweets from the 2019 dataset and 40,973 from the 2020 dataset.

### 3.3. Sentiment Analysis

In order to assess the sentiment expressed by each tweet in our datasets, we employed the SentiStrength analysis tool, which was introduced by Thelwall and Buckley in 2010 [41]. SentiStrength is a lexicon-based sentiment tool, which calculates the sentiment score of each tweet based on a predefined set of manually labeled dictionaries. SentiStrength attributes each tweet with two scores, assessing the positive and negative sentiment based on scale from 1 to 5 and a scale from -1 to -5. The overall sentiment for each tweet derives from the sum of the individual scores associated with each of the tweet's positive and negative tokens. Tokens that are assessed as neutral are attributed with a score of 0. In order to estimate more accurately the correct sentiment from each tweet, SentiStrength exploits a list of booster words that augment the sentiment (either positive or negative) of subsequent words by one (1) or two (2) points depending on the strength of the word or decrease it by one (1) point if the word is more moderate. Additionally, the sentiment expressed by words can be inverted through the exploitation of negating words (e.g., sad, not sad).

The classification of tweets based on their attributed sentiment scores is done through the following rules. If  $p$  refers to the positive sentiment score (1–5 scale) and  $n$  refers to the negative sentiment score (-1 to -5 scale), then:

If  $p > n$ , the tweet is classified as “Positive”.

If  $p < n$ , the tweet is classified as “Negative”.

If  $p = -n$  and  $p < 4$ , the tweet is classified as “Neutral”.

If  $p \geq 4$  and  $n \geq -4$ , the tweet is not classified and should not be deemed fit for further analysis due to the significant deviation in the sentiment score.

SentiStrength was chosen as the preferred sentiment analysis tool for this paper, based on its proven performance according to various benchmarking tests [42–44]. Additionally, the fact that SentiStrength was initially developed based on text from social media (MySpace) adds to its suitability in the present analysis.

### 3.4. Word Clouds

Predominant keywords in a corpus of tweet tokens can indicate the top topics discussed by users. Word clouds are a form of visual representation of the frequency of  $n$  number of keywords. Bolder and bigger words in a word cloud indicate keywords that are used more often by users, while on the other hand, smaller words are associated with a smaller frequency. While word clouds are an easy method of initial assessment of topics in a tweets dataset, its results have to be further analyzed, since single keywords cannot easily provide a clear sense of a topic.

### 3.5. Latent Dirichlet Allocation (LDA)

The Latent Dirichlet allocation (LDA) is a Bayesian model for topic detection, which was proposed by Blei et al. in 2003 [45]. The underlying principle of LDA is that each topic consists of similar words, and as a result, latent topics can be identified by words inside a corpus that frequently appear together in documents or, in our case, tweets. Additionally, LDA is also based on the assumption that topics are formulated on the probability distributions of words, something that differentiates it from other topic modeling methods that rather rely on word frequencies. Furthermore, unlike other clustering models that associate tweets with a single topic, LDA assumes that each document contains different topics.

For the implementation of the LDA model in our paper, we applied the LDavis method through the gensim Python library [46]. LDavis associates each word with a singular id, while the collection of all words forms the corpus. After the user has defined the desired number of topics, which LDavis should identify, the user can visualize the topics as well as the words that are associated with them.

## 4. Results

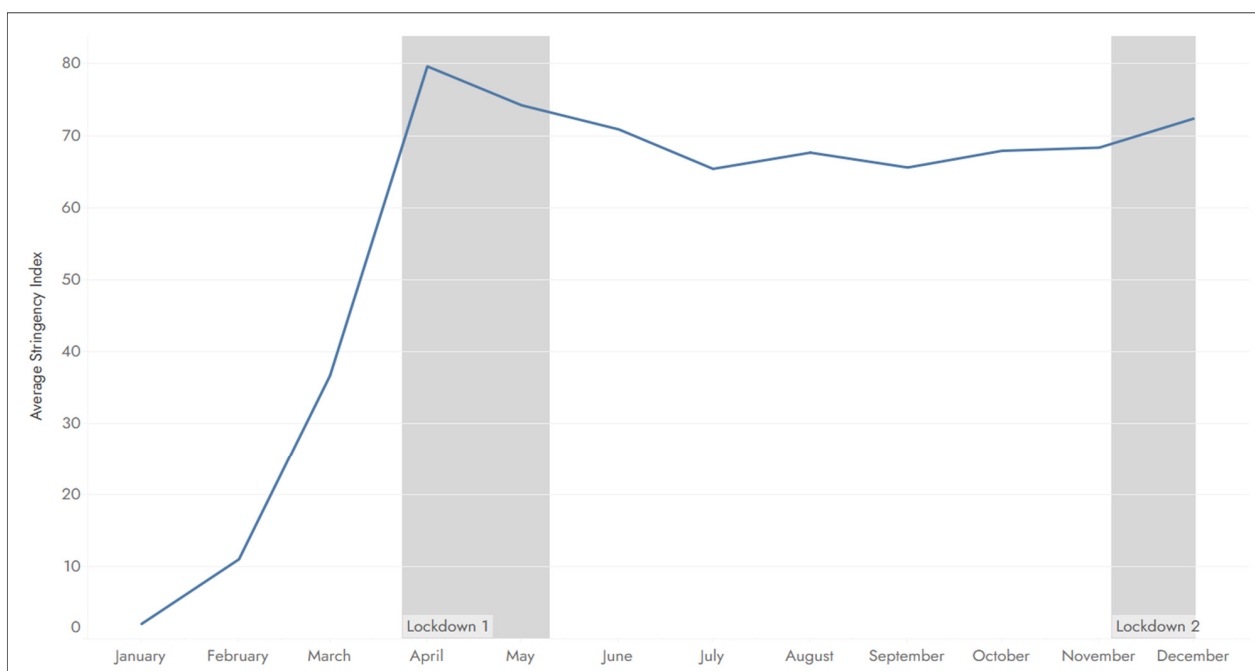
In total, 539,375 posts were collected and processed in the present study for the total data collection period. After removing tweets that did not meet the criteria of the data filtering process, the final sample consisted of 418,624 tweets for the two years. Table 1 shows descriptive statistics measures that were calculated for the final sample per year.

**Table 1.** Descriptive statistics of Twitter data sample per year.

	2019	2020
Initial number of tweets	285,123	254,252
Tweets with English language	267,421	237,463
Final number of tweets	222,136	196,488
Number of unique users	98,886	90,293
Number of tweets per user	2.25	2.18
Tweets with geolocation	494	478
Average number of tweets (per day)	609	538
Average number of tweets (per month)	18,511	16,374

Based on the results presented in Table 1, the social media activity of users regarding public transport matters in London is slightly decreased for the year 2020. More specifically, the total number of tweets published in 2020 is 12% smaller compared to 2019. Respectively, the percentage of the number of tweets per user for 2020 is also smaller in comparison to 2019.

For the year 2020, which corresponds to the pandemic period, the total number of posts was grouped according to the particular conditions prevailing at each pandemic phase. Pandemic phases were based on the stringency index as calculated for the UK. Stringency index is a composite measure based on nine response indicators, including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest) [47]. Figure 2 presents the evolution of the UK's average stringency index per month for 2020.



**Figure 2.** COVID-19 stringency index (UK).

Average number of tweets is slightly decreased during lockdown periods, while during periods when all kinds of restrictions begin to be lifted, there is an increase in users' social media activity (Figure 3). This may be due to the fact that during lockdowns, especially lockdown 1, public transport services were reduced to the minimum, and only core public transport services were provided. In addition, the public was strongly advised to avoid using public transport and use their car instead, if possible. As passenger traffic decreased, social media posts were also decreased.

#### 4.1. Sentiment Analysis

Results from the sentiment analysis in 2019 (Figure 4) show a relatively uniform trend during the months. Neutral tweets are more frequent overall, while positives are less frequent throughout the year.

The sentiment analysis for the 2020 dataset, as illustrated in Figure 5, shows some differences when compared to 2019. Initially, in April, a sudden decrease in the overall number of tweets is observed, which, however, can be explained by the fact that underground ridership reached a minimum this month [48]. Although neutral tweets are again the majority in most months, with positive tweets being less frequent, negative tweets are more in May and in October. The first spike of negative tweets, which falls inside the time period of the first lockdown, could be allocated to the measures taken by the British

government on 11 May, which focused on a gradual return to normality but with an urge to avoid the public transport system [49]. This is reflected by tweets collected in May 2020 (e.g., “boris johnson announces lockdown restrictions eased england sadiq khan pleaded londoners use tfl unless last resort bid maintain social distancing public transport”, “must avoid traveling public transport underground platforms packed london morning”). The spike in sentiment is also reflected by the data of the stringency index (Figure 2).

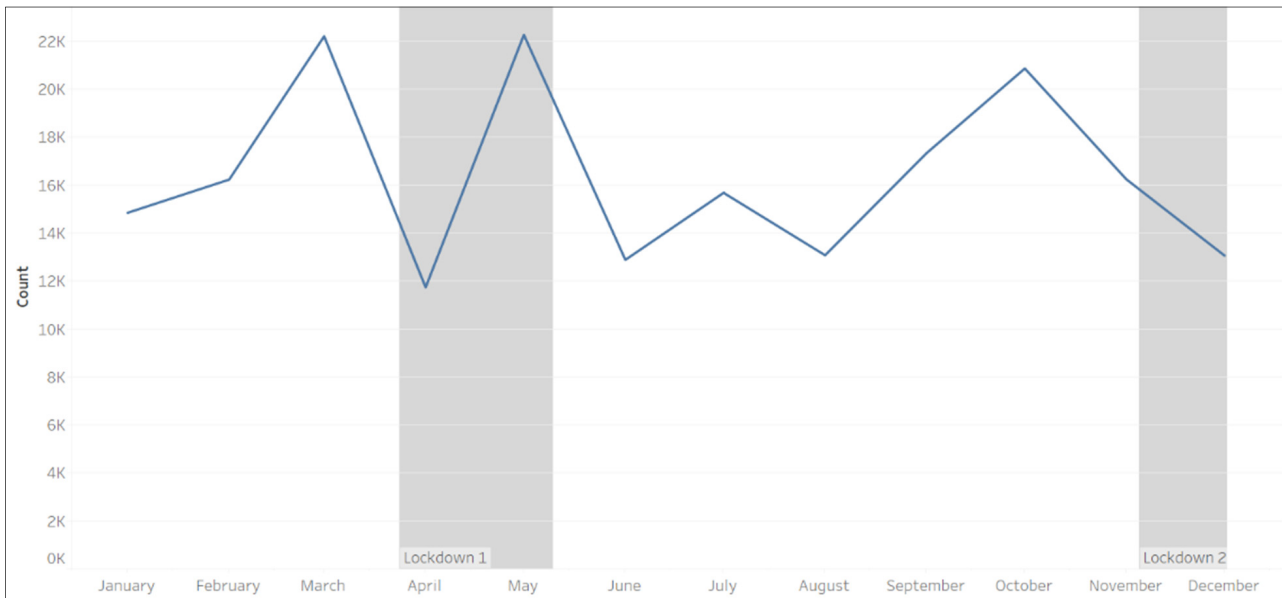


Figure 3. Number of tweets per pandemic phase.

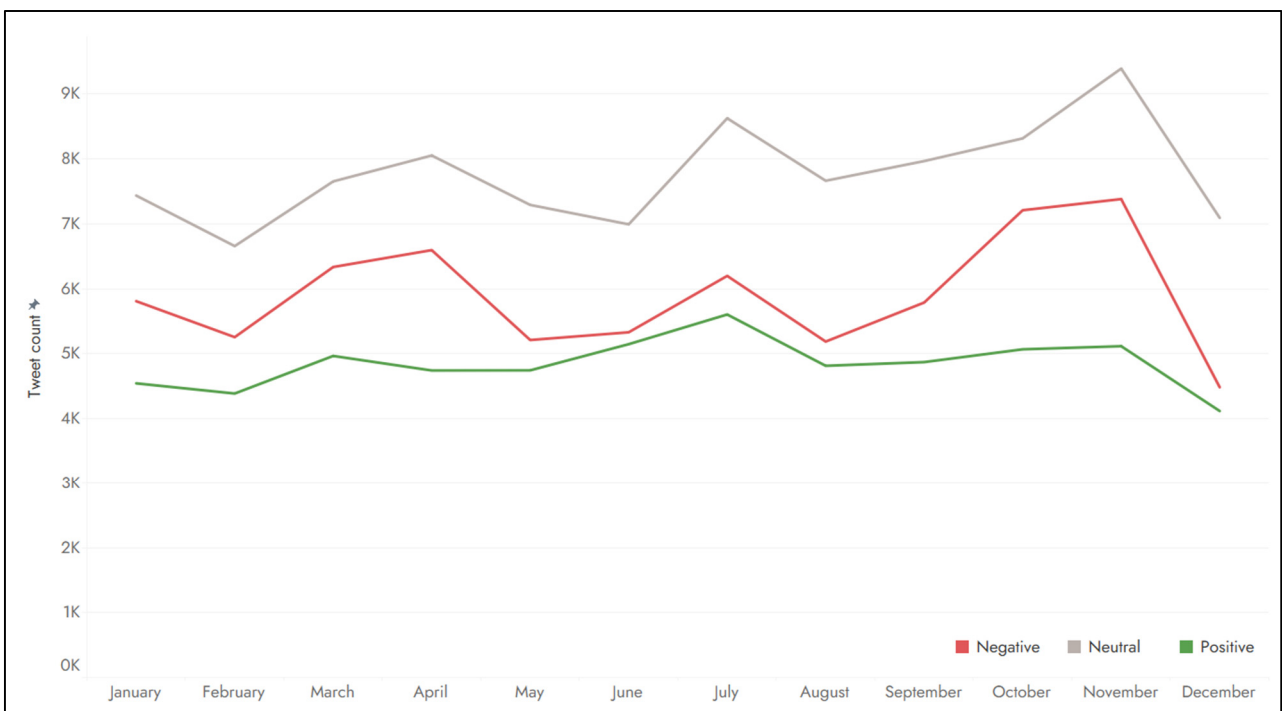


Figure 4. Evolution of sentiment classes over 2019.



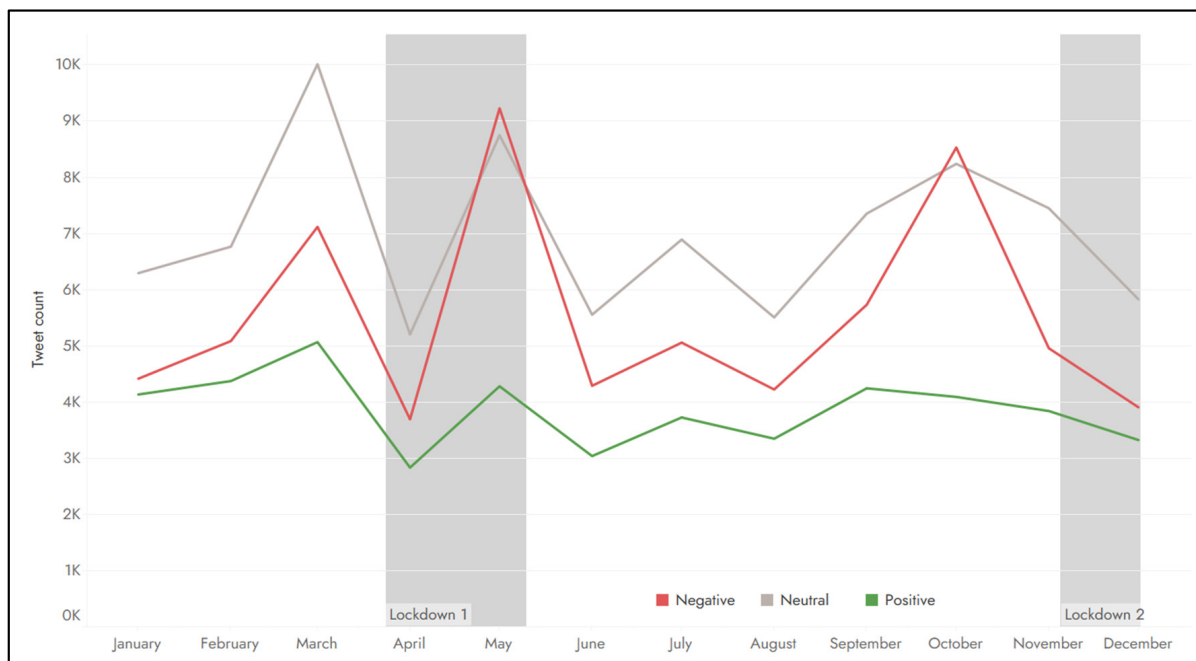
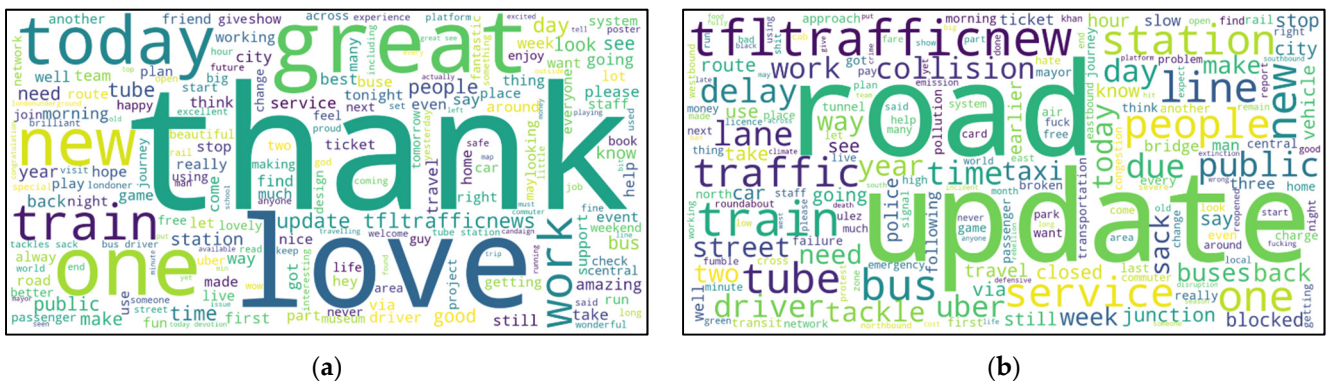


Figure 5. Evolution of sentiment classes over 2020.

The second spike of negative tweets appears in October 2020 and coincides with the stricter measures imposed by the British government in response to the pandemic from 17 October onwards. More in detail, among the measures taken was the encouragement towards citizens to reduce the number of journeys to the absolutely necessary and avoid using the public transport network [50]. Furthermore, during this time period, there were active negotiations regarding the bailout of the TfL underground from the British government in order to keep the network active during the second lockdown of the city [51]. This is validated by the content of tweets from this period (e.g., “walking miles day avoid public transport london getting quad bike”, “transport london secures bn bailout”, “happening tfl bailout”). As with the spike in May 2020, the spike in October is reflected by the data of the stringency index (Figure 2).

#### 4.2. Word Clouds

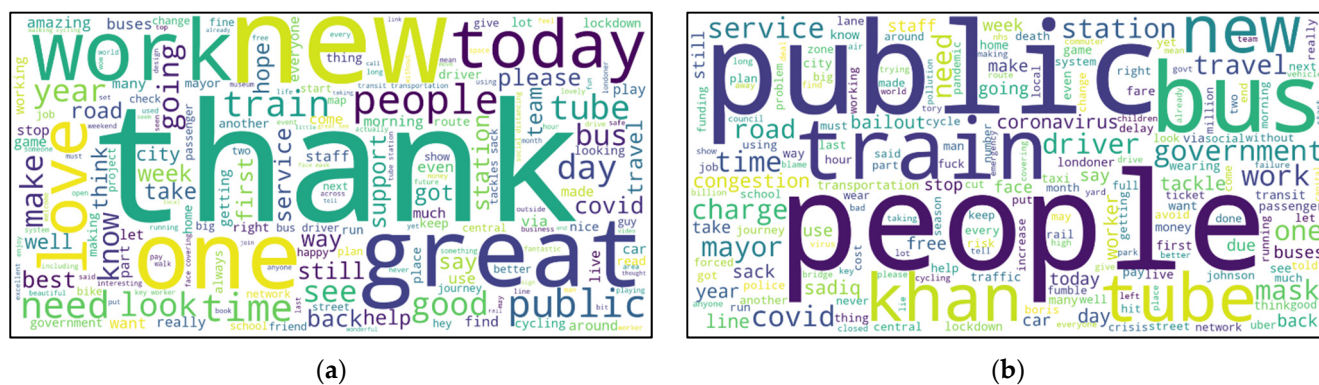
Figures 6 and 7 present the word clouds of the 50 most frequent keywords mentioned in positive and negative tweets in the 2019 and 2020 datasets, respectively.



(a)

(b)

Figure 6. (a) Word Clouds for positive sentiment in the 2019 dataset; (b) Word Clouds for negative sentiment in the 2019 dataset.



**Figure 7.** (a) Word clouds for positive sentiment in the 2020 dataset; (b) word clouds for negative sentiment in the 2020 dataset.

By observing the keywords emerging in the positive sentiment word cloud for 2019 (Figure 6), we can see keywords such as “thank”, “great”, “love”, “tube”, “station”, “train”, and “public” highlighted with bolder font, indicating the increased frequency associated with these terms. From this, it can be assumed that positive tweets focus on the positive aspects of the provided public transport services.

On the other hand, among the negative keywords in the corresponding word cloud, the most frequent terms are “update”, “road”, “train”, “station”, “tube”, “service”, and “tfltrafficnew”, which are accompanied by terms such as “delay”, “traffic”, “time”, and “collision”. From these observations, we can conclude that users focus on tweeting about incidents that happened on the public transport network (e.g., train delays or line closures, etc.) or the road network (e.g., congestion, road accidents, etc.).

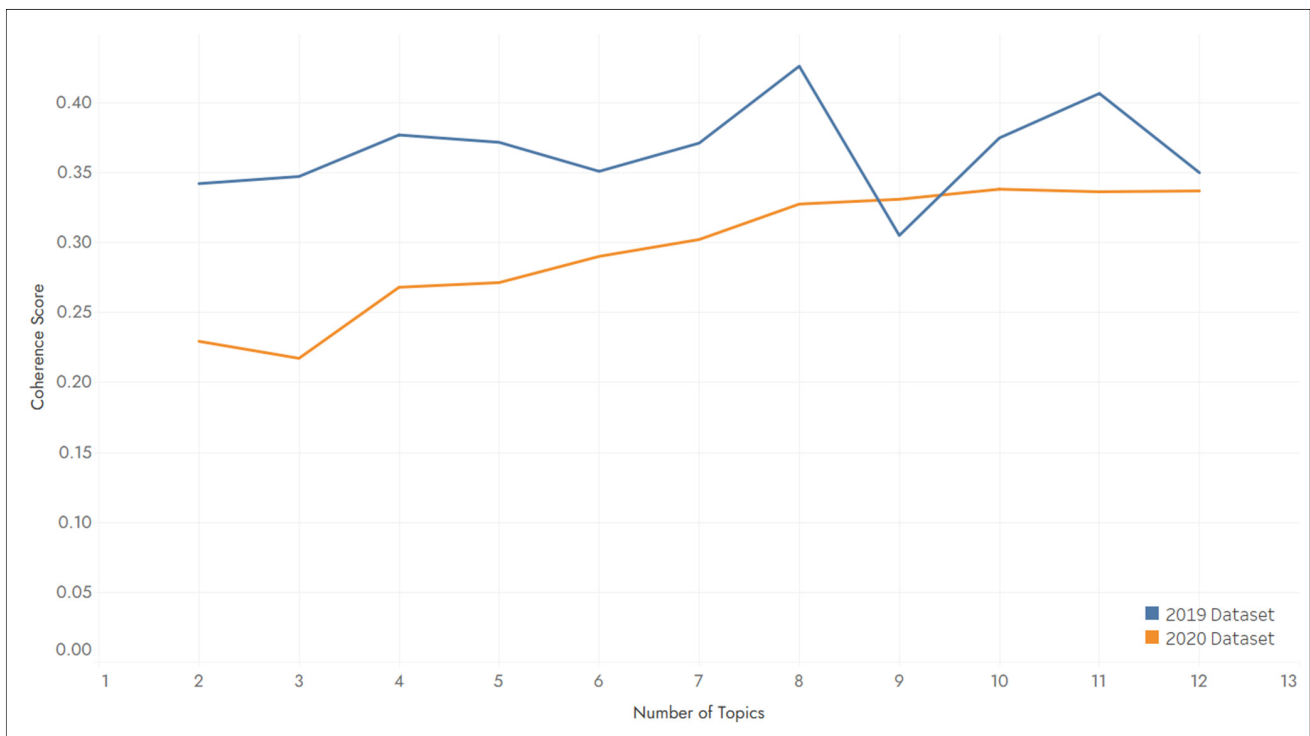
According to the terms appearing on the word cloud associated with positive sentiment in the 2020 dataset (Figure 7), we can observe that the higher frequency is linked with keywords such as “thank”, “new”, “great”, “today”, “love”, “tube”, and “train”, which appear on the dataset of the previous year as well. However, the difference lies in the appearance of less frequent terms such as “COVID” and “lockdown”, which indicate a potential slight shift regarding the topics tweeted about.

Although differences between the positive word clouds of 2019 and 2020 are not so apparent, differences between the corresponding negative word clouds are more straightforward due to the negative context in which the keywords “COVID”, “coronavirus”, “lockdown”, and “mask” are used. As a result, terms such as “tube”, “people”, “bus”, “public”, and “train” are followed by pandemic related keywords like “COVID”, “coronavirus”, and “mask”, indicating a potential reference to the measures taken by public transport authorities as a response to the pandemic. Furthermore, the terms “khan”, “mayor”, “bailout”, and “government” highlight a critical stance by users in terms of the policies adopted during the pandemic period.

#### 4.3. Topic Modeling

In order to investigate the differences regarding the topics that are associated with negative sentiment before and during the pandemic, we analyzed the results derived from the application of the LDA method for only the negative tweets from each of the two datasets.

The first step is to define the optimal number of topics through a clustering process. A small number of topics could result in the merge of two different topics in terms of content, while on the other hand, a large number could lead to fragmented and unclear topics. The optimal number of topics is determined by the coherence score, which highlights the overall coherence achieved in each topic. In order to determine the number of topics that achieve the best coherence score, we tested different values and assessed the results (Figure 8). For the 2019 dataset, the optimal number of topics selected was 8, while for the 2020 dataset, the corresponding value was 10.



**Figure 8.** Coherence score per number of topics.

In order to determine the content of each of the clustered topics, we examined the plots from the LDAvis method used in the analysis (Figure 9). The distance map on the left side of the plot illustrates the clustered topics. The size of the circle reflects the frequency of the terms associated with this topic. As a result, topics that appear larger are linked with terms (called tokens) that appear more frequently on the corpus and thus are more popular among Twitter users. Overlapping circles indicate terms that are similar between topics. The range of each topic is also reflected by the percentage of the tokens included, as appears on the right side of the plot. Topics with a higher percentage of tokens are more popular since they include more terms from the corpus.

The right side of the plot presents the top-30 most relevant terms associated with each selected topic (in Figure 9, topic 2). The red bar indicates the frequency of each term in this particular topic, while the grey bar indicates the overall frequency of the term in the corpus.

The most predominant topics from the clustering process were selected based on the percentage of the tokens that they included, the total number of topics, and the coherence of each topic. Table 2 presents the most prevalent topics from each of the two datasets that were analyzed. For the 2019 dataset we selected four of the eight topics as they include 61.9% of the total tokens. Regarding the 2020 dataset, the four most prevalent topics include 50.3% of the total tokens. Due to the low token percentages of each of the remaining six topics, they were not deemed as representative for the analysis since they refer to topics that are not so popular among users. For each of the selected topics, we include the 30 most salient terms associated with the topic as they derived from the LDA model. This number was chosen for a clearer conception of a topic's content, despite the commonly chosen value of 15 according to the literature [52]. In order to emphasize emotional terms that are related to each topic, we highlight them in bold font. Furthermore, each topic is labeled with a short description that encompasses the semantics of the topic.

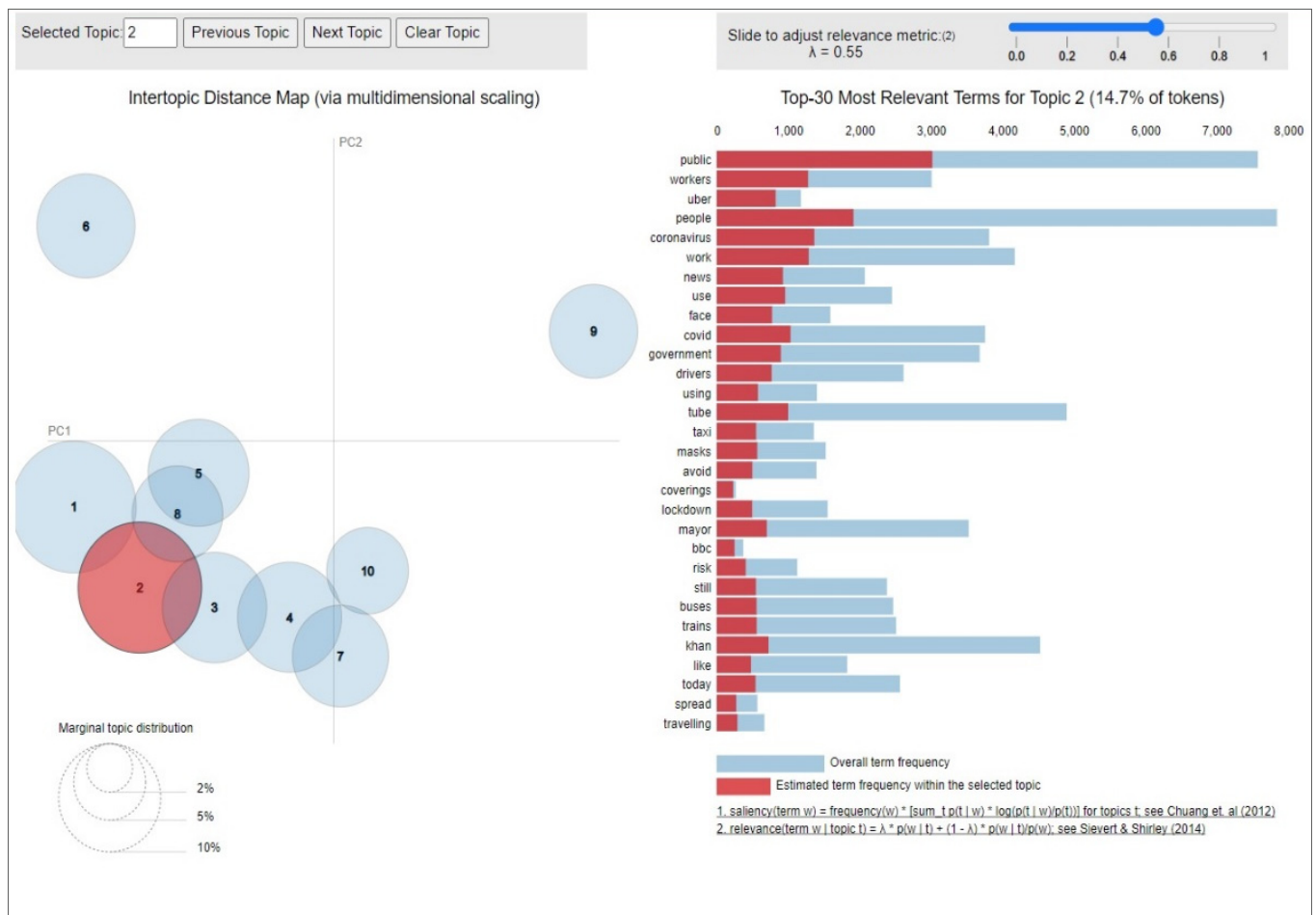


Figure 9. Illustration of LDA’s results for the most salient terms associated with topic 2 of the 2020 dataset.

Table 2. Most prevalent topics in the 2019 and 2020 negative sentiment tweet datasets.

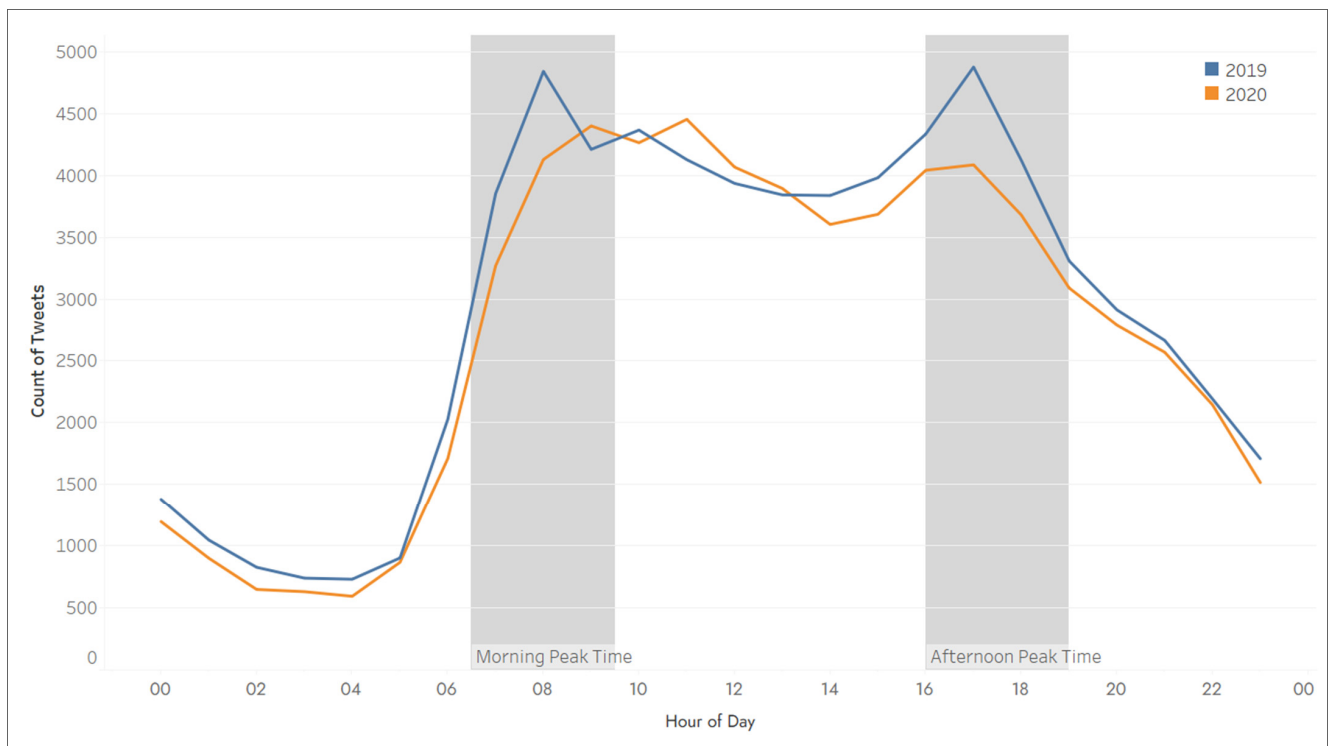
Topics	Keywords	Description	Percentage of Tokens
<b>2019 Dataset</b>			
Topic 1	bus, get, train, time, service, like, people, tube, minutes, hate, journey, driver, stop, money, home, every, hour, could, day, line, pay, know, worst, need, morning, use, card, shit, back, even	Public transport quality issues	18.4
Topic 2	people, public, one, day, another, man, need, tube, think, packages, demonstration, mayor, black, car, woman, get, want, cars, ulez, cab, ultra, children, khan, uber, use, low, around, anti, tax, health	TfL general announcements (news, advertisements, etc.)	15.5
Topic 3	update, tfltrafficnews, road, collision, traffic, lane, due, junction, blocked, closed, slow, earlier, approach, following, street, reopened, delays, westbound, eastbound, emergency, southbound, northbound, broken, lanes, roundabout, one, expect, circular, fully, north	Traffic and public transport incident reports	14.7
Topic 4	pollution, air, caution, ulez, new, today, charge, blackwall, congestion, roads, taxis, use, drivers, devices, alistair, beg, hubs, vehicle, vehicles, liste, uber, god, zone, times, quality, emission, cars, sent, poor, public	Traffic incidents and environmental impact	13.3

Table 2. Cont.

Topics	Keywords	Description	Percentage of Tokens
<b>2020 Dataset</b>			
Topic 1	people, bus, public, buses, work, social, get, transportation, transit, <b>stop</b> , distancing, <b>packed</b> , many, home, take, going, travel, drivers, think, road, mylondon, staff, <b>die</b> , tube, like, need, driver, service, new, cars	Pandemic measures in public transport	14.9
Topic 2	public, workers, uber, people, coronavirus, work, news, use, face, COVID, government, drivers, using, tube, taxi, masks, <b>avoid</b> , coverings, lockdown, mayor, bbc, <b>risk</b> , still, buses, trains, khan, like, today, spread, travelling	Health and safety protocols	14.7
Topic 3	free, charge, public, congestion, new, get, travel, use, standard, tube, bailout, evening, one, <b>avoid</b> , bus, transit, increase, pay, transportation, COVID, work, car, cycling, peak, monday, <b>mad</b> , day, possible, government, news	Traffic and public transport operational conditions	10.4
Topic 4	khan, sadiq, banksy, mayor, bailout, outbreak, boris, belly, money, jonhson, coronavirus, mujinga, cycle, wildlife, system, government, ban, city, years, blame, says, govt, <b>waste</b> , emergency, one, debt, lanes, <b>poor</b> , artist, award	Government and social consequences of the pandemic	10.3

As we can observe from the results of Table 2, all of the four topics in the 2019 dataset are associated with the operation of London's transport system. Topic 1 focuses on quality issues linked with the provided public transport services. Topic 2 focuses on events that are related to the TfL in general, such as a demonstration that was held in 2019 that disrupted transport services [53]. Topic 3 focuses on reporting on incidents that occurred in the transport network and on upcoming delays. Topic 4 includes tweets that refer to the measures that are taken towards mitigating the environmental impact of transport in London, such as the Urban Low Emission Zones (ULEZ) and the congestion charge. The focus of the topics on the operational conditions of the transport system is portrayed by the tweeting activity of the users who are more active during the two peak time periods of the day when the public transport system is at its full capacity and delays are more usual (Figure 10).

Regarding the results for the most predominant topics in the 2020 dataset, we can observe that although tweets referring to traffic conditions remain relatively high in the ranking (topic 3), topics that refer to the effect of the coronavirus on transport are emerging and are in the two first places in terms of token percentage. More in detail, topic 1 focuses on the social distancing measures taken in the public transport network and on crowding cases that are observed in buses or underground trains. The concern expressed by users is supported by literature indicating that public transport vehicles and stations are potential hotbeds for the spread of the virus, and thus, users are discouraged from choosing them [30,48]. Topic 2 is related to the health and safety protocols that are issued in response to the coronavirus and on the compliance with the measures by public transport personnel and users. Finally, topic 4 is associated with the response of government bodies against the pandemic. The bailout agreed between the British government and the city of London to ensure the financial sustainability of London's underground system was a popular theme among the tweets, while the case of Belly Mujinga was also a popular theme [54].



**Figure 10.** Number of negative sentiment tweets per time of day in 2019 and 2020.

By comparing the list of the most predominant topics from the two datasets, it can be clearly observed that the negativity of Twitter users has shifted from topics related strictly to the operation of the transport network to topics related to the effect of the pandemic on transport. The difference in the contents of the topics can also be supported by the difference in tweeting activity during peak time periods between 2019 and 2020 when the public transport system is at its full capacity and delays are more usual (Figure 10). The fact that the topics analyzed in Table 2 are mentioned in tweets expressing a negative sentiment, led us to investigate the hypothesis that perhaps the pandemic has prevented users from using public transport.

In detail, we examined the frequency of the appearance of three terms related to the pandemic in the negative tweets of 2020. The terms chosen for this analysis, as illustrated in Figure 11, were “COVID”, “lockdown”, and “distancing”. In addition, in order to assess the potential effect of the pandemic in public transport use, we correlated the derived frequencies with the reduction in ridership of the London underground in relation to the equivalent ridership of 2019, according to Department for Transport data [55]. Pearson r test indicated statistical importance correlation between all three terms. The term “distancing” achieved a correlation of  $r = -0.297$ ,  $n = 304$ ,  $p = 0.000$ , while the terms “lockdown” and “COVID” achieved correlations of  $r = -0.260$ ,  $n = 304$ , and  $p = 0.000$  and  $r = -0.162$ ,  $n = 304$ , and  $p = 0.005$ . respectively. The negative correlation for all three terms indicates that ridership decreases as the frequency of these keywords increases. These results support our initial hypothesis that the pandemic and the warnings of officials have potentially deterred transport users from choosing public transport, while at the same time, the number of tweets with negative sentiment referring to public transport was being increased.

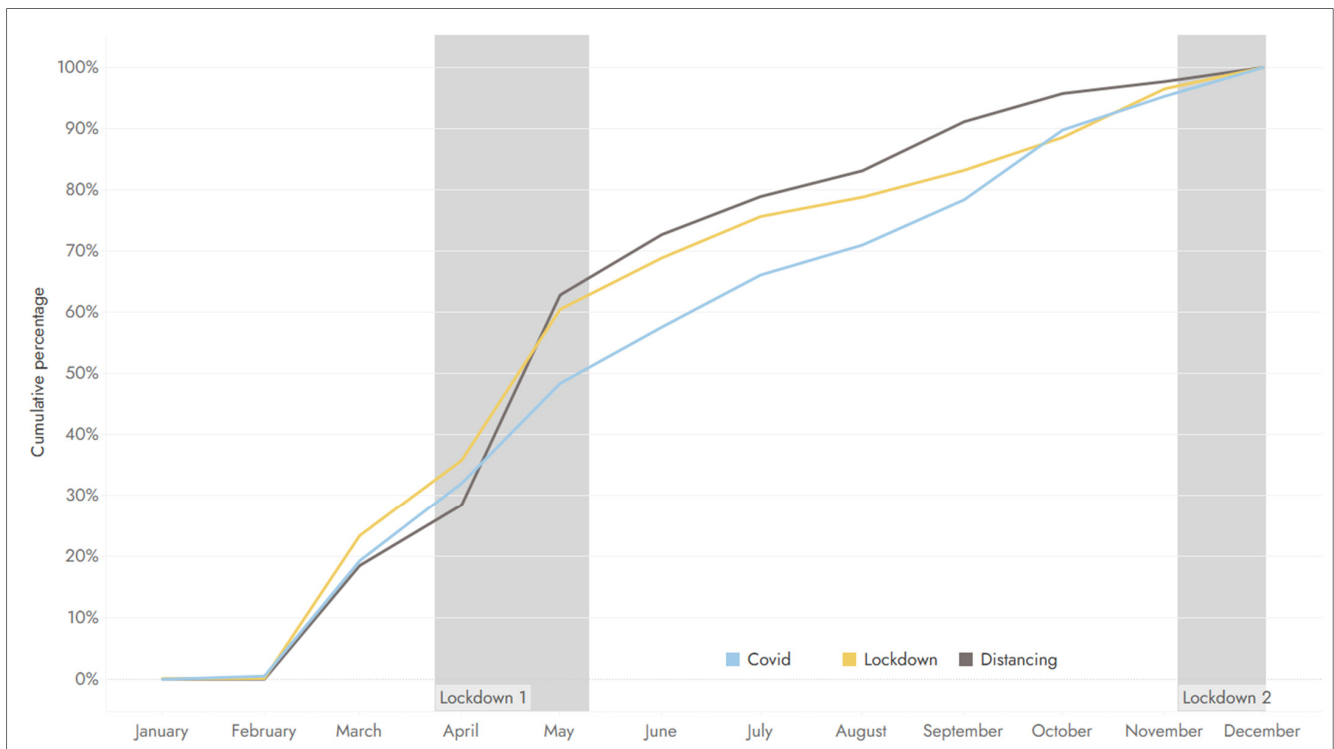


Figure 11. Evolution of keywords in 2020.

## 5. Conclusions

This research attempts to identify the effect of the COVID-19 pandemic on public transport use in the city of London through the analysis of twitter data for the years 2019 and 2020. For the needs of the analysis, we examined the sentiment of the collected tweets through the application of the SentiStrength tool and the most frequently discussed topics by applying the LDA model.

Sentiment analysis results indicate a uniform pattern in terms of the sentiment of tweets in the baseline year of 2019. However, in 2020, an increase of negative tweets is observed mainly during periods of measure imposition in response to the spread of the pandemic. In response to our first research question, results indicate that the pandemic has affected sentiment in certain time periods, but overall sentiment trends are similar to 2019.

Although results from the assessment of the word clouds of positive tweets show no clear distinction between the 2019 and 2020, the corresponding analysis for negative tweets shows the emergence of terms related to the pandemic, such as “COVID”, “coronavirus”, and “mask”. The effect on users’ negativity is more apparent from the results of the LDA model. More in detail, although topics from the analysis of the negative tweets of 2019 show a focus on the operational conditions and the provided quality of service from the transport system, negative tweets on 2020 focus more on the pandemic and its impact on the public transport system. Furthermore, correlations between pandemic-related terms and ridership data indicate that the coronavirus crisis had a negative effect on public transport demand. Additionally, the comparison of the number of terms expressing emotion among the most relevant terms in the detected topics shows an increase in 2020, although a more detailed methodology should be employed in order to better assess the expression of emotion.

The tremendous growth of social media platforms such as Twitter provides transport operators an opportunity to continuously gather and analyze customer feedback. Customers’ viewpoints and their comments on the quality of the products and services can help operators make successful business decisions.

Harvesting transport information from social media is a relatively new field, with major potential in improving understanding of users’ needs and perspectives, and it can

be used during transport planning, management, and control, as well as for supporting the achievement of transport policy goals. More specifically, it can be used for improving the quality of a public transport service based on users' opinions, taking actions for solving complex problems based on real-time reports for incidents, and dissemination of traffic and other types of information in a more direct way. Social media monitoring allows handling customer complaints and monitoring customers' responses so that those comments and feedback can be evaluated to improve the quality of transport services.

Furthermore, social media have particular advantages compared to traditional information systems in communicating during emergency situations, such as the COVID-19 pandemic. Social media platforms like Twitter can enable real-time, two-way communication between large groups of people and a transport agency. During extreme and emergency circumstances, there is a need for accurate, timely knowledge provision. Social media can allow public transport agencies to provide information to affected commuters quickly, which in turn allows them to change plans and/or avoid travel and, most importantly, reduces the anxiety, anger, or even potential danger associated with not complying with health protocols and social distancing measures.

Despite their numerous advantages, social media are also characterized by disadvantages, which apparently affect our analysis, such as selective bias, underrepresentation or overrepresentation of certain population groups, fake information, and others. Another issue that should be further addressed is the use of appropriate text-mining techniques that are required for the extraction of valuable transport-related information from social media. A transport-oriented lexicon should be constructed that would include a set of words, terms, and phrases commonly used in the transport domain. This set should be able to tag words according to their sentiment for more accurate sentiment analysis. Furthermore, the validity of sentiment analysis is influenced by the presence of irony and sarcasm in users' posts, feelings that tend to occur strongly when users report something negative, especially in exceptional circumstances such as the COVID-19 pandemic.

The present paper paves the way for future research on the determination of the effect of COVID-19 on public transport use. The next steps could focus on the exploitation of more sophisticated data mining tools, such as artificial intelligence, for the detection of emotion. Additionally, a similar methodology can be applied to other geographical areas (such as the United States of America), which followed different policies to counter the dissemination of the pandemic and where users might have different levels of conformity against government directives. New findings could provide valuable insight in terms of the suitability of the proposed methodological framework for the determination of public transport users' response to the pandemic.

**Author Contributions:** Conceptualization, I.P., G.G., A.K., A.N. and P.P.; methodology, A.N., A.K. and I.P.; software, A.K. and A.N.; validation, A.N., A.K. and I.P.; formal analysis, A.K. and A.N.; investigation, A.K., A.N. and I.P.; data curation, A.N. and A.K.; writing—original draft preparation, A.N., A.K. and I.P.; visualization, A.K.; supervision, I.P.; project administration, A.N., A.K. and I.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** Part of this study was funded within the framework of the C.H.A.N.G.E. project: "Enhancing the bicycle-sharing Economy through Innovative Services & Applications". This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the operational program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE (project code: T1EDK-04582).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.



## References

1. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard with Vaccination Data. Available online: <https://covid19.who.int/> (accessed on 29 July 2021).
2. Abdullah, M.; Dias, C.; Muley, D.; Shahin, M. Exploring the Impacts of COVID-19 on Travel Behavior and Mode Preferences. *Transp. Res. Interdiscip. Perspect.* **2020**, *8*, 100255. [[CrossRef](#)]
3. Campisi, T.; Basbas, S.; Skoufas, A.; Akgün, N.; Ticali, D.; Tesoriere, G. The Impact of COVID-19 Pandemic on the Resilience of Sustainable Mobility in Sicily. *Sustainability* **2020**, *12*, 8829. [[CrossRef](#)]
4. Politis, I.; Georgiadis, G.; Nikolaidou, A.; Kopsacheilis, A.; Fyrogenis, I.; Sdoukopoulos, A.; Verani, E.; Papadopoulos, E. Mapping Travel Behavior Changes during the COVID-19 Lock-down: A Socioeconomic Analysis in Greece. *Eur. Transp. Res. Rev.* **2021**, *13*, 21. [[CrossRef](#)]
5. Budd, L.; Ison, S. Responsible Transport: A Post-COVID Agenda for Transport Policy and Practice. *Transp. Res. Interdiscip. Perspect.* **2020**, *6*, 100151. [[CrossRef](#)] [[PubMed](#)]
6. Nikitas, A.; Tsigdinos, S.; Karolemeas, C.; Kourmpa, E.; Bakogiannis, E. Cycling in the Era of COVID-19: Lessons Learnt and Best Practice Policy Recommendations for a More Bike-Centric Future. *Sustainability* **2021**, *13*, 4620. [[CrossRef](#)]
7. Politis, I.; Georgiadis, G.; Papadopoulos, E.; Fyrogenis, I.; Nikolaidou, A.; Kopsacheilis, A.; Sdoukopoulos, A.; Verani, E. COVID-19 Lockdown Measures and Travel Behavior: The Case of Thessaloniki, Greece. *Transp. Res. Interdiscip. Perspect.* **2021**, *10*, 100345. [[CrossRef](#)]
8. Social Media & COVID-19: A Global Study of Digital Crisis Interaction among Gen Z and Millennials. Available online: <https://www.who.int/news-room/feature-stories/detail/social-media-COVID-19-a-global-study-of-digital-crisis-interaction-among-gen-z-and-millennials> (accessed on 29 July 2021).
9. Nikolaidou, A.; Papaioannou, P. Utilizing Social Media in Transport Planning and Public Transit Quality: Survey of Literature. *J. Transp. Eng. Part A Syst.* **2018**, *144*, 128. [[CrossRef](#)]
10. Nikitas, A.; Michalakopoulou, K.; Njoya, E.T.; Karampatzakis, D. Artificial Intelligence, Transport and the Smart City: Definitions and Dimensions of a New Mobility Era. *Sustainability* **2020**, *12*, 2789. [[CrossRef](#)]
11. Rathore, A.K.; Kar, A.K.; Ilavarasan, P.V. Social Media Analytics: Literature Review and Directions for Future Research. *Decis. Anal.* **2017**, *14*, 229–249. [[CrossRef](#)]
12. Apostolidis, L.; Papadopoulos, S.; Liatsikou, M.; Fyrogenis, I.; Papadopoulos, E.; Keikoglou, G.; Alexiou, K.; Chondros, N.; Kompatsiaris, I.; Politis, I. I-CHANGE: A Platform for Managing Dockless Bike Sharing Systems. In Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Cagliari, Italy, 1–4 July 2020; Volume 12250.
13. Gal-Tzur, A.; Grant-Muller, S.M.; Kuflik, T.; Minkov, E.; Nocera, S.; Shoor, I. The Potential of Social Media in Delivering Transport Policy Goals. *Transp. Policy* **2014**, *32*, 115–123. [[CrossRef](#)]
14. Jahanbin, K.; Rahmanian, V. Using Twitter and Web News Mining to Predict COVID-19 Outbreak. *Asian Pac. J. Trop. Med.* **2020**, *13*, 378.
15. Cinelli, M.; Quattrocchi, W.; Galeazzi, A.; Valensise, C.M.; Brugnoli, E.; Schmidt, A.L.; Zola, P.; Zollo, F.; Scala, A. The COVID-19 Social Media Infodemic. *Sci. Rep.* **2020**, *10*, 16598. [[CrossRef](#)]
16. Zhao, Y.; Cheng, S.; Yu, X.; Xu, H. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. *J. Med. Internet Res.* **2020**, *22*, 18825. [[CrossRef](#)] [[PubMed](#)]
17. Sanders, A.C.; White, R.C.; Severson, L.S.; Ma, R.; McQueen, R.; Paulo, H.C.A.; Zhang, Y.; Erickson, J.S.; Bennett, K.P. Unmasking the Conversation on Masks: Natural Language Processing for Topical Sentiment Analysis of COVID-19 Twitter Discourse. *medRxiv* **2020**. [[CrossRef](#)]
18. Andreadis, S.; Antzoulatos, G.; Mavropoulos, T.; Giannakeris, P.; Tzionis, G.; Pantelidis, N.; Ioannidis, K.; Karakostas, A.; Gialampoukidis, I.; Vrochidis, S.; et al. A Social Media Analytics Platform Visualising the Spread of COVID-19 in Italy via Exploitation of Automatically Geotagged Tweets. *Online Soc. Netw. Media* **2021**, *23*, 134. [[CrossRef](#)]
19. Yang, T.; Shen, K.; He, S.; Li, E.; Sun, P.; Chen, P.; Zuo, L.; Hu, J.; Mo, Y.; Zhang, W.; et al. CovidNet: To Bring Data Transparency in the Era of COVID-19. 2020. Available online: <https://arxiv.org/pdf/2005.10948.pdf> (accessed on 29 November 2021).
20. Technical Analytics White Paper | OmniSci. Available online: [https://www2.omnisci.com/resources/technical-whitepaper/lp?\\_ga=2.192127720.316702718.1564495503-925270820.1564495503](https://www2.omnisci.com/resources/technical-whitepaper/lp?_ga=2.192127720.316702718.1564495503-925270820.1564495503) (accessed on 29 July 2021).
21. Li, C.; Chen, L.J.; Chen, X.; Zhang, M.; Pang, C.P.; Chen, H. Retrospective Analysis of the Possibility of Predicting the COVID-19 Outbreak from Internet Searches and Social Media Data, China, 2020. *Eurosurveillance* **2020**, *25*, 2000199. [[CrossRef](#)]
22. COVID-19: The First Study to Look at Whether Fake News Actually Changes People's Behaviour. Available online: <https://theconversation.com/COVID-19-the-first-study-to-look-at-whether-fake-news-actually-changes-peoples-behaviour-144819> (accessed on 29 July 2021).
23. Sharma, K.; Seo, S.; Meng, C.; Rambhatla, S.; Liu, Y. COVID-19 on Social Media: Analyzing Misinformation in Twitter Conversations. 2020. Available online: <https://arxiv.org/pdf/2003.12309v4.pdf> (accessed on 29 November 2021).
24. Wang, T.; Lu, K.; Chow, K.P.; Zhu, Q. COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model. *IEEE Access* **2020**, *8*, 138162–138169. [[CrossRef](#)]
25. De las Heras-Pedrosa, C.; Sánchez-Núñez, P.; Peláez, J.I. Sentiment Analysis and Emotion Understanding during the COVID-19 Pandemic in Spain and Its Impact on Digital Ecosystems. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5542. [[CrossRef](#)]

26. Naseem, U.; Razzak, I.; Khushi, M.; Eklund, P.W.; Kim, J. COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 1003–1015. [CrossRef]
27. Kruspe, A.; Häberle, M.; Kuhn, I.; Zhu, X.X. Cross-Language Sentiment Analysis of European Twitter Messages during the COVID-19 Pandemic. 2020. Available online: <https://arxiv.org/pdf/2008.12172.pdf> (accessed on 29 November 2021).
28. Jang, H.; Rempel, E.; Roth, D.; Carenini, G.; Janjua, N.Z. Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis. *J. Med. Internet Res.* **2021**, *23*, e25431. [CrossRef] [PubMed]
29. de Vos, J. The Effect of COVID-19 and Subsequent Social Distancing on Travel Behavior. *Transp. Res. Interdiscip. Perspect.* **2020**, *5*, 100121. [CrossRef]
30. Basbas, S. COVID-19 and Public Transport Demand Trends in Sicily: Analyzing External Factors and Governmental Recommendations. *Eur. Transp./Trasp. Eur.* **2021**, *83*, 9. [CrossRef]
31. Przybylowski, A.; Stelmak, S.; Suchanek, M. Mobility Behaviour in View of the Impact of the COVID-19 Pandemic—Public Transport Users in Gdansk Case Study. *Sustainability* **2021**, *13*, 364. [CrossRef]
32. Aparicio, J.T.; Arsenio, E.; Henriques, R. Understanding the Impacts of the COVID-19 Pandemic on Public Transportation Travel Patterns in the City of Lisbon. *Sustainability* **2021**, *13*, 8342. [CrossRef]
33. Troko, J.; Myles, P.; Gibson, J.; Hashim, A.; Enstone, J.; Kingdon, S.; Packham, C.; Amin, S.; Hayward, A.; Van-Tam, J.N. Is Public Transport a Risk Factor for Acute Respiratory Infection? *BMC Infect. Dis.* **2011**, *11*, 16. [CrossRef] [PubMed]
34. Carteni, A.; di Francesco, L.; Henke, I.; Marino, T.V.; Falanga, A. The Role of Public Transport during the Second COVID-19 Wave in Italy. *Sustainability* **2021**, *13*, 11905. [CrossRef]
35. COVID-19 Data Visualisation. Available online: <https://pandemic.internationalosos.com/2019-ncov/COVID-19-data-visualisation> (accessed on 29 July 2021).
36. Mortality Analyses—Johns Hopkins Coronavirus Resource Center. Available online: <https://coronavirus.jhu.edu/data/mortality> (accessed on 29 July 2021).
37. Pak, A.; Paroubek, P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta, 17–23 May 2010.
38. Twitter: Number of Users Worldwide 2019–2020 | Statista. Available online: <https://www.statista.com/statistics/303681/twitter-users-worldwide/> (accessed on 29 July 2021).
39. Twitter API Academic Research. Available online: <https://developer.twitter.com/en/products/twitter-api/academic-research> (accessed on 29 July 2021).
40. Twitter API Academic Research Resources. Available online: <https://developer.twitter.com/en/use-cases/do-research/academic-research/resources> (accessed on 29 July 2021).
41. Thelwall, M.; Buckley, K.; Paltoglou, G.; Cai, D.; Kappas, A. Sentiment in Short Strength Detection Informal Text. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 2544–2558. [CrossRef]
42. Abbasi, A.; Hassan, A.; Dhar, M. Benchmarking Twitter Sentiment Analysis Tools. In Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, 26–31 May 2014.
43. Jongeling, R.; Sarkar, P.; Datta, S.; Serebrenik, A. On Negative Results When Using Sentiment Analysis Tools for Software Engineering Research. *Empir. Softw. Eng.* **2017**, *22*, 2543–2584. [CrossRef]
44. Ribeiro, F.N.; Araújo, M.; Gonçalves, P.; André Gonçalves, M.; Benevenuto, F. SentiBench—A Benchmark Comparison of State-of-the-Practice Sentiment Analysis Methods. *EPJ Data Sci.* **2016**, *5*, 23. [CrossRef]
45. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022. [CrossRef]
46. Sievert, C.; Shirley, K. LDAvis: A Method for Visualizing and Interpreting Topics. 2015. Available online: <https://aclanthology.org/W14-3110.pdf> (accessed on 29 November 2021).
47. Hale, T.; Angrist, N.; Goldszmidt, R.; Kira, B.; Petherick, A.; Phillips, T.; Webster, S.; Cameron-Blake, E.; Hallas, L.; Majumdar, S.; et al. A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* **2021**, *5*, 529–538. [CrossRef] [PubMed]
48. Vickerman, R. Will COVID-19 Put the Public Back in Public Transport? A UK Perspective. *Transp. Policy* **2021**, *103*, 95–102. [CrossRef] [PubMed]
49. Coronavirus Lockdown: What Are the New Rules Announced by Boris Johnson? | Coronavirus | The Guardian. Available online: <https://www.theguardian.com/world/2020/may/10/uk-coronavirus-lockdown-what-has-boris-johnson-announced> (accessed on 28 July 2021).
50. Local COVID-19 Alert Level Update: 15 October 2020—GOV.UK. Available online: <https://www.gov.uk/government/news/local-COVID-19-alert-level-update-15-october-2020> (accessed on 29 July 2021).
51. Coronavirus: TfL Secure Two-Week Extension of Bailout—BBC News. Available online: <https://www.bbc.com/news/uk-england-london-54568920> (accessed on 29 July 2021).
52. Medford, R.J.; Saleh, S.N.; Sumarsono, A.; Perl, T.M.; Lehmann, C.U. An “Infodemic”: Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the Coronavirus Disease 2019 Outbreak. *Open Forum Infect. Dis.* **2020**, *7*, ofaa258. [CrossRef] [PubMed]

- 
53. Extinction Rebellion Rush-Hour Protest Sparks Clash on London Underground | Extinction Rebellion | The Guardian. Available online: <https://www.theguardian.com/environment/2019/oct/17/extinction-rebellion-activists-london-underground> (accessed on 29 July 2021).
  54. Belly Mujinga's Death: Searching for the Truth—BBC News. Available online: <https://www.bbc.com/news/uk-54435703> (accessed on 29 July 2021).
  55. Transport Use during the Coronavirus (COVID-19) Pandemic—GOV.UK. Available online: <https://www.gov.uk/government/statistics/transport-use-during-the-coronavirus-COVID-19-pandemic> (accessed on 29 July 2021).