Original Article

A regression analysis of trends in population changes in tourist destinations:
Using keyword search volume and statistical population data

Masahide Yamamoto (Faculty of Foreign Studies, Nagoya Gakuin University, myama@ngu.ac.jp)

Abstract

This study uses "Mobile Kukan Toukei™" (MOBILE SPATIAL STATISTICS) to identify the number of visitors in different periods at specific tourist destinations. Mobile Kukan Toukei is statistical population data created by the operational data of mobile phone networks. The service makes it possible to estimate the population structure of a region by gender, age, and residential area. In addition, it attempts to demonstrate an alternative method to infer the number of visitors in specific areas more accurately. The various interests of tourists influence their keyword search before or during travel, and ultimately emerge as some kind of trend in a specific keyword's search volume. By connecting the Mobile Kukan Toukei and keyword search volume, linear equations could be derived. These findings could lead to a model to forecast tourism demand in a destination.

Keywords

mobile phone, statistical population data, keyword search volume, cross-correlation analysis, regression analysis

1. Introduction

The tourism industry, being labor intensive, was expected to absorb some of the labor force from different regions of Japan. Therefore, various economic enterprises and local governments had been struggling to promote the industry. Since Prime Minister Koizumi's first term, the government of Japan had been attempting to give the tourism industry a boost, because authorities have come to recognize the importance of promoting tourism in order to stimulate sluggish regional economies in Japan. For example, the government launched the Visit Japan Campaign (VJC) in 2003, which is a promotional effort of the government to activate inbound tourism with the objective of uniting the public and private sectors.

Due to these efforts including the VJC, Japan's tourism industry has been enjoying a rapid increase in the number of incoming tourists from other countries. According to the Japan National Tourism Organization, the number reached approximately 31.2 million people in 2018 (see Figure 1).

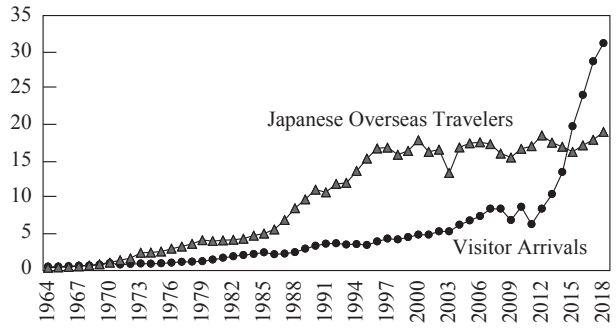

Figure 1: The annual number of visitor arrivals and Japanese overseas travelers (Unit: million people)

Notes: These figures are compiled from the data of the Ministry of Justice in Japan.

However, the industry still faces three challenges. The first involves overcoming a regional gap. Recently, several famous tourist destinations such as Kyoto have come to suffer from so-called "over-tourism," whereas most places still have room to accept more tourists. Therefore, attracting visitors to not too popular places should create a win-win situation.

The second challenge is the seasonal gap. For some years, Japan's tourism industry has been suffering from significant volatility in demand depending on the season and day of the week. Furthermore, there has been a significant loss of business opportunities because of congestion during the busy season. To cope with such volatility, tourism facilities, such as inns and hotels, have been trying to level the demand through daily and/or seasonal pricing adjustments. For example, room rates on the days before holidays are usually more expensive than they are on other days. Despite these efforts, the differences between on-season and off-season occupancy rates of rooms and facilities are still large. In other words, attracting customers in the off-season is an important challenge for tourism.

The third challenge involves reviving regional economies that have incurred huge capital losses from frequent natural disasters such as earthquakes, typhoons, and floods since 2011. Tourism is expected to revitalize those disaster-hit areas.

There must be several solutions to the challenges. Of course, enhancing the attractions of the regions to appeal to tourists is key in the long run. However, sometimes the approach takes time and therefore it is not easy to examine the effect.

On the other hand, holding events to draw visitors could be an alternative in the short term, especially for rural areas devoid of well-known tourist attractions. The method could also eliminate the seasonal gap. Numerous events including newly launched ones are currently held in Japan. To date, it has been difficult to accurately grasp the extent to which these events attract visitors and the types of people who visit them.

Basically, it is important to verify the number of visitors to effectively market a tourist destination regardless of events being held there. Although measuring the exact number of

visitors to open tourist areas has been quite difficult so far, it has now become possible to estimate the number by obtaining some population data.

In this study, "Mobile Kukan Toukei™" provided by NTT DOCOMO, Inc. and DOCOMO Insight Marketing, Inc. has been used to count the number of visitors at specific tourist destinations.(1) Mobile Kukan Toukei is statistical population data created by operational data from mobile phone networks.(2)

This study also attempts to demonstrate an alternative method to more accurately infer the number of visitors in specific areas.

2. Related studies

Other related studies also need to be considered. First, let us look at studies on using mobile phone location data for tourism surveys. These can be traced back to 2008. Ahas et al. [2008] introduced the applicability of passive mobile positioning data for studying tourism. They used a database of roaming location (foreign phones) and call activities in network cells: the location, time, random identification, and country of origin of each called phone. Using examples from Estonia, their study described the peculiarities of the data, data gathering, sampling, the handling of the spatial database, and some analytical methods to demonstrate that mobile positioning data have valuable applications for geographic studies. Japan Tourism Agency conducted a similar study using international roaming service in December 2014 [Japan Tourism Agency, 2014].

Several studies employing location data have emerged since the work of Ahas et al. [2008]. Liu et al. [2013] investigated the extent to which behavioral routines could reveal the activities being performed at mobile phone call locations captured when users initiate or receive voice calls or messages. Gao and Liu [2013] attempted to examine the methods used to estimate traffic measures using information from mobile phones. Steenbruggen et al. [2015] used mobile phone data to provide new spatio-temporal tools for improving urban planning and reducing inefficiencies in current urban systems. They addressed the applicability of such digital data to develop innovative applications to improve urban management.

The studies described above could be characterized as similar to Ahas et al. [2008]. Some are based on results obtained by analyzing data roaming activity. However, recent studies have used a much larger volume of data. Japan's Okinawa Prefecture analyzed data provided by NTT DOCOMO, Inc., which is the largest mobile phone service provider in Japan. Therefore, their data should be more reliable since the sample is quite large. The Project Report that Okinawa Prefecture published [2013] is of a study that used location data obtained from a domestic mobile phone network. The aim of the project was to survey the characteristics and behavior of tourists who were visiting Okinawa Prefecture. Okinawa pref. conducted the survey in order to grasp the trends and needs of repeat customers. The survey revealed the composition of tourists to Okinawa Prefecture by residence, gender, and age. They examined how the number of travelers changes depending on the month (October 2012 and January 2013) and the day of the week.

Yamamoto [2019] used the location data (i.e. Mobile Kukan Toukei) to identify the number of visitors in each period and their characteristics. The study sites of the survey are tourist destinations in Ishikawa Prefecture and Toyama city in Japan, including Kanazawa city, which became nationally popular after the Hokuriku Shinkansen opened in 2015. Yamamoto compared the effect of the opening of the Shinkansen in each city.

Next, studies concerning keyword searches related to tourism are covered. As Sheldon [1997] notes, tourism is an information intensive industry. The size of the tourism industry alone suggests that it generates large volumes of information to be processed and communicated. The Internet has fundamentally changed the manner in which tourism-related information is distributed and how people plan for travel. Thus, keyword search marketing has become important in tourism as well as in other industries.

Xiang and Pan [2011] pointed out that search engine marketing is gaining the status of a major online marketing strategy for many destinations. Search queries are perhaps the most important behavioral aspect of the use of search engines. Keywords in travelers' queries reflect their knowledge about the city and its competitors. Xiang and Pan attempted to identify the patterns in online travel queries across tourist destinations, and offered insights for the manner in which tourism destinations are searched online and implications for search engine marketing for destinations.

Pan and Li [2011] examined the linguistic structure of destination image. They attempted to demonstrate the importance of niche keywords in search engine marketing, in order to establish the importance of niche phrases for tourism destination image (TDI).

Xiang and Gretzel [2010] investigated the extent to which social media appear in search engine results in the context of travel-related searches. The study employed a research design that simulates a traveler's use of a search engine for travel planning by using a set of pre-defined keywords in combination with nine U.S. tourist destination names. The analysis of the search results showed that social media constitute a substantial part of the search results, indicating that search engines likely direct travelers to social media sites.

Pan et al. [2007] analyzed 701 Excite.com accommodation search queries and suggested that travelers most often search for their accommodations simultaneously with other aspects of their travel, such as destinations, attractions, transportation and dining; and that most commence their search by seeking specific hotels in conjunction with their destination city.

Ayanso and Karimi [2015] used a unique cross-sectional dataset of the top 500 internet retailers in North America and empirically investigated the moderating effects of keyword competition on the relationship between ad position and its determinants in the sponsored search market. The empirical analysis indicated that the position of ads for web-only retailers

is dependent on bid values and ad relevancy factors, whereas multi-channel retailers are more reliant on their bid values.

Although there are a large number of studies on online marketing today, there are still limited articles on the topic of tourism. Yamamoto [2018] conducted a survey for attracting tourists online and measured its effect. He displayed ads on keyword search results related to regional tourism and used these to attract participants. Then he measured the percentage of visitors who visited a download (PDF brochure) site through the keyword advertising. Studies on tourism-related keyword search emerged after 2010; a few have attempted to analyze data to identify the number of visitors. However, no study using mobile phone location data has attempted to combine the data with keywords search volume so far.

3. Methods

As stated above, this study used the Mobile Kukan Toukei to count the number of visitors at specific tourist destinations. The service makes it possible to estimate the population structure of a region by gender, age, and residence. The locations and characteristics of the individuals obtained herein are derived through a non-identification process, aggregation processing, and concealment processing. Therefore, it is impossible to identify specific individuals.

This survey was conducted from December 27, 2015, to January 14, 2017. The sites surveyed are tourist destinations in Ishikawa Prefecture. The survey areas are presented in Figure 2.

There are two reasons why the area around two hot springs was chosen. First, both hot springs are close to Kanazawa City and thus compete with each other. Second, they are also popular enough to obtain sufficient search volume of keywords related to the areas. Kenrokuen Park, one of the most famous gardens in Japan, is located in Kanazawa City as well and was added to the survey area to see the difference of trends in population change from those at the hot springs.

When selecting these areas, it was essential to identify their "regional mesh codes." A regional mesh code is a code for identifying the regional mesh. It stands for an encoded area that is substantially divided into the same size of a square (mesh) based on the latitude and longitude in order to use it for statistics. With regard to regional mesh, there are three types of meshes: primary, secondary, and tertiary. The length of one side of a primary mesh is about 80 km, and those of secondary and tertiary meshes are about 10 km and 1 km respectively.

In addition, split regional meshes also exist, which are a more detailed regional division. A half-regional mesh is a tertiary mesh that is divided into two equal pieces in the vertical and horizontal directions. The length of one side is about 500 m. Furthermore, the length of one side of a quarter and 1/8 regional meshes is about 250 m and 125 m respectively.

For example, the mesh code of Wakura Hot Springs, which is one of the survey areas, is a third order code 5536-5703 (Figure 3). If the survey area cannot be covered in one mesh, it is possible to combine multiple meshes, like Kenrokuen whose codes are 5436-6572, 5436-6573-1, and 5436-6573-3 (Figure 4). Kenrokuen Park was added to the survey areas to compare with the hot springs and enable observation of their distinctions. Those of Yamanaka Hot Springs are composed of 5436-22995 and 436-2390 (see Figure 5).

This research also uses Google Trends, which is a website provided by Google which indicates the search volume of specific keywords in Google Search over time and also analyzes the popularity of top search queries across various regions and languages. The trends in the keyword search volume are expected to suggest tourists' behaviors and routes. Their various interests might influence their keyword search before or during the travel, and ultimately emerge as some kind of trend in a specific keyword's search volume.

Figure 2: Survey areas

Figure 3: The regional mesh of Wakura Hot Springs

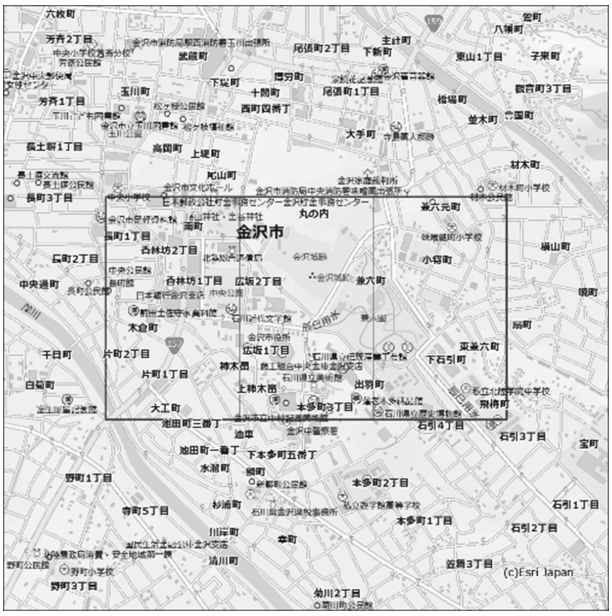

Figure 4: The regional mesh of Kenrokuen Park

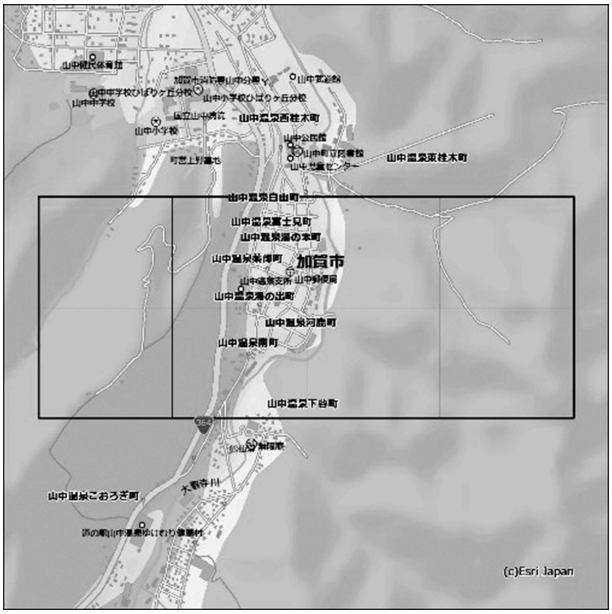

Figure 5: The regional mesh of Yamanaka Hot Springs

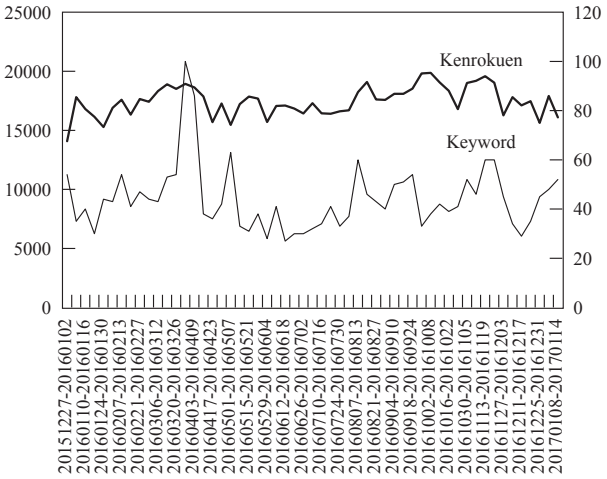

Figure 6: Trends in population change at Kenrokuen and keyword search volume (Weekly populations of Kenrokuen from Dec. 27, 2015, to Jan. 14, 2017)

Note: Left axis indicates population at Kenrokuen Park; right axis indicates keyword search volume.

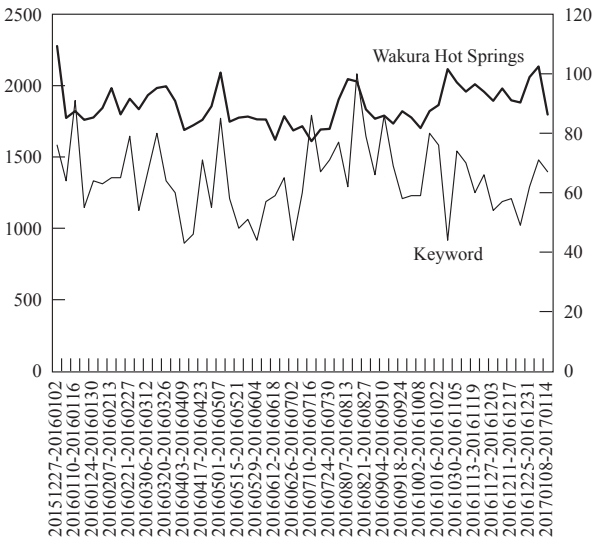

Figure 7: Trends in population change at Wakura Hot Springs and keyword search volume (Weekly populations of Wakura Hot Springs from Dec. 27, 2015, to Jan. 14, 2017)

Note: Left axis indicates population at Wakura Hot Springs Park; right axis indicates keyword search volume.

4. Analysis of the data

Here onwards, the paper presents an analysis by connecting the Mobile Kukan Toukei and keyword search volume. Figure 6 shows the transitions of the population at Kenrokuen and that of the search volume of the keyword "Kenrokuen." Although it seems as if the two trends fluctuate approximately in accord with each other as well as Wakura and Yamanaka Hot Springs (see Figures 7 and 8), the correlation coefficient between them is merely 0.265 (Table 1).

It should be noted that the transitions of population at Yamanaka Hot Springs indicated a relatively higher correlation coefficient (i.e. 0.444) with search volume of Kenrokuen than the volume of Yamanaka Hot Springs itself (0.395). What could this mean?

Similarly, the population at Wakura Hot Springs also showed a higher correlation coefficient (0.438) with the search volume for Kenrokuen. From these results, it could be assumed that tourists search for information not only about their current destination but also about neighboring places including where they are going to.

Next, this survey obtained search volume data for 42 keywords through Google Trends, provided by Google Inc., and attempted correlation analysis between the Mobile Kukan Toukei and the tourism related keywords' (e.g., name of tourism site)

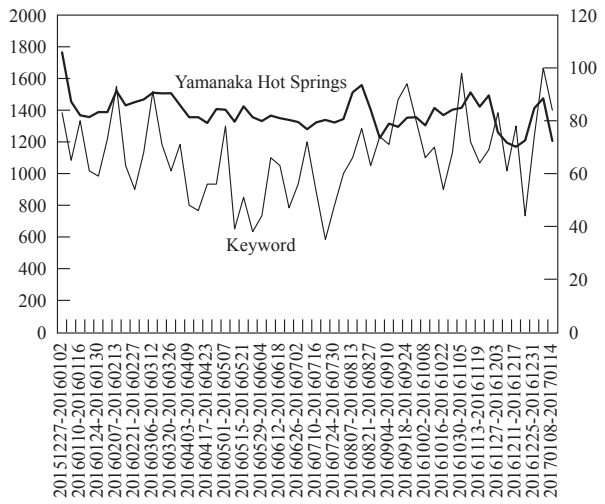

Figure 8: Trends in population change at Wakura Hot Springs and keyword search volume (Weekly populations of Wakura Hot Springs from Dec. 27, 2015, to Jan. 14, 2017)

Note: Left axis indicates population at Wakura Hot Springs Park; right axis indicates keyword search volume.

search volume data (see Table 2), from November 11th, 2015 to January 28th, 2017.

Table 3 shows that the transitions of the population at two hot springs indicated relatively higher correlation with the search volume of several keywords, whereas that of Kenrokuen Park barely displayed correlation except for a phrase, 21st Century Museum of Contemporary Art. As could be easily imagined, travelers do not necessarily search keywords of what they are currently enjoying but normally of what they are going to visit later.

4.1 Cross-correlation analysis

Then, what kind of process should be done when faced with the case of the Kenrokuen? This study also attempted cross-correlation analyses among them. With regard to the word Kenrokuen, its correlation coefficient rises when it comes to one week later or before (Table 4). As for other keywords, Kanazawa Tourism showed higher correlation three to five weeks before, and Kanazawa Station from one week to three weeks before as well. In this way, people search each keyword at different times.

Similar findings, which include negative correlations, were observed for Wakura and Yamanaka Hot Springs (see Tables 5 and 6). However, the tendency of the correlations to become more apparent in different periods is not as distinctive as in the case of Kenrokuen.

4.2 Regression analysis

Using the above data, linear equations with regard to the population at Wakura and Yamanaka hot springs could be derived (see Tables 7, 8, 9, and 10). The independent variables are chosen based on their correlation coefficients and thus forced entry was adopted. Those significant at beyond 5 % are excluded. Although both of the R^2 values are around 0.6, the adjusted R2 values are less than 0.6. Thus it is likely to be difficult to

Table 1: The correlation coefficients among the transitions of the population at tourist destinations and keyword search volume ($N = 55$)

		Population			Keyword search volume		
		Kenrokuen	Wakura	Yamanaka	Kenrokuen	Wakura	Yamanaka
Kenrokuen (pupulation)	Correlation coefficient	1	.011	.087	.265	–.079	.045
	Significance probability		.937	.528	.051	.567	.746
Wakura (pupulation)	Correlation coefficient	.011	1	.586**	.438**	.109	.401**
	Significance probability	.937		.000	.001	.427	.002
Yamanaka (pupulation)	Correlation coefficient	.087	.586**	1	.444**	.181	.395**
	Significance probability	.528	.000		.001	.187	.003

Note: ** The correlation coefficient is significant at the 1 % level.

Table 2: The list of the keywords surveyed

Area	Keywords
Kanazawa	Kenrokuen, Kanazawa Tourism, Kanazawa Station, Kanazawa Castle, Higashichayagai, Korinbo, 21st Century Museum of Contemporary Art (21st CMCA), Omicho Market, Itaru (restaurant), Kazuemachi, Yuwaku Hot Springs, Toshiie Maeda
Wakura Hot Springs	Wakura Hot Springs, Wajima, Wajima Market, Notojima Island, Notojima Aquarium, Noto Tourism, Chirihama Beach, Suzu, Wakura Station, Noto Airport
Yamanama Hot Springs	Yamanama Hot Springs, Katayamazu Hot Springs, Yamashiro Hot Springs, Kaga Hot Springs, Kaga Station, Awazu Hot Springs, Komatsu Airport
Others	Hokuriku Tourism, Hokuriku Shinkansen, Gokayama, Shirakawago, Fukui Tourism, Fukui Station, Eiheiji Temple, Tojinbo, Awara Hot Springs, Toyama Tourism, Tateyama, Toyama Station, Unazuki Hot Springs

Table 3: The Correlation coefficients between transitions of the population at tourist destinations and the keyword search volume ($N = 55$)

		Kenrokuen	Kanazawa Tourism	Kanazawa Station	Kanazawa Castle	Higashi chayagai	Korinbo
Population of Kenrokuen	Correlation coefficient	.265	–.023	.168	.085	.143	–.003
	Significance probability	.051	.868	.221	.539	.297	.985
Population of Wakura	Correlation coefficient	.438**	.161	.450**	.118	.442**	.434**
	Significance probability	.001	.240	.001	.391	.001	.001
Population of Yamanaka	Correlation coefficient	.444**	.369**	.449**	.253	.456**	.269*
	Significance probability	.001	.006	.001	.062	.000	.047
		21st CMCA	Omicho Market	Itaru	Kazue machi	Yuwaku Hot Springs	Toshiie Maeda
Population of Kenrokuen	Correlation coefficient	.403**	.039	.140	–.094	.131	–.179
	Significance probability	.002	.780	.310	.497	.342	.191
Population of Wakura	Correlation coefficient	.080	.635**	.200	–.029	.175	.011
	Significance probability	.560	.000	.143	.836	.202	.936
Population of Yamanaka	Correlation coefficient	.087	.619**	.269*	–.021	.250	.040
	Significance probability	.530	.000	.047	.879	.065	.771
		Wakura	Wajima	Wajima Market	Notojima	Notojima Aquarium	Noto Tourism
Population of Kenrokuen	Correlation coefficient	–.079	–.170	.244	.012	–.027	–.075
	Significance probability	.567	.213	.072	.930	.846	.589
Population of Wakura	Correlation coefficient	.109	–.140	.155	–.147	.109	.000
	Significance probability	.427	.308	.258	.284	.429	.999
Population of Yamanaka	Correlation coefficient	.181	.035	.116	.025	.152	.102
	Significance probability	.187	.799	.399	.854	.268	.461
		Chirihama	Suzu	Wakura Station	Noto Airport	Yamanaka	Katayamazu
Population of Kenrokuen	Correlation coefficient	.111	.184	–.047	.091	.045	–.061
	Significance probability	.420	.180	.734	.510	.746	.660
Population of Wakura	Correlation coefficient	–.078	–.063	.296*	.029	.401**	.229
	Significance probability	.571	.647	.028	.832	.002	.093
Population of Yamanaka	Correlation coefficient	.026	–.067	.228	.029	.395**	.375**
	Significance probability	.851	.629	.094	.835	.003	.005
		Yamashiro	Kaga	Kaga Station	Awazu	Hokuriku Tourism	Hokuriku Shinkansen
Population of Kenrokuen	Correlation coefficient	.037	.087	.102	.097	–.106	–.229
	Significance probability	.790	.526	.458	.480	.440	.093
Population of Wakura	Correlation coefficient	.421**	–.478**	.302*	.461**	–.247	.232
	Significance probability	.001	.000	.025	.000	.069	.088
Population of Yamanaka	Correlation coefficient	.224	–.262	.324*	.275*	.153	.379**
	Significance probability	.100	.054	.016	.042	.265	.004
		Komatsu Airport	Gokayama	Shirakawago	Fukui Tourism	Fukui Station	Eiheiji
Population of Kenrokuen	Correlation coefficient	.097	.032	–.180	–.070	.183	–.076
	Significance probability	.483	.817	.190	.610	.181	.582
Population of Wakura	Correlation coefficient	–.219	.329*	.378**	–.024	.263	.457**
	Significance probability	.109	.014	.004	.862	.053	.000
Population of Yamanaka	Correlation coefficient	–.108	.375**	.240	.143	.348**	.493**
	Significance probability	.431	.005	.078	.296	.009	.000
		Tojinbo	Awara	Toyama Tourism	Tateyama	Toyama Station	Unazuki
Population of Kenrokuen	Correlation coefficient	.103	–.062	–.072	.027	.172	–.041
	Significance probability	.454	.653	.602	.846	.211	.764
Population of Wakura	Correlation coefficient	–.003	.266	.040	–.287*	.208	.005
	Significance probability	.984	.050	.772	.033	.128	.970
Population of Yamanaka	Correlation coefficient	–.026	.342*	.249	–.116	.192	.198
	Significance probability	.851	.011	.067	.397	.161	.148

Note: * The correlation coefficient is significant at the 5 % level. **. The correlation coefficient is significant at the 1 % level.

Table 4: The cross-correlation coefficients between the transition of the population at Kenrokuen and the keyword search volume ($N = 55$)

		Kenrokuen	Kanazawa Tourism	Kanazawa Station	Higashi chayagai	21st CMCA	Omicho Market
The same week	Correlation coefficient	.265	–.023	.168	.143	.403**	.039
	Significance probability	.051	.868	.221	.297	.002	.780
1 week later	Correlation coefficient	.366**	–.129	.230	.185	.312*	.239
	Significance probability	.006	.346	.091	.180	.020	.079
2 weeks later	Correlation coefficient	.236	–.249	–.005	–.081	.197	.007
	Significance probability	.082	.067	.974	.565	.150	.959
1 week before	Correlation coefficient	.354**	.208	.460**	.314*	.474**	.188
	Significance probability	.008	.128	.000	.020	.000	.169
2 weeks before	Correlation coefficient	.225	.249	.552**	.358**	.506**	.346**
	Significance probability	.099	.067	.000	.007	.000	.010
3 weeks before	Correlation coefficient	.183	.432**	.495**	.266*	.465**	.254
	Significance probability	.182	.001	.000	.050	.000	.062
4 weeks before	Correlation coefficient	–.040	.355**	.318*	.177	.375**	–.008
	Significance probability	.774	.008	.018	.196	.005	.953
5 weeks before	Correlation coefficient	–.163	.401**	.261	.134	.250	–.071
	Significance probability	.234	.002	.054	.329	.066	.605
6 weeks before	Correlation coefficient	–.081	.272*	.281*	.148	.451**	.045
	Significance probability	.556	.044	.038	.282	.001	.746

Notes: * The correlation coefficient is significant at the 5 % level. ** The correlation coefficient is significant at the 1 % level. The table only includes part of the results.

Table 5: The cross-correlation coefficients between the transition of the population at Wakura Hot Springs and the keyword search volume ($N = 55$)

		Notojima Aquarium	Noto Tourism	Yamashiro	Kaga	Awazu	Hokuriku Tourism
The same week	Correlation coefficient	.109	.000	.421**	–.478**	.461**	–.247
	Significance probability	.429	.999	.001	.000	.000	.069
1 week later	Correlation coefficient	–.193	–.265	.027	–.398**	.279*	–.227
	Significance probability	.158	.051	.845	.003	.039	.096
2 weeks later	Correlation coefficient	–.436**	–.458**	.242	–.263	.280*	–.414**
	Significance probability	.001	.000	.075	.052	.039	.002
1 week before	Correlation coefficient	.030	–.178	.382**	–.334*	.189	–.351**
	Significance probability	.825	.192	.004	.013	.167	.009
2 weeks before	Correlation coefficient	–.192	–.319*	.433**	–.316*	.212	–.348**
	Significance probability	.160	.018	.001	.019	.121	.009
3 weeks before	Correlation coefficient	–.226	–.306*	.364**	–.002	.346**	–.258
	Significance probability	.097	.023	.006	.990	.010	.057
4 weeks before	Correlation coefficient	–.215	–.278*	.291*	.174	.315*	–.241
	Significance probability	.114	.040	.031	.203	.019	.077
5 weeks before	Correlation coefficient	.053	–.153	–.143	–.277*	–.159	.095
	Significance probability	.702	.265	.299	.040	.246	.491
6 weeks before	Correlation coefficient	.105	–.221	–.094	–.215	–.139	–.146
	Significance probability	.447	.106	.496	.116	.312	.288

Notes: * The correlation coefficient is significant at the 5 % level. ** The correlation coefficient is significant at the 1 % level. The table only includes part of the results.

Table 6: The cross-correlation coefficients between the transition of the population at Yamanaka Hot Springs and the keyword search volume ($N = 55$)

		Kenrokuen	Yamanaka	Katayamazu	Yama shiro	Hokuriku Shinkansen	Goka yama
The same week	Correlation coefficient	.444**	.395**	.375**	.224	.379**	.375**
	Significance probability	.001	.003	.005	.100	.004	.005
1 week later	Correlation coefficient	.328*	.386**	.374**	–.068	.241	.256
	Significance probability	.015	.004	.005	.621	.079	.059
2 weeks later	Correlation coefficient	.216	.169	.095	.119	.251	–.017
	Significance probability	.113	.217	.492	.387	.070	.903
1 week before	Correlation coefficient	.113	.402**	.364**	.374**	.305*	.376**
	Significance probability	.410	.002	.006	.005	.023	.005
2 weeks before	Correlation coefficient	–.050	.138	.263	.328*	.302*	.201
	Significance probability	.718	.316	.052	.014	.025	.142
3 weeks before	Correlation coefficient	–.099	.081	.068	.131	.197	.172
	Significance probability	.473	.559	.620	.342	.150	.209
4 weeks before	Correlation coefficient	–.010	–.012	.170	.090	.191	.131
	Significance probability	.940	.932	.216	.512	.163	.340
5 weeks before	Correlation coefficient	.045	–.007	.002	.163	.235	.112
	Significance probability	.742	.957	.988	.234	.084	.414
6 weeks before	Correlation coefficient	.158	–.014	.126	.076	.320*	.234
	Significance probability	.250	.916	.361	.582	.017	.086

Notes: * The correlation coefficient is significant at the 5 % level. ** The correlation coefficient is significant at the 1 % level. The table only includes part of the results.

Table 7: The summary of the model (Wakura and Yamanaka Hot Springs)

Model	R	R^2	Adjusted R^2	Standard error of estimated value
Wakura	.819a	.671	.533	95.23022
Yamanaka	.765a	.586	.395	79.88023

Note: "a" is predicted value (constant).

adopt these models.

In this manner, the linear equation for Kenrokuen Park can also be obtained. However, fewer correlation coefficients are significant at the 1 % level for Kenrokuen.

Next, this study incorporates the concept of time lag, as shown in Table 4. The value of R^2 is 0.782 and the adjusted R^2 is 0.664 (see Tables 11, 12, and 13). Table 13 contains several keywords indicating low correlations with the population of Kenrokuen at the same time, such as "Kanazawa tourism" and "Unazuki." Those keywords showed higher correlations with the population when a different time was selected.

The linear equations for Wakura and Yamanaka Hot Springs can be modified by also incorporating such keywords.

5. Conclusion and future challenges

This study demonstrated an analysis by connecting the Mobile Kukan Toukei and keyword search volume. The various interests of the tourists influence their keyword search before or during travel, and ultimately emerge as some kind of trend in a specific keyword's search volume. This study also attempted cross-correlation analyses among them. As a result, a

Table 8: The analysis of the variance (Wakura and Yamanaka Hot Springs)

Model		Sum of squares	Degree of freedom	Average square	*F* value	Significance probability
Wakura	Regression	704171.774	16	44010.736	4.853	.000[b]
	Residual error	344614.226	38	9068.795		
	Total	1048786.000	54			
Yamanaka	Regression	333845.316	17	19637.960	3.078	.002[b]
	Residual error	236091.521	37	6380.852		
	Total	569936.836	54			

Note: "b" is predicted value (constant).

Table 9: The factors of the model (Wakura Hot Springs)

	Model	Non-standardization factor		Standardization factor β	t value	Significance probability
		B	Standard error			
Wakura	(constant)	1164.225	297.384		3.915	.000
	PopulationYamanaka	.236	.184	.174	1.284	.207
	Kenrokuen	–.537	1.365	–.054	–.393	.696
	KanazawaStation	.029	3.154	.002	.009	.993
	Higashichayagai	1.921	1.504	.198	1.277	.209
	Korinbo	1.460	1.466	.124	.996	.326
	Omicho	1.060	1.969	.110	.538	.593
	WakuraStation	.551	.874	.074	.630	.532
	Yamanaka	–.202	1.580	–.019	–.128	.899
	Yamashiro	1.592	1.613	.127	.987	.330
	Kaga	–2.005	1.049	–.232	–1.910	.064
	KagaStation	–.559	1.455	–.049	–.384	.703
	Awazu	.945	.909	.123	1.039	.306
	Gokayama	–.149	1.382	–.017	–.108	.915
	Shirakawago	.124	1.444	.014	.086	.932
	Eiheiji	2.636	1.776	.211	1.484	.146
	Tateyama	–1.322	.819	–.227	–1.615	.115

Note: A dependent variable is the population at Wakura Hot Springs.

Table 10: The factors of the model (Yamanaka Hot Springs)

	Model	Non-standardization factor		Standardization factor β	t value	Significance probability
		B	Standard error			
Yamanaka	(constant)	624.041	242.581		2.573	.014
	PopulationWakura	.239	.127	.325	1.885	.067
	Kenrokuen	–.107	1.284	–.015	–.083	.934
	KanazawaTourism	1.229	1.876	.151	.655	.516
	KanazawaStation	–.610	3.006	–.050	–.203	.840
	Higashichayagai	.357	1.384	.050	.258	.798
	Korinbo	–1.035	1.293	–.119	–.800	.429
	Omicho	.999	1.717	.141	.582	.564
	Itaru	.590	.830	.095	.711	.482
	Yamanaka	.009	1.211	.001	.007	.994
	Katayamazu	–.147	1.252	–.019	–.118	.907
	KagaStation	.691	1.240	.083	.557	.581
	Awazu	–.925	.818	–.163	–1.130	.266
	HokurikuShinkansen	1.858	1.104	.241	1.683	.101
	Gokayama	.818	.998	.130	.819	.418
	FukuiStation	–.561	1.702	–.061	–.330	.744
	Eiheiji	3.258	1.547	.355	2.107	.042
	Awara	–.099	1.427	–.012	–.069	.945

Note: A dependent variable is the population at Yamanaka Hot Springs.

Table 11: The summary of the model (Kenrokuen)

Model	R	R^2	Adjusted R^2	Standard error of estimated value
Kenrokuen	.884a	.782	.664	720.10244

linear equation could be derived. These findings could lead to a model to forecast tourism demand in a destination.

The linear equations illustrated above can be made more accurate by incorporating the average. For example, that of Ken-

Table 12: The analysis of the variance (Kenrokuen)

Model		Sum of squares	Degree of freedom	Average square	*F* value	Significance probability
Kenrokuen	Regression	65161023.47	19	3429527.551	6.614	.000[b]
	Residual error	18149163.51	35	518547.529		
	Total	83310186.98	54			

Table 13: The factors of the model (Kenrokuen)

	Model	Non-standardization factor		Standardization factor β	*t* value	Significance probability
		B	Standard error			
Kenrokuen	(constant)	9720.875	1740.55		5.585	0
	Kenrokuen (+1)	8.908	9.094	0.1	0.98	0.334
	Kanazawa Tourism (–3)	9.094	15.01	0.096	0.606	0.549
	Kanazawa Station (–2)	27.271	25.828	0.193	1.056	0.298
	Higashichayagai (–2)	–2.773	11.5	–0.032	–0.241	0.811
	21st CMCA (–2)	-2.35	12.498	–0.029	–0.188	0.852
	Omicho (–2)	0.731	12.365	0.009	0.059	0.953
	Kazuemachi (+2)	–15.87	5.335	–0.275	–2.974	0.005
	Yuwaku (–6)	16.655	7.329	0.238	2.273	0.029
	ToshiieMaeda (–6)	14.418	8.88	0.178	1.624	0.113
	Wakura (–3)	–2.986	11.76	–0.03	–0.254	0.801
	Wajima Market (–1)	11.086	9.627	0.14	1.152	0.257
	Notojima (–6)	–16.375	12.104	–0.227	–1.353	0.185
	Chirihama (–3)	–13.94	11.129	–0.239	–1.253	0.219
	Suzu (–6)	21.253	11.699	0.259	1.817	0.078
	Komatsu Airport (–3)	–8.343	13.956	–0.078	–0.598	0.554
	Fukui Station (–1)	20.786	11.77	0.187	1.766	0.086
	Tojinbo (–3)	17.45	14.145	0.192	1.234	0.226
	Toyama Station (–6)	14.146	15.776	0.118	0.897	0.376
	Unazuki (–3)	22.361	9.794	0.289	2.283	0.029

Note: A dependent variable is population at Kenrokuen Park. "+1" means "1 week later," while "-1" stands for "1 week before."

Table 14: The cross-correlation coefficients between the transition of the population at Kenrokuen and the keyword search volume ($N = 55$)

		The same time	1 week before	1 week later	Average (previous 2 weeks)	Average (successive 2 weeks)	Average (3 weeks)
Kenrokuen (population)	Correlation coefficient	.265	.354**	.366**	.356**	.361**	.426**
	Significance probability	.051	.008	.006	.008	.007	.001

Note: ** The correlation coefficient is significant at the 1 % level.

rokuen used the search volume of the keyword "Kenrokuen" in one week before the transition of the population. The resulting correlation coefficient was 0.366 (Table 14). Taking the average volume of three weeks (previous, present, and successive), the correlation coefficient would be slightly higher (0.426).

Of course, these equations still have room to be made more reliable by adopting several other factors such as the population of the same week in the previous year. Further search for better factors would be a future challenge of this study as well as attempting the same approach in different areas. In addition, building linear equations by using tourists' characteristics such as their gender, age, and residence should also be considered as the Mobile Kukan Toukei can examine them.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP15K01970.

Notes

(1) "Mobile Kukan Toukei" is a trademark of NTT DOCOMO, Inc.

(2) (*) NTT DOCOMO's "Mobile Kukan Toukei" services are only available to subscribers in Japan.

References

Ayanso, A. and Karimi, A. (2015). The moderating effects of keyword competition on the determinants of ad position in sponsored search advertising, *Decision Support Systems*, Vol. 70, 42-59.

Ahas, R., Aasa, A., Roose, A., Mark, Ü., and Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, Vol. 29, No. 3, 469-485.

Gao, H. and Liu, F. (2013). Estimating freeway traffic measures from mobile phone location data. *European Journal of Operational Research*, Vol. 229, No. 1, 252-260.

Japan Tourism Agency. (2014). *Keitaidenwa kara erareru ichijouhou tou wo katuyousita hounichi gaikokujin doutaichousa houkokusho* [Foreign visitors' dynamics research report utilizing mobile phone location information]. Retrieved September 28, 2018, from http://www.mlit.go.jp/common/001080545.pdf.

Liu, F., Janssens, D., Wets, G., and Cools, M. (2013). Annotating mobile phone location data with activity purposes using machine learning algorithms. *Expert Systems with Applications*, Vol. 40, No. 8, 3299-3311.

Okinawa Prefecture Culture, Sports, and Tourism Department. (2013). *Senryakuteki repeater souzou jigyou houkokusho* [Report on strategic creation of repeaters]. Retrieved October 5, 2019, from http://www.pref.okinawa.jp/site/bunka-sports/kankoseisaku/kikaku/report/houkokusixyo/documents/07dairokusiyou.pdf.

Pan, B. and Li, X. R. (2011). The long tail of destination image and online marketing. *Annals of Tourism Research*, Vol. 38, No. 1, 132-152.

Pan, B., Litvin, S. W., and O'Donnell, T. E. (2007). Understanding accommodation search query formulation: The first step in putting 'heads in beds'. *Journal of Vacation Marketing*, Vol. 13, 371-381.

Sheldon, P, J. (1997). *Tourism information technology*. CABI Publishing.

Steenbruggen, J., Tranos, E., and Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities? *Telecommunications Policy*, Vol. 39, No. 3-4, 335-346.

Xiang, Z. and Gretzel, U. (2010). Role of social media in online travel information search original. *Tourism Management*, Vol. 31, No. 2, 179-188.

Xiang, Z. and Pan, B. (2011). Travel queries on cities in the United States: Implications for search engine marketing for tourist destinations. *Tourism Management*, Vol. 32, No. 1, 88-97.

Yamamoto, M. (2019). Furthering big data utilization in tourism, in F. García Márquez and B. Lev (eds.), *Data science and digital business*. Springer, 157-171.

(Received November 12, 2019; accepted December 25, 2019)