

Article

Twitter Data Mining to Map Pedestrian Experience of Open Spaces

Milena Vukmirovic ^{1,*} , Miroslava Raspopovic Milic ² and Jovana Jovic ²

¹ Department of Landscape Architecture and Horticulture, Faculty of Forestry, University of Belgrade, 1 Kneza Višeslava 1, 11000 Belgrade, Serbia

² Faculty of Information Technology, Belgrade Metropolitan University, Tadeuša Košćuška 63, 11000 Belgrade, Serbia; miroslava.raspovic@metropolitan.ac.rs (M.R.M.); jovana.jovic@metropolitan.ac.rs (J.J.)

* Correspondence: milena.vukmirovic@sfb.bg.ac.rs; Tel.: +381-11-50-33-913

Abstract: This research investigated the classification and visualisation of Twitter user-generated data. Twitter data were classified based on their sentiment relating to pedestrian experience of the quality of open spaces, based on their content. The research methodology for Twitter data collection, processing and analysis included five phases: data collection, data pre-processing, data classification, data visualisation and data analysis. The territorial focus was on Oxford Street, London, UK. Special attention was placed on the questions regarding the potential of using Twitter data for extracting relevant topics for the public space and investigating whether the sentiment for these topics can relate to urban design and improvement of pedestrian space. The proposed research model considered amount and relevance, its possibilities regarding the interpretation of the collected sample, the potential of the data for the purpose of the analysis of pedestrian space quality, the precision of sentiment determination and the usability of data in relation to a particular open public space.

Keywords: social network data; Twitter; pedestrian experience; open public spaces; Oxford Street



Citation: Vukmirovic, M.; Raspopovic Milic, M.; Jovic, J. Twitter Data Mining to Map Pedestrian Experience of Open Spaces. *Appl. Sci.* **2022**, *12*, 4143. <https://doi.org/10.3390/app12094143>

Academic Editor: Federico Divina

Received: 7 February 2022

Accepted: 7 April 2022

Published: 20 April 2022

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



Copyright: © 2022 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

Research published in *The Lancet* in 2018 warned that “if current trends continue, the 2025 global physical activity target (i.e., a 10% relative reduction in insufficient physical activity) will not be met,” urging the implementation of policies to increase population levels of physical activity worldwide [1,2]. Considering the overall percentage of insufficient physical activity, high-income Western countries register the rate of 36.8%, whereas Central and Eastern Europe fares better at 23.4%.

According to the the World Health Organization (WHO), walking is “a great way to get the physical activity needed to obtain health benefits” and strategies to promote it should therefore be pursued [3]. Walking does not require any specific skills, which means it is performable by most of the population. This is defined by Objective 3.1 of the Physical Activity Strategy for the WHO European Region 2016–2025 as reduction of car traffic and increase of walking and cycling suitability. However, if the aim is to motivate people to walk, it is necessary to create a supportive pedestrian environment with walkable features (quality sidewalks and destinations) and places that are felt to be safe, convenient, and empowering physical activity.

This is in line with the Sustainable Development Goals (SDG) for good health and well-being (SDG3) and sustainable cities and communities (SDG11), and with the Health 2020 European Policy for Health and Well-being. Effective policies include improving provisions for cycling and walking infrastructure, improving road safety and creating more opportunities for physical activity in public open spaces and parks, in workplaces and in other local community settings [4]. That means citizens’ opinions about the quality of the pedestrian space and their improvement to meet users’ expectations must be heard.

Over the last ten years, digital tools, social networks, and applications have played a main role in our everyday lives [5–7]. This new way of communicating has already sharpened urban life through more dynamic exchange of information [8,9]. Furthermore, virtual performances posted by the users of social networks such as Facebook, Instagram, Twitter and other open-generated data may help urban designers to collect necessary information about cognitive and perceptive impressions of the users.

Since we are living in a time of smartphone revolution, a new field of research possibilities has opened, due to its technological features. This can be seen in the growing interest across urban design in understanding the role played by new location-based media and the impact of the increasing availability of urban digital data from various sources [10]. With the development of social media and social networks, online platforms [11] have become a place for sharing opinions and not just a place for sharing details about personal daily lives. Social network users use these platforms to express their opinions about services, products, events, public individuals, public locations, etc. The experience of citizens and visitors in cities can now be shared through pictures, videos, posts, and check-ins, while social media has become a vital information source through which different urban phenomena can be explored [12,13] and can help to explain a variety of spatial-temporal phenomena [14]. One aspect of the shared urban environment which is particularly important for planners and policy makers is pedestrian space. Analysis of social media data can give insights into the quality of this space, which can then be used to improve or support its use.

For several years now, much attention has been paid to the transformation of the main city streets of major European cities. Many streets are in capitals such as Paris, Madrid, Brussels, Oslo, Vienna, Copenhagen, Ljubljana, London, etc., and these world-famous streets attract many users, visitors and tourists. On the other hand, the main city street is much more than a place; it is also a state of mind and a set of values [15]. Transformation of the main city street can have a significant effect in relation to future urban practices and values but also in relation to the quality of life of the citizens of that city.

Having in mind several initiatives and plans concerning the transformation of Oxford Street in London [16,17] and the very nature of the changes envisaged for it [18,19], this street was taken as an appropriate example to investigate the potential of user-generated data related to public space. In addition to the above, the Oxford Street District is a major part of London's West End and among the world's most visited destinations. Approximately 200 million people visit the district each year, around 70% of whom arrive by public transport [20]. It is also home to over 38,000 residents and employs over 155,000 people. This location was chosen because the planning phase of the transformation of Oxford Street has just begun, and the implementation phase has not started yet at the moment of writing this work. This gives a good basis to be able to collect data from social networks while having a baseline and comparable research conducted using traditional methods of data collection (i.e., surveys and interviews). It is important to mention that a similar study was done between 1 July 2015 and 29 February 2016 for the predefined places of interest in Belgrade, Serbia, when 2872 tweets were collected [21], but the focus was on tweets with geotags in order to measure the attractiveness of the locations in this city.

In accordance with the above-mentioned factors, this research investigates the user-generated data from Twitter to classify and visualize Twitter data sentiment relating to the pedestrian experience of the quality of open spaces, based on their content. It is one of the most popular data sources for research and offers an opportunity to study human communication and social networks [22] because of its open network, allowing access to information published through this platform.

Considering the character of social networks and user-generated content, the study reflects tertiary communication as knowledge that is distributed across different individuals and social groups but can be less easily retrieved/usable and operationalised. To find a unique focus for the use of this data, in this research, a sentiment analysis approach was used on Twitter data to map and evaluate user opinions and sentiments related to the specific location, with the goal of classifying them in information clusters about the quality

of the pedestrian environment. Classification of Twitter data was done by determining the sentiment of each tweet and classifying them in one of three categories: positive, neutral, and negative sentiments. The results from this analysis were used for consideration of ideas for urban planning of the public space. The location of interest was Oxford Street in London, UK.

Accordingly, this work proposes a basis for identifying issues dealing with a certain public space characteristic that can be further analysed in more depth. Therefore, the research question (RQ1) focuses on extracting common topics from Twitter data that are related to certain public space and investigating whether the collective sentiment on this public space can relate to urban design and improvement of pedestrian space. RQ1 aims to extract semantics (topics of interest for pedestrian space) and sentiments (public's opinions about environmental elements relating to public pedestrian space). The following question (RQ2) seeks to validate the finding from RQ1 by comparing it with publicly available data on the community opinions about the public space. The overall goal of this work is to map user experience and opinions about the pedestrian space and to detect sentiment clusters around environmental elements that can be further used for urban and public space design. Given the time- and cost-effectiveness of the proposed model, many of the overlooked opinions and elements can be collected and analysed.

2. Materials and Methods

The research methodology for Twitter data collection, processing and analysis is shown in Figure 1. The flowchart shows several key steps in the research methodology, including data collection, data pre-processing, data classification, data visualisation and data analysis. *Data collection* presumes data retrieval from Twitter. *Collected data* is retrieved in original format, and it requires processing and “cleaning” of data, which is done during the *Data pre-processing phase*. *Cleaned data* is further classified in the *Data Classification* phase using sentiment analysis methods, which is then visualized (*Data Visualisation*) using different statistical and visualisation methods. Once the visualisations are ready, the last phase of this research was conducted through *Data Analysis*. Each phase of this research method is further described in the following sections.

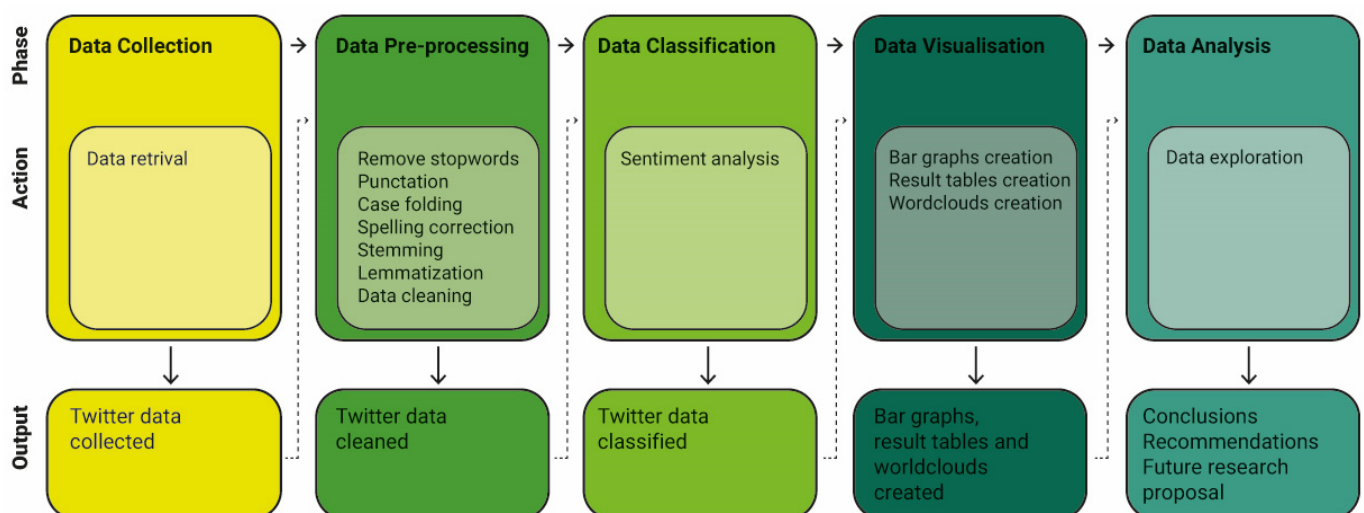


Figure 1. Flowchart of research methodology.

2.1. Data Collection

Data used in this study was collected from Twitter in the period from 5 May 2021 to 24 January 2022. Data collection was focused on the location of Oxford Street, London, UK. Data was collected based on geotag and related hashtags. Data was acquired through Twitter API (Application Programming Interface). For *Academic Research*, this API allows for collection of 10 million tweets per month, with the limitation of 100 requests per

15 min [8,23]. During this period, 24,821 tweets were collected, but after the data “cleaning” process, the final sample was reduced to 23,587 tweets. It is necessary to state that data that was collected during the COVID-19 measures at times included lock down, which may to some extent have influenced opinions, keeping in mind that the citizens wanted to walk down the street and stay in open spaces. This includes certain restrictions on the presence of visitors and tourists, while it can be assumed that during this period there were mostly locals.

During the collection of data certain assumptions were made:

- Only tweets in English were collected.
- Tweets containing hashtags #oxford, #oxfordstreet, #oxfordst were collected. 738 tweets were collected using these hashtags.
- Tweets containing *oxford* and *oxford street* in the tweet text were also collected.
- Geotagged tweets containing *Oxford Street* as a place obtained from Twitter’s reverse geocode endpoint in the REST API were collected.
- Retweets were removed, as they were considered to be duplicate content.

22,530 tweets were collected using defined hashtags and keywords. Only 327 collected tweets contained geotag, while 27 tweets were collected based on geotagging. Compared to the research conducted during 2015 and 2016 [21], it can be concluded that Twitter’s modifications of privacy settings, which included significant reinforcements of user data protection and user location, led to smaller samples of tweets with geotags.

2.2. Data Pre-Processing

The phase of *data pre-processing* includes several steps that eliminate parts of tweets that contain elements that may unnecessarily affect the sentiment score. The goal is to reduce the “noise” that comes with social networks data, such as links, slang, stop words, etc., as much as possible so that data is in a text format that can be parsed and classified. For these steps, Python’s String methods and the library Natural Language Toolkit (NLTK) were used [24]. NLTK is a suite of libraries and programs that helps with work in natural language processing (NLP) for English.

Data pre-processing steps include the following:

- Capitalization/case-folding—converting upper case to lower case. When case-sensitive analysis is used, two same words can be treated as different.
- Punctuation—Removing punctuation, digits and special characters that don’t convey any sentiment.
- Remove stop words—removing stop words which do not affect the meaning of the sentence, removing short words that contain 3 or less characters (*his, all, are, at*, etc.).
- Spelling correction—correcting spelling conducted because it could affect the sentiment.
- Stemming—normalizing words from the inflected form to be able to analyse words in their root dictionary form. For example, *loves, loving, lovable* are often used in the same context, and in stemming affixes are eliminated to the root word “love.”
- Lemmatization—normalizing words using vocabulary and morphological analysis of words. As in stemming, the word is converted to its root dictionary form; however, unlike in stemming, lemmatization considers context and normalizes words to its meaningful root word (i.e., changing “worse” to “bad”).
- Data cleaning—removing URL and username, as this data will not provide any additional information needed for this stage of research.

Examples of tweets before and after data pre-processing are given in Table 1.

2.3. Data Classification

Data classifications were conducted using a sentiment analysis tool developed by CJ Hutto and Eric Gilbert, VADER (Valence Aware Dictionary and sEntiment Reasoner) [25]. VADER is a lexicon- and rule-based sentiment analysis tool that is specifically oriented to analyse sentiments in social media data. Each tweet was categorized in one of the sentiment

categories: positive, neutral, or negative. VADER sentiment analysis uses a sentiment lexicon: a list of words that are mapped to emotion intensities known as sentiment scores. Sentiment scores are normalized between -1 (most extreme negative) and $+1$ (most extreme positive). The sentiment score of a text is calculated as a sum of the intensity of each word in the text. Based on the sentiment score, tweets are classified as positive (sentiment score higher than 0), neutral (sentiment score equals to 0) and negative (sentiment score lower than 0). The sentiment score for each tweet is stored in order to make this data available for further analysis of the level of negative or positive sentiments.

Table 1. Example of tweets before and after the pre-processing.

Tweets BEFORE the Pre-Processing		Tweets AFTER the Pre-Processing
Think will be hitting Regent Street more often . . . Happy to say people have started coming back out, Oxford Circus, shops, restaurants, and theatres were buzzing and busy! Loving it! Street photography in Oxford. Waiting for the bus can be boring especially when you are tired from walking around a city as a tourist. This is part of a series of candid shots of people waiting at a bus stop.	→	think will be hit Regent Street more often happy to say people have started coming back out oxford circus shop , restaurant and theatre were buzz and busy love it
#streetphotography #oxford #bustop #waiting . . . https://t.co/N3obbtjHED https://t.co/2gzYy8Uoa5	→	street photography in oxford wait for the bus can be boring especially when you are tire from walk around a city as a tourist
Good news, they are adding turn lanes at the laundromat in Oxford. Bad news, the stadium across the street is holding graduation and people ignored the no parking signs.	→	good news they are add turn lane at the laundromat in oxford
https://t.co/g18Cix0Qb8 https://t.co/DfU9prgY29	→	bad news the stadium across the street is hold graduation and people ignore the no parking sign

2.4. Data Visualisation and Presentation

Data was presented and visualized using tables, bar graphs, word cloud and histograms. Histograms were used to visually represent data distribution and to do finer analysis of sentiments by analysing more deeply within the sentiment classifier. For instance, looking into negative sentiments, it is important to analyse what tweets are more negative than others. Bar graphs were used to visually present the quantity of the specific tweets and their sentiments. A word cloud was used to represent used words based on their frequency. The more that a certain word appears in collected tweets, the larger the font size used to visually present that word in the word cloud.

3. Results

By processing the collected data, using the previously mentioned methodology, the results were obtained, which are presented below. Figure 2a,c represent the words that appear with the highest frequency in tweets, with positive, negative, and neutral sentiment, respectively.



Figure 2. Word cloud visualisation of (a) tweets with positive, (b) negative and (c) neutral sentiment.

Besides the keywords used for data collection, such as “oxford street”, most frequently used words in positive tweets are *love*, *today*, *look*, *good*, *work*, *life*, *smile* and *feel*. Similarly,

for negative tweets the most frequently used words are *London, people, attack, Jewish, protest, stop, fire, hate*; while in neutral tweets the most frequently used words are *London, shop, flagship, Christmas, park, place* and *Selfridge* (Table 2).

Table 2. Most frequently used words in tweets with positive, negative, or neutral sentiment.

Word	Count in Positive Tweets	Count in Negative Tweets	Count in Neutral Tweets
street	7604	6250	8870
oxford	7058	5677	8042
love	1541	0	0
today	1221	0	0
look	842	0	0
good	801	0	0
work	772	0	0
life	766	0	0
smile	659	0	0
feel	651	0	0
London	0	1000	1550
shop	0	0	1001
walk	0	0	544
flagship	0	0	393
Christmas	0	0	382
park	0	0	374
place	0	0	303
Selfridge	0	0	288
people	0	579	0
attack	0	471	0
Jewish	0	461	0
protest	0	350	0
stop	0	291	0
fire	0	253	0
hate	0	160	0

As data was collected based on the assigned keywords and hashtags, it was of interest to analyse what were other related hashtags showing up in the collected tweets. Table 3 provides this information based on the tweet sentiment, giving the number of tweets in which related hashtags appear and the percentage of the total number of positive, negative and neutral tweets (Figure 3).

Table 3. Most frequently used related hashtags in tweets.

Hashtags (Positive)	No.	%	Hashtags (Negative)	No.	%	Hashtags (Neutral)	No.	%
london	186	1.95%	london	76	1.16%	london	163	2.17%
oxford	161	1.69%	oxfordstreet	63	0.96%	oxford	120	1.60%
oxfordstreet	126	1.32%	oxford	39	0.60%	oxfordstreet	102	1.36%
retail	97	1.02%	antisemitism	30	0.46%	streetphotography	58	0.77%
retailnews	39	0.41%	londontraffic	25	0.38%	google	57	0.76%
londonprotest	29	0.30%	mefire	20	0.31%	photography	43	0.57%
selfridges	31	0.33%	ldnont	19	0.29%	street	42	0.56%
christmas	23	0.24%	defundthebbc	18	0.28%	retail	42	0.56%
fashion	21	0.22%	extinctionrebellion	17	0.26%	ldnont	28	0.37%
realestate	19	0.20%	londonprotest	15	0.23%	realestate	27	0.36%

Considering the overall sentiment of the collected tweets, the analysis shows that 40.40% of tweets carry positive sentiment, while 27.73% are negative and 31.87% are neutral (Table 4).

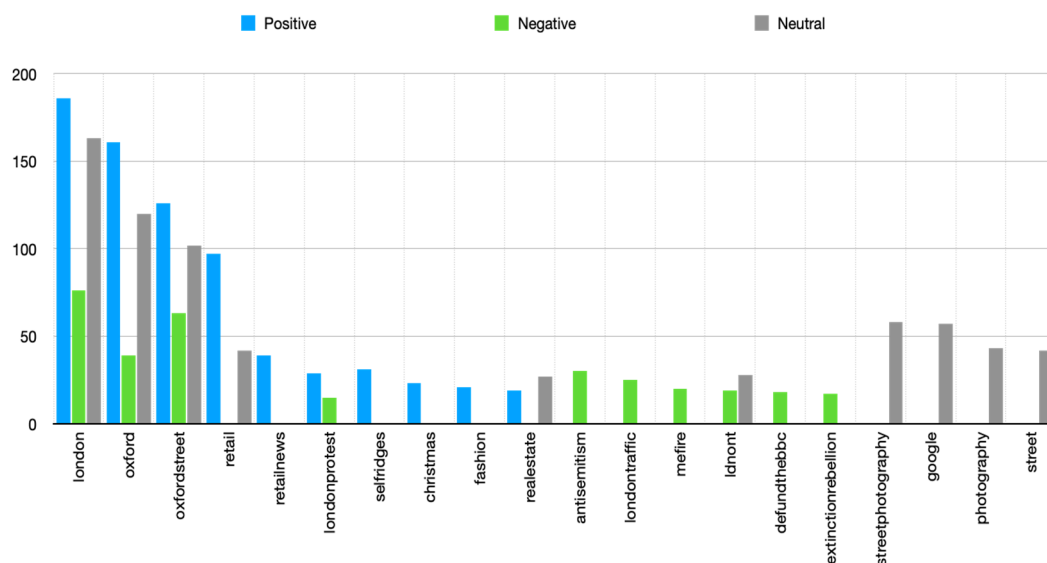


Figure 3. Number of tweets with most frequently used hashtags by positive, negative, and neutral sentiments.

Table 4. Overall sentiment values for the collected data.

	Number of Tweets	Percentage of Tweets
positive sentiment	9528	40.40%
negative sentiment	6541	27.73%
neutral sentiment	7518	31.87%

Sentiment for all tweets was further analysed to determine the level of positive and negative sentiment. Figure 4 shows the number of tweets in specified intervals of sentiment scores. Intervals were assigned in 0.25 increments, while neutral sentiment tweets are positioned at “0.” It can be seen that “weak positive” tweets that follow in the range between 0–0.25 are similar to the tweets that are classified as “strong positive”, 0.75–1.

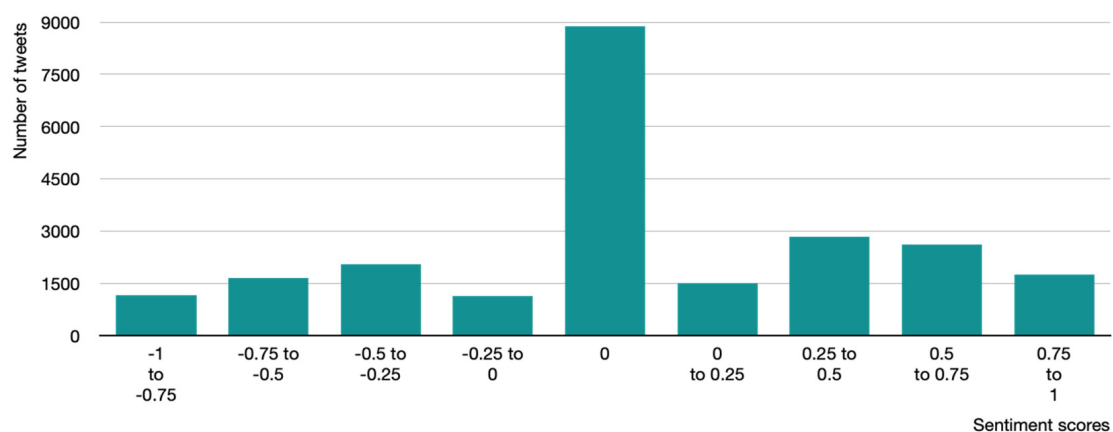


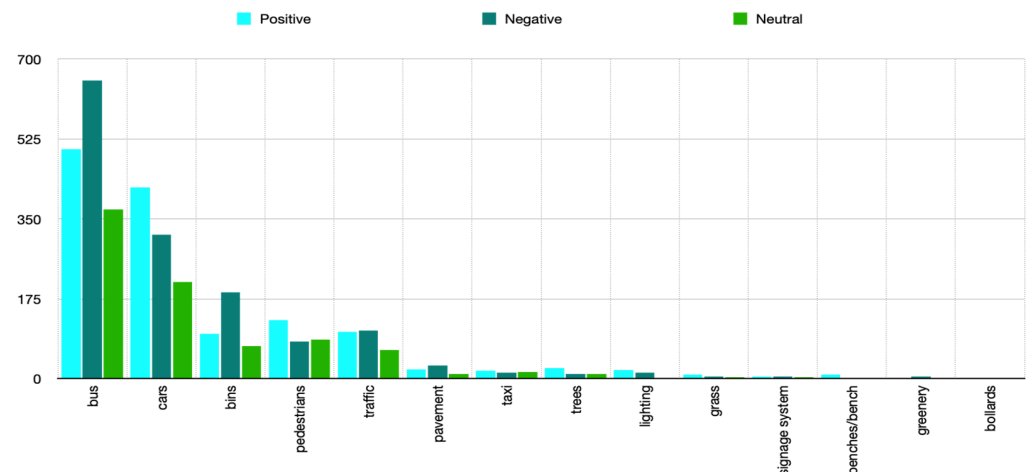
Figure 4. Distribution of tweets in the intervals for specific sentiment scores.

However, most of the positive tweets are either in the sentiment score range of 0.25–0.5 or 0.5–0.75 (Table 5). Similar trends are seen with negative sentiment scores as well, since “weak negative” and “strong negative” do not differ much, whereas the middle range scores are where most of the negative sentiment scores fall.

Words relating to the environment elements that can contribute to pedestrian experience were analysed separately (Figure 5).

Table 5. Number of positive, negative, and neutral tweets in 0.25 intervals.

Intervals	−1 to −0.75	−0.75 to −0.5	−0.5 to −0.25	−0.25 to 0	=0	0 to 0.25	0.25 to 0.5	0.5 to 0.75	0.75 to 1
No.	1160	1661	2051	1127	8877	1513	2833	2614	1751

**Figure 5.** Specific word occurrences in tweets by type.

Words relating to environment elements appear in 3638 tweets, most of which deal with expressions related to traffic or some modality of transport, such as *bus* (1527 tweets), *cars* (947 tweets), *pedestrians* (294 tweets), *traffic* (271 tweets) and *taxi* (47 tweets). The sentiment of *pedestrians*, *cars* and *taxi* carry positive sentiments, whereas *bus* and *traffic* carry negative sentiments.

Other terms, such as *bins* (358 tweets), *pavement* (60 tweets), *trees* (45 tweets) and *lighting* (35 tweets), appear on a much smaller scale. Further analysis determined that elements of urban equipment such as *bins* and *pavement* carry negative sentiment, whereas *trees*, *grass* and *lighting* appear with positive sentiment.

4. Discussion

Considering the aim of this research, special attention was paid to the questions regarding the potential of using Twitter data. Results were analysed having in mind the amount and relevance of the collected data, as well as the possibilities regarding data interpretation. The potential of the data for the purpose of the analysis of pedestrian space quality, precision of sentiment determination and the usability of data were considered in relation to a particular open public space—Oxford Street, London, UK.

4.1. Amount and Relevance of the Collected Twitter Data

During the nine months, 24,821 tweets were collected, which related to the research area of Oxford Street in London. After removing the retweets, the sample included 22,530 tweets collected using defined hashtags and keywords (*#oxford*, *#oxfordstreet*, *#oxfordst*, *oxford* and *oxford street*), and 327 collected tweets contained geotag, while only 27 tweets were collected based on geotagging at Oxford Street. It was noticed that the intensity of certain tweets was higher at the time when an event such as the London protests or the anti-Semitic act on the bus took place. In addition to the above, there were a very small sample of tweets that have geotags. That share is 1.39%. It is assumed that the reason is that most of the social network users switch off geolocation, both on the application and on the phone.

The amount and relevance of this sample can be compared with the data related to the conducted public consultations related to the future appearance and use of Oxford Street and its contact zone (Oxford Street District—OSD). The first consultations were held in

the period from 6 November 2017 until 3 January 2018, shortly after the announcement of the Mayor of London that “about half a mile of the street from Oxford Circus to Orchard Street could become a “traffic-free pedestrian boulevard” [26]. In the specified period, around a million people were directly contacted for the public consultation [18]. During the consultation, 14,377 responses were collected, “with just over 9000—about 64%—either supporting the project outright or backing the plans with some concerns about certain elements” [27].

Public consultations on the draft Oxford Street District Place Strategy were held from 6 November to 16 December 2018, which was preceded by the presentation of the Strategy to the residents, employers and those visiting or working in the area. Consultations were conducted through online questionnaires (on the website and by email) and direct surveys. During the 6 weeks of consultation, 1800 online questionnaires were completed, while 354 were directly surveyed [28]. Considering the respondent type, the largest group were regular visitors to the OSD (61%), followed by residents of Westminster (34%) and District workers (20%).

Considering the results of the official public consultation reports, it can be interpreted that the data collected from Twitter can be characterized as relevant, having in mind both the amount (2757 tweets per month) and the thematic framework of interest, determined by the choice of content that contains specific terms and hashtags. An increase in the sample and more concrete answers and discussions could be expected after the implementation of a social network campaign that would be focused on a specific topic.

4.2. Twitter Data Mining

Insight into the results of the research could easily lead to the dominant character of Oxford Street—trade and retail, which is characterized as positive. This is confirmed by the number of tweets with positive sentiment that included hashtags such as *#retail* (97 tweets with positive sentiment and 42 tweets with neutral sentiment), *#retailnews* (39 tweets with positive sentiment), *#Selfridges* (31 tweets with positive sentiment) and *#fashion* (21 tweets with positive sentiment), see Table 3.

The factual situation according to which Oxford Street (Figure 6) is predominantly characterised as “the retail spine of London’s West End” [20] speaks in favour of the above. This means that department stores, flagship and high-street retailers predominantly occupy the ground floor.

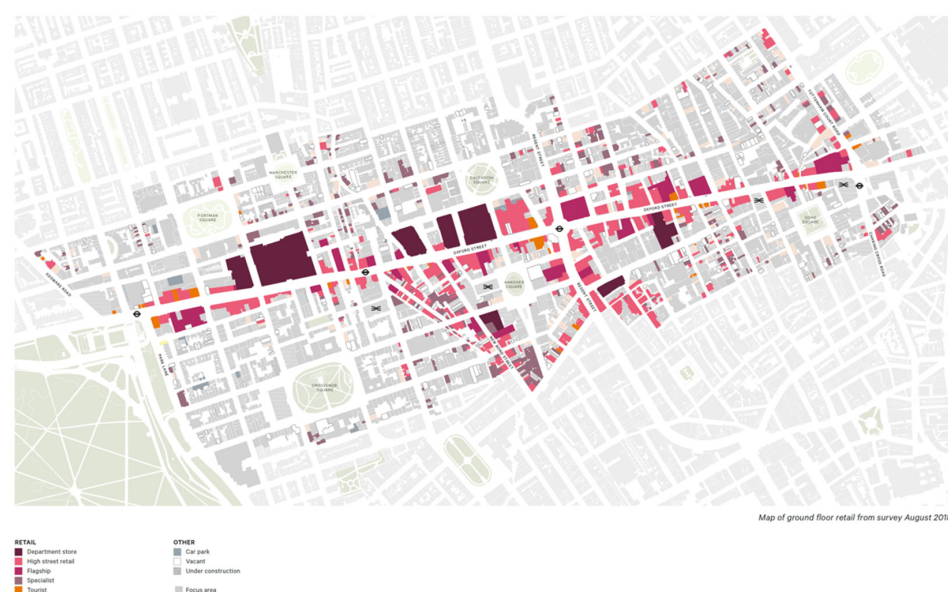


Figure 6. OSD ground-floor retail (Reprinted with permission from ref. [28]. Copyright 2019 Publica.) [28].

On the other hand, as the main problem, if individual events that receive a lot of attention in a short period of time are left out [29], the traffic can be clearly singled out (*#londontraffic*, 25 tweets with negative sentiment). When this is compared with the results of the public consultation, the traffic problem clearly stands out as one of the ten “most frequently raised issues” [30] (p. 26). On this occasion, the participants in the public consultations pointed out the need for banning the motorised traffic from Oxford Street and supported greater restriction to traffic within the OSD [30].

The above overview of the results indicates that Twitter data can give a clear overall picture of a particular location and its dominant character and issues, viewed from the urban perspective.

4.3. Twitter Data Mining in Pedestrian Space Quality Analysis

In this research, special attention was paid to the elements of the pedestrian environment, which were analysed using specific words related to this micro scale such as *bench*, *bin*, *bollard*, *signage system*, *lighting*, *pavement* (urban equipment and furnishing), *grass*, *trees*, *greenery* (landscaping elements), *pedestrian*, *car*, *bus*, *taxi* and *traffic* (modes of transport). These terms appear in 3638 tweets, which is 15.42% of the total sample. The most common are tweets related to traffic (76.1%), where buses and traffic in general are mentioned as an issue, unlike pedestrians and taxis, which carry positive sentiment (see Figure 5). In relation to urban equipment, the notion of *bins* is most prevalent in a negative context, while other elements occur to a lesser extent. It should also be noted that the words *trees* (1.2%), *grass* (0.5%), and *lighting* (1.0%), although appearing on a significantly smaller scale, carry positive sentiment. In this way, issues can be clearly separated from desirable elements of pedestrian spaces.

If these results are compared with the results of public consultations on the issue of respondents’ interest in certain topics, a certain analogy can be observed (see Table 6). Thus, the most interest arises in relation to transport (63.6%), landscape (4.6%), amenity (3.5%), lighting (3.4%) and materials (2.1%) [30] (p. 34). Further, specific recommendations from the first public consultations in 2018 [27] referred to pedestrianisation of Oxford Street, levelling the pavement of the road and sidewalks to improve accessibility and extension of the taxi ranks.

Table 6. Comparison of the results obtained from Twitter and by using traditional tools in relation to an individual topic.

Results Obtained from Twitter Analysis		Results obtained Using Traditional Tools	
Element	%	Element	%
Traffic	76.1%	Transport	63.6%
Trees			
Grass	1.9%	Landscape	4.6%
Greenery			
Bench			
Bins	11.5%	Amenity	3.5%
Bollards			
Lighting	1.0%	Lighting	3.4%
Pavement	1.6%	Materials	2.1%

This comparison with the results of targeted research suggests that Twitter can be considered a resource channel in the analysis of pedestrian areas. The assumption is that some other social networks, such as Instagram, which is also popular in Europe [31], will be an even better and richer basis, and that data collected from several social networks can be combined with research. For these purposes, some other methods for sentiment research should be considered to get a better insight of the sentiment of data that are in formats other than text.

Observing positive and negative sentiment related to individual elements of the pedestrian space can contribute to future decisions of urban designers related to transformation

proposals. What users have described as positive should be retained and nurtured, while what is characterized as negative should find solutions and improvements. For example, bins are characterized negatively, due to accumulated garbage, irregular maintenance, and high leakage between elements. Accordingly, more regular maintenance and emptying, greater volume of manure and closer distance between individual elements should be proposed. On the other hand, what is characterized positively, such as trees or grass, should be emphasized, the area and number of seedlings should be increased, or some additional alternative solutions in the form of green walls and roofs should be given.

4.4. Precision of Sentiment Determination

In the context of this work, it was noticed that sentiment classification of individual tweets does not always perform classification accurately. For instance, the tweet “I wonder if London holiday guides still refer to Oxford Street as the “ultimate shopping destination”. Cos . . . it’s not. Just bare American candy shops, traffic, and Zara’s. It’s so yuck now” was classified as positive, even though by examination it was identified that this tweet carries negative sentiment. Something similar can be seen with the tweet “I never thought I’d love walking down a heaving Oxford Street, but it was bloody brilliant! Normality”, which was classified as negative, and it carries positive sentiment. Previous research identifies VADER classification of social media for all normalized sentiment scores between -1 and $+1$, at F1 Classification Accuracy of 0.96 [25]. A possible reason for this is due to the lack of sentiment-oriented text, often expressed by emoticons, slang, initials and abbreviations [25]. VADER does not detect irony and phrases, but will rate individual words, which may lead to the wrong classification of the sentiment [32].

Different authors have analysed and compared VADER’s precision and accuracy. Accuracy is defined as a proportion of total number of correctly classified tweets and total number of analysed tweets, while precision represents the measure of correctly classified positive tweets only. VADER’s accuracy rate is 83%, while the precision is 90% (F1 score is 89%) [33]. The high accuracy and precision rate of the VADER method are not the only advantages of this method. Its computational efficiency and quickness are also an advantage. Furthermore, VADER is transparent and easily accessible, with its use and easy interpretation, making it useful for non-computer scientists. As different researchers report much lower precision and accuracy when VADER is applied to different topics and domains, future research should analyse further the precision and efficiency of other sentiment analysis approaches that may provide better performance when analysing social media on the topic of urban planning of public spaces.

4.5. Usability of Data in Relation to a Particular Open Public Space

It should be considered here that the research of the potential of Twitter data for the purpose of evaluating the quality of pedestrian space was done in relation to one of the most famous and most visited open public urban spaces—Oxford Street in London. Therefore, the question arises to what extent the use of Twitter data would be possible for the needs of another pedestrian space, which is not so well known and popular, in relation to the sample size itself. There are also important cultural factors regarding the intensity of the use of social networks in a certain territory, as well as the language in which certain topics are tweeted, which can be considered a minor challenge. Also, the analysis cannot fully indicate the character of the users of the space, such as the age of the users, their gender, occupation, or place of residence, which is important for the needs of urban design.

In addition, this work proposes a combined research approach, social network sentiment data analysis and survey of public opinions for mapping user experience and their opinions about environmental elements of the pedestrian or open public space. This Twitter data sentiment analysis demonstrates that certain patterns of user experience and opinions can be clearly mapped from the tweets’ sentiments. Twitter data information is compared with research conducted using traditional methods (interviews and surveys). Research about public opinions of public space that was derived from traditional data collection

was noticeably labour- and time-consuming, while being able to collect smaller amounts of data from specific demographics. On the other hand, collecting data from social networks is much faster, less labour-intensive and less time-consuming. Real-time availability of data-driven analysis is not slow and costly, as are traditional methods. In the case of the used case study of Oxford Street, it was noticed that certain events triggered a higher volume of data collection. Having this in mind, and other surveys conducted about the public space design of this location, it would be of interest to combine not only these sources but also to combine similar efforts at the same time, to aggregate a larger amount of usable and relevant data.

5. Conclusions

This study indicates the fact that users of pedestrian areas observe certain phenomena in detail and characterize them in terms of values, and that social networks give them the opportunity to express them and to inform others about them. If these channels are observed carefully, valuable data can be obtained, based on which one can get an impression of the quality and needs of a certain pedestrian space. This research explored how topics related to environmental elements of public pedestrian space can be extracted from Twitter data, and the sentiment for each topic was classified based on the sentiment scores for each tweet. The extracted topics and their sentiments were compared to the research conducted using traditional methods for data collection through surveys and interviews. Our research shows that an analogy was found between analysing sentiments of Twitter data and traditional research. While data collection and analysis are efficient using data collection from social media, it also shows that complementing efforts with traditional data collection can contribute to the context of urban planning and design. It is interesting that data from social networks, which do not represent targeted research such as interviews and surveys, give similar results, especially bearing in mind that users of open spaces react to some situations on their own initiative and record their observations and impressions. In that sense, urban planners and designers get an overview of the positive and negative properties of a certain space in relation to which they can give an appropriate proposal for transformation—emphasizing and nurturing characteristics that are singled out as advantages and changing and proposing solutions for what is singled out as an issue. However, the relevance of this tool is possible only when there is a sufficient sample, which in the presented Oxford Street study can be considered acceptable compared to the sample obtained using traditional methods. With regards to this, the behaviour of this tool in relation to less visited open spaces, i.e., those of a local character, should be checked. Also, bearing in mind that the work focused on comparing the results obtained through social networks and the results obtained by traditional methods, future research should be directed to other locations such as Paris, Brussels, or Milan, to verify the obtained results.

On the other hand, to explore how cities can create more quality space for people, social media is considered as an important tool [34–40] for urban planners and designers to reach the community and to advance the conversation about urban issues. Future research should explore defining a framework for extracting data about public space from other systems such as social networks, web-based systems, and other related open data. Several challenges should be addressed in future research: (1) identify data sources that are adequate for extracting common and relevant topics for urban planning of pedestrian space (i.e., TripAdvisor, Foursquare, Facebook, etc.), (2) design software architecture that can integrate different external heterogeneous data sources while identifying proper methods for data transformation that can optimize topic extraction efficiently and (3) identify methods to automatically classify topics from collected data only related to public pedestrian spaces. In accordance with the above, further research will go in the direction of applying some of the current methods [41] of researching the quality of open public urban spaces, such as the Placemaking method [42], PEDS (Pedestrian Environment Data Scan [43]), PERS [44] or the SACLAV [45] method that is being developed at the University of Belgrade.

Author Contributions: Conceptualization, M.V. and M.R.M.; methodology, M.V. and M.R.M.; software, M.R.M. and J.J.; validation, M.V., M.R.M. and J.J.; formal analysis, M.R.M. and J.J.; investigation, M.V., M.R.M. and J.J.; resources, M.V., M.R.M. and J.J.; data curation, M.V. and M.R.M.; writing—original draft preparation, M.V.; writing—final version, review and editing, M.V., M.R.M. and J.J.; visualisation, M.V. and J.J.; supervision, M.V. and M.R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study does not require ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study did not report any data.

Acknowledgments: The work presented here was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Project No. III44006 and Project No. 451-03-68/2021-14/200169. Also, our research team wants to thank Alenka Temeljotov Salaj from NTNU for support and guidance on preparing the paper for publication.

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

References

- Guthold, R.; Stevens, G.A.; Riley, L.M.; Bull, F.C. Worldwide trends in insufficient physical activity from 2001 to 2016: A pooled analysis of 358 population-based surveys with 1.9 million participants. *Lancet Glob. Health* **2018**, *6*, E1077–E1086. [\[CrossRef\]](#)
- Gascon, M.; Götschi, T.; de Nazelle, A.; Gracia, E.; Ambròs, A.; Márquez, S.; Marquet, O.; Avila-Palencia, I.; Brand, C.; Iacorossi, F. Correlates of Walking for Travel in Seven European Cities: The PASTA Project. *Environ. Health Perspect.* **2019**, *127*, 097003. [\[CrossRef\]](#) [\[PubMed\]](#)
- Walker, J.; Schwartz, S.; Lick Rehorova, J.; Vlcek, J. (Eds.) *Urban Mobility Partnership, Promoting Mobility Behaviour Change*; Walk21 Foundation: Cheltenham, UK, 2019.
- Pompe, A.; Temeljotov Salaj, A. Qualitative criteria of urbanism and brands: A comparative analysis. *Urbani Izziv* **2014**, *25*, 5–23. [\[CrossRef\]](#)
- Bredl, K.; Hunniger, J.; Jensen, J.L. Methods for Analyzing Social Media: Introduction. In *Methods for Analyzing Social Media*; Routledge: London, UK, 2017.
- Bruns, A.; Stieglitz, S. Quantitative approaches to comparing communication patterns on Twitter. *J. Technol. Hum. Serv.* **2012**, *30*, 160–185. [\[CrossRef\]](#)
- Zafarani, R.; Abbasi, M.A.; Liu, H. *Social Media Mining*; Cambridge University Press: Cambridge, UK, 2014.
- Castells, M. *The Rise of the Network Society*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
- Torres, Y.Q.A.; Costa, L.M.S.A. Digital Narratives: Mapping Contemporary Use of Urban Open Spaces through Geo-social Data. *Procedia Environ. Sci.* **2014**, *22*, 1–11. [\[CrossRef\]](#)
- Vukmirovic, M.; Vansita Lazarevic, E. Place Competitiveness Expressed through Digital Data. In *Keeping up with Technologies to Improve Places*; Vanista Lazarevic, E., Vukmirovic, M., Krstic-Furundzic, A., Djukic, E., Eds.; Cambridge Scholars Publishing: Newcastle upon Tyne, UK, 2015; pp. 24–38.
- Raspopovic Milic, M.; Vukmirovic, M.; Cvetanovic, S. Information System Supporting Heterogeneous Sources for Promoting Destination of Rural Areas. In *Handbook of Research on Urban-Rural Synergy Development through Housing, Landscape, and Tourism*; Krstic-Furundzic, A., Djukic, A., Eds.; IGI Global: Hershey, PA, USA, 2020; pp. 139–154. [\[CrossRef\]](#)
- Vukmirovic, M.; Raspopovic, M. Vulnerability of Public Space and the Role of Social Networks in Crisis. In *Keeping up with Technologies to Create Cognitive City*; Vanista Lazarevic, E., Vukmirovic, M., Krstic-Furundzic, A., Djukic, E., Eds.; Cambridge Scholars Publishing: Newcastle upon Tyne, UK, 2019; pp. 60–72.
- Raspopovic Milic, M.; Vukmirovic, M.; Banovic, K. Sentiment Analysis of Twitter Data of Historical Sites. In *Proceedings of the Places and Technologies 2019, the 6th International Academic Conference on Places and Technologies*, Pecs, Hungary, 9–10 May 2019.
- Furtado, A.S.; Fileto, R.; Renso, C. Assessing the Attractiveness of Places with Movement Data. *J. Inf. Data Manag.* **2013**, *4*, 124–133.
- Kostof, S. *America by Design*; Oxford University Press: New York, NY, USA, 1987.
- Mayor of London. *The London Plan 2016*; London Assembly: London, UK, 2016.
- City of Westminster. *City Plan 2019–2040*; City of Westminster: City of Westminster, UK, 2021.
- Vukmirovic, M.; Miljkovic, E. Green transformations of the main street in European capital cities. In *Book of Proceedings of the 6th Conference of Interdisciplinary Research on Real Estate, Enschede, The Netherlands, 17–18 September 2020*; Institute for Real Estate Studies: Ljubljana, Slovenia, 2021.
- Vukmirovic, M.; Miljkovic, E. Main city street green transformation framework. *Real Estate Res. Q.* **2022**; in the publishing process.
- City of Westminster. *Oxford Street District Framework*; City of Westminster: City of Westminster, UK, 2021.

21. Djukic, A.; Vukmirovic, M.; Jokovic, J.; Dinkic, N. Tweeting in open public space: Case study Belgrade. In *Enhancing Places Through Technology*; Zammit, A., Kenna, T., Eds.; Edições Universitárias Lusófonas: Lisbon, Portugal, 2017.
22. Miller, G. Social scientists wade into the tweet stream. *Science* **2011**, *333*, 1814–1815. [CrossRef] [PubMed]
23. Twitter. Twitter API. 6 February 2021. Available online: <https://developer.twitter.com/en/docs/twitter-api> (accessed on 10 February 2021).
24. NLTK Project. Natural Language Toolkit. Available online: <https://www.nltk.org> (accessed on 4 February 2021).
25. Hutto, C.; Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June 2014.
26. BBC News. London's Oxford Street Could be Traffic-Free by December 2018, Says Mayor, 6 November 2017. Available online: <https://www.bbc.com/news/uk-england-london-41878406> (accessed on 10 January 2022).
27. BBC News. 'Traffic-Free' Oxford Street Consultation Backed by Public, 13 March 2018. Available online: <https://www.bbc.com/news/uk-england-london-43384939> (accessed on 10 January 2022).
28. City of Westminster. *Oxford Street District. Place Strategy and Delivery Plan, London*; City of Westminster: City of Westminster, UK, 2019.
29. Kunza, J. Why are Jews Protesting the BBC? Unpacking the Hanukkah Bus Incident. 20 December 2021. Available online: <https://jewishunpacked.com/why-are-jews-protesting-the-bbc-unpacking-the-hanukkah-bus-incident/> (accessed on 10 January 2022).
30. Steer. *Oxford Street District—Place Strategy and Delivery Plan Consultation Results and Analysis Report*; Westminster City Council: London, UK, 2019.
31. Statcounter. Mobile Social Media Stats Europe Jan 2021–Jan 2022, January 2022. Available online: <https://gs.statcounter.com/social-media-stats/mobile/europe> (accessed on 2 February 2022).
32. Tymann, K.; Lutz, M.; Palsbroker, P.; Gips, C. GerVADER-A German Adaptation of the VADER Sentiment Analysis Tool for Social Media Texts. *LWDA* **2019**, 178–189.
33. Nguyen, H.; Veluchamy, A.; Diop, M.; Iqbal, R. Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches. *SMU Data Sci. Rev.* **2018**, *1*, 7.
34. Gehl People. Managing Public Space in the 'New Normal', 23 April 2020. Available online: <https://gehlpeople.com/blog/managing-public-space-in-the-new-normal/> (accessed on 10 January 2022).
35. Lonas, A. 'The Joys of Walking': Pro-Pedestrian Groups Hope Increase in Strolling Will Last beyond Pandemic, 30 May 2020. Available online: <https://www.washingtonexaminer.com/news/the-joys-of-walking-pro-pedestrian-groups-hope-increase-in-strolling-will-last-beyond-pandemic> (accessed on 10 January 2022).
36. Agryzkov, T.; Marti, P.; Nolasco-Cirugeda, A.; Serrano-Estrada, L.; Tortosa, L.; Vicent, J.F. Analysing successful public spaces in an urban street network using data from the social networks Foursquare and Twitter. *Appl. Netw. Sci.* **2016**, *1*, 1–15. [CrossRef] [PubMed]
37. Nieto Ferreira, R.; Mohtadi, T. People-Centered Engagement, Even If Not In-Person, 6 May 2020. Available online: <https://gehlpeople.com/blog/people-centered-engagement-even-if-not-in-person/> (accessed on 10 January 2022).
38. Stewart, B. Twitter as Method: Using Twitter as a Tool to Conduct Research. In *The SAGE Handbook of Social Media Research Methods*; SAGE Publications: London, UK, 2016; pp. 251–265.
39. Ahmed, W. Using Twitter as a Data Source: An Overview of Social Media Research Tools (2019), 18 June 2019. Available online: <https://blogs.lse.ac.uk/impactofsocialsciences/2019/06/18/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-2019/> (accessed on 22 December 2021).
40. Temeljotov Salaj, A.; Gohari, S.; Senior, C.; Xue, Y.; Lindkvist, C. An interactive tool for citizens' involvement in the sustainable regeneration. *Facilities* **2020**, *38*, 859–870. [CrossRef]
41. Kaparias, I.; Bell, M.; Gosnall, E.; Abdul-Hamid, D.; Dowling, M.; Hemnani, I.; Mount, B. Assessing the pedestrian experience in public spaces. In Proceedings of the 91st Annual Meeting of the Transportation Research Board, Washington, DC, USA, 22–26 January 2012.
42. Project for Public Spaces. What Is Placemaking? 2007. Available online: <https://www.pps.org/article/what-is-placemaking> (accessed on 10 January 2022).
43. Clifton, K.J.; Smith, A.D.L.; Rodriguez, D. The development and testing of an audit for the pedestrian environment. *Landsc. Urban Plan.* **2007**, *80*, 95–110. [CrossRef]
44. Allen, D. PERS v2: Auditing public spaces and interchange spaces. Walk21-VI. In Proceedings of the 6th International Conference on Walking in the 21st Century, Zurich, Switzerland, 22–23 September 2005.
45. Vukmirovic, M.; Djukic, A.; Antonic, B. *Place Networks. Experience the City on Foot*; University of Belgrade—Faculty of Architecture: Belgrade, Serbia, 2018.