Journal of Big Data Analytics in Transportation
https://doi.org/10.1007/s42421-021-00036-1

ORIGINAL PAPER

Analyzing Parking Sentiment and its Relationship to Parking Supply and the Built Environment Using Online Reviews

Zhiqiu Jiang[1] · **Andrew Mondschein**[1]

Received: 17 September 2020 / Revised: 26 January 2021 / Accepted: 8 February 2021

Abstract
This study examines positive and negative sentiments associated with parking experiences reported in online Yelp reviews for four metropolitan areas in North America, leveraging large location-based social network (LBSN) data to understand parking sentiment as a measure of parking search or post-parking experiences. Demand from travelers and business owners for more parking is a significant issue for local transportation planners and decision-makers, but to date, there has been little study of how local parking management strategies or built environment characteristics modify parking experiences and sentiments. To understand this relationship, we first conduct a sentiment analysis (SA) to identify the emotional, affective content of parking-related reviews embedded in the Yelp reviews. We then use generalized mixed effects (GLME) models to examine the associations between parking sentiment and (a) parking management practices, and (b) characteristics of the built environment. The SA results show that positive and negative parking sentiments are significantly spatially clustered in each metropolitan area. GLME models show that sentiments are significantly associated with destination activity types, parking management strategies, and several built environment factors. The results of this study indicate how different interventions advocated by transportation policies may influence perceptions of parking in commercial and mixed-use districts with implications for overall support for neighborhood transportation planning best practice. Furthermore, the findings represent that emerging data mining and statistical methods can successfully leverage big data to reveal travel experiences and their relationship to urban contexts, providing an effective solution to obtaining experiential transportation information.

Keywords Sentiment analysis · Parking · Non-work activity · Travel behavior · Built environment · Text mining

Introduction

Transportation planners acknowledge the complex relationship between parking and issues, such as traffic congestion, mode choice, economic activity, and development patterns. They understand that simply providing more parking can be counterproductive, and searching for parking in commercial and mixed-use areas can waste fuel, contribute to traffic congestion, and overload local parking supplies and spill into adjacent neighborhoods (Shoup 2006, 2011). While planners seek to manage urban parking, driver perceptions of parking availability are a critical component of the choice to park and demand for additional parking supplies. INRIX reports that sixty-one percent of US respondents reported feeling stressed looking for parking, and sixty-three percent stated that they avoid destinations because of expected difficulty finding parking (INRIX 2017). Furthermore, customer perceptions of parking availability are a serious concern for business owners, who frequently see driving as the primary means of access to their establishments (Bureau of Transportation Statistics 2018). Collectively, driver sentiments and their effects on business owners can place serious pressure on local planners and political leaders to provide more parking. However, contemporary planning best practice encourages planners to manage access to commercial and mixed-use destinations by providing shared or priced parking and by designing built environments amenable to alternative modes. Given this tension, a better understanding of driver attitudes and perceptions towards parking may inform planners seeking to foster multimodal, sustainable transportation and urbanization.

✉ Zhiqiu Jiang
zj3av@virginia.edu

Andrew Mondschein
mondschein@virginia.edu

[1] Department of Urban and Environmental Planning, University of Virginia, Charlottesville, VA, USA

Published online: 05 March 2021

In this study, we examine whether parking supply and the built environment modify parking sentiment. In particular, we investigate how parking sentiment is associated with the provision of parking and built environment characteristics, using content from Yelp online reviews, a large location-based social network (LBSN) dataset, for four metropolitan areas in North America. We use text mining and statistical methods to understand parking experiences and their relationship to urban/suburban contexts, as related in social media content. An ever-increasing part of the population makes use of social media, writing "tips" and "reviews" when having visited a destination. LBSNs have become an important means for acquiring experiential data and travel information that can be analyzed to understand travel behavior and transportation systems (Kambele et al. 2015; Sekar et al. 2017). LBSNs can reflect aspects of daily travel, such as information search, decision-making, and post-travel evaluation that are not typically addressed by travel surveys (Mondschein 2015).

We review relevant literature, focusing on transportation research using social media, and research examining the relationship between vehicle parking and the built environment. We then introduce our data, Yelp online reviews with parking content for non-work activities in four North American metropolitan areas. With sentiment analysis, the subjective experience of parking can be measured and analyzed at fine geographic scales across multiple metropolitan areas. The analysis proceeds through three stages: (a) extracting parking sentiment by business from Yelp reviews; (b) assessing the relationship between parking sentiment and parking management practices; and (c) assessing the relationship between parking sentiment and the built environment. Using spatial analysis and mixed effects models, significant relationships are found between parking sentiment, parking management, and several built environment factors. The results, linked to local geographies, serve as powerful indicators of how parking supply and built environment factors are associated with traveler experiences and sentiments. Findings show that parking management and built environment factors have significant effects on traveler perceptions of their parking experience. Methodologically, emerging data mining and statistical methods can successfully leverage big data to reveal travel experiences and their relationship to urban contexts, providing an effective solution to obtaining experiential transportation information.

Literature Review

Relationships Among Parking Choices, Facilities, and the Built Environment

The provision of parking has become an important component of suburban and even urban accessibility (Manville and Shoup 2005), and parking availability can significantly affect the probability of choosing automobile travel mode option (Pandhe and March 2012). North American cities have long included parking requirements for new urban development, but particularly in older areas, widespread automobility combined with relatively dense development has resulted in parking shortages, as perceived by drivers (Shoup 2011). Today, as planners seek to facilitate multimodal, transit-oriented development in both cities and suburbs, parking is again being limited in many commercial and mixed-use areas (Dittmar and Ohland 2012).

Rather than build more parking, transportation planners use both pricing and built environment strategies for reducing parking demand and encouraging mode shift from driving to more sustainable travel modes. For instance, charging for parking has become a widely used approach to managing parking demand (Millard-Ball et al. 2014). Pricing is an important mechanism for controlling automobile use because (a) people are sensitive to parking cost, as well as parking search and walk times in choosing destinations and mode, and (b) parking supply and price are at least partially controllable through policy levers, such as zoning, regulation, and taxation (Inci 2015). At the same time, planners and designers consider built environment factors to be important mechanisms for encouraging mode shift. Elements, such as the "5D's": density, diversity, distance to transit, destination access, and design (Ewing and Cervero 2001, 2010), can reduce reliance on cars and parking, and to increase non-motorized modes' attractiveness. For example, Christiansen et al. (2017) found that higher density around destinations is associated with lower likelihood of using the car, and the odds also decrease when the end destination is closer to the city centre. Stevens (2017) also found that compact development does make people drive less, even though the impact on reduction of vehicle miles traveled (VMT) appears to be small in magnitude.

Pricing and built environment-based parking demand management follow a microeconomic framework, modifying relative costs of driving and other modes (Marsden 2006; Weinberger et al. 2010). However, attitudes and affective states also influence transportation decision-making (Griffioen-Young et al. 2004). The Theory of Planned Behaviour (Ajzen 1991) and the Theory of Reasoned Action (Fishbein and Ajzen 1975) posit that a positive attitude leads to the formation of a greater behavioral intention (motivation), which is more likely in turn produce the behavior (Verplanken and Aarts 1999). Parkany et al. (2004) reviewed literature in social psychology and transportation and found that attitudes are very important in travel mode choice. Parking behavior may be determined by attitudes and intentions. For example, Bamberg et al. (1999) found that attitudinal factors toward parking fees, parking space availability, and gas tax rises affect travel mode. The decision of whether and where to park is based on perceived impedances as well as affective qualities of travel, such as the stress of finding parking (INRIX 2017). In previous research, we found that parking supply has a limited relationship with parking sentiment but that the way parking is provided, such as in shared lots, may affect sentiment (Mondschein et al. 2020b).

For transportation planners, an equally important relationship is that between transportation experiences and attitudes towards specific planning interventions. Support for road building, for example, is associated with more driving (Börjesson et al. 2015) and increasing regional congestion (Rose 1990). Parking management strategies, such as pricing and parking maximums, also elicit public, political responses that can make or break a plan or policy (King et al. 2007; Mondschein et al. 2020b). Therefore, a better understanding of the relationships between positive and negative attitudes towards parking and factors, such as parking management strategies and built environment characteristics, may inform planners seeking to foster multimodal, sustainable transportation and urbanization as well as help shed light not just on the behavioral effects of those strategies but their political feasibility. In addition, this approach can identify general best practices for parking management strategies and built environment approaches, as well as local variations in sentiment that can be used to identify specific issues or unexplained areas of positive or negative parking experiences.

Transportation Research Using Location-Based Social Network Data

In recent years, large LBSN datasets, such as Yelp, Twitter, Tripadvisor, and Facebook, have expanded rapidly, attracting an increasing number of users, who often use these services to help make destinations and route choices for travel (Evans and Saker 2017). With text mining methods, transportation researchers are able to extract travel information from online text reviews and connect it with specific locations (Sekar et al. 2017), investigate travel mode choice to non-work destinations (Jiang and Mondschein 2019), and use travel-related reviews to implement a planning decision support system (Zhou et al. 2017). These data have the potential to address documented limitations with traditional travel surveys: declining sample sizes (Stopher and Greaves 2007), under-reporting of trips (Forrest and Pearson 2005), imprecision or absence of locations and times (Arribas-Bel and Bakens 2019; Stopher et al. 2005), and infrequently updated content (Chen et al. 2010). Compared to traditional survey data, textual analysis methodologies can provide distinctive insights from LBSNs and supplement existing travel analysis, as well as allow investigation of variability in travel attitudes linked with destinations across neighborhoods, cities and countries, at high volume and spatial precision (Sekar et al. 2017).

Sentiment analysis (SA) estimates people's opinions, attitudes, and emotions from written language. SA is a component of natural language processing (NLP) and is also widely utilized in text mining and machine learning (Liu 2012). The development of SA methods has allowed LBSN sentiment mining to estimate attitudes in geographic contexts. Specifically, with the help of improved NLP techniques (Aggarwal and Zhai 2012), the text extracted from LBSNs can be analyzed to identify the emotional content of behaviors in urban environments (Roberts et al. 2019). However, the analysis of sentiment for "big" textual data is challenging due to the fact that human interpretation of each observation would be too time-consuming to be useful. This challenge can be addressed by means of automated SA techniques focusing on determining the polarity—positive or negative—of natural language text. Among these techniques, lexicon-based SA methods for classifying the polarity of texts have gained attention in recent work and their performance has been shown to be robust across domains and texts (Ding et al. 2008; Taboada et al. 2011).

In addition, a majority of SA literature primarily focuses on broad geographic scales, such as cities or regions. For example, Caragea et al. (2014) performed sentiment classification of user posts in Twitter during the Hurricane Sandy and visualized these sentiments at global and regional scale. Mitchell et al. (2013) investigate correlations between individuals' posts and a wide range of emotional, geographic, demographic, and health characteristics using geo-tagged Twitter data. However, these prior analyses lack more geographically specific analysis of factors that may affect travel attitudes and behavior. In this study, we include geographically fine-scaled transportation and built environment data to investigate the localized relationships between parking sentiments at non-work businesses and transportation and land use factors.

Characteristics of Yelp Data

Yelp is an LBSN where reviewers rate "businesses," including a variety of destination types, and contribute long-form text reviews so that users can make more informed non-work activity choices. The online text of Yelp reviews contains relatively rich information about the travel experiences of a variety of travel modes (Jiang and Mondschein 2019; Mjahed et al. 2017; Mondschein 2015). We use the 2018 release of the *Yelp Academic Dataset* (Yelp 2018b), which provides full-text reviews (4,326,942 reviews) and the precise latitude and longitude of each reviewed business in selected cities in North America and Europe. Each review is timestamped in terms of when the review as submitted (not when the activity took place). Besides the spatial location and timestamp information, each business in the Yelp dataset is originally categorized using a multi-label classification approach (Tung 2015) with nearly 1000 categories (Yelp 2018a). We reclassify these categories into Yelp's reported "10 big categories," which are *Active Life, Arts, Automotive, Health, Hotels & Travel, Nightlife, Other, Restaurants, Service*, and *Shopping*, transforming each business from multi-label to single-label using an identification algorithm to match the business within the 10 big categories. Because each business in the raw dataset includes multiple category labels with the first being the main category, the algorithm selects the first label from the raw dataset and assigns a single category to the business (Jiang and Mondschein 2019).

In addition to aggregated review information, most businesses have associated parking attributes, a set of binary categories (*True/False*) indicating the availability of five parking attributes at each business, such as "parking garage," "parking lot," "street parking," "parking valet," or "validated parking." The parking attributes provide a means of ground-truthing the type of parking supplied in different neighborhoods across our study cities, though they do not indicate the absolute quantity of parking supplied. In addition, they represent specific strategies used by businesses, business collectives, such as business improvement districts, and planners to more effectively manage parking supplies in commercial and mixed-use districts.

One limitation of using social media data, such as Yelp, is it lacks embedded demographic and socio-economic information about each reviewer. For this analysis, our population of interest is patrons of establishments in urban commercial and mixed-use districts. While these patrons may not represent all urban residents, the sentiments of this self-selected group are likely to have significant impacts on local parking demand as well as local planning and decision-making. Still, Yelp users themselves may not be representative of all patrons of establishments in urban commercial and mixed-use districts. Therefore, we use empirical methods, where possible, to determine demographic characteristics from available data to assess whether these factors are likely to have a significant impact on our outcome variable. These approaches, described here, include comparison of Yelp users to aggregate population data, imputation of demographic characteristics, and the use of proxies for demographic information.

First, we consider Yelp users in the aggregate relative to the population as a whole. As of 2019, Yelp has an average of over 36 million monthly unique users (Yelp 2019). Comparing the demographics of Yelp users from a Quantcast survey (Quantcast 2017) to US Census (U.S. Census Bureau 2016) and Canada Statistics (Statistics Canada 2016) data on the general population, Yelp users are more female (61% of users) than US and Canada census respondents. Yelp users' households are also slightly more educated and wealthier on average than households in the US and Canada (Yelp 2019).

Given the aggregate differences between Yelp user demographics and the population as a whole, we assessed whether a significant relationship may be observable between our outcome variable of interest, parking sentiment, and specific demographic factors. Note that we describe the sentiment analysis methodology further in "Sentiment Analysis of Parking Reviews in Yelp" section below. We considered gender as a demographic characteristic of the Yelp users that may influence parking sentiment, using a name-based prediction method to predict the users' gender using the "username" variable for each review. This variable is the username chosen by users when they register. Most of these usernames (~ 90%) were a standard name word (the first name), and we used the R package "gender" (Mullen 2020) to predict each user's gender based on their username. The prediction of the gender package is based on first names using historical datasets. After prediction, we found that the percentage of the predicted female group was 55.6% (# of count: 30,440) and the predicted male was 44.4% (# of count: 24,307). 54,747 of 61,776 (89%) usernames were used to conduct prediction. 7029 usernames were unpredictable inputs, since they were a single letter or character combinations that cannot be found as a name in the historical name datasets. Using the names with assignable male or female genders, we linked the predicted gender information with the parking reviews, and the Spearman correlation analysis (Schober et al. 2018), a method that can measure the association between the continuous data and ordinal data, shows near-zero correlations between parking sentiment (values are from the SA results) and predicted gender being male in all four focus cities (Charlotte: $r = 0.002$; Las Vegas: $r = 0.005$; Phoenix: $r = 0.006$; Toronto: $r = 0.008$). Thus, we do not find a significant correlation between the parking sentiment and users' gender groups.

To test whether a significant relationship may exist between income and sentiment, we used the cost of restaurants (included in the Yelp dataset at 4 levels) as a proxy for patron income, with the reasonable assumption that more expensive restaurants will be patronized by higher income Yelp reviewers, *ceteris paribus*. While restaurant price is a reasonable proxy for the income of its patrons, on average, this approach would still not reveal specific interactions between individual income and restaurant price. The Spearman correlation analysis shows near-zero correlations between parking sentiment (values are from the SA results) and restaurant cost in all four focus cities (Charlotte: $r=0.017$; Las Vegas: $r=0.026$; Phoenix: $r=0.028$; Toronto: $r=0.043$). Thus, we find no significant correlation found between the parking sentiment and this proxy for income levels.

While we were able utilize Yelp data to examine gender and income associations with our outcome variable, other demographic factors are more difficult to assign to reviewers. For example, we sought approaches to investigate the race and ethnicity of Yelp users. However, the Yelp dataset refers only to users' first (given) names. While a few R packages, e.g., predictrace (Kaplan 2020), wru (Khanna 2020), can use the last name (surname) to predict race or ethnicity, this approach is not validated for first names. Similarly, we are unable to determine reviewer age from available data. We discuss how these limitations could affect interpretation of the results in "Discussion and Conclusion".

Finally, we also examined whether the parking sentiment is correlated with overall review stars, given the concern that parking sentiment and overall activity experience may be correlated. Because we extract only the text segment that specifically describes the parking experience (Mondschein et al. 2020a), we expect that the parking sentiment should be isolated to the parking experience itself. When we test this association, the Spearman correlation analysis shows near-zero correlations between parking sentiment (values are from the SA results) and the review stars in all four focus cities (Charlotte: $r=0.009$; Las Vegas: $r=0.029$; Phoenix: $r=0.015$; Toronto: $r=0.004$).

Study Area and Research Questions

Geographically, we use a subset of the full dataset, focusing on four North American cities: "Charlotte, North Carolina" "Las Vegas, Nevada" and "Phoenix, Arizona" in the US, and "Toronto, Ontario" in Canada. Each city requires some amount of parking for commercial developments in their zoning codes (Charlotte Planning, Design, & Development Department 2017; City of Toronto Council 2018; Clark County Department of Comprehensive Planning 2017; Phoenix City Council 2015). As a variety of urban activities and the associated travels occur in big cities, there is an increasing need to establish a highly functional and efficient parking management solution that ensures resident satisfaction and utilizes the existing parking facilities throughout the city.

The top three business categories in these four cities, by percent of all businesses in the Yelp dataset, are Restaurants, Service, and Shopping, with an average percentage of 31%, 21%, and 17%, respectively.[1] The combined number of reviews in the four cities is approximately 4.3 million. Among all the reviews, about 46% of them were about "Restaurants" category.[2] We examine the commercial and mixed used districts of these four North American cities for answering our key empirical research questions:

- How are positive or negative parking experiences associated with parking provision?
- How do business parking management strategies and built environment characteristics shape parking sentiments?

The study proceeds in three stages. First, we conduct a SA to (1) identify the emotional content of reviews referring to parking and analyze the spatial distribution of parking sentiments across the four cities. Then, using generalized mixed-effect (GLME) models, we examine (2) the relationship between parking sentiment and parking management strategies, and (3) the relationship between parking sentiment and other factors in the built environment in downtown Las Vegas, Nevada.

Sentiment Analysis of Parking Reviews in Yelp

Yelp reviews frequently include transportation content (Jiang and Mondschein 2019; Mjahed et al. 2017; Mondschein et al. 2020b). The average word count of a Yelp review containing parking keywords from the four cities is 205 words, potentially including information of individual users' opinions and their parking experiences, or the reasons why they choose or do not choose parking when they travel to certain businesses. Examples from the dataset:

> "The parking is *free* and *easy*. That is *awesome*."
> "Limited menu a hard place to find using GPS, parking can be a little *hectic* too".

[1] In terms of the variety of Yelp business types, we calculated the percentage of each business category for all the businesses. The result shows that the percentages of *Active Life, Arts, Automotive, Health, Hotels & Travel, Nightlife, Other, Restaurants, Service,* and *Shopping* are 2.45%, 3.03%, 5.38%, 7.25%, 2.86%, 8.65%, 1.58%, 30.52%, 20.92%, 17.36%. "Restaurants" category is the category with the largest number of businesses (45.51%).

[2] The percentage of each business category in all the reviews is: *Active Life* 1.36%, *Art* 2.33%*s*, *Automotive* 3.32%, *Health* 3.32%, *Hotels & Travel* 4.35%, *Nightlife* 20.68%, *Other* 0.49%, *Restaurants* 45.51%, *Service* 10.37%, and *Shopping* 8.27%.

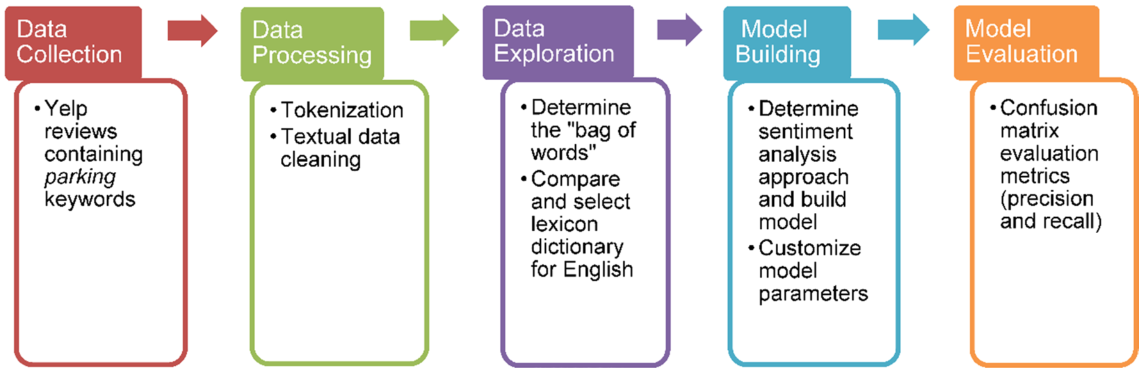

Fig. 1 SA framework for Yelp parking reviews

Table 1 Statistics of parking reviews in Yelp dataset

Parking reviews in Yelp dataset					
Metropolitan area	Total number of reviews	Total number of parking reviews[a]	Percentage of parking reviews (%)	Total number of parking reviews with parking attributes[b]	Percentage of parking reviews with parking attributes (%)
Charlotte	276,570	13,217	4.8	12,919	97.7
Las Vegas	1,812,400	45,801	2.5	32,835	71.7
Phoenix	1,606,907	44,969	2.8	39,969	88.9
Toronto	631,065	19,774	3.1	19,735	99.8
Total	4,326,942	123,761	2.9	105,458	85.2

[a]Parking reviews refer to reviews that mention keywords, such as "parked" or "parking" in the Yelp dataset

[b]Parking attributes refer to the parking information provided for businesses in the Yelp dataset. It includes the availability of street parking, parking lot, parking garage, parking valet or validated parking service, stored in a binary format using "true" or "false" index

"… just *frustrated* in trying to find a parking spot."
"… as always, parking is a little *tough* uptown."
"One downfall is that parking is *horrible*, with *narrow* spaces and not a lot available."
"Located in a very busy intersection, plenty of commuters and parking is pretty *bad*."
"It's a good place for quick meets with *easy* parking and easy access along Dundas."
"My only complaints, it was pricey and parking was *challenging*."

Various methods can be used for conducting a sentiment analysis. In this study, we use a lexicon-based approach to measure the emotional content of the large number of reviews with parking experience. A summary of SA steps is shown in Fig. 1.

Data Collection and Preprocessing

First, we use a text mining approach to identify parking reviews. To focus the search for parking reviews, we only use keywords "parking" or "parked" as search criteria within a given review, generating the frequency of parking reviews. Reviews with parking content from these four cities are 2.9% of all reviews, (see Table 1). Note that this might be an underestimate, since not all possible terms related to parking may be included in the selected set of terms. 85.2% of parking reviews are associated with businesses providing parking attribute information. We use the 105,458 parking reviews that can be linked to businesses with parking attribute information for our sentiment analysis.

Data preprocessing includes textual data tokenization and data cleaning. Tokenization splits long strings of text into smaller pieces, or tokens. To find the best token to represent the parking reviews, we tokenize each review as paragraphs, sentences, and smaller word chunks first. Paragraph is defined by a new line in the review, sentence is defined by the ending sentence punctuation, and word chunk is defined by punctuation in the middle of a sentence. Each of these tokens must contain at least one parking keyword. Then, we compare and determine the best tokenization for the SA. We cleaned the data for each tokenized string of text, using the 'tm' package in R statistical programming language (Feinerer 2018). The cleaning process involves a sequential process for each tokenized string: making a corpus of words, converting into lowercase, removing punctuation, numbers, and URLs, stripping whitespace, and removing words irrelevant to SA, such as "the" or "an."

Data Exploration

The word chunk is chosen as our token unit since it shows most appropriate representation of parking experience information. Specifically, neither paragraph nor sentence is good enough for our case. Sentences with parking terms may be very long since some people use multiple commas instead of periods. We cut sentences into word chunks that can actually describe parking experience. The analysis uses the Harvard IV dictionary, a general-purpose psychology-based dictionary. It includes greater than 11,000 words with 1915 positive and 2291 negative sentiment words (Stone et al. 2007). This dictionary is able to capture sentiment through different sets of words associated with quantified sentiments (Saxena et al. 2018).

Model Building

To estimate the sentiment scores of parking reviews represented as word chunks, we use *analyzeSentiment()* function in the SentimentAnalysis package (Feuerriegel and Proellochs 2018) in R to generate the initial sentiment scores. This approach is a lexicon-based approach that can classify the sentiment, returning the sentiment scores for each selected dictionary. The scores range from − 1 to + 1 with − 1 showing an extremely negative sentiment and + 1 representing most positive, with 0 being a "neutral" parking experience. Our best model fits each parking review with an estimated sentiment score based on the degree of positivity and negativity in the bag of words, including assigning weights that are most predictive in the context of the observed corpus (dictionary corpus).

Model Evaluation

In terms of the nature of the large dataset, the total number of parking experience word chunks is more than 100,000, making reading each review and assigning a manual score impossible. Instead, we adopt a two-step performance evaluation for the model results. In Step 1, we read and check all of the predicted min sentiment scores and max sentiment scores for each city and for each business type. It produces a 2 (sentiments) * 4 (city) matrix, results are listed below:

Charlotte:

review (high): "Plenty of parking."
review (low): "Parking is a NIGHTMARE."

Las Vegas:

review (high): "It offers good parking options."
review (low): "Hard to grab a parking spot."

Phoenix:

review (high): "Plenty of free parking available as well!"
review (low): "Parking is a wee bit of a pain."

Toronto:

review (high): "Free parking is a nice bonus."
review (low): "I lost to parking."

We have a clear general impression of the classification from step 1's results. Then, in Step 2, we review a random sample 500 (0.5% of entire dataset) of estimated sentiment scores and the corresponding original parking reviews. We manually read them one by one and create a confusion matrix (using three categories: "positive," "neutral," and "negative") to compare the precision and recall between the model results and the human-judged results. The percentage of accurate categorization is 80%. According to the precision and recall metrics, TF (positive sentiment categories erroneously classed as non-positive) = 5%, FT (non-positive sentiment categories erroneously classed as positive) = 15% (Sokolova et al. 2006). 80% accuracy, while introducing error into the subsequent analysis, is favorable for sentence-based SA using current methods. Importantly, the error is distributed across both positive and negative predictions, with some bias toward over-prediction of positive sentiment. The results of the subsequent analysis should be understood keeping this potential bias in mind (Taboada et al. 2011).

Spatial Hotspots of Parking Sentiments

Business-Level Parking Sentiments

Using the average sentiment score for each business, the distribution of positive, neutral and negative reviews is similar across the four cities (see Appendix 1). We found that positive sentiments are the majority across the four cities. The results are consistent with the count values of sentiment classification outputs as well. Most of the sentiment scores are distributed around 0, given the average amount of parking reviews per business, this is not surprising, since the sentiment score is normalized by the analysis model ranging from − 1 to 1.

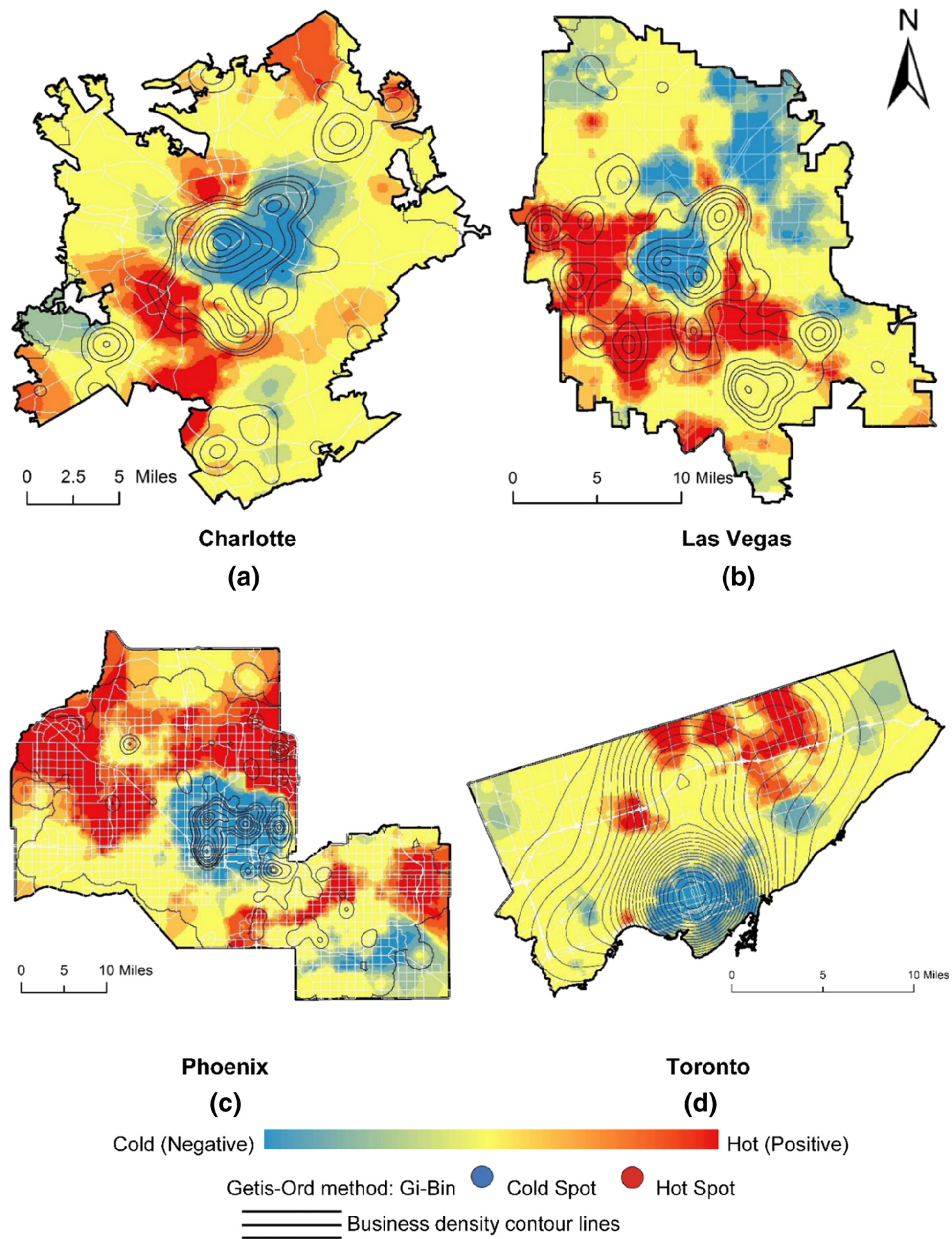

Fig. 2 Heatmap of "positive", "neutral" and "negative" parking sentiments. This map is better visualized in color

Hotspot Analysis of Business Parking Sentiments

Geographic Information System (GIS) analysis enables a hotspot analysis of the spatial pattern of parking sentiments, in terms of the sentiments distribution of business itself and its surrounding businesses. We use the Getis-Ord Gi* statistic (Ord and Getis 2010), a spatial statistical approach, to determine the clustering pattern of parking sentiments. Getis-Ord Gi* finds where high and low sentiment ratios cluster spatially. The GIS Gi* statistic is estimated for each business with a *z*-score. The larger the *z*-score is, the more intense the hot spot clusters of high sentiment scores. The smaller the *z*-score is, the more intense the cold clusters of low sentiment scores.

To have a better visualization, after obtaining the *z*-score of each business, we use Inverse Distance Weighted interpolation (ESRI 2018) to map the clustering patterns from Gi* hot and cold spots, shown in Fig. 2. In addition, business density is also illustrated using contour lines for visual comparison, ranging from 0 to 200 businesses per sqkm, in increments of 5 businesses per sqkm (Fig. 2). Overall, the map shows distinct patterns, and negative sentiment clusters are clearly associated with the central business districts of all four cities, evident from the densest business clusters.

However, not all business districts are clusters of negative parking sentiment. Suburban commercial areas, such as those on the north side of Phoenix and south and west sides of Las Vegas actually show clusters of positive parking sentiment.

Analysis of the Effect of the Provision of Parking on Parking Experiences

We use generalized linear mixed-effect (GLME) models to further evaluate how the built environment affects parking sentiment across cities and activity types (Mcculloch and Neuhaus 2001; Zhang et al. 2016). A GLME model is an extension of classical linear regression models. The standard form of a GLME model is

$$y_i|b \sim \mathrm{Distr}\left(\mu_i, \frac{\sigma^2}{\omega_i}\right) \qquad g(\mu) = X\beta + Zb + \delta \tag{1}$$

where y is the response variable, the sampling unit i in our model represents the ith business (each business has a unique business id). The ith response variable y_i corresponds to the averaged sentiment score for this business. In our case, the response variable is the averaged sentiment score of a business, showing an overall parking experience of the parking environment of a business. β is fixed-effects, representing destination activity types, parking management strategies, and several built environment factors. b is random-effects, which is associated with individual experimental units drawn at random from the population and account for variations between groups. In our case, the random-effects variable refers to the city where the business is located in. Distr is the distribution of y given by b, which assumes the distribution of the response variable conditioned on the random-effects variable belongs to the exponential family (Mcculloch and Neuhaus 2001). μ is the mean of y given by b, σ^2 denotes the dispersion parameter, and ω_i represents the weight for observation i. g denotes the link function that describes the relationship between μ and a linear combination of the predictors. Therefore, the mean response μ is given by

$$\mu = g^{-1}(\eta), \tag{2}$$

where g^{-1} is the inverse of the link function, and η is the linear predictor of the mixed effects. We use the function *glmer()* from the R package *lme4* (Bates et al. 2014) for fitting the generalized linear mixed-effects models. In particular, we set "family = binomial(link = 'logit')" in *glmer()*, which specifies the conditional distribution to be binomial. The *glmer()* allows us to fit a generalized linear mixed model incorporating both our fixed-effects parameters (business categories, parking supply, neighboring parking attributes, and built environment variables) and random effect variable (city) in a linear predictor, via maximum likelihood (Bates et al. 2017). By introducing both fixed and random effects, GLME models are useful for cross-sectional data where the response variable may be other than normally distributed (McCulloch and Neuhaus 2005). Detailed settings of each model can be found in the following sections.

In Table 2 a detailed description of variables used in GLME models is shown, showing four categories of variables: (1) business categories; (2) business parking supply; (3) neighborhood parking attributes; and (4) built environment characteristics. The variables in the built environment category are obtained from the US Environmental Protection Agency's Smart Location Database (SLD) (United States Environmental Protection Agency 2014). Because the SLD is limited to the United States and no similar database exists for Canada, Model 3, which includes built environment variables is limited to the 3 US cities, excluding Toronto.

Model 1: The Relationship of Parking Sentiment to Parking Supply

Model 1, which estimates business-level parking sentiment scores, split the dataset 90/10 into training and testing data. In this model, we seek to understand how the types of parking supplied by a business predicts parking sentiment. Independent variables include business parking attributes, activity types, and cities of training dataset to predict the sentiment score in the testing dataset. We tested a series of combinations of independent variables and random effects, such as 1. Parking attributes (5 variables) + activity type + one level of grouping (city); 2. Parking attributes and two levels of grouping (activity types, city); 3. Parking attributes + activity types + their interaction factors + random effects (city). RMSEs (root mean square errors) of these alternative models are used as metrics to measure model performance.

The results of final fitted Model 1 with the minimum RMSE are shown in Table 3. The coefficients are log-odds scaled, shown with standard errors, test statistics (z values) and p-values. Effects of business categories can be more readily interpreted through effect plots in Fig. 3a. We observe that business categories, such as Restaurants and Shopping, and most of the parking attributes except the parking garage significantly explain business sentiment scores. If a business reports that "street parking" is available, its parking sentiment score is significantly lower. All else equal, street parking is an indicator of a more traditional commercial environment, which results in a more challenging parking search, occurring in traffic and across a wider area (Wijayaratna and Wijayaratna 2016). Conversely, if a business has parking validation or its own parking lot, parking experiences will be more positive. Garage parking itself

Table 2 Descriptions of variables

Category	Variable abbreviation	Description	Data source	Data type
Business category	Active Life	Business category is Active Life	Yelp	Binary
	Arts	Business category is Arts	Yelp	Binary
	Automotive	Business category is Automotive	Yelp	Binary
	Health	Business category is Health	Yelp	Binary
	Hotels and Travel	Business category is Hotels and Travel	Yelp	Binary
	Nightlife	Business category is Nightlife	Yelp	Binary
	Other	Business category is Other	Yelp	Binary
	Restaurants	Business category is Restaurants	Yelp	Binary
	Service	Business category is Service	Yelp	Binary
	Shopping	Business category is Shopping	Yelp	Binary
Business parking availability attributes	Parking valet	Business parking valet is available	Yelp	Binary
	Parking lot	Business parking lot is available	Yelp	Binary
	Street parking	Business street parking is available	Yelp	Binary
	Parking garage	Business parking garage is available	Yelp	Binary
	Validated parking	Business validated parking is available	Yelp	Binary
Neighboring parking attributes[a]	Neighbor valet	Neighboring valet parking index	Yelp	Numeric
	Neighbor lot	Neighboring parking lot index	Yelp	Numeric
	Neighbor street	Neighboring street parking index	Yelp	Numeric
	Neighbor garage	Neighboring parking garage index	Yelp	Numeric
	Neighbor validated	Neighboring validated parking index	Yelp	Numeric
Built environment[b]	Population density	Population density	SLD[c]	Numeric
	Business density	Yelp business density	Yelp	Numeric
	Job density	Gross employment density (jobs/acre)	SLD	Numeric
	Land use-mix index	Household workers per job equilibrium index	SLD	Numeric
	Road density	Total road network density	SLD	Numeric
	Auto access to jobs index	Jobs within 45 min auto travel time	SLD	Numeric
	Transit access to jobs index	Jobs within 45-min transit commute time	SLD	Numeric
	Distance to transit stop	Distance from population weighted centroid to nearest transit stop	SLD	Numeric

[a]Neighboring parking attributes measure the same modalities as individual business parking attributes, but are the distance-weighted average of those factors for all businesses within 1.5 km of each business. They are normalized from [0,1] when fitting the models

[b]All the built environment variables are calculated at census block group level, and their basic statistics is shown in Appendix 2

[c]SLD is the Smart Location Database (United States Environmental Protection Agency 2014)

is insignificant, possibly because it is easier to find but less convenient to access and costly. Valet parking has a negative relationship to parking sentiment, implying that drivers view valet parking as time-consuming, expensive, and risky. Model 1 does not directly measure parking demand or traffic congestion, but it confirms the intertwined relationship between the type of parking available at businesses and affective experience.

Model 2: Parking Sentiment Relation to Neighboring Parking Attributes

Model 2 adds parking attributes for the neighborhood surrounding each business as independent variables in the model. Neighboring parking attributes measure the same modalities as individual business parking attributes, but are the distance-weighted average of those factors for all businesses within 1.5 km of each business:

Similar to Model 1, we built our Model 2 by testing different combinations of the independent variables, and the best-fitting Model 2 is shown in Table 4. Figure 3b gives a more direct comparison of the effects of business categories. Compared to Model 1, the "Service" type becomes insignificant. Controlling for other factors, only two neighborhood parking management strategy variables are significant. Intriguingly, for both validated and valet parking, increased supply in the neighborhood have the inverse effect that they do for an individual business. The provision of valet parking by neighboring businesses increases parking sentiment, suggesting it may alleviate neighborhood parking demand.

Table 3 GLME modeling results of Model 1

Model 1: The relationship of parking sentiment to parking supply
Dependent variables: parking sentiment score
Random effect: City
Data: Yelp data for Charlotte, Las Vegas, Phoenix, Toronto

Weights: review count per business				
AIC	BIC	logLik	Deviance	df.resid
95,133	95,208	−47,556	95,113	13,855
Scaled residuals (Model 1)				
Min	1Q	Median	3Q	Max
−13.047	−0.718	−0.152	0.632	15.205
Random effects (Model 1)				
Groups	Name	Variance	Std. Dev.	
City	(Intercept)	0.0004908	0.02215	
Number of obs: 23,512				
Groups: City, 4				

Fixed effects (Model 1):

Term	Estimate (p-value)	Std. error	Statistic (z value)
(Intercept)	0.050***	0.018	2.779
Nightlife	−0.039***	0.014	−2.755
Restaurants	−0.049***	0.014	−3.523
Service	0.041**	0.017	2.415
Shopping	−0.068***	0.015	−4.469
Parking valet TRUE	−0.040***	0.006	−7.230
Parking lot TRUE	0.012***	0.004	3.198
Street parking TRUE	−0.046***	0.005	−9.514
Validated parking TRUE	0.045***	0.015	3.032

Model 1 validation	RMSE
Best model (Model 1)	0.152
Alternative Model a1	0.155
Alternative Model b1	0.158

Alternative Model a1: including interaction factors of business categories and cities

Alternative Model b1: including interaction factors of business categories and business parking availability attributes

GLME fit by maximum likelihood (Laplace Approximation) ['glmerMod'] in R package *lme4* (Bates et al. 2017, p. 4)

***Significant at the 99% level

**Significant at the 95% level

*Significant at the 90% level

Increased validated parking for neighboring businesses reduces parking sentiment, suggesting that even if validated parking is beneficial for an individual, as a whole it is an indicator of a more costly, limited parking supply. The AIC of model 2 is lower than model 1, showing that the addition of the neighborhood variables provides a better fit to the response variable overall.

Spatial Distribution of Model Residuals

We use a Gi* hotspot analysis to assess the spatial distribution of the parking sentiment residuals from Model 2, the best-fitting model. The residual is the difference between the predicted sentiment score and the actual score in testing dataset. Residual distribution plots for each city are shown in Fig. 4. Blue points show clusters of residuals with larger negative values, suggesting sentiment "underprediction", while the red points are clustered residuals with larger positive values, showing the clusters of sentiment "over-prediction". Put another way, blue areas are places where the parking supply model predicts parking sentiment is worse than it actually is, and red areas are where the model predicts sentiment is better than it really is. In general, positive or negative residual clusters are infrequent in all cities, including in

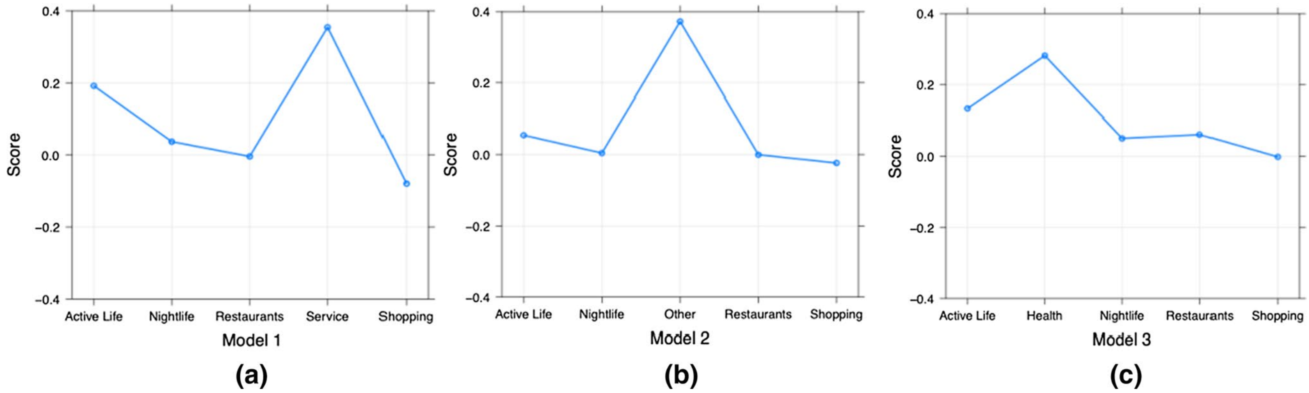

Fig. 3 Business category effect plots for Model 1, Model 2, and Model 3. *The *y*-axis label "score" is the predicted business parking sentiment score

central business districts. However, sentiments in much of central Charlotte are actually better than the model predicts, while areas north and east of central Phoenix have worse sentiments than predicted. These spatial clusters suggest that other factors besides the type of parking supply may be influencing parking experiences in these areas. Planners may be able to use clustered residuals to identify neighborhoods where parking experiences deviate from expectations.

Analysis of the Effect of Other Built Environment Factors on Parking Experiences

Model 3: Understanding Built Environment Effects on Parking Experience

Models 1 and 2 use qualitative parking supply measures to predict parking sentiment, but they do not include the information about the built environment. Model 3 adds several built environment variables from the Smart Location Database (United States Environmental Protection Agency 2014) at Census Block Group (CBG) level. It includes built environment variables (Ewing and Cervero 2001, 2010), such as population density, business density, a land use-mix index and street density, and transportation accessibility variables (auto access and public transportation access to jobs, and distance to transit stops). The description of the selected variables is shown in Table 2. We spatially join the built environment variables (calculated at the CBG level) to each business. The fitted GLME Model 3 results are shown in Table 5.

The parameter estimates with significant p-values for Model 3 are shown in Table 5 and effect sizes of business categories illustrated in Fig. 3c. For business-specific parking attributes, street parking is significantly negatively correlated with sentiment, as is valet parking at the business. Conversely, availability of garage and validated parking are positively associated with sentiment. For neighborhood-wide measures, neighborhood street parking has a strong negative effect on parking sentiment. Unlike their business-specific counterparts, the neighborhood parking garage index and validated parking index are negative. This suggests that these strategies are generally effective for businesses, but as more businesses take advantage of those strategies in the area, the benefit to sentiment is reduced.

Most of the built environment factors are statistically significant, other than population density and road density, which are therefore excluded from Model 3. Unsurprisingly, business density has significantly negative effects on parking sentiment score, as more businesses result in relative parking scarcity. Perhaps relatedly, the land use-mix index has a significantly negative effect on parking sentiment. Planners have emphasized the importance of promoting a mixed-use development in dense areas. These results bear out the idea that when there is a more diverse set of uses in a neighborhood, controlling for all else, finding parking is more difficult and less appealing. The auto access to jobs index is negatively associated with parking sentiment. This reinforces that auto-based accessibility, based on network measures of impedance, explicitly does not account for where cars will park at those destinations.

Most intriguingly, distance to the nearest transit stop is a statistically significant predictor of negative parking sentiment with a small effect. The longer the distance between a business and the nearest transit facility, the lower the sentiment score. This effect suggests that cross-modal effects on sentiment are possible, and that transit-based mobility may alleviate perceived scarcity, and negative sentiment, for parkers. Future research can investigate how cross-modal travel choices vary in places with limited parking supplies but higher or lower levels of transit access.

Table 4 GLME modeling results of Model 2

Model 2: Parking sentiment relation to neighboring parking attributes
Dependent variables: parking sentiment score
Random effect: City
Data: Yelp data for Charlotte, Las Vegas, Phoenix, Toronto

Weights: review count per business				
AIC	BIC	logLik	Deviance	df.resid
86,695	86,791	−43,334	86,669	12,542
Scaled residuals (Model 2)				
Min	1Q	Median	3Q	Max
−13.521	−0.719	−0.152	0.645	14.638
Random effects (Model 2)				
Groups	Name	Variance	Std. Dev.	
City	(Intercept)	0.0004046	0.02011	
Number of obs: 23,512				
Groups: City groups: 4				

Fixed effects (Model 2):

Term	Estimate	Std. error	Statistic (z value)
(Intercept)	0.058***	0.018	3.303
Nightlife	−0.050***	0.014	−3.478
Other	0.318*	0.184	1.731
Restaurants	−0.055***	0.014	−3.858
Shopping	−0.078***	0.016	−5.010
Parking valet TRUE	−0.027***	0.006	−4.601
Parking lot TRUE	0.008*	0.005	1.784
Street parking TRUE	−0.055***	0.005	−10.643
Parking garage TRUE	−0.016***	0.006	−2.799
Validated parking TRUE	0.090***	0.015	5.890
Neighborhood parking valet	1.157***	0.254	4.550
Neighborhood validated parking	−0.218***	0.079	−2.741

Model 2 validation	RMSE
Best model (Model 2)	0.154
Alternative Model a2	0.157
Alternative Model b2	0.157

Alternative Model b1: including interaction factors of business categories and cities

Alternative Model b2: including interaction factors of business categories and business parking availability attributes

Business category is a dummy variable, the base category is “Active Life”

GLME fit by maximum likelihood (Laplace Approximation) [’glmerMod’] in R package *lme4* (Bates et al. 2017, p. 4)

***Significant at the 99% level

**Significant at the 95% level

*Significant at the 90% level

Discussion and Conclusion

This study uses online Yelp reviews to evaluate how parking supply and the built environment shape parking experiences in four North American cities through text mining and statistical methods. The results show that transportation system management and the built environment have significant impacts on how individuals experience daily travel. The spatial hotspot analysis shows that negative parking sentiment clusters are associated with central business districts (CBDs) in all four cities but are not always associated with business clusters outside of CBDs. Model 1 suggests that activity type is an important predictor of parking sentiment, so for example, dining and shopping significantly explain sentiment. The type of parking available is also significantly associated with parking sentiment.

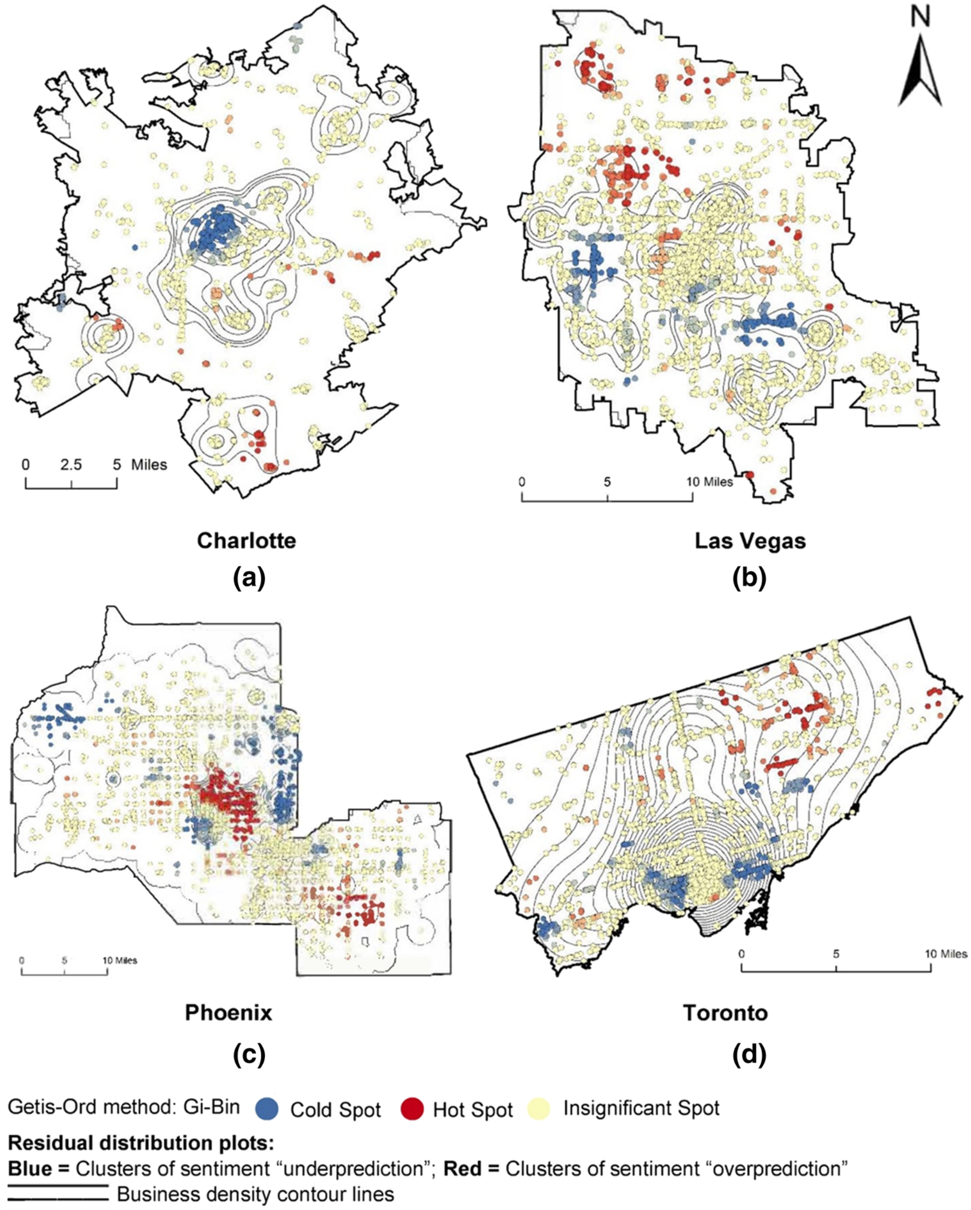

Fig. 4 Model 2 residual spatial distribution plots

Specifically, if a business has parking validation or its own parking lot, parking experiences will be more positive. When we include the impacts of neighborhood-wide parking options on parking sentiment, we find that the provision of valet parking by neighboring businesses increases parking sentiment at a local business, and increased validated parking for neighboring businesses reduces parking sentiment. The final model examines how built environment factors affect parking sentiment for the three US cities. The result shows that most of the built environment factors are statistically significant to parking experience, other than population density and road density. In particular, auto access to jobs index has a stronger significantly negative effect on parking sentiment score. Conversely, proximity to transit has a significantly positive effect on parking sentiment, indicating that provision of alternative access modes can enhance experiences for drivers as well.

To what degree should we be concerned with these findings? Most straightforwardly, localized sentiment measures can be used as indicators of how well transportation systems are functioning, in the eyes of their users, whether for parking or other modes, such as transit or walking. Planners working in the public interest do not typically frame system performance in terms of sentiment. However, in each of the four cities, we observe that parking sentiment varies significantly between central and peripheral business clusters. This represents a potential competitive advantage among neighborhoods that can be directly quantified using crowdsourced information. The analysis demonstrates that the provision of parking and built environment factors affect

Table 5 Results of Model 3 built environment effects on parking experience

Model 3: Results of built environment effects on parking experience
Dependent variable: business parking sentiment score
Data: Yelp data for Charlotte, Las Vegas, Phoenix; Census data; Smart Location Database
Variables' description: Table 2

Weights: review count per business				
AIC	BIC	logLik	Deviance	df.resid
86,975	87,100	−43,470	86,941	11,484
Scaled residuals				
Min	1Q	Median	3Q	Max
−13.124	−0.791	−0.189	0.672	15.434
Random effects				
Groups	Name	Variance	Std. Dev.	
City	(Intercept)	0.0008349	0.02889	
Number of obs: 11,498; City groups: 3				

Fixed effects:

Variable category	Variable abbreviations	Estimate (*p*-value)	Std. error	Statistic (*z* value)
	(Intercept)	0.201***	0.023	8.787
Business category	Health	0.075**	0.035	2.133
	Nightlife	−0.041**	0.020	−2.028
	Restaurants	−0.035*	0.020	−1.782
	Shopping	−0.067**	0.022	−3.069
Business parking availability attributes	Parking valet TRUE	−0.038***	0.007	−5.638
	Street parking TRUE	−0.049***	0.007	−6.960
	Parking garage TRUE	0.015**	0.006	2.361
	Validated parking TRUE	0.065***	0.017	3.821
Neighborhood parking attributes	Neighborhood garage index	−0.356*	0.212	−1.680
	Neighborhood street index	−0.927*	0.562	−1.650
	Neighborhood validated parking index	−0.149*	0.079	−1.883
Built environment	Business density	−0.015***	0.005	−2.919
	Land use-mix index	−0.047***	0.012	−3.800
	Auto access to jobs index	−0.136***	0.016	−8.283
	Distance to transit stop	−0.028**	0.011	−2.454

Model validation

Model	RMSE
Best model (Model 3)	0.145
Alternative Model a3	0.151
Alternative Model b3	0.154

Alternative model a3: including interaction factors of business categories and cities; Alternative model b3: including interaction factors of business categories and business parking availability attributes

Business category is a dummy variable, the base category is "Active Life"

GLME fit by maximum likelihood (Laplace Approximation) ['glmerMod'] in R package *lme4* (Bates et al. 2017, p. 4)

***Significant at the 99% level

**Significant at the 95% level

*Significant at the 90% level

parking sentiment. Parking sentiment, as a part of travelers' activity experiences of the businesses in cities, might affect the overall sentiment of their experiences. Business owners also would like to know more about the ease or difficulty that people have getting to them. Shoup (2018) observes that parking research is most underdeveloped in its understanding of the "political calculus" of parking management strategies. How planners and locality can gain acceptance for parking best practices, including built environment strategies that modify demand, remains little understood. This study suggests that everyday parking experiences may play an important role in support for parking management and built environment planning.

While travelers' positive sentiments and satisfaction would appear, at first glance, to be desirable outcomes in the aggregate, we fully recognize that increasing parking satisfaction could function at cross-purposes with broader transportation goals: providing ample, low-cost parking makes drivers happier, and activity-rich neighborhoods with a dense mix of uses decrease driver satisfaction. We do not propose that planners actively seek to increase dissatisfaction with driving and parking to encourage mode shift, but we do observe that to be competitive with suburban districts, planners will need to counterbalance dissatisfaction with car-based access with other advantages, whether related to transportation or other neighborhood features. The results showing positive parking sentiment associated with transit proximity suggest that multimodal commercial districts where travelers of on multiple modes are satisfied with their experiences are possible. Further research can examine how sentiments toward different travel modes, as well as toward other local factors, interact in a neighborhood.

Methodologically, results suggest that parking sentiment scores and affective experience measures estimated from textual data, such as Yelp reviews, can be tied to specific geographic locations and further be analyzed with spatial statistic methods. The Getis-Ord Gi* statistic, in our case, successfully detects spatial clusters of parking sentiment. The study demonstrates that emerging data mining and statistical methods can successfully leverage big data to reveal travel experiences and their relationship to urban contexts, suggesting an effective way to obtain useful transportation information. Opportunities to explore different approaches to textual data analysis, such as topic modeling, could supply additional information regarding how transportation system performance is correlated with the built environment.

This analysis uses multiple approaches to deal with issues of self-selection and representativeness in Yelp data, which are an issue more broadly with all self-selecting LBSNs. Addressing these issues requires a combination of empirical methods to diagnose or control for potential biases, as well as clear caveats and recognition of potential effects of biases that cannot be controlled. For gender and income, two demographic factors that could reasonably be imputed or analyzed by proxy, we found that neither has a significant relationship to parking sentiment. However, we were not able to impute or assess the effects of age, race, or ethnicity of reviewers. For age, increased need for comfort during travel and a reduced desire to walk longer distances for utilitarian purposes has been documented (Hess 2012; Keadle et al. 2016). Therefore, we might expect that parking sentiment among older adults would be lower for parking strategies that require more utilitarian walking. For race and ethnicity, parking experiences that require interaction with individuals, such as valet parking or even parking validation, could potentially result in distinctively racist or alienating experiences. These experiences are well documented in the context of transit, taxi, and rideshare travel, but so far undocumented for parking experiences (Purifoye 2015; Sarriera et al. 2017). While this study's research design would not be able to address this question, further research could potentially mine Yelp's extensive reviews to identify whether parking experiences are perceived specifically as racist or alienating.

Broadly, working with LBSNs requires clear understanding of each dataset's strengths and limitations. Ultimately, some research questions may continue to require purpose-built survey efforts that reach populations that do not participate as readily in LBSNs, or where LBSNs do not supply critical information to answer those questions. Additional factors, such as time of day, or day of week, may affect the experience of parking as well (Litman 2006; Millard-Ball et al. 2014b). Still, while a survey or online poll can provide detailed information about parking sentiment and parking behaviors, the reality for urban planners is that these surveys are rarely undertaken, and usually only in specific neighborhoods where there has been demand and funding for a parking study. The LBSN-based approach allows for a much broader look across a city, to allow for better comparisons across neighborhoods and even between cities, allowing for more empirically robust understandings of parking management and build environment strategies that result in positive parking experiences that are compatible with broader goals towards reduced total parking, densification, and multimodal travel in commercial and mixed-use areas. Additionally, in future research, spatially precise parking utilization data could be integrated into analysis of parking experiences to understand how supply, utilization, and travel experiences co-vary in different locations.

Appendix 1

See Fig. 5.

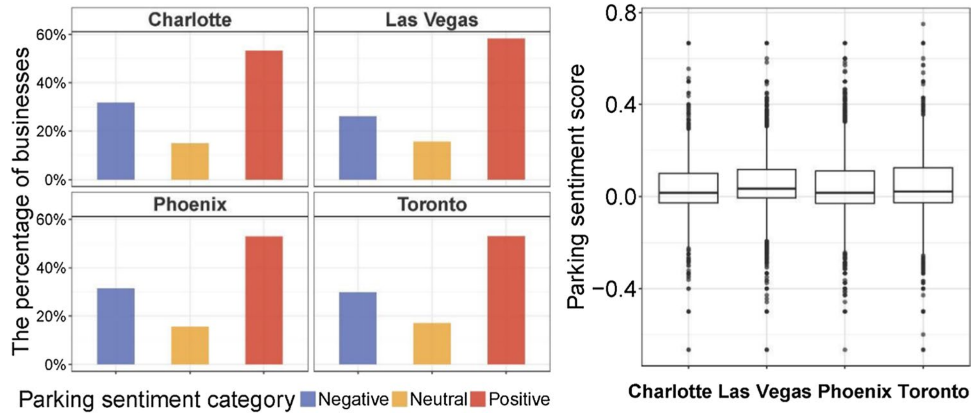

Fig. 5 Parking review sentiment category and score at city-level

Appendix 2

See Table 6.

Table 6 Summary statistics of built environment variables

Variable	Min	1st Quantile	Median	Mean	3rd Quantile	Max
Population density	0	2.613	5.199	6.700	8.905	62.338
Business density	0	0.123	0.274	0.435	0.476	3.279
Job density	0	1.442	3.781	12.946	9.569	158.146
Land use-mix index	0	0.369	0.418	0.445	0.596	1
Road density	0.167	13.65	17.75	17.895	22.31	40.42
Auto access index	516.7	141,850.5	187,169.6	194,411.7	249,124.6	358,133.7
Transit access index 1: jobs within 45-min transit commute time	0	0	0	4698	4638	37,657
Transit access index 2: distance from population weighted centroid to nearest transit stop (meter)	2.32	284.93	489.40	525.48	749.12	1199.23

All the built environment variables are calculated at census block group level, except the business density, all other variables are from the Smart Location Database (United States Environmental Protection Agency 2014)

Author Contributions The authors confirm contribution to the paper as follows: ZJ, ASM: study conception and design; ZJ, ASM: analysis and interpretation of results; ZJ, ASM: draft manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials Some or all data, models, or materials that support the findings of this study are available from the corresponding author upon request.

Code Availability Some or all code that support the findings of this study are available from the corresponding author upon request.

Compliance with Ethical Standards

Conflicts of Interest The authors declare that they have no known competing financial interests or personal relationships that would influence the work reported in this paper.

References

Aggarwal CC, Zhai C (2012) A survey of text clustering algorithms. In: Aggarwal CC, Zhai C (eds) Mining text data. Springer, New York, pp 77–128. https://doi.org/10.1007/978-1-4614-3223-4_4

Ajzen I (1991) The theory of planned behavior. Organ Behav Hum Decis Process 50(2):179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Arribas-Bel D, Bakens J (2019) Use and validation of location-based services in urban research: an example with Dutch restaurants. Urban Stud 56(5):868–884. https://doi.org/10.1177/0042098018779554

Bamberg S, Kühnel SM, Schmidt P (1999) The impact of general attitude on decisions. Ration Soc 11(1):5–25. https://doi.org/10.1177/104346399011001001

Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. ArXiv: 1406.5823 [Stat]

Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B, Grothendieck G, Green P (2017) lme4: linear mixed-effects models using "Eigen" and S4. https://cran.r-project.org/web/packages/lme4/index.html

Börjesson M, Hamilton CJ, Näsman P, Papaix C (2015) Factors driving public support for road congestion reduction policies: congestion charging, free public transport and more roads in Stockholm, Helsinki and Lyon. Transp Res A Policy Pract 78:452–462. https://doi.org/10.1016/j.tra.2015.06.008

Bureau of Transportation Statistics (2018) Transportation statistics annual report. Bureau of Transportation Statistics. https://www.bts.dot.gov/sites/bts.dot.gov/files/docs/browse-statistical-products-and-data/transportation-statistics-annual-reports/Preliminary-TSAR-Full-2018-a.pdf

Caragea C, Squicciarini A, Stehle S, Neppalli K, Tapia AH (2014) Mapping moods: geo-mapped sentiment analysis during hurricane sandy. In: ISCRAM 2014 conference proceedings—11th international conference on information systems for crisis response and management, pp 642–651. https://pennstate.pure.elsevier.com/en/publications/mapping-moods-geo-mapped-sentiment-analysis-during-hurricane-sand

Charlotte Planning, Design, & Development Department (2017) City of Charlotte Zoning Ordinance

Chen C, Gong H, Lawson C, Bialostozky E (2010) Evaluating the feasibility of a passive travel survey collection in a complex urban environment: lessons learned from the New York City case study. Transp Res A Policy Pract 44(10):830–840. https://doi.org/10.1016/j.tra.2010.08.004

Christiansen P, Engebretsen Ø, Fearnley N, Usterud Hanssen J (2017) Parking facilities and the built environment: impacts on travel behaviour. Transp Res A Policy Pract 95:198–206. https://doi.org/10.1016/j.tra.2016.10.025

City of Toronto Council (2018) City of Toronto—zoning by-law

Clark County Department of Comprehensive Planning (2017) Comprehensive master plan for Clark County, Nevada

Ding X, Liu B, Yu PS (2008) A holistic lexicon-based approach to opinion mining. In: Proceedings of the international conference on web search and web data mining—WSDM '08, p 231. https://doi.org/10.1145/1341531.1341561

Dittmar H, Ohland G (2012) The new transit town: best practices in transit-oriented development. Island Press, Washington D.C

ESRI (2018) How IDW works

Evans L, Saker M (2017) Location-based social media: space, time and identity. Springer, Cham

Ewing R, Cervero R (2001) Travel and the built environment: a synthesis. Transp Res Rec J Transp Res Board 1780(1):87–114. https://doi.org/10.3141/1780-10

Ewing R, Cervero R (2010) Travel and the built environment: a meta-analysis. J Am Plan Assoc 76(3):265–294. https://doi.org/10.1080/01944361003766766

Feinerer I (2018) Introduction to the "tm" Package Text Mining in R

Feuerriegel S, Proellochs N (2018) Package "SentimentAnalysis"

Fishbein M, Ajzen I (1975) Belief, attitude, intention, and behavior: an introduction to theory and research. Addison-Wesley Pub. Co., Boston

Forrest T, Pearson D (2005) Comparison of trip determination methods in household travel surveys enhanced by a global positioning system. Transp Res Rec J Transp Res Board 1917:63–71. https://doi.org/10.3141/1917-08

Griffioen-Young HJ, Janssen HJW, Van Amelsfoort DJC, Langefeld JJ (2004) The psychology of parking. In: Proceedings of the ECOMM 2004 conference

Hess DB (2012) Walking to the bus: perceived versus actual walking distance to bus stops for older adults. Transportation 39(2):247–266. https://doi.org/10.1007/s11116-011-9341-1

Inci E (2015) A review of the economics of parking. Econ Transp 4(1–2):50–63. https://doi.org/10.1016/j.ecotra.2014.11.001

INRIX (2017) The impact of parking pain in the US, UK and Germany. http://www2.inrix.com/research-parking-2017

Jiang Z, Mondschein A (2019) Examining the effects of proximity to rail transit on travel to non-work destinations: evidence from Yelp data for cities in North America and Europe. J Transp Land Use 12(1):303–326. https://doi.org/10.5198/jtlu.2019.1409

Kambele Z, Li G, Zhou Z (2015) Travelers' information-seeking behaviors. J Travel Tour Mark 32(1–2):141–152. https://doi.org/10.1080/10548408.2014.986017

Kaplan J (2020) predictrace: predict the race of a given surname using census data (1.2.1) [Computer software]. https://CRAN.R-project.org/package=predictrace

Keadle SK, McKinnon R, Graubard BI, Troiano RP (2016) Prevalence and trends in physical activity among older adults in the United States: a comparison across three national surveys. Prev Med 89:37–43. https://doi.org/10.1016/j.ypmed.2016.05.009

Khanna K (2020) Wru: who are you? Bayesian prediction of racial category using surname and geolocation (0.1–10) [Computer software]. https://CRAN.R-project.org/package=wru

King D, Manville M, Shoup D (2007) The political calculus of congestion pricing. Transp Policy 14(2):111–123. https://doi.org/10.1016/j.tranpol.2006.11.002

Litman T (2006) Parking management best practices. American Planning Association, Chicago

Liu B (2012) Sentiment analysis and opinion mining. Synth Lect Hum Lang Technol 5(1):1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Manville M, Shoup D (2005) Parking, people, and cities. J Urban Plan Dev 131(4):233–245

Marsden G (2006) The evidence base for parking policies—a review. Transp Policy 13(6):447–457. https://doi.org/10.1016/j.tranpol.2006.05.009

Mcculloch C, Neuhaus J (2001) Generalized linear mixed models. Wiley, New York

McCulloch CE, Neuhaus JM (2005) Generalized linear mixed models. In: Armitage P, Colton T (eds) Encyclopedia of biostatistics. Wiley, New York, p b2a10021

Millard-Ball A, Weinberger RR, Hampshire RC (2014) Is the curb 80% full or 20% empty? Assessing the impacts of San Francisco's parking pricing experiment. Transp Res A Policy Pract 63:76–92

Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM (2013) The geography of happiness: connecting Twitter sentiment and expression, demographics, and objective characteristics of place. PLoS ONE 8(5):e64417. https://doi.org/10.1371/journal.pone.0064417

Mjahed B, Lama AM, Archak E, Amr M, Hani S, Chen Y (2017) Exploring the role of social media platforms in informing trip planning. Transp Res Rec J Transp Res Board 2666:1–9. https://doi.org/10.3141/2666-01

Mondschein A (2015) Five-star transportation: using online activity reviews to examine mode choice to non-work destinations. Transportation 42(4):707–722. https://doi.org/10.1007/s11116-015-9600-7

Mondschein A, King DA, Hoehne C, Jiang Z, Chester M (2020a) Using social media to evaluate associations between parking supply and parking sentiment. Transp Res Interdiscip Perspect. https://doi.org/10.1016/j.trip.2019.100085

Mondschein A, King DA, Hoehne C, Jiang Z, Chester M (2020b) Using social media to evaluate associations between parking supply and parking sentiment. Transp Res Interdiscip Perspect 4:100085. https://doi.org/10.1016/j.trip.2019.100085

Mullen L (2020) gender: predict gender from names using historical data (R package version 0.5.4) [Computer software]. https://github.com/ropensci/gender

Ord JK, Getis A (2010) Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 27(4):286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x

Pandhe A, March A (2012) Parking availability influences on travel mode: Melbourne CBD offices. Aust Plan 49(2):161–171. https://doi.org/10.1080/07293682.2011.616177
Parkany E, Gallagher R, Viveiros P (2004) Are attitudes important in travel choice? Transp Res Rec J Transp Res Board 1894(1):127–139. https://doi.org/10.3141/1894-14
Phoenix City Council (2015) PlanPHX general plan update
Purifoye GY (2015) Nice-nastiness and other raced social interactions on public transport systems. City Community 14(3):286–310. https://doi.org/10.1111/cico.12116
Quantcast (2017) Yelp audience insights and demographic analytics. https://www.quantcast.com/yelp.com/demographics/WEB?country=US
Roberts H, Sadler J, Chapman L (2019) The value of Twitter data for determining the emotional responses of people to urban green spaces: a case study and critical evaluation. Urban Stud 56(4):818–835. https://doi.org/10.1177/0042098017748544
Rose MH (1990) Interstate: express highway politics, 1939–1989 (Rev. ed). University of Tennessee Press, Knoxville
Sarriera JM, Álvarez GE, Blynn K, Alesbury A, Scully T, Zhao J (2017) To share or not to share: investigating the social aspects of dynamic ridesharing. Transp Res Rec J Transp Res Board 2605(1):109–117. https://doi.org/10.3141/2605-11
Saxena A, Chaturvedi KR, Rakesh S (2018) Analysing customers reactions on social media promotional campaigns: a text-mining approach. Paradigm 22(1):80–99. https://doi.org/10.1177/0971890718759163
Schober P, Boer C, Schwarte LA (2018) Correlation coefficients: appropriate use and interpretation. Anesth Analg 126(5):1763–1768. https://doi.org/10.1213/ANE.0000000000002864
Sekar A, Chen RB, Cruzat A, Nagappan M (2017) Digital narratives of place: learning about neighborhood sense of place and travel through online responses. Transp Res Rec J Transp Res Board 2666(1):10–18. https://doi.org/10.3141/2666-02
Shoup DC (2006) Cruising for parking. Transp Policy 13(6):479–486. https://doi.org/10.1016/j.tranpol.2006.05.005
Shoup DC (2011) The high cost of free parking (updated). Planners Press, American Planning Association, Chicago
Shoup D (2018) Parking and the city. Routledge, New York
Sokolova M, Japkowicz N, Szpakowicz S (2006) Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In: Sattar A, Kang B (eds) AI 2006: advances in artificial intelligence, vol 4304. Springer, Berlin, pp 1015–1021. https://doi.org/10.1007/11941439_114
Statistics Canada (2016) Data products, 2016 census
Stevens MR (2017) Does compact development make people drive less? J Am Plan Assoc 83(1):7–18. https://doi.org/10.1080/01944363.2016.1240044
Stone PJ, Bales RF, Namenwirth JZ, Ogilvie DM (2007) The general inquirer: a computer system for content analysis and retrieval based on the sentence as a unit of information. Behav Sci 7(4):484–498. https://doi.org/10.1002/bs.3830070412
Stopher PR, Greaves SP (2007) Household travel surveys: where are we going? Transp Res A Policy Pract 41(5):367–381. https://doi.org/10.1016/j.tra.2006.09.005
Stopher P, Jiang Q, FitzGerald C (2005) Processing GPS data from travel surveys. Australasian Transport Research Forum (ATRF), 28th, 2005, Sydney, New South Wales, Australia, p 28
Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. Comput Linguist 37(2):267–307. https://doi.org/10.1162/COLI_a_00049
Tung E (2015) Automatically categorizing Yelp businesses
United States Environmental Protection Agency (2014) Smart location database. https://edg.epa.gov/data/Public/OP/SLD/SmartLocationDB.zip
U.S. Census Bureau (2016) U.S. Census Bureau: American community survey, 2016 5-year estimates
Verplanken B, Aarts H (1999) Habit, attitude, and planned behaviour: is habit an empty construct or an interesting case of goal-directed automaticity? Eur Rev Soc Psychol 10(1):101–134. https://doi.org/10.1080/14792779943000035
Weinberger R, Kaehny J, Rufo M (2010) US parking policies: an overview of management strategies. The TRIS and ITRD database
Wijayaratna S, Wijayaratna KP (2016) Quantifying the impact of on-street parking on road capacity: a case study of Sydney arterial roads. In: TRB 95th annual meeting compendium of papers
Yelp (2018a) API 2.0: all category list. Yelp for Developers
Yelp (2018b) Yelp dataset challenge
Yelp (2019) Yelp Factsheet. https://www.yelp.com/factsheet
Zhang X, Zhou Y, Ma Y, Chen BC, Zhang L, Agarwal D (2016) GLMix: generalized linear mixed models for large-scale response prediction. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 363–372. https://doi.org/10.1145/2939672.2939684
Zhou X, Wang M, Li D (2017) From stay to play—a travel planning tool based on crowdsourcing user-generated contents. Appl Geogr 78:1–11. https://doi.org/10.1016/j.apgeog.2016.10.002