



Quantifying the relationship between public sentiment and urban environment in Barcelona

Liya Yang^{a,b}, Carlos Marmolejo Duarte^{b,*}, Pablo Martí Ciriquián^c

^a *Guangdong Guodi Planning Science Technology Co., Ltd., Guangzhou, 510650, China*

^b *Department of Architectural Technology, Polytechnic University of Catalonia, Av. Diagonal, 649, 08028 Barcelona, Spain*

^c *Department of Building Sciences and Urbanism, University of Alicante, Carretera San Vicente del Raspeig s/n, 03690 San Vicente del Raspeig, Alicante, Spain*

ARTICLE INFO

Keywords:

Public sentiment
Twitter
Urban environment
Sentiment analysis

ABSTRACT

Public sentiment provides an important social reference for urban management and planning. The relationship between public sentiment and a single type of land use has yielded stable results in previous studies. Hitherto, there has been relatively little research on the correlation of the entire urban environment with public sentiment. Based on the unit of statistical area in Barcelona city, this research uses Twitter sentiment to represent public sentiment and develops a regression model for understanding the interrelationship of four layers: sociodemographic, built-environment, human mobility and socioeconomic activities. The result shows that: 1) The long-term spatial difference in public sentiment has correlations with the urban environment, though it is not decisive. 2) Regardless of disruptive events that are directly associated with public sentiments, the wealthier areas show a more positive correlation with higher public sentiment. 3) The distribution of sentiment tweets (non-neutral) has a close relationship with places where there is a high flow of human activities. This study contributes to the systematic literature of urban applications of sentiment analysis with new empirical observations and a transferable methodology.

1. Introduction

As Lewis Mumford pointed out, “the city... is the point of maximum concentration for the power and culture of a community” (Mumford, 1970). In this sense, a city is the product of an intricate relationship between the environment and human activities. Such complex relationships could be presented as routes of people’s mobility or spatial distribution of land uses. They could also be measured by the public sentiment and perception of the urban environment, which is constantly transformed and disturbed by human activities (Sénécal, 2007). Public sentiment refers to a combination of people’s feelings and perceptions, that is, “an attitude that is based on their thoughts and feelings”, according to the Collins English dictionary. It is “often the product of repeated place interactions and experience” (Dunlap et al., 2013). Therefore, the importance of understanding the relationship between public sentiment and the urban environment lies in the fact that it can reveal dynamic social contexts from a citizen perspective (Williams & Dunn, 2003), which can provide insights into which specific settings of the urban environment could promote people’s happiness and well-being. Furthermore, this understanding will be beneficial for future

urban governance and planning.

Location-based social network (LBSN) data, such as that of Twitter or Facebook, can overcome the limitation and provide a tangible vision to present “invisible” public sentiments in nearly real-time. The common feeling can be observed and aggregated through texts, emoticons and specific behaviour (e.g. by giving a “like” or forwarding). Therefore, sentiment analysis via location-based social network (LBSN) data has been a popular topic in urban studies on subjects such as work stress (Wang, Hernandez, et al., 2016), the sentiment of railway passengers (Collins et al., 2013) and mapping sentiment (Li et al., 2016).

Nevertheless, the relationship between public sentiment and the urban environment has not yet been fully investigated. Firstly, on a large geographical scale, several scholars argued that the density and size of cities has negatively contributed to people’s happiness (Morrison, 2011; Okulicz-Kozaryn & Mazelis, 2018). However, the variations in spatial difference of public sentiment within a city have not been studied. Secondly, for a single type of land use or urban factors, previous studies have yielded some globally compatible results, such as tourist attractions (Park et al., 2018) or green spaces (Chapman et al., 2018; Schwartz et al., 2019). Sociodemographic conditions (e.g. income, gender and

* Corresponding author.

E-mail addresses: li.ya.yang@upc.edu (L. Yang), carlos.marmolejo@upc.edu (C. Marmolejo Duarte), pablo.marti@ua.es (P. Martí Ciriquián).

<https://doi.org/10.1016/j.cities.2022.103977>

Received 19 December 2020; Received in revised form 30 June 2022; Accepted 6 September 2022

Available online 18 September 2022

0264-2751/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

unemployment) are also positively correlated with public sentiment (Ballas, 2013; Mitchell et al., 2013). However, few studies integrate all these environmental factors into a whole setting to evaluate the influence of the spatial difference in public sentiment. It is still not known to what degree the urban environment could influence public sentiment and which layer of the urban environment has a greater influence on public sentiment. In addition, most research has been limited to English texts or a single language. In fact, immigrants and visitors usually hold an important portion in international metropolises. An analysis based on a single language is not sufficient to reveal perceptions about the same city of people who use other languages.

Therefore, this research aims to offer an empirical quantitative study to provide some insights into the relationship between the urban environment and the variation in public sentiment at city level. It utilises thirty months of Spanish and English tweet data to quantify and classify the sentiment orientation of tweets, that is, positive, negative and neutral. Public sentiment is measured by two aspects: the spatial distribution of sentiment tweets (that is, tweets with positive and negative sentiment) and the spatial variation of sentiment scores. Based on the division of the basic statistical area (AEB) of Barcelona, the urban environment is conceived as a polymer with four layers: sociodemographic, built environment, human mobility and socioeconomic activities. A multivariate regression model is used to explore the variation in Twitter sentiment across AEBs with different urban environment characteristics. In addition, we analyse the temporal variation in Twitter sentiment to complete the research.

In summary, this research has unequivocal implications as it analyses the interaction between public sentiment and the urban environment and enriches urban governance.

- 1) From a methodological perspective, our model systematically inspects the spatial patterns of public sentiment at city level. We also measure the sentiment score and the density of sentiment tweets separately. The study indicates that the intensity of Twitter activity is not inevitably associated with a variation in public sentiment score, which has been overlooked by previous studies. In addition, our method offers an affordable approach to classify sentiments from multi-language texts. It would be beneficial to expand the scale of international samples in future urban research.
- 2) Theoretically, our results also contribute to empirical observations of emotional geography, which inspect the public sentiment in a combined urban setting. At city level, it suggests that the influence of socioeconomic indicators is stronger than other factors. The spatial variation in public sentiment score could be a long-period representation of socioeconomic condition, that is, wealthier areas show a more positive correlation with higher public sentiment. By contrast, the influence of the built environment on public sentiment is sensitive to the geographical scale of the study. Lastly, although the occurrence of disruptive events can disturb public sentiment immediately, it does not offset the long-term spatial distribution and orientation of public sentiment in Barcelona.
- 3) By extracting the key information that Twitter sentiment conveys, a further contribution of this paper is that it develops several scenarios of universal urban applications for policymakers and urban designers.

The remainder of the paper is organised as follows. Section 2 reviews three cornerstones of the research: a) the empirical studies and theoretical explanations of the relationship between public sentiment and urban environment, b) a critical review of Twitter sentiment analysis and c) a structured review of urban applications of sentiment analysis. Section 3 presents the research framework and calculation details. Section 4 reveals how public sentiment varies across the urban space. The theoretical and practical implications are discussed in Section 5.

2. Literature review

2.1. The relationship between public sentiment and urban environment

Studies of the relationship between urban environments and public sentiment could date back to the 1960s. This is a research field that has had a multidisciplinary character from the start (Craik, 1973). From the perspective of urban studies, Lynch (1960) introduced the concept of “mental map” to represent people’s perceptions of the surrounding built environment. The subjective perception of locations and connections between places can be considered a kind of neutral sentiment. Based on this framework, Kuipers (1978) drew up a theoretical model to state a person’s cognitive map (i.e. how people store their spatial surroundings in their mind). Such cognitive-centred theories only focus on the ability of human beings. They were challenged by interactive theory in the 1990s, which considered that space is a material that the body engages in and works with (Lupton, 1998), rather than something with a mere objective existence. Therefore, “emotions can be conceptualised as the felt and sensed reaction that arise in the midst of the (inter) corporal exchange between self and world” (Davidson et al., 2012). In simple words, the environment also influences the emotions of human beings, and then such emotions can be conveyed by sentiment expressions. Since the 1980s, collective evidence from emotional geography and environmental psychology has shown the interrelationship between public sentiment and the environment. For example, people are happier in the natural environment (MacKerron & Mourato, 2013). Likewise, different environmental conditions influence people’s perception of public policies. For example, people who live in an area where resources are extracted are more likely to support environmental protection (Blake & Behavior, 2001).

Numerous quantitative results regarding such relationships were generated until the appearance of social media data (Section 2.2). These studies yielded some solid results relating to specific urban places and the corresponding public sentiments. Individuals feel happier in scenic places in urban areas (Seresinhe et al., 2019). Larger and greener parks delivered more happiness to people according to case studies in San Francisco (Schwartz et al., 2019), Massachusetts (Cao et al., 2018) and New York (Bertrand et al., 2013; Plunz, Zhou, Vintimilla, Mckeown, Yu, Ugucioni, & Sutto, 2019). Gallegos et al. (2016) concluded that places with more amenities, such as restaurants, gyms and beaches, tended to be happier than other places in Los Angeles. Conversely, transportation hubs and industrial manufacturing places tended to show negative sentiment (Bertrand et al., 2013; Cao et al., 2018). In addition, subjective well-being is negatively correlated with spatial distance to leisure places, such as the beach or recreational places, which decrease as the distance to the place increases (Brereton et al., 2008). In conclusion, leveraging large datasets, specific urban places have a clear impact on life satisfaction and positive sentiment, though the impact may not be decisive. However, it is not known how a combined urban environmental setting can influence public sentiment in the long-term. For example, as Mouratidis (2019) stated regarding urban density, there is a contrast between subjective well-being in the city centre and in the suburban area. However, they failed to provide a complete analysis to explain how this contrast occurred.

Indeed, at large geographic scale, some scholars have found that higher neighbourhood environmental quality, in aspects such as safety (Hu et al., 2019), cleanness (Mouratidis, 2019) and the degree of accessibility to urban facilities (Kyttä et al., 2016), tends to have positive influences on the subject’s well-being and happiness. Several studies have observed that socioeconomic condition has a statistically stronger positive impact on happiness at individual (Wang, Wang, & Society, 2016), community (Quercia et al., 2012) and city level (Mitchell et al., 2013) across countries. Despite these achievements, few studies analyse the relationship between human sentiment and environment under the perspective of an integrated urban environment. Hu et al. (2019) studied people’s perceptions of urban neighbourhoods using online reviews and

eight indicators of the urban environment. However, all these environmental indicators were analysed separately without a comprehensive explanation at city level. Hajrasouli et al. (2018) develop a four-dimensional theoretical model to summarise the possible environmental elements that impact human mental health: objective environment, perceived environment and physical and social environment. However, a quantitative model that could decide which urban factors are most influential during the accretion of public sentiment at city level is still pending.

To address the research gap, the aim of this paper was to put urban indicators of socioeconomic, built environment, and human mobility and socioeconomic activities into one model to discuss the relationship between the urban environment and sentiments embedded in tweets. It was expected to provide a combined perspective to examine the spatiotemporal variance in public sentiment quantitatively and qualitatively.

2.2. Twitter sentiment analysis

2.2.1. The representativeness of Twitter

The representativeness of Twitter data has been under debate for a long time. Demographic characteristics, such as age, gender and race, and the use of social media may distort the result of investigations and affect their representativeness (Murthy et al., 2016). However, “geo-located Twitter data still yields generalizable results when studies are restricted to populated metropolitan areas with a high percentage of smartphone users” (Plunz, Zhou, Vintimilla, Mckeown, Yu, Uguccione, & Sutto, 2019). Lenormand et al. (2014) compared mobility patterns based on three sources: Twitter, census and cell phone data. They concluded that the results obtained with the three data sources were comparable, despite the representativeness of Twitter being lower than that of the other sources. Twitter is a useful source to unveil urban dynamics in an aggregated way rather than at individual level (Luo et al., 2019). In summary, using Twitter to investigate urban dynamics in a metropolis has the potential to generate reliable results when the dataset is large enough.

Undeniably, Twitter sentiment analysis faces more criticisms regarding accuracy, representativeness and privacy (Murthy et al., 2016; Wily, 2014). However, the accumulation of studies and applications will lead to some solid conclusions, as summarised in the above section. Several urban domains have been able to reach high precision, such as the prediction of influenza using Twitter (Culotta, 2014; Paul & Dredze, 2011; Sinnenberg et al., 2017) and disaster management (Horita et al., 2013; Neppalli et al., 2017).

2.2.2. Twitter sentiment analysis

Traditionally, personal questionnaires have been used to understand the relationship between places and people’s feelings, such as attitudes toward green areas (Balram et al., 2005) or waste disposal facilities (Bacot et al., 1994). Such a method is usually limited to the number of sampling units and a certain spatiotemporal range. As a widely used micro blog platform, the analysis of Twitter sentiments can provide enormous and valuable information on social relationships and daily life, which were barely detected by traditional statistical data. This information has been applied to the prediction of stock market movements (Pagolu et al., 2016), political elections (Paul et al., 2017; Yaqub et al., 2020), disaster management (Neppalli et al., 2017) and disease surveillance (Sinnenberg et al., 2017). The essence of sentiment analysis is to investigate the rationale of the expression of sentiments in texts and estimate the sentiment orientation (that is, positive, neutral or negative) toward the subject of the text (Nasukawa & Yi, 2003). Therefore, the framework of sentiment analysis has two main parts: the extraction of emotional expressions and the algorithm that quantifies the sentiment score of texts based on certain rules.

Current, corpus-based and lexicon-based approaches are two major methods used to conduct sentiment analysis. Corpus-based sentiment

analysis uses specific corpora as the training data to extract the feature of sentiments and classifies the polarity of an input text (Li et al., 2016). The feature of a sentence, for example, can be defined as the frequency of positive and negative terms. It measures the similarity of the entire sentence rather than single words (Candelieri & Archetti, 2015). The corpus-based approach can perform well for specialised domains, such as medical articles or political speech, because the source of the corpus points to that domain. However, it lacks sufficient evaluation of the quality of the algorithm and requires model training for each domain.

The lexicon-based method (Mitchell et al., 2013; Thelwall et al., 2010) contains a list of words that are predefined in terms of sentiment polarity and sentiment strength. It requires high quality of lexical resources for good performance. Based on dictionaries, the machine sentiment classifier searches for the matched emotion expression in texts and assigns sentiment scores to them. For instance, Gilbert and Hutto (2014) established a lexicon using three sentiment lexicons, Linguistic Inquiry and Word Count (LIWC), Affective Norms for English Words (ANEW) and General Inquirer (GI), and supplemented the lexicon with commonly used expressions in social media data, such as acronyms, slang and emoticons. They applied heavy human inspection in the production of the lexicon to improve the quality. Compared with the corpus-based method, the algorithm of the lexicon-based method is easier to understand and adopt.

Despite these advantages, the shortfalls are similar to the corpus-based method. It is difficult to evaluate which method is better as semantic data are complex and heterogeneous, although all approaches claim that they are better than the others. In addition, with the increase of commercial potential in sentiment analysis, little software with excellent performance is free to the public (Kumar & Jaiswal, 2020). Therefore, to increase the accuracy and balance the cost, this study tries to utilise the intersection of two lexicon-based algorithms to measure the sentiment score from Twitter (Section 3.4).

2.3. Urban applications of sentiment analysis in future urban planning

Many studies have investigated the role of semantic analysis in enhancing future urban planning. However, single analysis studies can hardly be connected to a systematic view of the potential of sentiment analysis. Therefore, we propose a structured summary (Fig. 1) of the potential urban applications of sentiment analysis, which could be beneficial to connect the understanding of public sentiment with urban practices.

The unique contribution of sentiment analysis to urban planning and policies is twofold: 1) it serves as valuable social-spatial reference data to monitor and investigate various human-environment relationships in the urban space. The “communal matters of concern” (Nold, 2009) expressed by public sentiment could support better decision-making of urban policies and deal with urban issues. 2) The indigenous knowledge provided by Twitter and other social network data is beneficial for disclosing the local daily life and the embedded social and cultural contexts that were usually overlooked by authorised official data. For example, the spatiotemporal distribution of work stress in a long period (Wang, Hernandez, et al., 2016) and real-time public experiences during events (Balduini et al., 2013) could be monitored through tweets’ texts directly. Therefore, it could be a useful vehicle for the advanced analytic framework in future urban planning.

Specifically, sentiment analysis can make definite contributions to theoretical and practical aspects of urban planning and management. Firstly, regarding the theoretic framework of urban management, sentiment analysis can add a novel social layer to the governance of smart cities, which shift from a single government-centric paradigm to a data-driven, multi-stakeholder mode (Meijer & Bolívar, 2016). A smart city contains a collection of sentient infrastructure based on various information and communication technologies (Kandt & Batty, 2021) and a system of urban governance in which governments, organisations and citizens participate (Gao et al., 2020). Such an intersection of

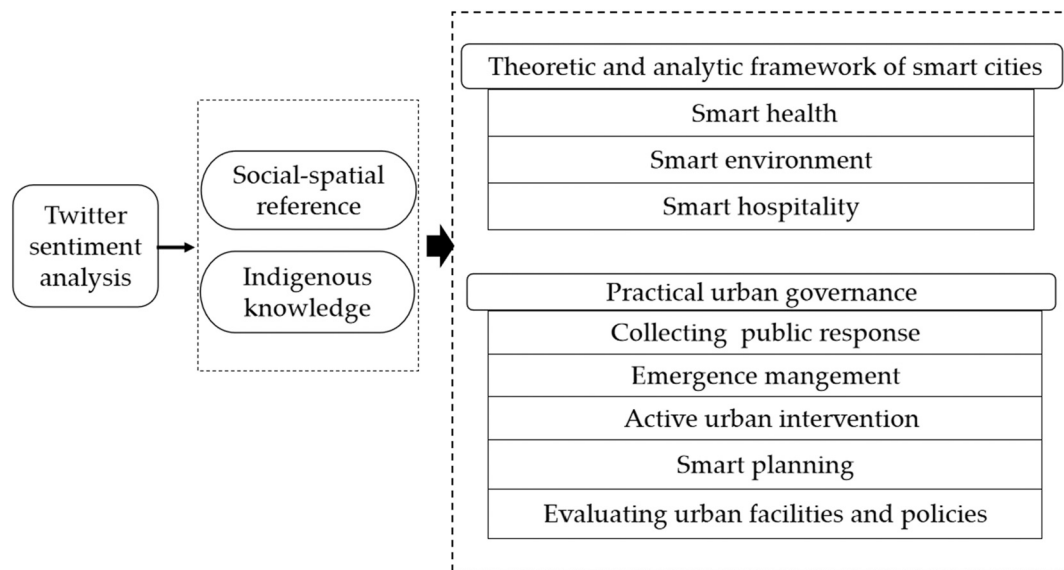


Fig. 1. Potential applications of Twitter sentiment analysis.

diverse elements (Lim et al., 2018) indicates that the governance of a smart city looks like a structure of dynamic collaborations rather than regulation (Cattacin & Zimmer, 2016). During the process, public sentiment, as a proxy of public well-being (Li et al., 2016), could be involved in many smart applications (Lim et al., 2018), such as a smart environment, smart health and smart hospitality, among others.

Secondly, at practical level, it can be an active instrument for urban governance in monitoring public response, evaluating urban facilities and well-being and promoting smart planning. Twitter is a comparable tool to monitor the public response during disasters (Crooks et al., 2013) and manage social events intentionally (Hubert et al., 2018). It is also an inexpensive method for checking public attitudes toward a specific policy or urban intervention (Sutoyo and Almaarif, 2020). For example, Marquez et al. (2019) research showed that inter-group communication between refugees and non-refugees was helpful to alleviate online negative sentiment toward refugees, as the probability of communication was positively correlated with public sentiment. Nik-Bakht and El-Diraby (2016) designed an online social media game to detect people's perspectives on the sustainability of urban infrastructure. The mission of players is to annotate infrastructure-related tweets using a set of indicators of sustainability. The attitude of sustainability is produced spontaneously by players when they play the game.

Furthermore, sentiment analysis can be applied to the assessment of the urban environment. For example, Hollander and Shen (2017) detected people's attitudes toward cycling services in Washington DC using related tweets. Collins et al. (2013) utilised Twitter to study the sentiment of passengers of suburban trains near Chicago city. They found that dissatisfaction with incidents can be detected by the variation in Twitter sentiment during a 24-hour timeline. LBSN sentiment analysis is also valuable for smart planning. Quercia et al. (2014) proposed more emotional-route options (beautiful, quiet and happy) for users rather than the shortest routes on map services, according to the user's online opinions of places.

To sum up, sentiment analysis has shown its potential in urban applications, which can stimulate citizens' participation in urban governance actively or passively. In this sense, our study, as a city-level observation, can contribute to comprehensive understanding of a complete urban environment (Section 4), and thus it could provide insight into the design of master plans and the analytic framework of a city (Section 5).

3. Methods and materials

3.1. Research design

A 30-month Twitter dataset was utilised to calculate public sentiment and several open/official datasets were used to construct indicators of the urban environment in Barcelona (Fig. 2). The research design followed three main steps:

- (1) The Twitter dataset was processed to reduce data noise. English and Spanish Tweets were extracted for further sentiment analysis. The representativeness of the dataset in Barcelona is also inspected in Section 3.2.
- (2) A hybrid method was introduced to quantify public sentiment using two algorithms (Section 3.4), whose effectiveness was validated by a sampling test (Appendix B). Furthermore, we used the density of sentiment tweets and the sentiment score to represent the variation in Twitter sentiment.
- (3) Combined with the urban environment indicators extracted from open point of interests (POIs) data, the official transport survey, and census data (Section 3.5), regression analysis was conducted to explore the relationship between Twitter sentiment and the urban environment.

3.2. Data collection and processing

The scope of the study was restricted to Barcelona, Spain. Barcelona is a typically compact city whose area is 101.9 km². It supports 1.6 million residents. The Twitter dataset that we acquired from Twitter Streaming API between September 2016 and April 2019, contained 1,100,244 tweets in total (data source: Martí et al. (2019)). Retweets were not included. The speed of retrieving data was 450 requests every 15 min. The returned data was a representative sample within Barcelona from the Twitter server.

To reduce non-active users and improve the data quality, repeated tweets and users who only appeared once in the studied area were excluded from the dataset, which accounted for 44.82 % of all 123,437 Twitter users during the period. Meanwhile, non-individual accounts and their tweets, such as weather information, companies, websites, etc. were also removed (see Appendix A: Table A.1). Most of the high-frequency users were commercial and information accounts and were excluded from the dataset. The cleaned Twitter dataset contained

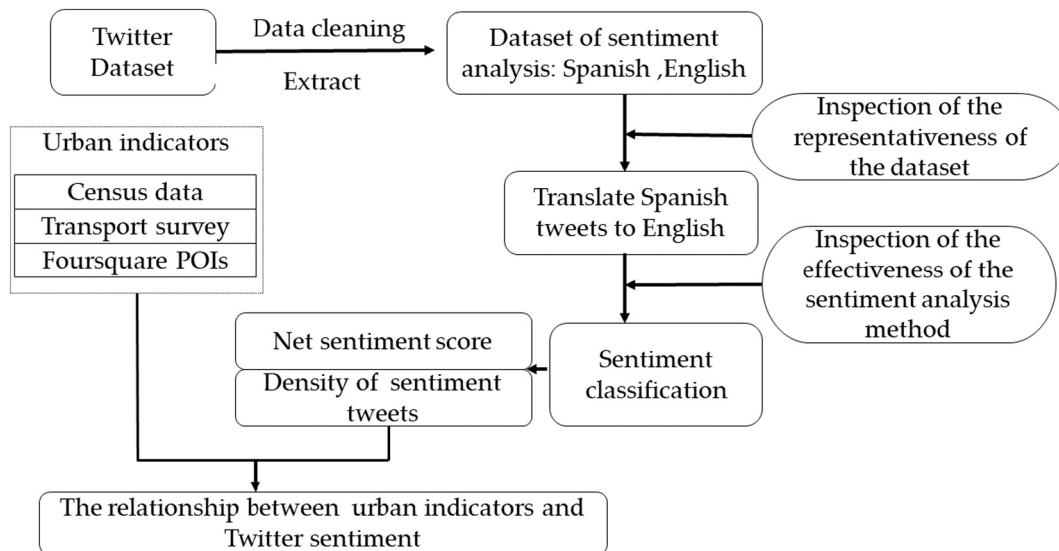


Fig. 2. Research framework.

707,549 tweets that were generated by 63,178 users. We manually checked the remaining users who had 1000 tweets, accounting for 3.5 % of all tweets. No dense publishing behaviour in a short period was found.

3.3. Representativeness of the analysed Twitter dataset

According to the automatic language detection of Twitter, Spanish, English and Catalan tweets accounted for nearly 80 % of the cleaned dataset. The high percentage of English tweets was caused by the huge volume of visitors. As a world-famous tourist attraction, according to the *Annual Tourism Sector of Barcelona Report 2017*, the total number of overnight tourists who stayed in hotel accommodation was nearly nine million. Catalan and Spanish are the official languages of Catalonia. The initial proposal was to analyse Twitter sentiments in the three languages. However, a human sampling inspection (under 95 % confidence level and 4.6 confidence interval) showed that over 50 % of 453 samples of Catalan tweets did not contain any valid text except geolocation information. This was largely caused by geolocation that refers to Catalan local place names. Firstly, half of these tweets were generated by the automatic geolocation of posted pictures or users' "check-ins". Secondly, text cleaning removed all hashtags, links, unrecognised characters and symbols. Therefore, tweets that only contained this type of content

would be cleaned except for the geo-information. Since Twitter's algorithm of language detection is not publicly available, we could only estimate that the machine detected them as Catalan tweets because the algorithm considered that hashtags and links could be used by all language users.

Therefore, considering the poor quality of Catalan tweets, the sentiment analysis only considered English and Spanish tweets. This should not greatly affect the effectiveness of the research. According to the investigation from the Institute of Statistics of Catalonia in 2018, 97.5 % of people in Catalonia understood and used Spanish, and 76.1 % wrote mobile messages in Spanish or in a combination of Spanish and other languages. Only 11.7 % of people only used Catalan to write mobile messages.

The final dataset of Spanish and English tweets contained 429,271 tweets in total. Fig. 3 shows the weekly distribution of the original cleaned dataset, which implies that the temporal distribution was not distorted by the cleaning process. Data loss occurred six times due to an unexpected technical problem. The overall decreasing trend of geo-tagged tweets is a global phenomenon mainly because of the change in the default geotag mode in Twitter and increasing privacy concerns (Tasse et al., 2017).

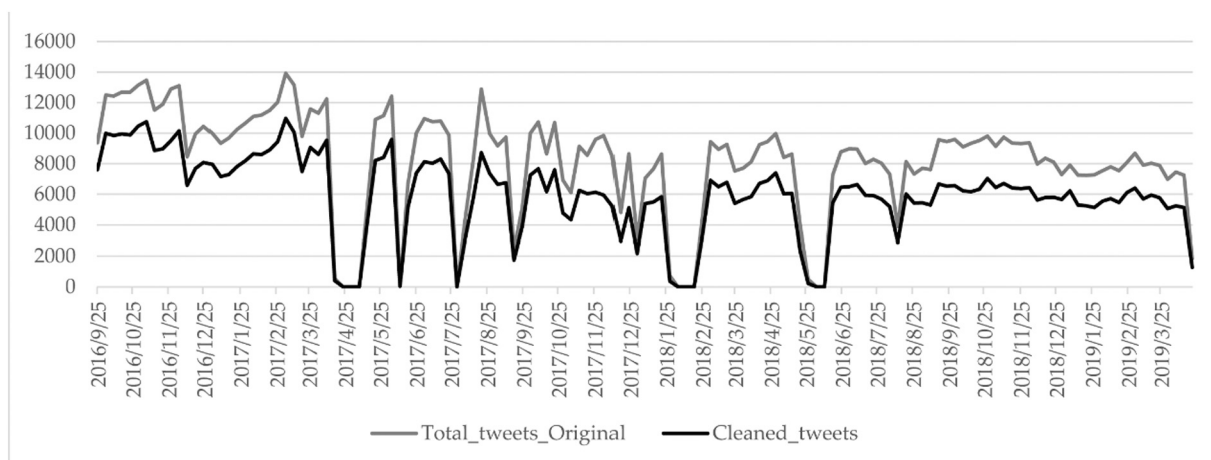


Fig. 3. Comparison of the temporal distribution of datasets. Source: own elaboration.

3.4. Sentiment analysis

The process of sentiment analysis includes two steps. Firstly, we translated Spanish tweets into English via Google Translate, due to a consideration of the quality and consistency of sentiment evaluation. To our knowledge, free, open tools of Spanish sentiment analysis have seldom been widely tested and used. Conversely, the availability of translated texts has been proved by [Balahur and Turchi \(2012\)](#). They concluded that translated Spanish texts could provide comparable results for sentiment analysis. Moreover, since few free open programs support sentiment analysis of multi-languages, the consistency of rating results is hard to guarantee if we adopt different tools in different languages.

Secondly, the sentiment analysis was conducted by two widely used lexicon-based programs ([Al-Shabi, 2020](#); [Bonta and Janardhan, 2019](#)): Vader that specifically focuses on social media texts ([Gilbert & Hutto, 2014](#)) and SentiStrength that focuses on the estimation of sentiments in short informal texts ([Thelwall et al., 2010](#)). Both programs can attain similarly high accuracy in Twitter texts ([Al-Shabi, 2020](#); [Ribeiro et al., 2016](#)) and informal online comments ([Bonta and Janardhan, 2019](#)).

The intersection of the results of both algorithms is the final classification, that is, a tweet is confirmed as positive only if it is in the positive category of Vader and SentiStrength. The classification of neutral and negative tweets follows the same rule. In addition, we extracted a sample dataset and manually inspected the effectiveness of our method and whether the translation greatly affected the sentiment detection (see the details in [Appendix B](#)).

Regarding the result of sentiment classification, Vader rated the sentiment of words based on sentiment orientation and the intensity of the emotion. The normalised compound score of a sentence, on a scale from -1 (maximum negative) to 1 (maximum positive), is the sum of scores of each word. Similarly, SentiStrength uses an integral range from -5 to 5 to indicate the sentiment orientation except for 0 . The 1 and -1 represent “no positive emotion” and “no negative emotion” separately. The binary sentence score is given by the maximum value of the positive score and the negative score. The sentiment orientation of a sentence is the sum of the two scores. [Table 1](#) lists thresholds of sentiment classification of both algorithms for a given sentence.

[Table 2](#) shows the result of classification of the entire English and Spanish dataset. The rate of agreement of the two algorithms is over 70% in both languages. The percentage of neutral tweets accounts for the largest proportion in both Spanish and English tweets; negative tweets are $<5\%$. The similar proportion of negative tweets could be observed in the sentiment classification of tourist destinations in Chicago ([Padilla et al., 2018](#)). Considering the pivotal role of tourism in Barcelona, this could imply that the percentage of negative tweets is lower in tourist places and cities. Moreover, Barcelona city is the core city of Barcelona Metropolitan Area, whose average income is higher than neighbouring municipalities. As previous studies suggested, this may also influence the low proportion of negative tweets that we have found.

3.5. Evaluation of the relationship between urban indicators and Twitter sentiment

Initially, the ideal model was to evaluate the influence of urban indicators on Twitter sentiments directly. However, we found that a logit

Table 1
Thresholds of sentiment classification.

Sentiment	Vader	SentiStrength
Positive	Compound score ≥ 0.05	Senti_sum > 0
Neutral	$-0.05 < \text{Compound score} < 0.05$	Senti_sum = 0
Negative	Compound score ≤ -0.05	Senti_sum < 0

Source: own elaboration.

Table 2
Results of sentiment classification.

Category of sentiment	English tweets	Percentage	Spanish tweets	Percentage
Positive	47,307	22.96 %	68,356	30.62 %
Negative	6756	3.28 %	11,039	4.94 %
Neutral	108,160	52.51 %	83,923	37.59 %
Total agreement tweets	162,223	78.75 %	163,318	73.15 %
Total tweets	205,997	100.00 %	223,274	100.00 %

Source: own elaboration.

model based on a single tweet level failed to generate reliable outcomes. The neuron network could measure the correlations between two variables by comparing the performance of the classification. However, it could not quantify which indicator is more relevant to the dependent variable and its calculation process is unexplainable ([Abiodun et al., 2018](#)). Therefore, we aggregated Twitter data using the unit of Basic Statistical Area (AEB) defined by Barcelona’s government and measured the relationship by the most commonly used multiple regression. The AEB is a territorial unit for statistical purposes in Barcelona city that has a homogeneous socioeconomic population within it, which divides Barcelona into 233 sectors whose average area is 0.44 km^2 ([Fig. 4](#)). To guarantee the representativeness, AEBs that were introduced into the analysis of the relationship should have both positive tweets and negative tweets in both languages. Following the criterion, 200 AEBs were entered in the analysis. Thirty-three excluded AEBs were mainly located in the peripheral areas of Barcelona.

Furthermore, as the goal was to explore where sentiment tweets would be more distributed, the studied tweets were only restricted to positive tweets and negative tweets because neutral tweets do not perform sentiment orientation. Moreover, the involvement of neutral tweets would twist the average sentiment score to close to the percentage of the neutral tweets of each AEB. Neutral tweets accounted for over 30% of the entire dataset. However, the score range of neutral tweets was between -0.05 to 0.05 . Therefore, the variation in the average sentiment score would be similar to the inverse of the percentage of neutral tweets (see [Appendix A Fig. A.1](#)).

Public sentiment was evaluated by two aspects: the density of sentiment tweets and the net sentiment score. The former measures the spatial distribution of positive and negative tweets (i.e., the sentiment tweets). The latter describes the net variation in Twitter sentiment among AEBs. Considering that tweets could be generated at various places by residents and visitors, the density of sentiment tweets is given by the average density of all positive and negative tweets on each AEB, that is:

$$\text{Density of sentiment tweets} = \frac{\text{Sum of positive and negative tweets}_i}{\text{Area}_i} \quad (1)$$

The logarithmic form is introduced to satisfy the requirement of normal distribution of the dependent variable in the regression model. Based on the intersecting result of sentiment classification, the sentiment score of a tweet is solely assigned by the Vader sentiment score because the score of SentiStrength does not produce a continuous scale ([Thelwall et al., 2010](#)). Considering the extremely uneven population of positive and negative tweets, the net sentiment score of each AEB is calculated by the average score of positive tweets and negative tweets:

$$\text{Net Sentiment}_i = \frac{\sum \text{Vader score}_{\text{posneg}}}{\sum \text{Tweets}_{\text{posneg}}} \quad (2)$$

[Fig. 4\(a\)](#) shows that sentiment tweets are mainly concentrated in the city centre that contains most of the famous tourist attractions, such as La Sagrada Família, the Ciutat Vella (the medieval neighbourhood), Passeig de Gràcia avenue and La Rambla street. Both of these streets are major commercial and tourist avenues in Barcelona and contain some of the most celebrated architectural works. Compared with the density, the

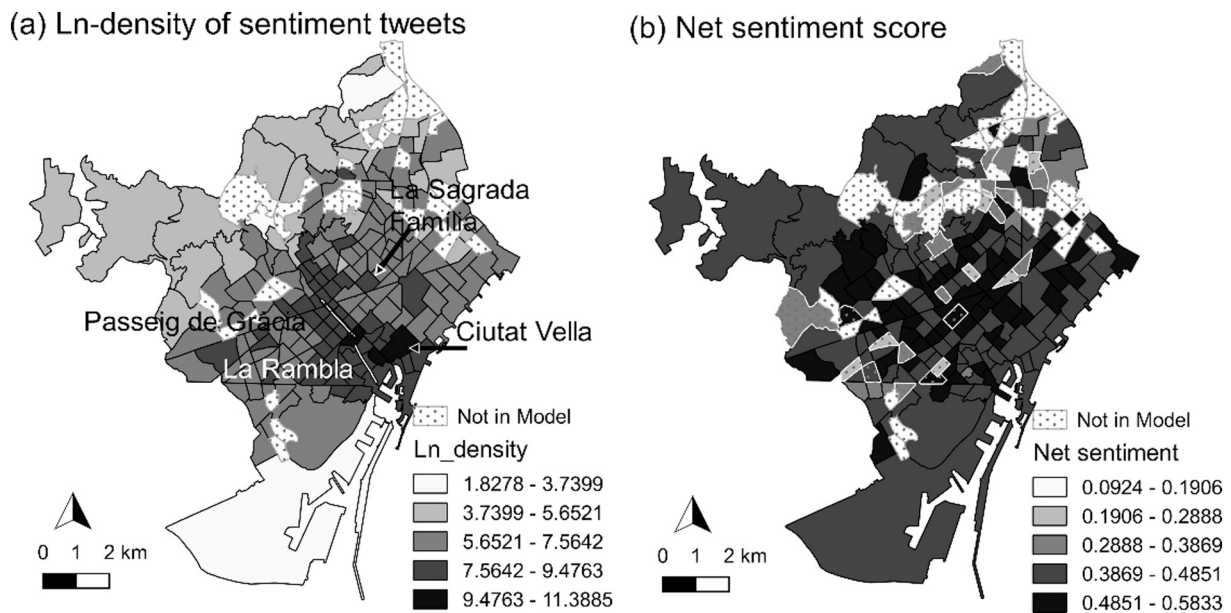


Fig. 4. Spatial distribution of the density of sentiment tweets and net sentiment. Source: own elaboration.

spatial variation of sentiment score (Fig. 4(b)) is dispersed, though the central area performs higher positive scores in general.

The urban environment indicators consist of four layers: socio-demographic, built environment, and human mobility and socioeconomic activities (Fig. 5). Sociodemographic variables could reflect the wealth of the AEB indirectly. The urban built environment contains specific land uses and residents' perceptions of surrounding environmental quality. The most visited places are also introduced into the model as categorical variables. Human mobility and socioeconomic activities are represented by two data sources. The first is the citizens' activities and socioeconomic indicators that were built by Marmolejo and Cerda Troncoso (2017) and Marmolejo-Duarte and Cerda-Troncoso (2020) using the origin-destination mobility data from the Metropolitan Transport Authority survey. The time density is the number of hours that citizens expended in a given transport zone. The diversity is computed considering the activities that citizens perform out of their home (e.g., working, shopping, visiting friends), and the socioeconomic diversity of the people performing the aforementioned activities. Additionally, the larger centrality index indicates the larger time density and diversities. Secondly, Foursquare POIs are exploited to indicate more detailed spatial information and places that people tended to use, whose categories of classification are based on Yang and Durarte (2019).

All the indicators are involved in the OLS regression model to predict the relationship between the urban environment and public sentiment:

$$LnDensity_i = \beta_0 + \gamma_i * S_i + \delta_i * B_i + \mu_i * H_i + \epsilon_i \tag{3}$$

where *i* is an AEB. β_0 is the regression constant; S_i , B_i , H_i represent the variables of socioeconomic, built environment and human activities respectively; γ , δ , μ are gradients associated with these three groups of variables separately; ϵ_i is the error term with the usual properties. Correspondingly, we use the same explanatory variables to estimate the variation of the net sentiment:

$$Net\ Sentiment_i = \beta_0 + \gamma_i * S_i + \delta_i * B_i + \mu_i * H_i + \epsilon_i \tag{4}$$

4. Results

4.1. Correlation between urban environment and Twitter sentiment

Fig. 6 visualises all paired Pearson's correlation values of indicators

that do not filter out statistically insignificant variables (*p*-value < 0.05). The first important observation is the weak correlation between the density of sentiment tweets and the sentiment score. This indicates that the intensity of Twitter activities does not greatly influence the orientation of public sentiment. Moreover, the density of sentiment tweets is highly correlated with human and socioeconomic activities and less associated with sociodemographic indicators. The total centrality index (Fig. 7(a)) has a higher positive correlation with the density, which indicates that these tweets are concentrated in the central area of Barcelona. Places that people tend to stay in longer (Fig. 7(c)(d)(e)), such as restaurants, outdoor resorts and workplaces, receive more sentiment tweets than transport and residential places. By contrast, the relationship between net sentiment and human activities is less intense, though they are all positively correlated. The variation of sentiment score has the weakest correlation with urban environment indicators.

Regarding sociodemographic indicators, our results not only show a similar wealth and happiness tendency as in previous studies (Mitchell et al., 2013; Quercia et al., 2012), but they also reflect the details of spatial difference between density of tweets and Twitter sentiments. In the case of Barcelona, wealthier AEBs have a higher density of sentiment tweets and positive sentiments. Areas with a higher percentage of high-income positions and highly educated people (Fig. 7(g)(f)) show a higher density of tweets and optimism. Conversely, AEBs with a higher percentage of lower-income positions (Fig. 7(i)), such as artisans, operators and clerks, perform a negative correlation with the density and the sentiment score. Professionals related to catering services (Fig. 7(h)), personal security and salesman have a positive association with density but a negative relationship with positive scores. Remarkably, these people do not live in wealthy areas. The density of the population also (Fig. 7(b)) appears to have a similar relationship.

The built environment indicators have a complex, subtle relationship with public sentiment due to spatial heterogeneity. Historical areas, the density of storefronts and beach areas are positively correlated with the density of sentiment tweets and sentiment scores. The percentage of railway areas is negatively correlated with the two sentiment variables because large empty railway areas only appear in suburban districts (Fig. 8(g)). The percentage of urban parks and gardens appears to be negatively correlated with public sentiment due to the fact that some centrally located AEBs lack such greenery (Fig. 8(h)). Similarly, the negative value of the commercial storefront is caused by its uneven

1. Sociodemographic	
Original statistical scale : Census tract	
<p>Density of population: habitants/km²</p> <p>1) Total population</p> <p>2) Total Spanish population</p>	<p>Education attainment: (University population/ Population above 16 years-old)</p> <p>3) People with university degree(%)</p>
<p>Percentage of job positions: (population of specific occupations/ the total population of occupations)</p> <p>4) Managers in companies and public administrations</p> <p>5) Scientific and intellectual technicians and professionals</p> <p>6) Technicians and support professionals</p> <p>7) administrative employees</p> <p>8) Workers of catering services, personal security, and salesman</p> <p>9) Skilled workers in agriculture and fishing</p> <p>10) Craftsman and skilled workers of the manufacturing</p>	
<p>11) Facility and machinery operators and assemblers</p> <p>12) Unskilled workers</p> <p>Socio-professional index: (principal components extracted from the percentage professional positions)</p> <p>13) High</p> <p>14) Medium</p> <p>15) Low</p>	
Data source: Padron INE. 2017; Census INE. 2001	
2. Built-environment	
Original statistical scale : 1)-5) ,11)-19): AEB; 6)-10): Census tract	
<p>Specific land use :</p> <p>1) Historical area (%)</p> <p>2) Urban parks and garden (%)</p> <p>3) Area of railway system (%)</p> <p>4) Area of water front and beach (%)</p> <p>5) Urban dense area (%)</p> <p>6) Density of storefront (storefronts/km²)</p> <p>7) Commercial equipment accounts for the total storefront (%)</p>	
<p>Percentage of household consider there is a pollution issue: (number of households / all census households)</p> <p>8) Noise</p> <p>9) Dirty</p> <p>10) Contamination</p>	
<p>Whether the AEB contains top visited places (buffered 0.4 km,polygon):</p> <p>11) Casa Milà and Casa Batlló</p> <p>12) La Rambla street</p> <p>13) Passeig de Gràcia avenue</p> <p>14) Camp Nou stadium</p> <p>15) La Sagrada Família temple</p> <p>16) Sants- Montjuïc hill</p> <p>17) Cosmo La Caixa museum</p> <p>18) El Born historical zone</p> <p>19) Park Güell</p>	
Data source: Statistical Department of Barcelona ; Census, INE. 2001 ; own elaboration	
3. Human mobility and socioeconomic activities	
Original statistical scale : 1)-6)Transport zones; 7)-15) AEB	
<p>1) Centrality index of activities</p> <p>2) Diversity (total/working days/ non-working days)</p> <p>3) Diversity of activities (total /working days/ non-working days)</p>	<p>4) Socioeconomic diversity (working days/ non-working days)</p> <p>5) Time density (total/working days/ non-working days)</p> <p>6) Density of people performing activities (total /working days/ non-working)</p>
<p>Density of Foursquare POI (points/km²):</p> <p>7) Bar</p> <p>8) Restaurant</p> <p>9) Outdoor resorts</p> <p>10) Workplace</p> <p>11) Residential place</p> <p>12) Hotel</p> <p>13) Transportation</p>	
Data source: Enquesta de Mobilitat Quotidiana (EMQ 2006) origin–destination household surveys, Autoritat del Transport Metropolità's (ATM) ; Yang & Marmolejo-Duarte(2019)	

Fig. 5. Urban environment indicators.
Source: own elaboration.

Sociodemographic			Built environment			Human mobility and socioeconomic		
	NetSentiment	LndenSentiment		NetSentiment	LndenSentiment		NetSentiment	LndenSentiment
D_population	-.1486*	.1994**	%Historical area	.0305	.4020**	Total Centrality	.2229**	.6165**
D_Spanish people	-.1502*	.1852**	%Park_garden	-.0934	-.0590	Total Diversity	.1876**	.4947**
% High_education	.3267**	.2234**	%Railway	-.0894	-.0214	Total_Diversity_wk	.1876**	.4947**
Pr_manager	.2747**	.1166	%Waterfront_beach	.0565	.0141	Total_Diversity_nw	.2137**	.4784**
Pr_scientific	.3173**	.2502**	%_Urbandense	.0756	.1299	Diversity_activities_wk	.1371	.4753**
Pr_technician	.0413	.1287	D_storefront	.2323**	.5432**	Diversity_activities_nw	.1830**	.4549**
Pr_adiminstrative	-.2626**	-.2203**	%Commercial_st	-.1329	.0999	Diversity_se_wk	.1363	.2081**
Pr_services	-.2025**	.0804	O_dirty	-.0431	.0903	Diverstiy_se_nw	.2177**	.1966**
Pr_agriculture	-.0011	-.0060	O_noise	.2772**	.5361**	Time density_total	.1315	.4475**
Pr_artisan	-.3387**	-.3662**	O_contamination	.2300**	.4218**	Time density_wk	.1577**	.4635**
Pr_operator	-.2849**	-.4216**	Casa_Mila_Batillo	.1519*	.3292**	Time density_nw	.0753	.3970**
Pr_noskill	-.1323	-.2098	LaRambla	.0492	.3339**	D_activepeople_total	.1213	.4784**
Pr_high	.3396**	.4559**	PgGracia	.0679	.3102**	D_activepeople_wk	.1446*	.4861**
Pr_mid	-.2633**	-.1701*	CampNou	-.0442	.0354	D_activepeople_nw	.0802	.4480**
Pr_low	-.2119**	.0142	SagradaFamilia	-.0006	.0542	D_FS_bar	.1521*	.6266**
NetSentiment	1	.2424**	Sants_Monjuic	.0338	-.0341	D_FS_restaurant	.2375**	.6851**
Lndensity_posneg	.2424**	1	El Born	.0652	.3308**	D_FS_transport	.1167	.3335**
D:Density	O: Opinion		Cosmolacaxia	.0726	-.1147	D_FS_outdoor	.1280	.6234**
Pr:Profession	FS:Foursquare		Park Güell	-.2081**	-.0916	D_FS_hotel	.1297	.5410**
						D_FS_residential	.0625	.2434**
						D_FS_workplace	.2089**	.5419**

st: storefront; wd: non-workingdays ; wk:workding days.

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Fig. 6. Pearson correlation between urban indicators and sentiment. Source: own elaboration.

spatial distribution. The percentage of commercial storefronts is lower in the urban centre because the other activities and person-oriented services, such as healthcare and culture services, dominate the central landscape.

Both noise and contamination are positively correlated with public sentiment because these two indicators are mainly concentrated in the city centre where various services, intense (despite noisy) street activities and amenities are located, except for the industrial zone in the southeast of Barcelona (Fig. 8(d), (e)). Thus, these indicators are actually proxies for such active environments. The perception of dirtiness is negatively correlated with positive sentiment because the opinion of dirty areas is greater in the industrial zone of Barcelona and in certain tourist-populated areas (Fig. 8(f)).

The location and characteristics of tourist attractions (Fig. 8(c)) affect their correlation with the two sentiment indicators. La Sagrada Família and Camp Nou are positively associated with density but negatively correlated with the sentiment score. La Sagrada Família is surrounded by a higher percentage of residents who might feel bothered by tourists. The public sentiment in Camp Nou was probably influenced by the results of soccer games. Cosmo Caixa Museum, a famous science museum, and Park Güell are far from the city centre. Therefore, they show a negative relationship with sentiment tweets comparatively and are capacity limited.

4.2. Regression analysis

The model of density of sentiment tweets statistically confirms the interrelationship discussed previously, whose R^2 reaches 0.689 (Table 3). The density of the population and the density of Foursquare POIs are introduced in their logarithmic form for fitting reasons. Spatial autocorrelation has a negligible effect on the model, according to the result of the Moran I test of spatial correlation in Geoda software (see Appendix C). Since urban settings vary in AEBs, the robust model is adopted to solve the problem of heteroscedasticity. For example, the historical area is a positive indicator in the model. However, several AEBs with a higher percentage of historical areas (Fig. 8(b)) contain less

sentiment tweets because they are in the peripheral districts of the city. This is the result of the urban aggregation of Barcelona that absorbed formerly independent towns nearby.

The most influential variable is population density. Functionally, indicators are related to leisure activities and are coherently correlated with a higher density of sentiment tweets, such as storefronts (Fig. 8(a)), urban parks, historical areas and commercial equipment. The density of outdoor resorts of Foursquare POIs is the second most dominant indicator, which includes famous tourist attractions, resting areas and scenic viewing places, among others. The urban parks and gardens become a positive indicator in the model because other variables compensate for the deficiency of parks and gardens in the central area of Barcelona (Fig. 8(h)). From the perspective of a combined urban setting, parks and gardens are attractive places for human activities. In addition, a high socio-professional index as a positive indicator entered in the model. This suggests that high-income AEBs are associated with a higher frequency of use of social networking software.

The model of net sentiment only contains 168 cases due to the OLS requirement of normal distribution. Some extreme positive and negative deviances were removed (Fig. 4(b)). After the test of heteroscedasticity and spatial correlation, the confirmed model (Table 4) shows that the urban environment has a lower impact on the variation of sentiment scores, as the adjusted R^2 is equivalent to 0.272. The regression model only confirmed a few statistical interrelationships between the sentiment score and the environment, which is partly caused by few tweets that contain contents that are directly related to the environment (Fuentes-Gamboa, 2020). This also implies that the urban environment is not the predominant impact factor on public sentiment.

The density of the population, low-income related professions, and the percentage of railway area are negatively correlated with the net sentiment score. However, as we stated in the correlation analysis, AEBs of dense population are not located in wealthy areas. AEBs with larger railway areas also have higher percentages of artisans who belong to the lower income class.

The most positive impact factor is the density of storefronts. The storefront consists of places for health services, education, social

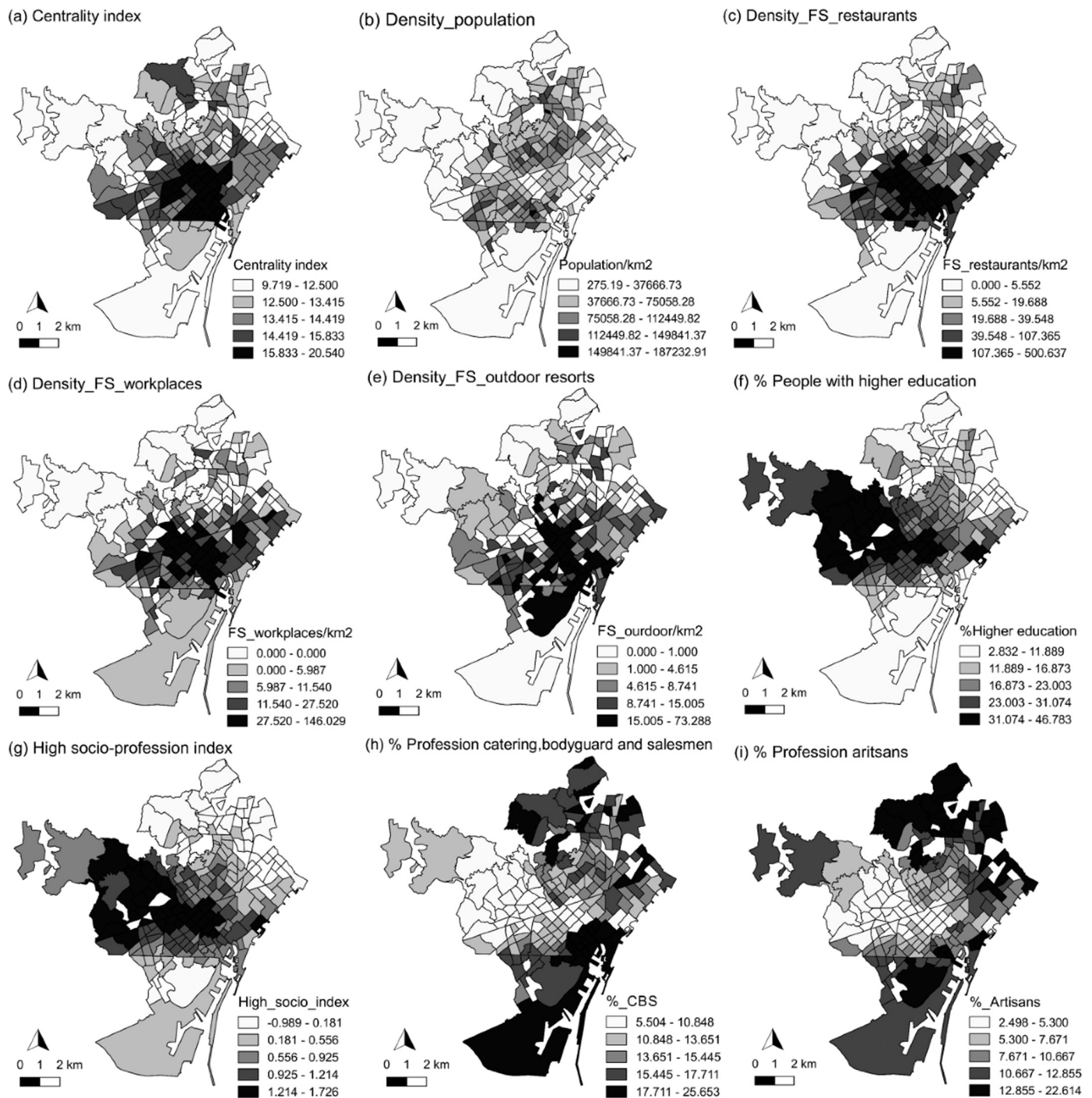


Fig. 7. Spatial distribution of sociodemographic, human mobility and socioeconomic activities. Source: own elaboration.

welfare, businesses, offices and culture, among others. It could be considered a proxy of service provision, jobs and bustling zones of the city. The commercial storefront became positive in the model because several AEBs in north-western areas have been eliminated from the model. The AEBs that were removed have a higher percentage of commercial storefront in the peripheral areas (Fig. 8(i)) where the net sentiment was comparatively low. The rest of the AEBs, with a high percentage of commercial storefronts, are mainly close to the city centre. Therefore, the commercial area may have a positive impact on the long-term public sentiment at a larger scale, though the influence is not very strong. Compared with other land uses, it may boost public sentiment

(Cao et al., 2018). To sum up, in general, a higher socioeconomic condition is associated with a higher net sentiment score. This result coincides with a study by Quercia et al. (2012) that revealed that the community's socioeconomic well-being and the sentiment score are highly correlated.

4.3. Temporal variation of sentiments

To explore the reason for a variation in sentiment, Fig. 9 visualises the temporal variation of public sentiments in the form of a proportion over the course of a day and week. On the daily scale, the most volatile

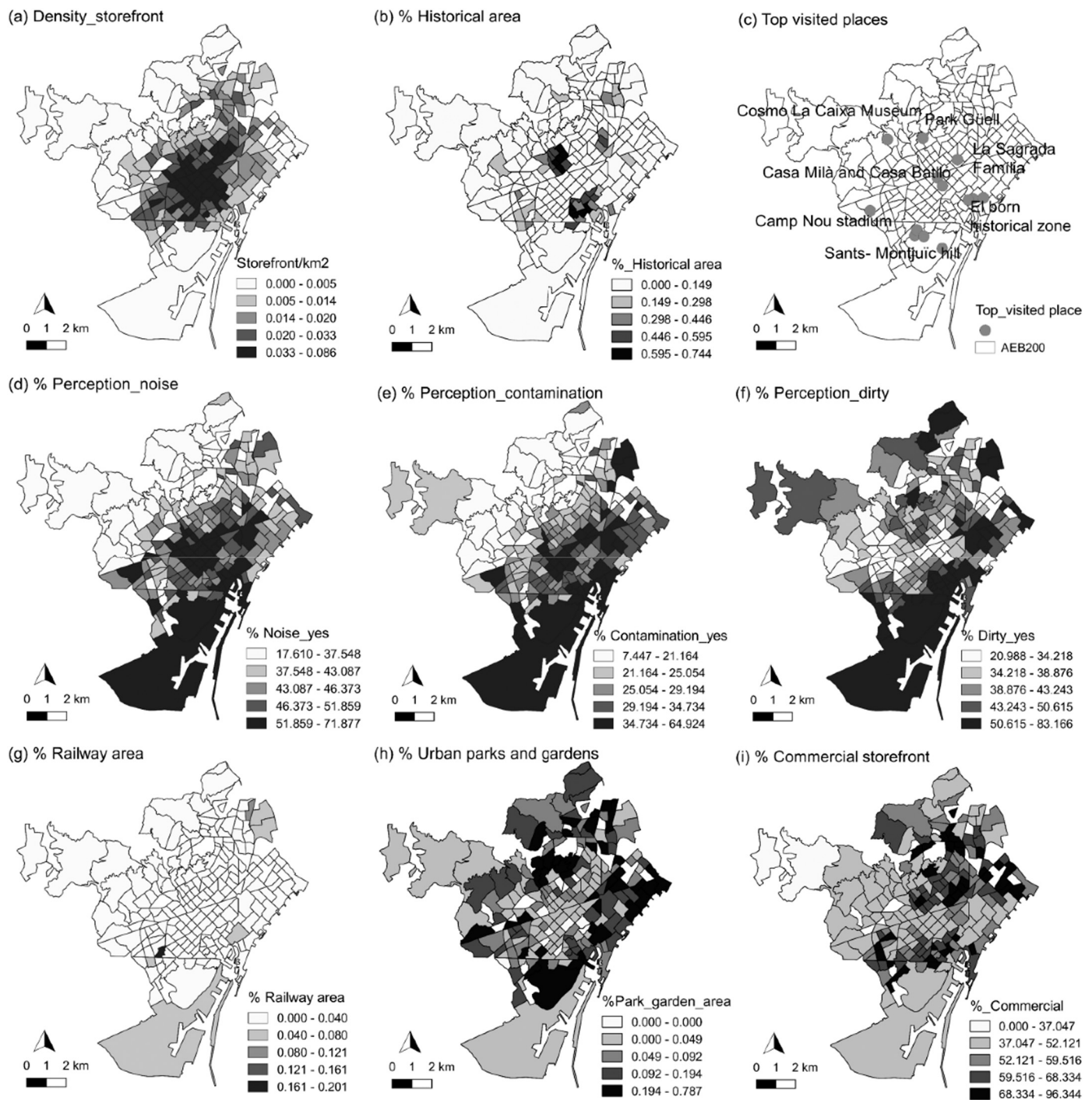


Fig. 8. Spatial distribution of the built environment. Source: own elaboration.

period is at midnight (01:00 am to 6:00 am). This fluctuation is understandable because users who were still awake and sent tweets at midnight probably encountered some disrupting events or issues. The nadir of sentiment at midnight was also observed in research by [Bertrand et al. \(2013\)](#) and [Gruebner et al. \(2017\)](#). On the weekly scale, the volume of sentiment tweets remained stable in general. However, weekly negative sentiment scores spiked on Thursday.

This was caused by the Barcelona terrorist attack that happened on Thursday 17 August 2017. [Fig. 10](#) depicts the weekly tendency of all Spanish and English negative tweets. It clearly shows the negative peak

of English tweets that appeared on the day. Several weeks did not have negative tweets, which were caused by the data loss of the original dataset. If we removed those null values, the variation of negative tweets did not exceed 0.5% in general. The negative peak of Spanish tweets happened on 1 November 2016 (Tuesday), when an important football game of UEFA Champions League was held: Manchester City beat FC Barcelona by 3–1. The importance of football in Spain could be reflected by the top five Twitter accounts with the largest number of followers – all of them are either football clubs or football players. This explains why the weekly variation in sentiment scores was lower on Tuesday and

Table 3
Model summary of the density of sentiment tweets.

	Robust	Durbin-Watson	
Linear regression			1.792
Number of observations	200	R-squared	0.689
F(10, 189)	35.070	Root MSE	0.760
Prob > F	0.000		

	Coef.	Std. Err.	Sig.
B			
Centrality index	0.112	0.031	0.000
Historical area (%)	1.110	0.256	0.000
LnD FS outdoor resort	0.029	0.009	0.001
LnD total population	0.371	0.089	0.000
Pg Gracia avenue	1.136	0.399	0.005
El Born historical area	0.975	0.253	0.000
Urban park garden	2.402	0.675	0.000
PC profession high	0.492	0.127	0.000
Contamination opinion	0.036	0.008	0.000
Density storefront	12.748	5.019	0.012
_Constant	-0.806	1.057	0.447

Source: own elaboration.

Table 4
Model summary of the net sentiments.

		Durbin-Watson	
Number of obs.	168		2.008
F(6, 161)	11.391	R-squared	0.298
Prob > F	0.000	Adjusted R square	0.272
		Root MSE	0.038

	Coef.	Std. Err.	Sig.
B			
Profession artisan	-0.004	0.0010	0.0000
% Railway area	-0.466	0.1550	0.003
Density population	-0.448	0.1070	0.000
Density storefront	0.729	0.230	0.002
% Commercial_st	0.001	0	0.010
_Constant	0.496	0.011	0.000

Source: own elaboration.

Thursday. In summary, the public sentiment can be disclosed from Twitter data and the temporal variation is more closely correlated to social or personal events.

5. Conclusions and policy implications

5.1. Conclusions and discussions

This research provides a panoramic perspective for examining the relationship between public sentiment and the urban environment using Twitter data and in terms of the spatial density of sentiment tweets and the variation of sentiment scores. Firstly, it reveals that sentiment score and the spatial distribution of sentiment tweets are not strongly correlated, a fact that has been ignored by previous studies. The higher density of sentiment tweets does not imply a higher sentiment score. The distribution of sentiment tweets has a closer relationship with human mobility and socioeconomic activities, which means that the density of Twitter activities can be a proxy variable to observe physical human activities. This conclusion has been proved by many investigations (Lenormand et al., 2014; Martí et al., 2019). Furthermore, in Barcelona, we found that sentiment tweets tend to assemble in places that are part of tourist attractions or leisure places, which indicates that some specific places could arouse people's sentiments more easily.

Regarding the score for Twitter sentiment, the variation in sentiment score is mainly influenced by disruptive events, as strongly suggested by

the analysis of negative sentiment. However, as the correlation and regression analysis imply, the long-term spatial difference in public sentiment among AEBs still exists and has correlations with the urban environment. Such a result can prove that the public sentiment is a "product of repeated place interactions and experience" (Dunlap et al., 2013).

The socioeconomic indicators could partially explain the variation in Twitter sentiment in a macro view. Firstly, the positive relationship between wealth and public sentiment has been observed in other countries (Mitchell et al., 2013; Quercia et al., 2012; Wang, Wang, & Society, 2016), just as the literature review mentioned. Therefore, the social and spatial inequality in wealth is always a noticeable issue for creating a "happy city" (Ballas, 2013), no matter how many green spaces are increased. Secondly, the higher the number of economic activities that exist, the more positive the Twitter sentiment. The density of storefronts and the percentage of commercial shops have a positive impact on sentiment score. Both indicators could be read as the active degree of economy and lively streets. The Pearson's correlation also reveals that AEBs with diverse urban activities tend to present a more positive sentiment.

The impact of the built environment on public sentiment is subtle. It is sensitive to the study scale, due to the spatial heterogeneity of built environment indicators in Barcelona. Such a difference increases the difficulty of pattern recognition at city level. For example, the contradictory impact of the percentage of commercial shops is presented between the correlation analysis and the model. After reducing the variance in the model, our model obtains a similar result to Cao et al. (2018). Although the coastal area, urban parks and density of storefronts do not appear in the model, they are positively associated with the sentiment score in Pearson's correlation analysis. A similar result has been found in many previous studies (Gallegos et al., 2016; Schwartz et al., 2019). This indicates that the built environment has a definite but limited impact on public sentiment at city level.

5.2. Policy implications

This research shows that the sophistication of public sentiment could be partially understood by Twitter sentiment analysis. It provides some reliable results that could depict some correlation between public sentiment and the urban environment. From a macro perspective, the correlation between wealth and public sentiment reminds us of the ultimate vision of smart cities and other future cities. It should highlight concerns about social equality and personal happiness rather than just the installation of advanced technologies. Under the framework introduced in the literature review (Section 3.5), knowledge drawn from our analysis could apply to urban applications in three ways.

Scenario 1 — monitoring social inequality between wealthy and poor. The result clearly shows that the orientation of Twitter sentiment could be a latent zonal indicator of social disintegration. Compared with official statistical data, Twitter sentiment would be a more direct, more rapid response to socioeconomic conditions in the pandemic period. This could also measure the diversity of general urban facilities and the active degree of economic activities.

Scenario 2 — the public response to an urban environment intervention. Our analysis proves that sentiment analysis based on anonymous Twitter users could provide useful information for sensing the city without privacy concerns. Therefore, this could be an economical method to evaluate the effect of urban projects at pre- and post-completion stages, especially the transformation of public space and the increase of urban green areas.

Scenario 3 — piloting public policy initiatives with understanding of special groups. The daily variation in Twitter sentiment (Fig. 9) may provide a new perspective to understand urban nightlife by analysing places where extreme sentiments aggregated. It would be useful to control alcoholic intemperance among young people.

It is undeniable that some flaws exist in the research. The lack of

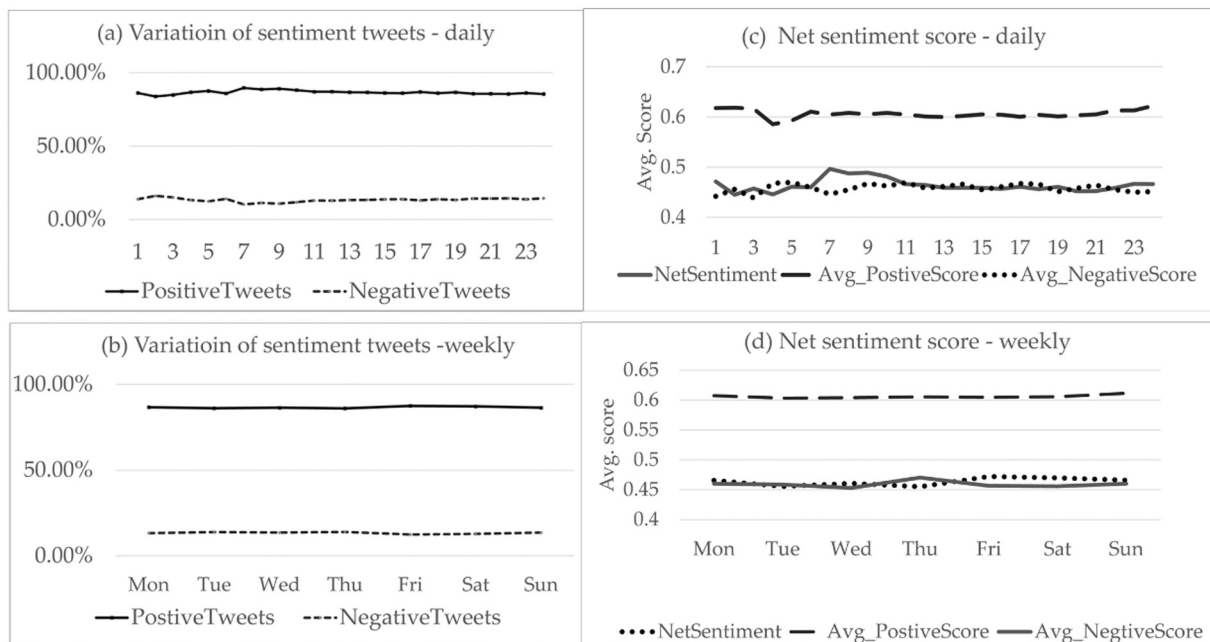


Fig. 9. Variation of percentage in sentiment tweets.
 Source: own elaboration. Note: The average negative score was multiplied by -1 for the sake of variation clearly.

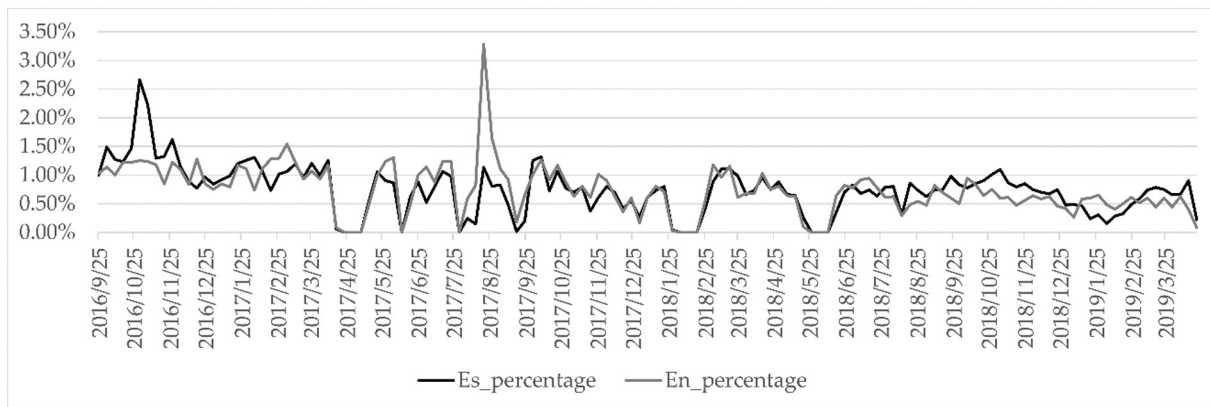


Fig. 10. Weekly variation in negative tweets.
 Source: own elaboration.

Twitter demography, such as gender, age and identity (tourist or residents), leads to the result having only scratched the surface of the connection between public sentiment and urban environment. Secondly, the dataset of Catalan tweets is not available to sentiment analysis, due to the poor quality of the data. This affects the representativeness of our results. In future research, the separation of tourists from locals will improve the accuracy of the analysis of public sentiments. Moreover, social and spatial inequalities were observed in our results to some degree. Such issues should be investigated further by combining with official statistical information.

CRedit authorship contribution statement

Liya Yang: Conceptualization, Methodology, Data Curation, Formal

analysis Writing – Original draft preparation, Investigation, Visualization.

Carlos Marmolejo Duarte: Supervision, Conceptualization, Methodology, Writing – Review & Editing.

Pablo Martí Ciriquián: Data acquisition, Draft Review.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Appendix A

Table A.1

Samples of non-individual accounts.

Types of non-individual accounts	Keywords of the filter	Examples of non-individual accounts
Commercial advertising	Bar, Restaurant, Shop, Club, Store, Café, Tatto, Barcelona, Pastry, Vintage, etc.	ジョルディ - Jordi ✨👉👈, Meteo UB Física, Bistrot Restaurant
Public information	Weather, News, school, jobs, Art, Film, etc.	TMJ-NJC Transp. Jobs, Local News, Trendsmap Barcelona, etc.
Website	.es, .com,	BRAVASBCN.COM

Source: own elaboration.

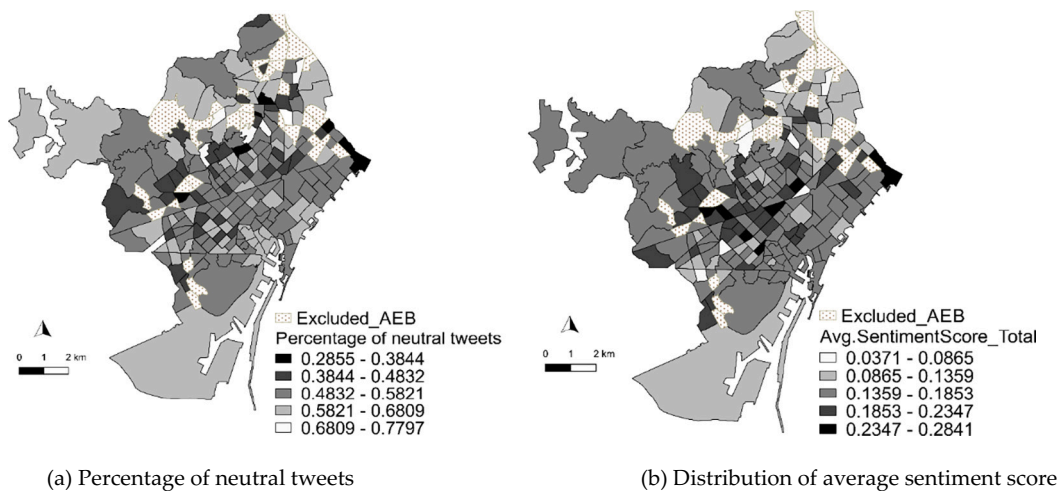


Fig. A.1. Spatial comparison between neutral tweets and average sentiment score.

(a) Percentage of neutral tweets, (b) Distribution of average sentiment score.

Source: own elaboration.

Appendix B. Human inspection of Spanish-English translation and sentiment classification

The aims of this inspection should be highlighted: 1) to measure the impact of the translation to the sentiment classification, rather than the quality of translation; 2) to verify the improvement in the mixed use of the two programs. Therefore, we measure the impact of translation by comparing the results of sentiment classifications between human and machine. Considering the expense of human evaluation, only the texts of tweets that are allocated in different categories of sentiments would be further investigated. A native Spanish and Catalan speaker with advanced English level was invited to classify the original Spanish tweets into positive, negative and neutral. Under the same statistical standard of the Catalan sample, a sample of 453 Spanish tweets was extracted from the dataset, which has a similar spatial representation to the total Spanish tweets (Fig. B.1).

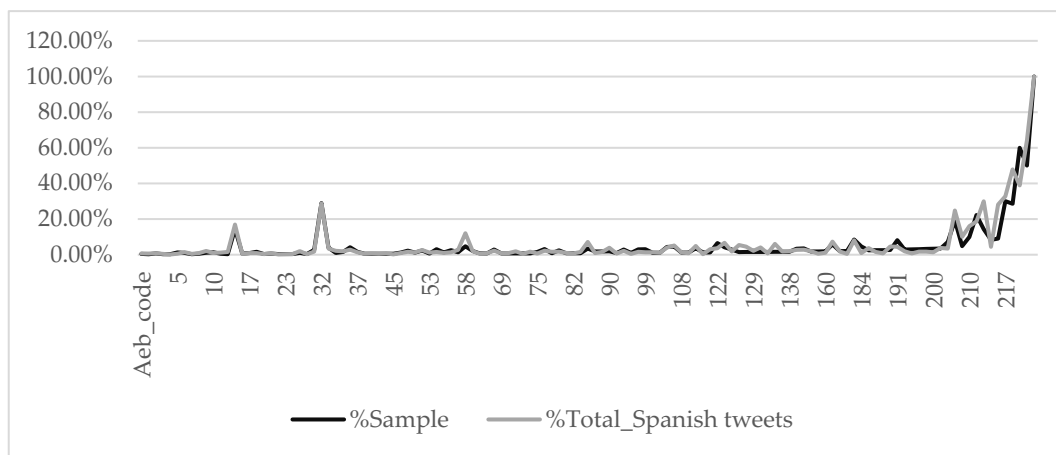


Fig. B.1. Spatial distribution of samples and total Spanish tweets.

Note: The Aeb_code is the code of each AEB. The percentage is the percentage of tweets of AEB considering the population.

Source: own elaboration.

Table B.1 lists all possible combinations of sentiment classifications. The total agreement is defined as the number of tweets that are classified in the same category of sentiment using different methods. The agreement level between the human and a single machine classification (Vader or SentiStrength) is about 66%. The total agreement of the two software programs reaches 75%. It is worth mentioning that the degree of agreement between human evaluations (made by different people) is about 80% (Ogneva, 2010). The total agreement among human, Vader and SentiStrength is 55.84%. However, based on the result of the intersection of Vader and SentiStrength, the agreement increases to 73.76%. This proves that the quality of sentiment classification is improved after intersecting the results.

Table B.1
Results of Spanish sampling sentiment evaluation.

Method of classification	Total number of agreement tweets	% agreement	Positive	%	Negative	%	Neutral	%
Human (original Spanish tweets)	453	–	204	45.03 %	17	3.75 %	230	50.77 %
Vader	453	–	187	41.28 %	38	8.39 %	228	50.33 %
SentiStrength	453	–	192	42.38 %	28	6.18 %	233	51.43 %
Human + Vader	301	66.45 %	127	42.19 %	8	2.66 %	166	55.15 %
Human + SentiStrength	298	65.78 %	127	42.62 %	8	2.68 %	163	54.70 %
Vader + SentiStrength	343	75.71 %	145	42.27 %	17	4.96 %	181	52.77 %
Human + Vader + SentiStrength	253	73.76 %	110	43.48 %	6	2.37 %	137	54.15 %

Source: own elaboration.

After investigating the specific texts of these tweets, we concluded that three reasons led to the unmatched classification (Table B.2). The biggest difference is between positive and neutral tweets. In the group of Neutral (H)-Positive (VS), 75 % of these tweets belong to commercial advertisements that only can be detected by manual examination. The advertising tweets are considered a neutral sentiment. Although the aforementioned cleaning process has already removed many commercial accounts, it was hard to identify them when the name of an account does not have any characteristics, such as “.com”, “studio” and “shop”. Understanding the texts, including idioms, lack of emotional words, metaphor, satire and jokes, is the biggest problem of sentiment classification. In fact, the understanding of texts is the core issue of sentiment analysis, regardless of whether human or software methods are used. However, such an issue is beyond the discussion of our research. Translation error is specifically defined as the wrong translation of word-to-word or untranslated words. Word translation error is the main cause of the different classification between positive and negative. However, the total number in this category was just 1 % of all samples. Hence, word translation does not greatly affect sentiment analysis.

Table B.2
Number and reasons for different sentiment classification.

Difference	Number of tweets	Reason for differences		
		Problem of understanding the meaning of text	Commercial advertisement	Word translation error
Neutral (VS) – Positive (H)	43	58.14 %	30.32 %	6.98 %
Neutral (H) – Positive(VS)	32	18.75 %	75.00 %	6.25 %
Positive (H) – Negative (VS)	5	60 %	0.00 %	40 %
Negative (VS) – Neutral (H)	6	100 %	0.00 %	0.00 %
Positive (VS) – Negative(H)	3	34 %	0.00 %	66 %
Negative (H) – Neutral (VS)	1	100 %	0.00 %	0.00 %

Source: own-elaboration.

Appendix C. Completed table of regression model

The Moran I test shows that there is no evident spatial dependence in the models (Table C.1), which is based on weighting matrices of the queen-contiguity. The spatial lag model still cannot improve the issue of heteroscedasticity (Table C.3). Therefore, we decided to keep results from the robust OLS model (Table C.2).

Table C.1
Moran I test for spatial dependence.

	MI/DF	VALUE	PROB
Ln sentiment tweets model			
Moran’s I (error)	–0.0011	0.6209	0.53467
Lagrange Multiplier (lag)	1	4.6768	0.03057
Robust LM (lag)	1	11.2023	0.00082
Lagrange Multiplier (error)	1	0.0006	0.98091
Robust LM (error)	1	6.526	0.01063
Lagrange Multiplier (SARMA)	2	11.2029	0.00369
Net sentiment model			
Moran’s I (error)	–0.0159	0.0714	0.94308
Lagrange Multiplier (lag)	1	0.8259	0.36345
Robust LM (lag)	1	9.8088	0.00174
Lagrange Multiplier (error)	1	0.091	0.76287
Robust LM (error)	1	9.0739	0.00259
Lagrange Multiplier (SARMA)	2	9.8998	0.00708

Source: own elaboration.

Table C.2
OLS model summary of sentiment tweets.

Linear regression	Robust	Durbin-Watson	1.792
Number of observations	200	R-squared	0.689
F(10, 189)	35.07	Root MSE	0.760
Prob > F	0		

(continued on next page)

Table C.2 (continued)

	Coef.	Std. Err.	T	Sig.	[95 % Conf. Interval]		VIF
	B				Lower	Upper	
Centrality index	0.112	0.031	3.650	0.000	0.051	0.173	1.950
Historical area (%)	1.110	0.256	4.350	0.000	0.607	1.615	1.240
LnD FS outdoor resort	0.029	0.009	3.470	0.001	0.012	0.045	1.130
LnD total population	0.371	0.089	4.170	0.000	0.195	0.546	1.440
Pg Gracia avenue	1.136	0.399	2.840	0.005	0.348	1.923	1.360
El Born historical area	0.975	0.253	3.850	0.000	0.475	1.475	1.190
Urban park garden	2.402	0.675	3.560	0.000	1.070	3.733	1.430
PC profession high	0.492	0.127	3.870	0.000	0.241	0.742	1.550
Contamination opinion	0.036	0.008	4.320	0.000	0.020	0.053	1.240
Density storefront	12.748	5.019	2.540	0.012	2.847	22.649	2.690
_Constant	-0.806	1.057	-0.760	0.447	-2.892	1.279	

Source: own elaboration.

Table C.3

Spatial-lag model summary of sentiment tweets.

R-squared	0.699	Log likelihood	-220.531
Sq. correlation	-	Akaike info criterion	465.062
Sigma-square	0.527	Schwarz criterion	504.642
S.E of regression	0.726		

Variable	Coefficient	Std. Error	z-Value	Probability
W_Lnd_posneg	0.194	0.081	2.394	0.017
_Constant	-1.238	0.852	-1.453	0.146
Centrality index	0.094	0.034	2.775	0.006
Historical area (%)	0.917	0.303	3.029	0.002
LnD FS outdoor resort	0.026	0.008	3.235	0.001
LnD total population	0.345	0.070	4.902	0.000
Pg Gracia avenue	1.126	0.352	3.199	0.001
El Born historical area	0.759	0.244	3.110	0.002
Urban park garden	2.248	0.490	4.585	0.000
PC profession high	0.382	0.117	3.262	0.001
Contamination opinion	0.029	0.007	4.185	0.000
Density storefront	10.249	4.985	2.056	0.040

Diagnostics for heteroscedasticity

Random coefficients

Test	Df.	Value	Prob.
Breusch-Pagan test	10.000	31.615	0.000

Source: own elaboration.

Table C.4

Model summary of the net sentiments.

Number of obs.	168	Durbin-Watson	2.008
F(6, 161)	11.391	R-squared	0.298
Prob > F	0	Adjusted R square	0.272
		Root MSE	0.038

	Coef.	Std. Err.	T	Sig.	[95 % Conf. Interval]		VIF
	B				Lower	Upper	
Profession artisan	-0.004	0.0010	-4.690	0	-0.006	-0.002	1.617
% Railway area	-0.466	0.1550	-3.003	0.003	-0.773	-0.160	1.056
Density population	-0.448	0.1070	-4.206	0	-0.659	-0.238	1.726
Density storefront	0.729	0.230	3.170	0.002	0.275	1.183	1.849
% Commercial_st	0.001	0	2.603	0.010	0	0.001	1.763
_Constant	0.496	0.011	45.460	0	0.474	0.517	

Source: own elaboration.

References

- Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11), Article e00938.
- Al-Shabi, M. J. I. (2020). In, 20(1). *Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining* (p. 1).
- Bacot, H., Bowen, T., & Fitzgerald, M. R. (1994). In, 22(2). *Managing the solid waste crisis: Exploring the link between citizen attitudes, policy incentives, and siting landfills* (pp. 229–244).
- Balahur, A., & Turchi, M. (2012). Multilingual sentiment analysis using machine translation?. In *Paper presented at the proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis*.
- Balduini, M., Della Valle, E., Dell'Aglio, D., Tsytsarou, M., Palpanas, T., & Confalonieri, C. (2013). Social listening of city scale events using the streaming linked data framework. In *Paper presented at the international semantic web conference*.
- Ballas, D. (2013). What makes a "happy city"? *Cities*, 32, S39–S50.
- Balram, S., Dragičević, S. J. L., & Planning, U. (2005). In, 71(2-4). *Attitudes toward urban green spaces: Integrating questionnaire survey and collaborative GIS techniques to improve attitude measurements* (pp. 147–162).
- Bertrand, K. Z., Bialik, M., Virdee, K., Gros, A., & Bar-Yam, Y. (2013). *Sentiment in new york city: A high resolution spatial and temporal view*. arXiv preprint arXiv:1308.5010.
- Blake, D. E. J. E., & Behavior. (2001). In, 33(5). *Contextual effects on environmental attitudes and behavior* (pp. 708–725).
- Bonta, V., Janardhan, N. K. N., & Technology. (2019). In, 8(S2). *A comprehensive study on lexicon based approaches for sentiment analysis* (pp. 1–6).
- Brereton, F., Clinch, J. P., & Ferreira, S. (2008). Happiness, geography and the environment. *Ecological Economics*, 65(2), 386–396.
- Candelieri, A., & Archetti, F. (2015). Detecting events and sentiment on twitter for improving urban mobility. In *Paper presented at the ESSEM@ AAMAS*.
- Cao, X., MacNaughton, P., Deng, Z., Yin, J., Zhang, X., & Allen, J. G. (2018). Using twitter to better understand the spatiotemporal patterns of public sentiment: A case study in Massachusetts, USA. *International Journal of Environmental Research and Public Health*, 15(2), 250.
- Cattacin, S., & Zimmer, A. (2016). Urban governance and social innovations. In *Social innovations in the urban context* (pp. 21–44). Cham: Springer.
- Chapman, L., Resch, B., Sadler, J., Zimmer, S., Roberts, H., & Petutschni, A. (2018). Investigating the emotional responses of individuals to urban green space using twitter data: A critical comparison of three different methods of sentiment analysis. *Urban Planning*, 3(1), 21–33.
- Collins, C., Hasan, S., & Ukkusuri, S. V. (2013). A novel transit rider satisfaction metric: Rider sentiments measured from online social media data. *Journal of Public Transportation*, 16(2), 2.
- Craik, K. H. J. A. R. O. P. (1973). *Environmental psychology*.
- Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). # earthquake: Twitter as a distributed sensor system. *Transactions in GIS*, 17(1), 124–147.
- Culotta, A. (2014). Estimating county health statistics with twitter. In *Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems*.
- Davidson, J., Smith, M. M., & Bondi, L. (2012). *Emotional geographies*. Ashgate Publishing Ltd.
- Dunlap, R., Harmon, J., & Kyle, G. (2013). Growing in place: The interplay of urban agriculture and place sentiment. *Leisure/Loisir*, 37(4), 397–414.
- Fuentes-Gamboa, C. L. (2020). *¿Qué nos pueden decir los tweets de las condiciones residenciales durante el confinamiento por la Covid-19? Un análisis semántico para Barcelona*. Retrieved from. Polytechnic University of Catalonia https://cpsv.upc.edu/ca/shared/tesis/tmbarch20presentacio_fuentes.pdf.
- Gallegos, L., Lerman, K., Huang, A., & Garcia, D. (2016). Geography of emotion: Where in a city are people happier?. In *Paper presented at the Proceedings of the 25th international conference companion on world wide web*.
- Gao, Z., Wang, S., & Gu, J. J. S. (2020). In, 12(20). *Public participation in smart-city governance: A qualitative content analysis of public comments in urban China* (p. 8605).
- Gilbert, C., & Hutto, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Available at (20/04/16). In *Paper presented at the eighth international conference on weblogs and social media (ICWSM-14)* <http://comp.social.gatech.edu/papers/icwsml4.vader.hutto.pdf>.
- Gruebner, O., Rapp, M. A., Adli, M., Kluge, U., Galea, S., & Heinz, A. (2017). Cities and mental health. *Deutsches Ärzteblatt International*, 114(8), 121.
- Hajrasouli, A., del Rio, V., Francis, J., Edmondson, J. J. J. O. U. D., & Health, M. (2018). *Urban form and mental wellbeing: Scoping a theoretical framework for action*. 5(10).
- Hollander, J. B., & Shen, Y. (2017). Using social media data to infer urban attitudes about bicycling: An exploratory case study of Washington DC. In *City networks* (pp. 79–97). Springer.
- Horita, F. E. A., Degrossi, L. C., de Assis, L. F. G., Zipf, A., & de Albuquerque, J. P. (2013, August 15-17). The use of volunteered geographic information (VGI) and crowdsourcing in disaster management: A systematic literature review. In *Paper presented at the proceedings of nineteenth Americas conference on information systems, Chicago, Illinois*.
- Hu, Y., Deng, C., & Zhou, Z. (2019). A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. *Annals of the American Association of Geographers*, 109(4), 1052–1073.
- Hubert, R. B., Estevez, E., Maguitman, A., & Janowski, T. (2018). Examining government-citizen interactions on twitter using visual and sentiment analysis. In *Paper presented at the proceedings of the 19th annual international conference on digital government research: governance in the data age*.
- Kandt, J., & Batty, M. (2021). Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities*, 109, Article 102992.
- Kuipers, B. (1978). Modeling spatial knowledge. *Cognitive Science*, 2(2), 129–153.
- Kumar, A., & Jaiswal, A. (2020). Systematic literature review of sentiment analysis on twitter using soft computing techniques. *Concurrency and Computation: Practice and Experience*, 32(1), Article e5107. <https://doi.org/10.1002/cpe.5107>
- Kyttä, M., Broberg, A., Haybatollahi, M., Schmidt-Thomé, K. J. E., Planning, P. B., & Design. (2016). In, 43(1). *Urban happiness: Context-sensitive study of the social sustainability of urban settings* (pp. 34–57).
- Lenormand, M., Picornell, M., Cantú-Ros, O. G., Tugores, A., Louail, T., Herranz, R., & Ramasco, J. J. (2014). Cross-checking different sources of mobility information. *PLoS one*, 9(8), Article e105184.
- Li, M., Ch'ng, E., Chong, A., & See, S. (2016). The new eye of smart city: Novel citizen sentiment analysis in twitter. In *Paper presented at the 2016 international conference on audio, language and image processing (ICALIP)*.
- Lim, C., Kim, K.-J., & Maglio, P. P. J. C. (2018). In, 82. *Smart cities with big data: Reference models, challenges, and considerations* (pp. 86–99).
- Luo, W., Wang, Y., Liu, X., & Gao, S. (2019). Cities as spatial and social networks: Towards a spatio-socio-semantic analysis framework. In *Cities as spatial and social networks* (pp. 21–37). Springer.
- Lupton, D. (1998). *The emotional self: A sociocultural exploration*. Sage.
- Lynch, K. (1960). *The image of the city* (Vol. 11). MIT press.
- MacKerron, G., & Mourato, S. J. G. E. C. (2013). In, 23(5). *Happiness is greater in natural environments* (pp. 992–1000).
- Marmolejo, C., & Cerda Troncoso, J. (2017). Spatiotemporal behavior of the population as an approach to analyze urban structure: The case of metropolitan Barcelona. *Cuadernos Geográficos*, 56(2), 111–133.
- Marmolejo-Duarte, C., & Cerda-Troncoso, J. (2020). Metropolitan Barcelona 2001–06, or how people's spatial-temporal behaviour shapes urban structures. *Regional Studies*, 54(4), 563–575.
- Marquez, N., Garimella, K., Toomet, O., Weber, I. G., & Zagheni, E. (2019). Segregation and sentiment: Estimating refugee segregation and its effects using digital trace data. In *Guide to mobile data analytics in refugee scenarios* (pp. 265–282). Springer.
- Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74, 161–174. <https://doi.org/10.1016/j.compenurbysys.2018.11.001>
- Meijer, A., & Bolívar, M. P. R. J. I. R. O. A. S. (2016). In, 82(2). *Governing the smart city: A review of the literature on smart urban governance* (pp. 392–408).
- Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS one*, 8(5), Article e64417.
- Morrison, P. S. J. R. S. (2011). In, 45(8). *Local expressions of subjective well-being: The New Zealand experience* (pp. 1039–1058).
- Mouratidis, K. J. C. (2019). In, 92. *Compact city, urban sprawl, and subjective well-being* (pp. 261–272).
- Mumford, L. (1970). *The culture of cities* (3rd ed.). Harcourt Brace Jovanovich.
- Murthy, D., Gross, A., & Pensavalle, A. (2016). Urban social media demographics: An exploration of twitter use in major american cities. *Journal of Computer-Mediated Communication*, 21(1), 33–49.
- Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Paper presented at the proceedings of the 2nd international conference on knowledge capture*.
- Neppalli, V. K., Caragea, C., Squicciarini, A., Tapia, A., & Stehle, S. (2017). Sentiment analysis during hurricane Sandy in emergency response. *International Journal of Disaster Risk Reduction*, 21, 213–222.
- Nik-Bakht, M., & El-Diraby, T. E. (2016). Sus-tweet-ability: Exposing public community's perspective on sustainability of urban infrastructure through online social media. *International Journal of Human-Computer Studies*, 89, 54–72.
- Nold, C. J. B. M. W. (2009). *Emotional cartography*.
- Ogneva, M. (2010). How companies can use sentiment analysis to improve their business. Retrieved from <http://mashable.com/2010/04/19/sentiment-analysis/>.
- Okulicz-Kozaryn, A., & Mazelis, J. M. J. U. S. (2018). In, 55(2). *Urbanism and happiness: A test of Wirth's theory of urban life* (pp. 349–364).
- Padilla, J. J., Kavak, H., Lynch, C. J., Gore, R. J., & Diallo, S. Y. (2018). Temporal and spatiotemporal investigation of tourist attraction visit sentiment on twitter. *PLoS one*, 13(6), Article e0198857.
- Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of twitter data for predicting stock market movements. In *Paper presented at the 2016 international conference on signal processing, communication, power and embedded system (SCOPES)*.
- Park, S. B., Kim, H. J., & Ok, C. M. (2018). Linking emotion and place on twitter at disneyland. *Journal of Travel & Tourism Marketing*, 35(5), 664–677.
- Paul, D., Li, F., Teja, M. K., Yu, X., & Frost, R. (2017). Compass: Spatio temporal sentiment analysis of US election what twitter says!. In *Paper presented at the proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*.
- Paul, M., & Dredze, M. (2011). You are what you tweet: Analyzing twitter for public health. In *Paper presented at the proceedings of the international AAAI conference on web and social media*.
- Plunz, R. A., Zhou, Y., Vintimilla, M. I. C., Mckeown, K., Yu, T., Ugucioni, L., & Sutto, M. P. (2019). Twitter sentiment in New York City parks as measure of well-being. *Landscape and Urban Planning*, 189, 235–246.
- Quercia, D., Ellis, J., Capra, L., & Crowcroft, J. (2012). Tracking "gross community happiness" from tweets. In *Paper presented at the proceedings of the ACM 2012 conference on computer supported cooperative work*.

- Quercia, D., Schifanella, R., & Aiello, L. M. (2014). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Paper presented at the proceedings of the 25th ACM conference on hypertext and social media*.
- Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. J. E. D. S. (2016). In , 5(1). *Sentibench—a benchmark comparison of state-of-the-practice sentiment analysis methods* (pp. 1–29).
- Schwartz, A. J., Dodds, P. S., O’Neil-Dunne, J. P. M., Danforth, C. M., & Ricketts, T. H. (2019). Visitors to urban greenspace have higher sentiment and lower negativity on twitter. *People and Nature*, 1(4), 476–485. <https://doi.org/10.1002/pan3.10045>
- Sénécal, G. (2007). Urban environment: Mapping a concept. Introductory note. *Environnement Urbain/Urban Environment* (Vol. 1).
- Seresinhe, C. I., Preis, T., MacKerron, G., & Moat, H. S. J. S. R. (2019). In , 9(1). *Happiness is greater in more scenic locations* (pp. 1–11).
- Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for health research: A systematic review. *American Journal of Public Health*, 107(1), e1–e8.
- Sutoyo, E., Almaarif, A. J. B. O. E. E., & Informatics. (2020). In , 9(4). *Twitter sentiment analysis of the relocation of Indonesia’s capital city* (pp. 1620–1630).
- Tasse, D., Liu, Z., Sciuto, A., & Hong, J. (2017). State of the geotags: Motivations and recent changes. In *Paper presented at the proceedings of the international AAAI conference on web and social media*.
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558.
- Wang, F., Wang, D. J. T. B., & Society. (2016). In , 5. *Geography of urban life satisfaction: An empirical study of Beijing* (pp. 14–22).
- Wang, W., Hernandez, I., Newman, D. A., He, J., & Bian, J. (2016). Twitter analysis: Studying US weekly trends in work stress and emotion. *Applied Psychology*, 65(2), 355–378.
- Williams, C., & Dunn, C. E. (2003). GIS in participatory research: Assessing the impact of landmines on communities in north-West Cambodia. *Transactions in GIS*, 7(3), 393–410.
- Wyly, E. J. D. I. H. G. (2014). In , 4(1). *The new quantitative revolution* (pp. 26–38).
- Yang, L., & Durarte, C. M. (2019). Identifying tourist-functional relations of urban places through foursquare from Barcelona. *GeoJournal*, 1–18.
- Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., & Vaidya, J. (2020). Location-based sentiment analyses and visualization of twitter election data. *Digital Government: Research and Practice*, 1(2), 1–19.
- Dr. Liya Yang** is a PhD of Urban and Architectonic Management and Valuations at Polytechnic University of Catalonia (Barcelona-Tech) and a senior researcher in Guangdong Guodi Planning Science Technology. Her research focus on urban management, spatial analysis of social media data and smart city.
- Dr. Carlos Marmolejo Duarte** is a full professor at Polytechnic University of Catalonia (Barcelona-Tech). His research is mainly developing in: urban management, real estate economics and environmental quality assessment.
- Dr. Pablo Martí Ciriquián** is a full professor in the Urban Design and Regional Planning Division, University of Alicante (UA). His expertise is in the field of urban design and urban dynamics in the Spanish Mediterranean area.