View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarWorks@UMass Amherst

University of Massachusetts Amherst

ScholarWorks@UMass Amherst

Travel and Tourism Research Association: Advancing Tourism Research Globally

Understanding Demographics and Experience of Tourists in Yellowstone National Park through Social Media

Yun Liang
The Pennsylvania State University

Junjun Yin
The Pennsylvania State University

Bing Pan
The Pennsylvania State University

Guangqing Chi
The Pennsylvania State University

Clio Andris
Georgia Institute of Technology

See next page for additional authors

Follow this and additional works at: https://scholarworks.umass.edu/ttra

Liang, Yun; Yin, Junjun; Pan, Bing; Chi, Guangqing; Andris, Clio; Miller, Zach; Jorgenson, Jack; and Nickerson, Norma, "Understanding Demographics and Experience of Tourists in Yellowstone National Park through Social Media" (2020). *Travel and Tourism Research Association: Advancing Tourism Research Globally*. 54.
https://scholarworks.umass.edu/ttra/2020/research_papers/54

This Event is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Travel and Tourism Research Association: Advancing Tourism Research Globally by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Author Information

Yun Liang, Junjun Yin, Bing Pan, Guangqing Chi, Clio Andris, Zach Miller, Jack Jorgenson, and Norma Nickerson

This event is available at ScholarWorks@UMass Amherst: https://scholarworks.umass.edu/ttra/2020/research_papers/54

Understanding Demographics and Experience of Visitors in Yellowstone National Park through Social Media

1 Introduction

National parks have been hailed as "America's best idea" (Wallace Stegner, 1983). In 2016, U.S. national parks attracted 331 million national and international visitors, a sharp increase from 2015 (U.S. Department of the Interior, 2017). These new visitors have put pressure on visitor management and may contribute to service failure (Kerlinger et al., 2013). To support visitor management and improve services, national parks and public land management agencies are interested in visitors' demographics and visitor experience.

At the same time, as a government agency, the National Park Service (NPS) faces limited, and potentially depleting resources for understanding its visitors. Surveys, field observations, automated counters have been the primary method for studying visitors (Mannell & Isoahola 1987; Hull & Stewart 1995). However, surveys are costly to implement, difficult to achieve high response rates, and post-hoc in nature. The ability of NPS to collect primary data from visitors is also bureaucratically challenging, as surveys are restricted considerably by federal requirements.

The emergence of social media platforms has provided a new opportunity for understanding visitor demographics and experience for better management of national parks with better availability and low cost of access (Di Minin, Tenkanen, & Toivonen, 2015). Real names and images of social media users could be utilized to identify visitors' gender, ethnicity and age (Yin, Chi, and Van Hook 2018). Crowdsourced data, such as time/date/location-stamped digital posts on social media, provides useful insight into visitor's spatial behavior (Tenkanen et al., 2017) that can be used by NPS and other federal land management agencies. Furthermore, text data in

social media could reveal visitor attitudes and experience (Wood, Guerry, Silver, & Lacayo, 2013; Sotiriadis & Van Zyl, 2013; Xiang, Schwartz, Cerdes Jr, & Uysal, 2015). Yet, detailed comparison between different datasets for validation has been rarely conducted. All data sources have limitations and require validation (Malik, Lamba, Nakos, & Pfeffer 2015).

Considering that there are not enough studies to validate social media textual data and survey data in visitor demographics, the purpose of this study is to compare the two different sources of data. The main question addressed is that: can social media be a good alternative for survey research in national parks? To answer this question, we want to address two different questions: can we learn the demographics of visitors from social media? What types of experience or service failures ca we learn from social media? Twitter is chosen social media platform due to provide high volume of real time data, including spatial-temporal patterns and experiences of lives of the public, which could be employed in tourism research (Kirilenko & Stepchenkova, 2017; Miah et al., 2017)

2 Literature Review

In this section, we first reviewed the literature of visitor experience in national parks, including definitions and components of visitor experience. Then, we discussed approaches to analyze social media textual data. Survey data and social media data were compared in the third part. Finally, we summarized the gaps between the literature and the goals of this study.

2.1 Visitor Experience in National Parks

According to Mannell and Isoahola (1987) and Hull and Stewart (1995), visitor experience in national parks involves cognitive appraisals of the degree to which the landscape meets expectations in fulfilling psychological needs and affective response. Satisfying experiences

indicated that visitors' expectations of experiences of national parks were met or exceeded and/or as experiences that result in positive emotions. Investigating visitor experience in national parks could understand visitor expectations, visitor motivations, effectiveness of management. In addition, understanding visitor experience could be helpful for designing visitor facilities, determining visitor satisfaction and dissatisfaction with particular facilities and services, and identifying use level of different regions of parks (Anderson, Lime & Wang, 1998; Cessford & Muhar, 2003; Eagles, McCool & Haynes, 2002: Hornback & Eagles, 1999). Therefore, understanding visitor experience is the first step to take further actions.

A study in Yosemite Valley revealed that natural landscape (waterfalls, natural scenery, mountains, and celestial features) was the most significant and attractive features of visitor experience. Apart from natural landscape, cultural landscape, social encounters, and learning experiences, including natural history, social and cultural history, bear safety, and NPS management, were mentioned by visitors (White, Youngs, Wodrich, & Borcherding, 2006). In addition, Hull and Stewart (1995) found that visitors in natural areas were attracted by natural landscapes and scenic beauty.

In summary, the research of visitor experience in national parks is significant for understanding visitor expectation, managing use level of different regions, etc. However, the majority of current researches have employed the survey to collect visitors' demographics and experiences.

2.2 Social Media Research and Approaches in Tourism

Twitter is an American microblogging and social networking service on which users post and interact with messages known as "tweets". Registered users can post, like, and retweet tweets, in addition, users can attach their locations when posting tweets. This social media platform has

become popular social media data source for tourism research (Kirilenko & Stepchenkova 2017). By using Twitter data, several previous studies have investigated spatial and temporal patterns of visitors (Chua, Servillo, Marcheggianin, & Moore, 2016), hot spots of destinations (Kisilevich et al., 2010; Garca-Palomares et al., 2015; Vu et al., 2015; Miah et al., 2017; Kaneko & Yanai, 2013), and visitors' characteristics (Donaire et al., 2014). In addition, Philander and Zhong (2016) explored customers' attitudes toward hospitality in Las Vegas, NV, by tweets.

According to the literature, the primary techniques to analyze unstructured textual data include sentiment analysis and text analytics, involving word frequency and distributions, information extraction, link and association analysis, visualization and predictive analytics (Batrinca & Treleaven 2015). Sentiment analysis was widely used in tourism social media data research. Kirilenko, Stepchenkova, Kim, and Li (2018) and Alaei, Becken, and Stantic (2019) assessed and compared different sentiment analysis approaches applied in tourism research. Generally, there are two approaches for sentiment analysis, namely lexicon-based and learning-based text classification (Kirilenko & Stepchenkova 2017). Several studies were utilized sentiment analysis to investigate visitors' emotions expressed in social media text data (Kirilenko & Stepchenkova 2017; Valdivia, Luzón & Herrera, 2017; Park, Kim & Ok, 2018).

To explore visitors' experience and satisfaction, text analytics was applied (Park, Ok & Chae 2016; Kirilenko & Stepchenkova 2017). Based on word or hashtag frequency, popular attributes of visitors' experiences could be identified. For example, Xiang, Schwartz, Gerdes Jr, and Uysal (2015) computed experience-related words based on guest reviews of hotels from Expedia, ranging from the core product, amenities, attributes, and staff-related experiences. In addition, text-link analysis was applied to identify patters or relationships among different words or hashtags. For instance, Berezina, Bilgihan, Cobanoglu, and Okumus (2016), a study of text

analytics of hotel reviews, showed that, there were “no child”, “no wait”, “no miles” and so on in positive reviews, while in negative reviews, the links included “no room/balcony/towels/fridge”, “no drain”, “no breakfast”, etc..

Therefore, text mining, including sentiment analysis, word or hashtag frequency and text-link analysis, could be applied to extract visitors’ emotions, experiences and satisfaction from social media text data. From practical perspective, text mining of social media text data could help visitor attractions or hotels to provide products or services to meet visitors’ demands. However, there are only limited studies to investigate topics discussed in Twitter data, especially in tourism research (Sotiriadis & Zyl, 2013).

In conclusion, past studies have investigated the approaches to analyze social media textual materials, including sentiment analysis, word/hashtag frequency, and text-link analysis. However, there are limited studies to investigate visitor experience in national parks by social media textual data. In addition, text classification were rarely employed for exploring visitor experience by social media data.

2.3 Comparison between Social Media Data and Survey Data

Although social media studies are applied widely, they still need validation from other datasets. First, users of social media may not represent the entire population (Jiang, Li & Ye, 2019). Malik, Lamba, Nakos, and Pfeffer (2015) indicated that users of geotagged tweets did not represent the US population accurately. Users of geotagged tweets were younger with higher median income, and in urban and coastal areas. In addition, there were a high population of Asian, Black or Hispanic/Latino users (Malik, Lambo, Nakos, & Pfeffer, 2015). Furthermore, various social media platforms may represent different user groups (Xiang, Du, Ma, & Fan,

2017). Secondly, social media data are unstructured and uploaded by users. When compared with survey data, social media data are hard to predict in format, amount and richness (Wang, Jin, Liu, Li & Zhang, 2018). Thirdly, social media data can exaggerate or miss key measurements of tourist behaviors. For example, Wang, Jin, Liu, Li and Zhang (2018) revealed that social media users were more likely to share more attractive and exciting sceneries with their friends and web viewers. In addition, the study showed that social media users were less likely to mention things considered shameful, such as free admission to the park.

Unlike general tourism studies, in nature-based tourism research, there are limited studies with which to compare social media data and traditional survey data's ability to reveal to the attractiveness of destinations and visitors' preference. For example, a case study, in Beijing Olympic Forest Park, compared attractiveness by social media data and survey data (Wang, Jin, Liu, Li & Zhang, 2018). The results revealed that the natural atmosphere, plants, water and recreational activities were important, but, in questionnaires, visitors were more likely to highlight cost and physical activities. In addition, Hausmann et al. (2018) explored visitors' preference for wildlife in Kruger National Park, South Africa, using photos from Instagram and Flickr and traditional survey data. Large-body mammals were the preferred category among Instagram, Flickr and survey. Apart from visitors' preferences on attractions and wildlife, social media data and survey data were compared for monitoring visitors' spatial and temporal patterns in a national park (Heikinheimo et al., 2017). Social media data better capture spatial and temporal patterns of visitors in the most popular sub-regions in the park than traditional survey data, while, in relative less popular areas measured by a survey, social media data had a discrepancy and less social media posts (Heikinheimo et al., 2017).

In conclusion, there are several gaps between our goals and the literature. First, limited studies validate and compare the similarity and differences between traditional survey data and social media data, especially in visitor demographics. Secondly, there are only few studies using social media data and the machine learning approaches to explore emotions and domains of visitor experiences. Thirdly, social media data is not applied widely in national parks research. Therefore, to better understand demographics and experiences of visitors in Yellowstone National Park, three research questions of this study are raised:

RQ1: What are the similarities and differences in visitors' demographic variables, including age, gender, ethnicity, and origins of residence, between the results of Twitter data and survey data?

RQ2: What are the spatial and temporal patterns of visitors in Yellowstone National Park based on geotagged tweets?

RQ3: What are the emotions and domains of visitor experience in Yellowstone National Park based on geotagged tweets?

3 Method

3.1 Data Collection

The data in this study were collected by a survey and from Twitter. The survey data was collected in summer 2016 in Yellowstone National Park by park managers. The day hikers in the Mt. Washburn trail and the Lonestar Geyser trail in Yellowstone National Park were intercepted while hiking in Yellowstone National Park. No overnight visitors were intercepted. The total sample size of the survey is 647. Variables of the survey data include visitors' gender, age and race. The geo-tagged Twitter data were collected from January to December 2016. The total

tweet sample size is 22,418 with the location of Yellowstone National Park and include user id, username, user image, post time of tweet, tweet and geo-location.

3.2 Data Analysis

3.2.1 Visitor Demographics

To answer the first research question, we followed the methods developed in (Yin, Chi, and Van Hook 2018) to estimate Twitter user demographics. Note that instead of using the Microsoft Azure facial recognition service to estimate the gender and age information from user' portrait, we utilized an open source implementation (Uchida 2019) based on convolutional neural networks, which trained a database with over 500k face images from IMDb and Wikipedia with age and gender labels (Rothe et al. 2016). The first and last names of Twitter users were used to estimate users' gender and race/ethnicity information. If the results of gender of a user by first name and image were different, the gender identified by first name was given priority. However, if the gender of a user was not identified by first name, the result of a user' image was used. Then, Chi-square test was used to compare visitors' demographics between survey data and Twitter data.

3.2.2 Spatial and Temporal Patterns

Spatial analysis could reveal visitors' concentrations and movement patterns in the Yellowstone National Park. In this study, spatial analysis includes geotagged tweets distribution, hot spot analysis and movement patterns were conducted in ArcGIS 10.7.1. Temporal pattern of visitation could reveal peak season and low season of a year in a park. Visitors' temporal distribution in the Yellowstone National Park over 2016 was graphed by Microsoft Excel.

3.2.3 Visitor Experience

Identify Tweets with Visitor Experiences

Although visitors were traveling in the Yellowstone National Park, tweets posted visitors may not include visitor experience of the Yellowstone National Park, therefore, it was necessary to separate tweets with visitor experience and without visitor experience. Four models, including Naïve Bayes, SVM, Logistic Regression and Random Forest, were utilized and compared.

To evaluate the accuracy of each model, 999 tweets were randomly selected from the entire Twitter dataset and labelled manually into binary (0/1) form (0 represents a tweet without visitor experience, while 1 represents a tweet with visitor experience). Then, 499 within these 999 tweets were used for model training and the rest of tweets were used for model testing. Scikit-learn, a machine learning package in Python, was utilized for the model training, testing and entire dataset prediction.

Hashtag Frequency Analysis

To identify topics discussed in Twitter text data in the Yellowstone National Park, types and frequencies of hashtags from Twitter text data were extracted. Collections, a Python library, was utilized to count frequency of each hashtag. Before counting hashtag frequency, all tweets were transferred to lowercases.

Sentiment Analysis

The sentiment of a tweet indicates the overall attitude articulated in a tweet. Generally, sentiment analysis may report three classes of attitudes, including positive, negative and neutral by assessing a text. Before sentiment analysis, several data preparation steps were applied for Twitter text data: 1) identifying English tweets; 2) removing URLs, punctuation signs and stop words from tweets; 3) transferring all uppercases into lowercases. To evaluation visitor

experience in the Yellowstone National Parks, VADER sentiment analysis, a lexicon-based sentiment analysis approach for text data in social media (Hutto & Gilbert, 2014), was employed to investigate visitor' emotion experience. The range of VADER sentiment analysis score is from -1 to 1. A score less than -0.05 represents negative attitudes toward experience, while positive attitude of sentiment analysis score is more than 0.05. And a score from -0.05 to 0.05 indicates neutral emotion toward experience in the Yellowstone National Park.

Domains of Visitor Experience

To investigate domains of visitor experience, text classification was applied to classify visitors' tweets into different categories. After identifying tweets with visitor experience of the Yellowstone National Park, 200 tweets from tweets with visitor experiences were selected to be label manually into different domains of visitor experience.

4 Results

4.1 Visitor Demographics

The genders of 1606 Twitter users were identified (Table 1). 46% of users were female and 54% of users were male. In the survey data, there were 642 visitors reported gender. 46.4% of visitors were female and 53.6% of visitors were male. The Chi-square statistic is 0.040 and the p-value is 0.84, revealing no significant difference in gender between survey data and Twitter data.

Table 2 showed the frequency of visitors by age groups. The ages of 1,157 users of Twitter were identified by users' portraits and 641 visitors were reported their ages in the survey. Since only visitors who were older than 18 years old were intercepted to participate in survey, therefore, the sample size of this age group is zero. However, in Twitter data, fourteen visitors who were

younger than 18 years old were identified. Compared to survey data, Twitter users were younger, 18-30 years old and 31-45 years old, occupying 93.5%. Only two visitors identified by Twitter data were older than 61 years old; however, in the survey data, there were 67 (10.4%) visitors who were older than 61. In addition, there were significant differences of visitors in 31-45 years old, 46-60 years old and older than 60 years old between Twitter data and survey data.

For this study, we examined the proportions of Twitter users' real last names among White/Black/Asian populations. Finally, there were 736 Twitter users that were extracted their real last names. After analyzing Twitter users' last names, the White occupied 68.4% of entire visitors, followed by the Black, occupying 10.3%. The Asian occupied 9.3% of entire visitors. In the survey, four races, including White, Asian, American Indian or Alaska, and Native Hawaiian or Pacific were employed and a category of 'more than one'. In the survey data, 624 visitors reported their ethnicity. The majority visitors in the survey were White, occupying 92.0%, followed by the Asian (6.1%). Visitors were American Indian or Alaska and Native Hawaiian or Pacific that only occupied 0.2% and 0.3%. In addition, 1.4% of visitors took that they had more than one race. Then, Chi-square tests were employed to compare percentage of the White and the Asian between the Twitter data and the survey data (Table 3). The results showed that there were statistical significances between the Twitter data and the survey data in the percentages of the White and the Asian.

Table 1 Frequency and Percentage of Visitors by Gender

Gender	Twitter		Survey	
	Frequency	Percentage	Frequency	Percentage
Female	738	46.0%	298	46.4%
Male	868	54.0%	344	53.6%
Total	1606		642	

Table 2 Frequency and Percentage of Visitors by Age Group

Age Group (years old)	Twitter		Survey		Chi-square statistic	p-value
	Frequency	Percentage	Frequency	Percentage		
>18	14	1.2%	0	0%	N/A	N/A
18-30	403	34.8%	208	32.1%	1.04	0.31
31-45	679	58.7%	185	28.6%	146.99	<0.001
46-60	60	5.2%	181	28.0%	188.83	<0.001
61+	2	0.2%	67	10.4%	118.11	<0.001
Total	1157		641			

df=1

Table 3 Percentage of Visitors by Ethnicity

Ethnicity	Twitter	Survey	Chi-square Statistic	p-value
White	68.4%	92.0%	175.37	<0.001
Black	10.3%	NA	NA	NA
Asian	9.3%	6.1%	28.91	<0.001
American Indian or Alaska	NA	0.2%	NA	NA
Native Hawaiian or Pacific	NA	0.3%	NA	NA
More than one	NA	1.4%	NA	NA
Total	736	624		

df=1

4.2 Spatial and Temporal Patterns

The main attractions in the Yellowstone National Park include the Old Faithful, Grant Village, Canyon, Mammoth Hot Spring, etc. (Figure 1). Figure 1 showed the spatial pattern of geotagged tweets in Yellowstone National Park. The red dots in the map represented the geo-locations of tweets. The spatial distribution of geotagged tweets reflected the road network in Yellowstone National Park. The concentration of geo-tagged tweets represented main attractions in the national park, matching the map of main attractions in the Yellowstone National Park. Therefore, geotagged tweets can reflect accurately tourists' hot spots in the Yellowstone National Park, while the survey data is difficult to collect tourists' real time geo-locations and lacks this type of data.

The temporal patterns of monthly Twitter data in Yellowstone National Park and average monthly visitation, Yellowstone National Park, 2014-2018 (U.S. Department of the Interior, 2019), are slightly different (Figure 2). The skewness and kurtosis of Twitter data are 1.3 and 0.5. The skewness and kurtosis of survey data are 0.6 and -1.5. The peak season of Twitter data is from July to September. However, the visitation statistics from Yellowstone National Park shows that the peak season of visitation is from May to September.

4.3 Visitor Experience

Table 4 presented the accuracy of each model. The SVM model had highest accuracy (0.806) in model testing. Therefore, the SVM model was employed to classify entire Twitter dataset to identify whether a tweet involved visitor experience in the Yellowstone National Park. Finally, totally 18,586 tweets with visitor experience were identified, occupying 82.90% of entire tweets dataset. The 18,586 tweets with visitor experience were utilized for further analyzing hashtag frequency, visitors' attitudes toward their experiences and the domains of visitor experiences.

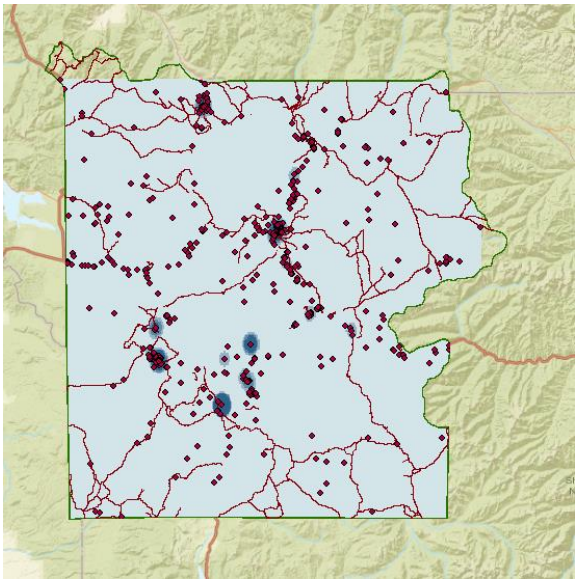

Figure 1 Spatial Distribution of Geotagged Tweets in Yellowstone National Park (left); The Main Attractions and Visitor Center in the Yellowstone National Park (right).

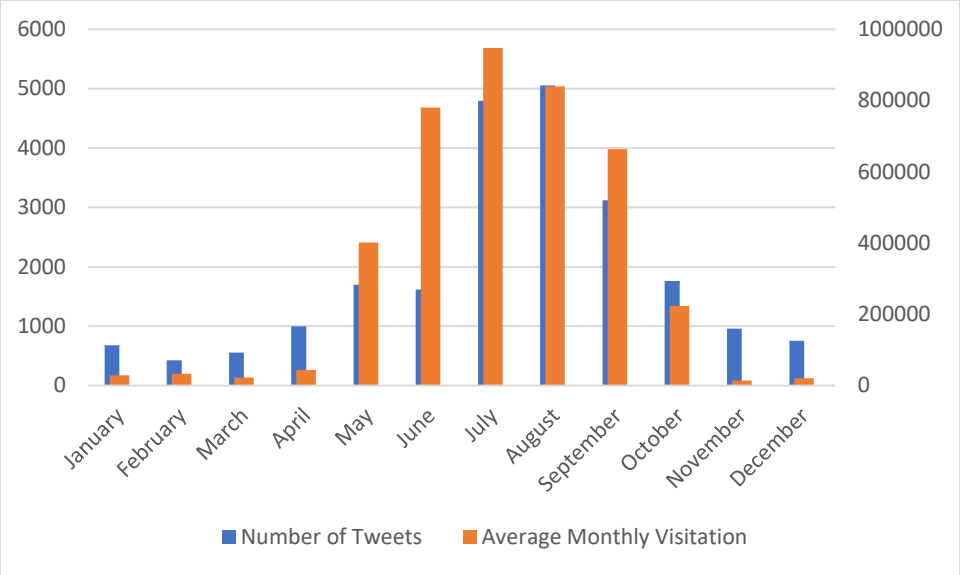

Figure 2 Monthly Tweets in Yellowstone National Park, 2016 & Average Monthly Visitation in Yellowstone National Park, 2014-2018. Adapted from "Visitation Statistics," by Yellowstone National Park, 2019. Retrieved from https://www.nps.gov/yell/planyourvisit/visitationstats.htm

Table 4 Accuracy of Models to Identify Whether Tweets involving Visitor Experience

Model	Accuracy
Naïve Bayes	0.732
SVM	0.806
Logistic Regression	0.746
Random Forest	0.746

4.3.1 Hashtag Frequency

There were 172 hashtags that were identified from 18,586 tweets with visitor experience. However, this list of hashtags did not include hashtags that were mentioned less than 10 times. Table 5 showed top 50 hashtags and their frequencies. The top 1 hashtag was '#yellowstone', suggesting that visitors wanted to let their Twitter followers know that they were at the Yellowstone National Park. There were several similar hashtags to '#yellowstone', such as '#yellowstonenationalpark', '#yellowst', '#yellowsto' etc. In addition, apart from 'yellowstone' and its similar hashtags, several specific sceneries, including '#oldfaithful',

‘#grandprismaticspring’, ‘#mammothhotsprings’ and so on, were extracted from tweets with visitor experience. Several names of wildlife, including ‘#bison’, ‘#bear’, and ‘#buffalo’, were also extracted from tweets with visitor experience. The several sceneries and wildlife were extracted from tweets, supporting a fact that visitors were attracted by the unique and well-known sceneries and wildlife.

Rank	Hashtag	Frequency	Rank	Hashtag	Frequency
1	yellowstone	1950	26	montana	59
2	yellowstonenationalpark	550	27	grandprismaticspring	59
3	oldfaithful	303	28	mammothhotsprings	57
4	wyoming	277	29	usa	51
5	bison	206	30	bear	51
6	tbt	186	31	buffalo	51
7	nature	145	32	yellowstonenational	51
8	roadtrip	133	33	yel	50
9	geyser	132	34	ynp	50
10	nofilter	126	35	findyourpark	47
11	yellowst	113	36	w	45
12	nps100	105	37	photography	39
13	travel	96	38	grandcanyon	39
14	yellowsto	92	39	national	36
15	nationalpark	90	40	grizzly	36
16	yellow	87	41	mammoth	36
17	y	83	42	elk	36
18	yellowston	81	43	yellowstonenationalpark2016	36
19	ye	77	44	earthquake	36
20	yellows	73	45	bigsky	33
21	yell	69	46	kenenwatu	33
22	yello	67	47	somedaysago	33
23	wildlife	64	48	summe	33
24	beautiful	63	49	yellowstonenati	33
25	park	61	50	americancultureis	33

Table 5 Top 50 Hashtags and Frequencies

4.3.2 Sentiment Analysis

The results of sentiment analysis showed that 40.6% (7,546 tweets) of tweets were positive, 52.7% (9,801 tweets) of tweets were neutral, and only 6.7% (1,239 tweets) were negative (Table 6). The results showed that only a small part of visitors had negative visitor experience in the Yellowstone National Park. Table 7 showed several examples of tweets in each attitude category. Positive tweets usually have positive adjective words, such as “favorite”, “beautiful” and so on. However, negative tweets expressed visitors’ disappointment because of missing some expected sceneries.

Table **6** Descriptive Statistics of Sentiment Analysis

Attitudes	Frequency	Percentage
Positive	7546	40.6%
Neutral	9801	52.7%
Negative	1239	6.7%
Total	18586	

Table 7 Examples of Tweets in Each Attitude Category

Category	Sentiment Score	Tweets
Positive	0.82	One of my favorite places in Yellowstone: Lewis Falls. So pretty and peaceful. We saw tracks\u2026 https://t.co/43PLpjfKZ1
	0.72	A peaceful tour though Yellowstone National Park along the Madison River. A true wonderland.\u2026 https://t.co/1LMIcUXKlb
Neutral	0	Nature is never finished. //Robert Smithson 1454831708.0 30 1454831708.0 0
	0	See our latest #Yellowstone
Negative	-0.57	Missing #Yellowstone and having no snow on the ground! \U0001f332\U0001f33b\U0001f333#takemethere @ Yellowstone National Park https://t.co/2PUhIBCCfD
	-0.44	dirty harry @ Yellowstone National Park https://t.co/1g3rFrZXOc

4.3.3 Domains of Visitor Experience

After identifying tweets with visitor experience, 200 tweets were randomly selected to be classified into different domains of visitor experiences in the Yellowstone National Park (YNP). Six domains of visitor experiences were categorized, namely landscape, wildlife, activities, posting photos/videos, infrastructure, and other (Liang, Kirilenko, Stepchenkova, & Ma, 2019). Here are the descriptions of six categories:

- Landscape. A tweet mentioned specific sceneries or landscapes in the YNP, such as the Old Faith, the Upper Fall, the Lower Fall, etc.
- Wildlife. A tweet mentioned wildlife, such as bison, bear, elk, etc., in the YNP.
- Activities. A tweet mentioned hiking, walking and other activities during the trip in the YNP.
- Posting Photos/Videos. A tweet mentioned that a Twitter user uploaded a picture(s) or a video(s).
- Infrastructures. A tweet mentioned facilities, infrastructures, campsites, accommodations, or other types of amenities.
- Other. A tweet only described general experiences in the YNP and did not mention specific landscape, sceneries, wildlife, activities, tourism facilities in the YNP.

Table 8 listed frequencies and percentages of each domain of visitor experience. There were 73 tweets, occupying 36.5%, that were classified into ‘Landscape’, which has the most tweets apart from ‘Other’. Then, ‘Wildlife’ ranked the second category, involving 19 tweets, occupying 9.5%. Totally, ‘Landscape’ and ‘Wildlife’ included 92 tweets and occupied 47% of entire tweets with visitor experience. This result supported a fact that visitors were attracted by landscape and

wildlife in the Yellowstone National Park. ‘Posting photos/videos’ had 13 tweets (6.5%), followed by ‘Activities’ (7 tweets, 3.5%). The least category was ‘Infrastructure’, including 4 tweets (2.0%). ‘Other’ w the largest category, including 84 tweets and occupying 42.0%. This was not surprising since social media textual data usually is unstructured, therefore, it is difficult to classify the majority tweets into certain categories.

Table 8 Domains, Frequency and Percentage of Visitor Experience

Label	Domains of Visitor Experience	Frequency	Percentage
0	Landscape	73	36.5%
1	Wildlife	19	9.5%
2	Activities	7	3.5%
3	Post photos/videos	13	6.5%
4	Infrastructure	4	2.0%
5	Other	84	42.0%

5 Conclusion and Discussion

This study, first, compared visitors’ demographics, including gender, age and ethnicity, between the Twitter data and the survey data. Secondly, visitor experiences were analyzed through geolocations and textual data of the Twitter dataset. For the first research question, there was no significant difference of the percentages of visitors’ gender between the Twitter data and the survey data. However, when comparing visitors’ age groups, only the group of 18-30 years old is no significant difference between the Twitter data and the survey data. The Twitter data indicated that the majority visitors (94.6%) were younger than 46 years old, while the survey data showed that there were 38.4% of visitors who were older than 45 years old. In addition, there were statistically significant difference in visitor ethnicity when comparing the Twitter data and the survey data. Therefore, only small part of the Twitter data can be validated with the survey data.

For the spatial patterns of visitors, the result by geotagged tweets indicated that visitors were attracted by main attractions in the Yellowstone National Park. Similarly, the temporal patterns of visitors in the Yellowstone National Park, the results revealed that the peak and low seasons of visitation extracted by the Twitter data followed the official average monthly visitation statistics. The potential reason to explain the number of visitors in the peak season and low season might be due to hiker versus general visitors: it is hard to hike during the winter months.

To examine visitor experience, hashtag frequency, sentiment analysis, and domains of visitor experiences were conducted. Both the analyses of hashtag frequency and domains of visitor experiences revealed that visitors were attracted by natural landscapes and wildlife of the Yellowstone National Park. Thirteen tweets mentioned that users uploaded photos or videos into Twitter and occupied 6.5%, indicating that photography is a popular activity by visitors in the Yellowstone National Park. In addition, visitors' activities in the Yellowstone National Park, including road trip, hiking and walking, were mentioned in the tweets. Eighty-four tweets (42.0%) were classified into the category of 'Other', supporting that textual data in social media are unstructured. The result of sentiment analysis revealed that only a few visitors (6.7%) had negative attitudes toward their experiences, indicating that most visitors in Yellowstone National Park had good experience.

Although, in this study, the Twitter data did not validate the survey data completely, the easy access and high volume of social media data will help park managers to understand demographics and experiences of visitors in national parks.

Reference

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research*, 58(2), 175–191. https://doi.org/10.1177/0047287517747753.

Anderson, D., Lime, D. and Wang, T. 1998. Maintaining the Quality of Park Resources and Visitor Experiences: a Handbook for Managers. University of Minnesota, St Paul.

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. *AI & SOCIETY*, 30(1), 89–116. https://doi.org/10.1007/s00146-014-0549-4.

Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing & Management*, 25(1), 1–24. https://doi.org/10.1080/19368623.2015.983631.

Cessford, G. and Muhar, A. 2003. Monitoring Options for Visitor Numbers in National Parks and Natural Areas. *Journal of Natural Conservation*. 11: 240–250.

Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*, 57, 295–310. https://doi.org/10.1016/j.tourman.2016.06.013

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. Frontiers in Environmental Science, 3. https://doi.org/10.3389/fenvs.2015.00063

Donaire, J. A., Camprubí, R., & Galí, N. (2014). Tourist clusters from Flickr travel photography. *Tourism Management Perspectives*, 11, 26–33. https://doi.org/10.1016/j.tmp.2014.02.003

Eagles, P., McCool, S. and Haynes, C. 2002. Sustainable Tourism in Protected Areas: Guidelines for Planning and Management. IUCN, Gland, Switzerland.

García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, 63, 408–417. https://doi.org/10.1016/j.apgeog.2015.08.002

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Minin, E. D. (2018). Social Media Data Can Be Used to Understand Tourists' Preferences for Nature-Based Experiences in Protected Areas. *Conservation Letters*, 11(1), e12343. https://doi.org/10.1111/conl.12343.

Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey. *ISPRS International Journal of Geo-Information*, 6(3), 85. https://doi.org/10.3390/ijgi6030085

Hornback, K. and Eagles, P. 1999. Guidelines for Public Use Measurement and Reporting at Parks and Protected Areas. IUCN, Gland, Switzerland.

Hull, R. B., & Stewart, W. P. (1995). The Landscape Encountered and Experienced While Hiking. *Environment and Behavior*, 27(3), 404–426. https://doi.org/10.1177/0013916595273007.

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *ICWSM*.

Jiang, Y., Li, Z., & Ye, X. (2019). Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. *Cartography and Geographic Information Science*, 46(3), 228–242. https://doi.org/10.1080/15230406.2018.1434834

Kaneko, T., & Yanai, K. (2013). Visual event mining from geo-tweet photos. *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 1–6. https://doi.org/10.1109/ICMEW.2013.6618407

Kerlinger, P., Burger, J., Cordell, H. K., Decker, D. J., Cole, D. N., Landres, P., O'Shea, T. (2013). Wildlife and recreationists: Coexistence through management and research. Washington, DC: Island.

Kirilenko, A. P., & Stepchenkova, S. O. (2017). Sochi 2014 Olympics on Twitter: Perspectives of hosts and guests. Tourism Management, 63, 54–65. https://doi.org/10.1016/j.tourman.2017.06.007.

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (Robert). (2018). Automated Sentiment Analysis in Tourism: Comparison of Approaches. *Journal of Travel Research*, 57(8), 1012–1025. https://doi.org/10.1177/0047287517729757.

Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., & Andrienko, G. (2010). Event-Based Analysis of People's Activities and Behavior Using Flickr and Panoramio Geotagged Photo Collections. *2010 14th International Conference Information Visualisation*, 289–296. https://doi.org/10.1109/IV.2010.94

Liang, Y., Kirilenko, A. P., Stepchenkova, S. O., & Ma, S. (David). (2019). Using social media to discover unwanted behaviours displayed by visitors to nature parks: Comparisons of

nationally and privately owned parks in the Greater Kruger National Park, South Africa. Tourism Recreation Research, 0(0), 1–6. https://doi.org/10.1080/02508281.2019.1681720.

Malik, M. M., Lamba, H., Nakos, C., & Pfeffer, J. (2015). Population Bias in Geotagged Tweets. *AAAI* 2015.

Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management*, 54(6), 771–785. https://doi.org/10.1016/j.im.2016.11.011

Park, S., Kim, H. J., & Ok, C. (2018). Linking emotion and place on Twitter at Disneyland. *Journal of Travel & Tourism Marketing*, 35(5), 664–677. https://doi.org/10.1080/10548408.2017.1401508.

Park, S., Ok, C., & Chae, B. (2016). Using Twitter Data for Cruise Tourism Marketing and Research. *Journal of Travel & Tourism Marketing*, 33(6), 885–898. https://doi.org/10.1080/10548408.2015.1071688.

Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, 55, 16–24. https://doi.org/10.1016/j.ijhm.2016.02.001

Rothe, R., R. Timofte, L. Van Gool, and L. Van Gool. 2016. Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks. International Journal of Computer Vision 126 (2–4):144–157.

Secretary Zinke Announces Record Visitation to America's National Parks. (2017). U.S. Department of the Interior. Retrieved from https://www.doi.gov/pressreleases/secretary-zinke-announces-record-visitation-americas-national-parks.

Sotiriadis, M.D. and Van Zyl, C. (2013) Electronic Word-of-Mouth and Online Reviews in Tourism Services: The Use of Twitter by Tourists. *Electronic Commerce Research*, 13, 103-124. https://doi.org/10.1007/s10660-013-9108-1.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports*, 7(1), 1–11. https://doi.org/10.1038/s41598-017-18007-4

Uchida, Y. 2019. Keras implementation of a CNN network for age and gender estimation. https://github.com/yu4u/age-gender-estimation.

Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment Analysis in TripAdvisor. *IEEE Intelligent Systems*, 32(4), 72–77. https://doi.org/10.1109/MIS.2017.3121555.

Visitation Statistics. (2019). U.S. Department of the Interior. Retrieved from https://www.nps.gov/yell/planyourvisit/visitationstats.htm

Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*, 46, 222–232. https://doi.org/10.1016/j.tourman.2014.07.003

Wang, Z., Jin, Y., Liu, Y., Li, D., & Zhang, B. (2018). Comparing Social Media Data and Survey Data in Assessing the Attractiveness of Beijing Olympic Forest Park. *Sustainability*, 10(2), 382. https://doi.org/10.3390/su10020382

White, D. D., & Hendee, J. C. (1999). Primal hypotheses: The relationship between naturalness, solitude and the wilderness experience benefits of development of self (DOS), development of community (DOC) and spiritual development (SD). In D. N. Cole, S. F. McCool, W. T.

Borrie & J. O'Loughlin (Eds.), Wilderness science in a time of change conference--Volume 3: Wilderness as a place for scientific inquiry (pp. 223-228). Missoula, MT: USDA Forest Service, Rocky Mountain Research Station.

White, D. D., Aquino, J. F., Budruk, M., & Golub, A. (2011). Visitors' Experiences of Traditional and Alternative Transportation in Yosemite National Park. *Journal of Park and Recreation Administration*, 29(1), 38–57.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, 3(1), 1–7. https://doi.org/10.1038/srep02976

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120–130. https://doi.org/10.1016/j.ijhm.2014.10.013.

Yin, J., G. Chi, and J. Van Hook. 2018. Evaluating the Representativeness in the Geographic Distribution of Twitter User Population. In Proceedings of the 12th Workshop on Geographic Information Retrieval - GIR'18, 1–2. ACM.