The Tourist's Image of the City: A comparative analysis of the visual features and textual themes of interest across three global metropolises

Rohit Priyadarshi Sanatani[1], Takehiko Nagakura[1], Daniel Tsai[1]

[1] Massachusetts Institute of Technology (MIT), USA
sanatani@mit.edu; takehiko@mit.edu; dtsai@mit.edu

Abstract. Tourist attractions play a major role in shaping 'mental images' of cities. The growing availability of urban big-data in recent years has opened up novel lines of inquiry into the nuances of urban imageability and sentiment. Drawing upon crowdsourced hybrid data in the form of both textual descriptions as well as photographs for 750 tourist attractions across Boston, Singapore and Sydney, this work compares the predominant themes of discussion and visual features of interest that shape tourist sentiment towards these cities. The study collects over 3500 user reviews and uses Latent Dirichlet Allocation (LDA) for the extraction of high-level topics of discussion. Object detection is also run on over 6000 photographs, and unsupervised clustering is carried out on extracted features to identify clusters of visual elements which capture tourist attention. The findings reinforce the popular identity of Boston as a city steeped in history, while strong perceptions of nature and greenery emerge from Singapore. Tourist interest in Sydney is dominated by specific anchors such as the Sydney Harbor Bridge.

Keywords: Urban Tourism, Topic Modeling, Sentiment Analysis, Unsupervised Clustering, Big Data

1 Introduction

Tourist destinations play a major role in defining the popular 'image' of a city. Be it the Statue of Liberty in New York, the London Bridge, or the Colosseum in Rome – these anchors often become synonymous with their respective urban centers themselves, and remain ingrained in everyday public consciousness. The tourists' image of a city often also comprises far more nuanced visual and experiential elements, which go beyond specific sites or structures. These elements become the building blocks of a city's 'identity', and are often considered to be ‘intangible’ elements that can only be experienced, but not easily analyzed.

While there have been numerous studies over the past decade inquiring into urban tourist sentiment, the exponential growth of big-data over the past few years have opened up valuable methodological approaches in this regard (Alaei et al. 2019). Social media data from platforms such as Twitter, Filckr and Tripadvisor has emerged as a popular data-source for such recent research directions (Galí et al. 2015, Valdivia et al. 2017, Cardone et al. 2021). Moreover, geolocated Point of Interest (POI) data available through platforms such as Google Places, and Foursquare provide valuable access to user-reviews, user-photographs, and visitation patterns, at a very high spatio-temporal resolution (Marti et al. 2019).

A number of methodological approaches have been employed in the past for inquiries into the experiential dimensions of the urban realm. While sentiment analysis is a popular method in this regard, topic modeling has been gaining popularity as a tool for exploring the different dimensions of subjective experience in cities (Kovacs-Gyori et al. 2018, Ghahramani et al. 2021). There have, however, been very few lines of inquiry that draw upon hybrid textual and photographic data to extract both visual features of interest as well as themes of discussion, as semantic content over and above quantitative sentiment scores. Such an approach can allow for the cataloging of structured high-level data with regards to the dimensions of tourist experience, and also has potential to be integrated into data-query and visualization platforms.

This body of research inquires into the major textual themes of discussion and visual features of interest contributing to popular sentiment amongst citizens/visitors towards urban tourist attractions in three cities – Boston, Singapore and Sydney. It further investigates the similarities and differences in these patterns across these urban centers, thus exploring the unique experiences that remain anchored to their respective perceptual identities.

2 Methodological Framework

2.1 Data collection

In order to extract meaningful content with regards to features of interest and themes of discussion, both visual as well as textual data becomes important. Visual data in the form of crowdsourced tourist photographs can provide valuable cues into the kinds of features within the city that the tourists were interested in. Textual data, on the other hand, comprise tourist generated descriptions or appraisals of urban anchors and environments. Such kind of data (in the form of reviews, comments etc.) provide insights into the themes of discussion emerging from different tourist destinations, along with the predominant sentiment associated with these themes.

While numerous platforms exist today that can serve as valuable repositories of both the above-mentioned data forms, this work set forth three primary criteria for selecting the data source. Firstly, the data needed to be of high spatial resolution. In other words, the data should be geo-located with an accuracy that is high enough for meaningful inferences with respect to its location to emerge. Secondly, the data should also be of high temporal resolution. In other words, the exact time and date of the review/photographs should also be available. Finally, the data should contain adequate and relevant semantic content with regards to the different qualitative aspects of tourist experience, which is often a challenge.

Based on these criteria, Point of Interest (POI) data sources emerged as the most appropriate in this regard. While there are numerous major POI data aggregators globally (such as Google Places, Foursquare and Yelp), this body of research adopted Google Places as the primary data source for analysis.

Data for points of interest listed as ‘tourist attraction’ was programmatically collected across the three cities using the Google Places API. The first round of data collection involved the collection of basic place details using the ‘Nearby Search’ API call. Query locations and radii were set at regular intervals across the cities, and places data queried from each location. The data included attributes such as location, place name place id, rating, total user ratings and the like. Next, for each of the places now listed in the dataset, 5 top reviews and 10 photograph ids were programmatically collected using the ‘Place Details’ API call. For each review, the query returned attributes such as author name, language, rating, review text and a timestamp. For each photo id, the query also returned author name, timestamp and the photo url. The final step in the data collection pipeline was to programmatically download each photograph using the photo url, and to store the files with meaningful filenames corresponding to the photo ids recorded in the dataset. **Table 1** below summarizes the volume of data collected through this pipeline.

Table 1. Summary of Google Places data collected

Boston	Singapore	Sydney
139 locations	430 locations	580 locations
603 reviews	1939 reviews	2721 reviews
1241 photos	4086 photos	5587 photos

2.2 Sentiment Analysis

The first step in the analysis pipeline was the extraction of overall sentiment from each review. This was carried out in Python, using the NLTK Vader pre-trained sentiment analyzer (Hutto and Gilbert 2014). This resulted in a single compound sentiment score (between -1 and +1) being associated with each review. While such a method of sentiment analysis does not yield high level understandings of the content of the reviews themselves, it is nevertheless an effective tool for grading the reviews on a linear scale, and thus separating the negative, positive and neutral reviews. The compound scores were then aggregated by location (tourist attraction), and visualized on a map to check for possible spatial dimensions to sentiment distribution (**Fig 1**).

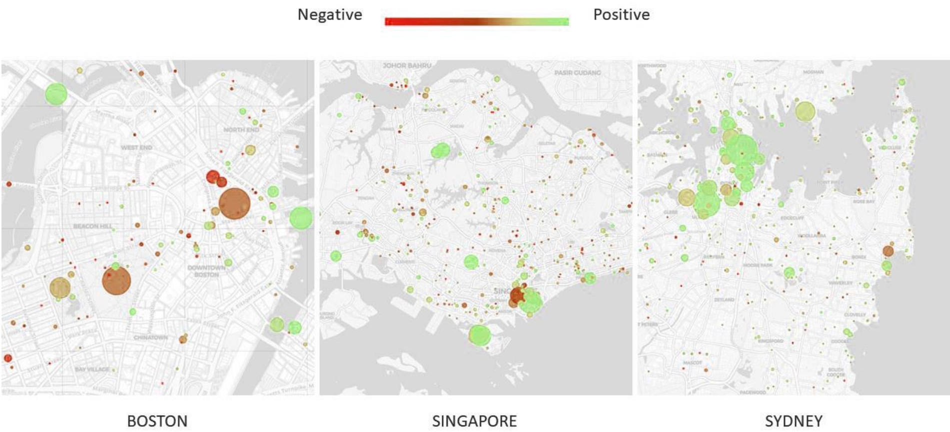

Figure 1. Overall compound sentiment scores for tourist locations across the three cities. Size indicates number of ratings (used here as a proxy for number of visits)

2.3 Topic Modeling

Extraction of overall sentiment, while useful for discriminating between positive and negative reviews, does not provide us with insights into the themes of discussion associated with each review. For this, an unsupervised Latent Dirichlet Allocation (LDA) (Blei et al. 2003) model was implemented on the city-specific datasets, in order to extract the most frequently occurring high level topics that were emerging from the reviews. An LDA approach was preferred to clustering, since a single review could contain more than one topic. Stop words were first removed from each review text, and the remaining words (tokens) were lemmatized (grouping together of inflected forms so that they can be analyzed as a single item). These words were converted into a dictionary of tokens which defined the feature space of the reviews themselves. The reviews were then encoded as bags of words (frequencies of each word in the dictionary). These formed the input vectors for the LDA model.

Based on iterative testing and evaluation, the 10 most common topic-clusters were extracted for each city using LDA **(Fig. 2)**. The clusters were in the form of collections of individual tokens, and corresponding coefficients which described the degree to which they contributed to the topic that defined that cluster. The high-level topics that defined each cluster thus needed to be manually inferred from the individual tokens and their coefficients.

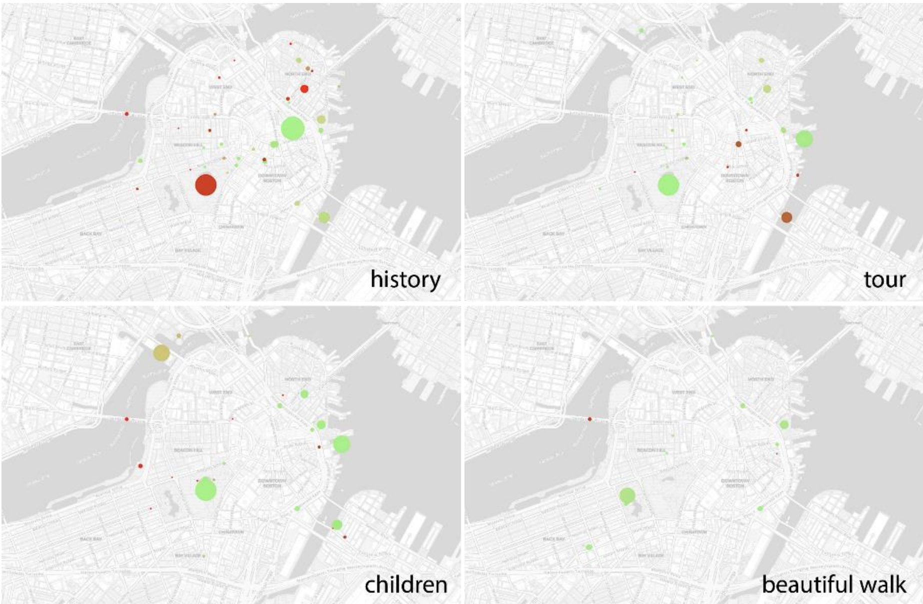

Figure 2. Topic modeling on review texts from Boston, along with their spatial distributions.

The larger aim of this body of research however was to investigate the themes of discussions emerging from tourist locations *in the context of sentiment*. It was necessary to infer not only the overarching topics emerging from these cities, but rather the topics that were consistently associated with negative sentiment, and those associated with positive sentiment. To this end, for each city, the individual reviews that had the highest sentiment ratings as well as those with the lowest were separated. Given the frequency distribution of review sentiment values, this was achieved by filtering all reviews with a sentiment score higher than 0.85, and those with scores lower than -0.25. This generated two separate datasets for each city, containing the reviews with best and worst sentiment scores respectively. The topic modeling workflow was then run independently on both the datasets for each city, in order to extract the high-level topics that were associated with positive and negative tourist sentiment respectively.

2.4 Visual feature extraction and unsupervised clustering

For extraction of visual features of interest, a faster-RCNN object detection model trained on the Open-Images V4 dataset (Kuznetsova et al. 2020) and provided by TensorFlow (Abadi et al. 2016) was run on the tourist photographs from each city. The model inferred the high-level semantic content contained within the images **(Fig. 3)**. This generated a dataset for each photograph containing the top 100 objects detected in them, along with the confidence scores for each detection. For analysis, only the features associated with confidence scores of 0.10 or higher were retained within the dataset.

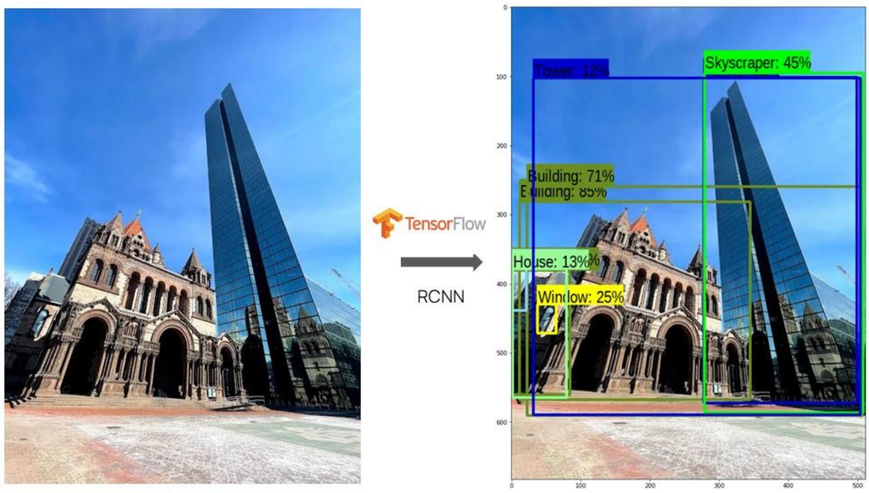

Figure 3. Faster RCNN object detection model for extraction of features of interest.

In order to infer the features that were consistently correlated with tourist interest within the cities, an unsupervised k-means clustering algorithm was used. Each data point associated with each photograph was converted into a bag of features, with the frequency of each feature occurrence recorded within the feature vector. Clustering was implemented in Python, using the scikit-learn library (Pedregosa et al. 2011). Based on iterative implementation, between 10 to 25 clusters were extracted from the photographs of each city.

3 Results – The dimensions of tourist experience

3.1 The Talk of the Town: Themes of tourist discussion

The results of the topic modeling exercise successfully mapped out several interesting dimensions of the unique themes of discussion that remain associated with tourist sentiment in different cities. They also provided valuable insights into the unique symbolic associations that are forged between the cities and the tourists through these seemingly innocuous everyday discussions.

In Boston, most common high-level topics revolved around historic buildings, churches and sites, conducted tours revolving around museums, aquariums, and statues, enjoyable experiences in parks such as beautiful and peaceful walks, and also enjoyable experiences with children and kids. Most of these topics were associated with strong positive sentiment, and overlapped greatly with the topics emerging from the positive review dataset. Topic modeling on the negative review dataset however revealed interesting themes such as complaints about Boston's weather (often in the context of tours being cancelled due to rain), discussions and debates around the history of slavery and the independence movement, consistent frustration with traffic at specific points in the city (such as its tunnels and bridges), and sporadic complaints about having to climbing stairs in some tourist locations.

The discussions emerging from Singapore were radically different. Unlike Boston, there was far less focus on history, and greater focus on nature, landscapes, and serenity. Common topics in this regard included numerous discussions surrounding birds and bird-watching in gardens, small and quiet playgrounds, beautiful and relaxing temples and churches, enjoyable tours in museums, good times with children and family, and very interestingly, numerous discussions surrounding staff at tourist establishments. The positive review dataset also revealed high-points in the tourist experience such as enjoyable rides in parks and jogging activities by the waterfronts. Interestingly, the negative discussions also surrounded very similar themes. There emerged multiple instances of frustration with long entrance queues in front of tourist establishments, bad staff, unpleasant experiences with children in parks, and general complaints about dirt and lack of cleanliness **(Fig. 4)**.

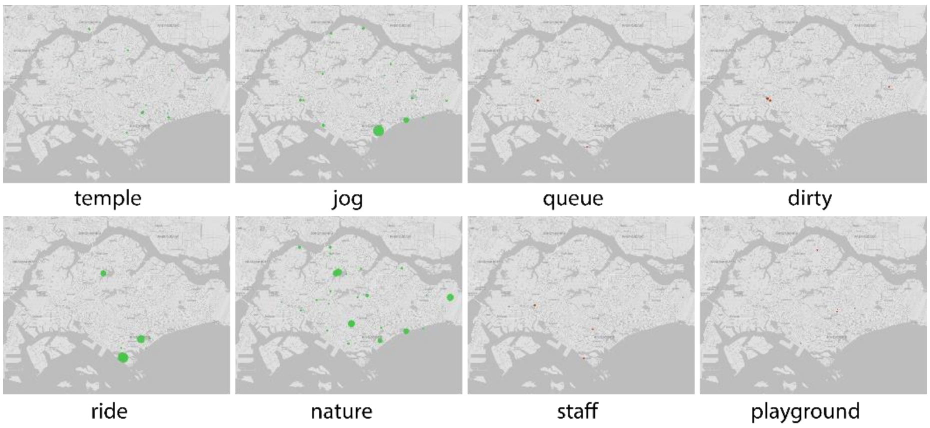

Figure 4. Spatial distribution of themes corresponding to extreme positive and negative tourist sentiment in Singapore.

Sydney, on the other hand, had a very different identity as emergent from tourist discussions. While the Sydney Harbor and specifically the Harbor Bridge recurred in multiple discussions, there was also a lot of talk about picnics by the waterfronts and also by the city's beaches. Discussions surrounding food was also a dominant theme in this context. Interestingly, there were multiple discussions surrounding toilets in the context of beach visits by tourists. Other topics were more in common with the other cities, such as talk surrounding children's playgrounds and churches. Unlike Boston and Singapore, the reviews with extreme good and bad reviews had topics which were different from the topics emerging from all reviews taken together. High-points associated with positive tourist sentiment revolved around specific sites such as the zoo and art galleries. Others were more generic and pertained to conducted tours and beaches. Low-points emerging from the city surrounded themes such as bad staff, poor equipment in parks, lack of proper public toilets and, interestingly, complaints about dogs.

Overall, the discussions and themes emerging from these three cities tie back strongly to the core experiential identities that they present to their tourists – that of Boston as a city steeped in history, Singapore as a city cradled within nature but with exciting activities, and Sydney as a city dominated by its harbor and its beaches. **Table 2** summarizes the key positive and negative topics emerging from these three cities.

Table 2. Summary of topics associated with highest (+ve) and lowest (-ve) tourist sentiment across the three cities.

	Boston		**Singapore**		**Sydney**	
+ve	Great guide/tour/ food	Great churches	Temple tour	Jogging/ safe	Fantastic tours	Zoo/ Food
	Aquarium/ Museum	Park/ skating	Amazing rides	Park/ nature	Art galleries	Nice beaches
-ve	Climbing stairs	Tunnel/ traffic	Entrance queues	Dirt/ cleanliness	Public toilets	Bad staff
	Rain/ruined tours	Slavery/ massacre	Bad staff	Unpleasant playground	Poor equipment	Dogs

3.2 The Image of the City: Visual features of tourist interest

As discussed at the very outset, every city presents very unique visual identities to tourists and visitors. These visual identities are often associated with nuanced configurations of specific visual features, which appear and disappear as tourists move around the city, but at the same time leave lasting impressions upon the subconscious mind of the visitor. The unsupervised clustering carried out as a part of this work was an attempt to empirically uncover some of these nuances, and the ways in which they forge different visual identities in different cities.

The overarching popular identity of Boston as a city steeped in history and historic sites was validated through this exercise. The most common cluster of visual features of interest surrounded those pertaining to historic brick facades that characterize many neighborhoods in Boston. Interestingly enough, a vast majority of these features appeared in photographs that were captured by tourists along Boston's famous 'Freedom Trail' – a 2.5-mile path through the city that weaves through important anchors of the history of the United States. In this context, another important set of features frequently capturing tourist attention were plaques, signage and information boards, which frequently accompany tourist destinations (**Fig. 5**). Urban parks also formed a major visual attractor, with a significant percentage of tourist photographs capturing such spaces. Selfies, group photographs and photos of other tourists formed another major cluster that captured tourist attention.

Figure 5. Plaques and signages around tourist attractions formed one of the strong features of tourist interest in Boston

The identity of Singapore as a city surrounded by nature emerged through the photographs as well, with a vast majority of features of interest pertaining to parks and semi-urban landscapes. These photographs were spread out finely across the length and breadth of the island city-state. The photos however also revealed how local vegetation characteristics can generate lasting visual identities. Elements such as palm trees were a recurring feature in tourist photographs and, in many cases, formed the primary subject matter of the photographs themselves. Other imageable features included the skyscrapers and skylines that characterize downtown Singapore **(Fig. 6)**. In many cases these elements were framed across waterfronts or the harbor. Finally, a very unique set of features of tourist interest emerging from the city were those surrounding fish – both in the form of living creatures in aquariums as well as in the form of food.

The visual features capturing tourist attention in Sydney tied back strongly to the themes emerging from the topics of discussion as discussed in the earlier section. The Sydney Harbor, and specific anchors such as the Harbor Bridge and the famous Opera House dominated tourist photographs. These appeared both as primary subjects, as well as backdrops in generic harbor front photographs. What was even more interesting was the prominence of boats (including ferries and yachts) as strong visual attractors. It is worth noting here that while both Boston and Singapore are harbor cities as well, boats did not contribute as much to the visual attention-field of the tourist. Moreover, images of fine dining, with features such as wineglasses and food in formal settings frequently recurred as a visual theme underlying the visitor's experience of Sydney.

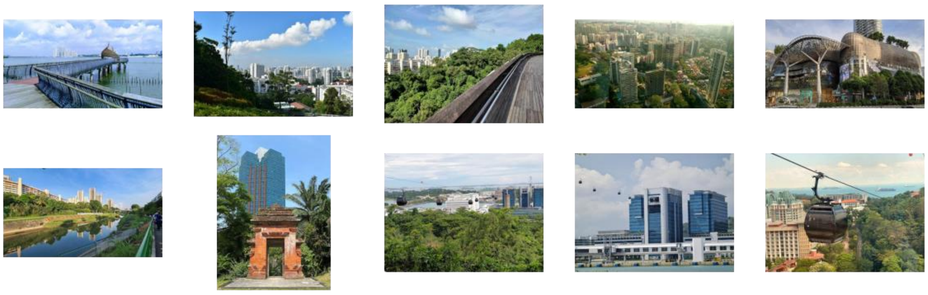

Figure 6. Skyscrapers and skylines – a popular subject of tourist interest in Singapore.

The features of visual interest emerging from the three cities opened up unique insights with regards to the specific elements within a tourist's visual field that actually capture tourist attention. It also revealed the unique ways in which low level visual features combine and recombine in different ways to generate high-level visual identities of cities. **Figure 7** summarizes the key feature-clusters emerging from the three cities.

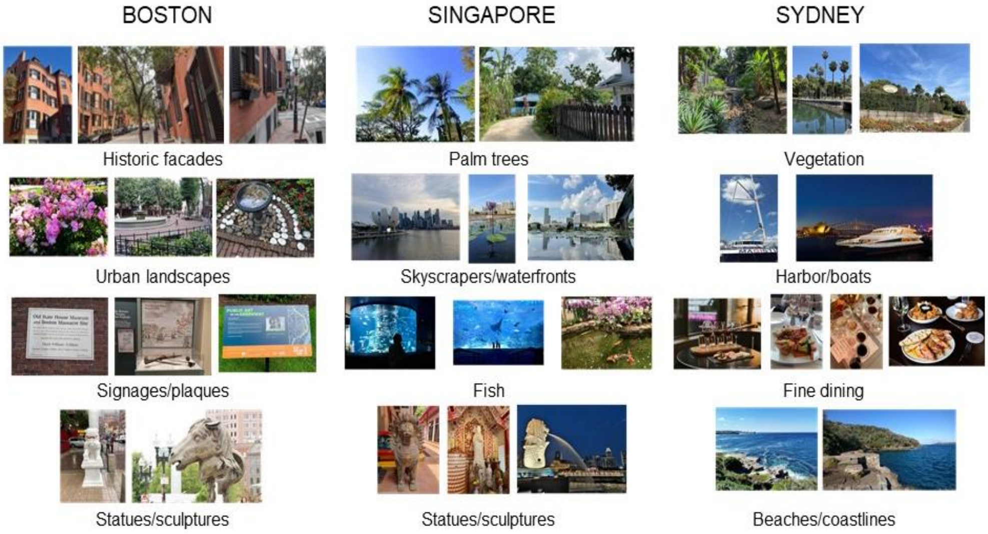

Figure 7. Summary of the top clusters for visual features of tourist interest across the three cities

4 Conclusion – Empirical approaches to urban identity

This work relied upon hybrid (textual and photographic) data to analyze the major themes of discussion and features of interest that contribute to overall sentiment towards urban tourist attractions in three cities. More importantly, it compared these patterns across cities, thus exploring the unique ways in which visual elements and textual themes come together to define urban identity.

However, there are a number of challenges and limitations with regards to such empirical approaches that are worth touching upon. Firstly, the platform from which such urban data is sourced is never neutral, and inevitably introduces a specific lens through which the city is understood and analyzed. Citizens and visitors talk about different aspects of urban life on different platforms, and do so in very different ways. The discussions generated about the same urban location on Twitter, Google Places and Flickr will be very different, and the nuances of these differences need to be kept in mind when working with such data. Secondly, most urban POI data aggregators are careful to mask sensitive personal information of the users that are generating the data, due to privacy considerations. While this is understandably an important necessity, it nevertheless does make it difficult for the analyst to discriminate between individuals visiting tourist locations as actual tourists, and those with a far greater degree of familiarity with the city visiting those same locations. Urban identity is a complex affair, and the tourist's identity can seldom be equated with the identity forged on the minds of long-term citizens. Moreover, the crowdsourced images and reviews analyzed for a given attraction were uploaded by different users, thus making it difficult to tease out meaningful correlations between visual and textual material. Thirdly, the current state of the art in object recognition as implemented in this work relies upon predetermined classes of visual elements that can be detected through the workflow. It thus relies upon a fixed symbolic vocabulary of elements that the model 'looks for', and is not equipped to comprehend different symbolic frameworks through which a tourist may interpret the very same visual content. Finally, and most importantly, the study relied on textual material in the English language. This inevitably introduces a specific research gaze, and excludes analyses of the urban experience of non-English speaking tourists in these cities.

Having said that however, the analytical methodology demonstrated through this work has potential to become a valuable tool for future studies. As discussed earlier, while topic modeling and sentiment analysis have been carried out for multiple urban studies in the past, the analysis of hybrid visual and textual data through unsupervised clustering is a novel approach. Extracted visual features and textual themes complement each other in the context of urban sentiment, and can provide nuanced and comprehensive insights which are often otherwise missed within mainstream sentiment analysis and topic modeling workflows. It is hoped that the methods outlined in this paper are taken forward to address future questions within an undeniably complex domain – that of subjective urban experience.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., & others. (2016). TensorFlow: A system for Large-Scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283.

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, *58*(2), 175–191.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Cardone, B., Di Martino, F., & Sessa, S. (2021). GIS-based fuzzy sentiment analysis framework to classify urban elements according to the orientations of citizens and tourists expressed in social networks. *Evolutionary Intelligence*, 1–10.

Galí, N., & Donaire, J. A. (2015). Tourists taking photographs: The long tail in tourists' perceived image of Barcelona. *Current Issues in Tourism*, *18*(9), 893–902.

Ghahramani, M., Galle, N. J., Ratti, C., & Pilla, F. (2021). Tales of a city: Sentiment analysis of urban green space in Dublin. *Cities*, *119*, 103395.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225.

Kovacs-Gyori, A., Ristea, A., Havas, C., Resch, B., & Cabrera-Barona, P. (2018). # London2012: Towards citizen-contributed urban planning through sentiment analysis of twitter data. *Urban Planning*, *3*(1), 75–99.

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., & others. (2020). The open images dataset v4. *International Journal of Computer Vision*, *128*(7), 1956–1981.

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.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment analysis in tripadvisor. *IEEE Intelligent Systems*, *32*(4), 72–77.