Bulletin of Applied Economics, 2025, 12(2), *135-147*
https://doi.org/10.47260/bae/1227

Tourist Sites and Visitor Numbers in Taiwan: An Online Buzz Analysis

Chien-Jung Ting[1*], Yi-Long Hsiao[2*] and Jia-Jhen Shen[3]

Abstract

This study highlights the influence of online buzz by examining the effect of tourism and recreational site-related keywords on inbound tourism to Taiwan. Using large-scale, real-time data from Google Trends, we assess how search interest in key attractions predicts tourist arrivals. Principal Component Analysis (PCA) is applied to construct a composite index of search activity, allowing us to explore its relationship with inbound tourist numbers. The results indicate that six keywords - Longshan Temple, Jiufen, Miramar Entertainment Park, Fushoushan Farm, Yangmingshan National Park, and Raohe Street Night Market - are significantly associated with tourist arrivals. The growing use of social media and real-time information platforms has enhanced tourists' awareness and interest in Taiwanese attractions, supporting the steady growth of inbound tourism. Government initiatives promoting smart tourism through digital technology and big data further amplify the visibility of Taiwan's tourism resources and contribute to tourism development. The study's principal contribution lies in leveraging real-time data; by providing richer and timelier information than conventional official statistics, these data yield more precise and statistically significant results. Moreover, because few studies use real-time data to examine the direct effect of online buzz on inbound tourist arrivals to Taiwan, this paper fills that gap.

JEL classification: C60, O11, R11.

Keywords: Principal Component Analysis (PCA), Tourism and Recreational Site, Vector Autoregressive (VAR).

[1] Department of Leisure, Recreation, and Tourism Management, Southern Taiwan University of Science and Technology, Taiwan. *Corresponding author.
[2] Department of Finance, National Dong Hwa University, Taiwan. *Corresponding author.
[3] Department of Leisure, Recreation, and Tourism Management, Southern Taiwan University of Science and Technology, Taiwan.

Article Info: Received: July 23, 2025. Revised: August 12, 2025. Published online: August 19, 2025.

1. Research Motivation

Taiwan's diverse tourism and recreational attractions, including ecotourism, have drawn increasing numbers of domestic and international tourists. Travelers often rely on online platforms, social media, and digital reviews to plan their trips, making tourism decisions more convenient and data-driven. In response, the Taiwanese government has promoted tourism through digital marketing, enhancing the visibility and diversity of travel experiences. Many of Taiwan's attractions have become international hotspots, recognized by global travel media such as *Wanderlust*, *Global Traveler*, and *BRUTUS*.[4]

According to a 2023 survey by Taiwan's Tourism Administration, MOTC([22]), natural and cultural experiences, along with culinary tourism, are key motivators for domestic travel. Taiwan's rich natural landscapes, cultural heritage, and gastronomy contribute to its appeal as a premier leisure destination. In 2023, 90% of residents traveled domestically, averaging 9.79 trips per person. While the pandemic disrupted tourism, inbound arrivals rebounded after border reopening in late 2022. According to data from the National Statistics, R.O.C. (Taiwan), the total number of inbound visitors to Taiwan reached 7.85 million in 2024, based on cumulative figures from January to December of that year.

Given the observed correlation between tourism site popularity and visitor numbers, this study investigates how online buzz - measured through real-time keyword search data - affects inbound tourism to Taiwan. Following recent big data research trends, we use high-frequency Google Trends data to analyze search interest in key attractions. We apply Principal Component Analysis (PCA) to extract key indices that reflect the primary tourism interests of travelers, and examine their relationship with inbound tourism using a Vector Autoregressive (VAR) model. Following the keyword classification method proposed by Matsumoto et al. (2013)([18]), we categorize recreational activity keywords into 14 major categories and 70 subcategories, then construct diffusion indices for each category. By evaluating the predictive power of these indices on inbound tourism, we aim to identify which categories of Taiwanese tourism and recreational sites are most strongly associated with tourist inflows.

Prior tourism studies using Google Trends largely fall into two strands: (i) single-country analyses that relate tourism-related search keywords to service consumption (Matsumoto et al., 2013[18]); (ii) methodological comparisons that deploy diverse econometric techniques to benchmark the predictive content of Google Trends - evaluating its strengths, limitations, and forecasting performance for single countries and across multiple variables and time points (Höpken et al., 2020[12]; Cebrián and Domenech, 2022[5]; De Luca and Rosciano, 2024[9]). Building on this literature, our study differs in several respects. We focus on inbound arrivals to Taiwan and ask whether online buzz strengthens the causal link to visitation. Leveraging high-frequency, real-time search data - which embed richer information than lagged official statistics - we obtain sharper, statistically significant estimates. Methodologically, we employ complementary econometric models and emphasize bidirectional causal dynamics. Substantively, we fill a gap by directly assessing how online buzz around tourist and recreational attractions affects arrivals to Taiwan, which constitutes a principal contribution of this study.

Ultimately, this study seeks to enhance forecasting accuracy of inbound tourism using rich, real-time data, and to better understand the extent to which interest in various tourism and recreational sites influences tourist arrivals in Taiwan.

---

[4] Taiwan has been recognized by several international travel magazines as a top destination. In 2024, the UK's Wanderlust magazine awarded Taiwan the Silver Prize for "The Most Desirable Island" for the second consecutive year. That same year, the U.S. magazine Global Traveler named Taiwan the "Best Leisure Destination in Asia" in its Leisure Lifestyle Awards. In 2023, Japan's BRUTUS magazine featured Taiwan under the theme "A Long-Awaited Overseas Journey Begins with Taiwan," showcasing its attractions and cuisine.

2. Literature review

The emergence of big data has prompted numerous studies to utilize high-frequency, real-time data as predictive variables for economic indicators and policy applications. The literature relevant to this study can be summarized as follows:

2.1 Internet Buzz and Forecasting GDP and Private Consumption

Numerous studies have adopted high-frequency data and dynamic factor models to enhance GDP forecasting accuracy (Barhoumi et al., 2010[3]; Boivin and Ng, 2005[4]; Evans, 2005[10]; Giannone et al., 2005[11]; Marcellino et al., 2003[17]). For example, Banbura (2011)([2]) integrated various types of high-frequency indicators and found improved GDP forecast performance. Similarly, Notini et al. (2012)([20]) used monthly data to support quarterly GDP predictions, outperforming official central bank figures. Liebermann (2012)([15]) applied bridge equations and factor models to forecast Ireland's GDP using domestic and international monthly indicators, and Luciani and Ricci (2013)([16]) employed a Bayesian Dynamic Factor Model to predict Norway's annual GDP growth, demonstrating superior performance compared to central bank estimates.

Furthermore, internet search data has also been used to forecast private consumption. Vosen and Schmidt (2011, 2012)([23],[24]) and Kholodilin et al. (2010)([13]) found that these alternative indicators often outperform traditional models.

2.2 Internet Buzz and Labor/Financial Markets

Askitas and Zimmermann (2009)([1]) identified a strong correlation between keyword search volume and unemployment in Germany. D'Amuri and Marcucci (2010)([8]) used Google job search trends as leading indicators for U.S. unemployment and found them superior to conventional models. In the financial domain, Takeda and Wakao (2014)([21]) found that while keyword searches for Japanese stock indices were not significantly correlated with stock prices, they were positively associated with trading volume.

2.3 Internet Buzz and Tourism

Matsumoto et al. (2013)([18]) analyzed the impact of tourism-related search keywords on Japanese consumption before and after the Great East Japan Earthquake. Choi and Varian (2012)([6]) demonstrated that incorporating Google search trends improved tourism demand forecasting compared to traditional models. Clark et al. (2019)([7]) used Google Trends search data to build a visitation-forecasting model and benchmarked it against a traditional autoregressive (AR) specification. For 2013–2017, their Google Trends model explained 97% of the variation in aggregate park visitation one year ahead and outperformed the AR model across all evaluation metrics, though accuracy varied markedly across individual park units. Extending this line of inquiry, Höpken et al. (2020)([12]) compared an autoregressive integrated moving average (ARIMA) model with a machine-learning approach based on artificial neural networks (ANN), showing that (i) Google Trends data, as a proxy for online search behavior, significantly improves forecasts relative to models using only lagged visitation, and (ii) ANN outperforms ARIMA. Taking a broader view, Li et al. (2021)([14]) reviewed tourism-forecasting studies using Internet data (2012–2019) and synthesized findings across four categories - search engines, web traffic, social media, and multi-source datasets - concluding that search-engine data are most frequently employed, time-series and econometric models remain dominant, and AI methods are still maturing. Addressing data quality, Cebrián and Domenech (2022)([5]) documented the widespread use of Google Trends for forecasting tourism demand and other economic variables, while highlighting key reliability issues by demonstrating variability across repeated queries on six different dates. More recently, De Luca and Rosciano (2024)([9]) proposed a transfer-function modeling framework that integrates Google Trends to forecast U.S. arrivals to Italy, offering decision-ready insights for industry stakeholders.

These studies suggest that high-frequency internet data offer richer real-time insights than conventional government statistics. Building on this foundation, our study employs real-time keyword search data related to tourism and recreational sites in Taiwan to forecast inbound tourist volume. Guided by prior literature and official data from Taiwan's Ministry of Transportation and Communications (MOTC), we identify popular keywords for empirical analysis. Keywords are categorized into 14 major types and 70 subcategories, representing Taiwan's most visited attractions:

1. **National Parks**: Yangmingshan, Taroko, Kenting, Shei-Pa, Yushan.
2. **National Scenic Areas**: Yehliu, Baisha Bay, Alishan, Sun Moon Lake, Penghu.
3. **Forest Recreation Areas**: Xitou, Taipingshan, Lala Mountain, Guanwu, Wuling.
4. **Leisure Farms and Agricultural Zones**: Cingjing Farm, Green World, Flying Cow Ranch, Janfusun, Fushoushan.
5. **Tourist Regions**: Chihkan Tower, Jiufen, Anping Old Street, Taipei 101, Confucius Temple.
6. **Museums**: National Museum of Marine Biology, Chiang Kai-shek Memorial Hall, Hakka Cultural Park, National Palace Museum, Chimei Museum.
7. **Religious Sites**: Longshan Temple, Dajia Jenn Lann Temple, Lukang Tianhou Temple, Beigang Chaotian Temple, Nankunshen Temple.
8. **Theme Parks**: Lihpao Land, Leofoo Village, Formosan Aboriginal Culture Village, Farglory Ocean Park, E-Da Theme Park.
9. **Fishing Ports**: Kezailiao, Zhuwai, Yong'an, Wuqi, Hsinchu.
10. **Natural Landscapes**: Shuinandong, Gaomei Wetlands, Shifen Waterfall, Thermal Valley, Qingshui Cliffs.
11. **Historic Streets**: Qishan, Daxi, Tamsui, Sanxia, Lukang.
12. **Cultural & Creative Sites**: Huashan 1914 Creative Park, Songshan Cultural Park, Pier-2 Art Center, National Kaohsiung Center for the Arts, Blueprint Cultural and Creative Park.
13. **Shopping Areas**: Hayashi Department Store, Miramar Entertainment Park, Yizhong Street, Ximending, Ruifeng Night Market.
14. **Night Markets**: Shilin, Ningxia, Fengjia, Kenting Street, Raohe.

This classification provides the basis for analyzing the relationship between internet search volume and inbound tourist flows, aiming to support evidence-based tourism policy through timely, data-driven insights.

3. Empirical Model

This study explores the dynamic relationship between internet search behavior and inbound tourism in Taiwan by integrating high-frequency online data with conventional tourism statistics. Drawing upon existing literature and the "Cumulative Statistics of Visitors to Major Tourist and Recreational Sites by Type" published by the MOTC([22]), the research identifies representative tourist and recreational destinations across the country. To quantify public interest in these destinations, the study utilizes search intensity data obtained from Google Trends, which serves as a proxy for real-time public attention.

The primary objective is to assess whether high-frequency online search trends related to major tourist attractions can effectively forecast tourism indicators, such as the number of annual inbound tourists. To operationalize this objective, the study employs Principal Component Analysis (PCA) to reduce the dimensionality of the keyword data and construct a series of diffusion indices that capture common movements in search activity. These indices are then incorporated into a Vector Autoregressive (VAR) model, which allows for the joint modeling of inbound tourist arrivals and online search dynamics, as well as the exploration of potential causal relationships among the variables.

The empirical analysis covers the period from 2004 to 2024, during which Google Trends data for selected keywords related to tourism and recreation in Taiwan were collected. To ensure the statistical robustness of the time series, standard preprocessing techniques are applied. The stationarity of each series is evaluated using the Augmented Dickey-Fuller (ADF) unit root test, and appropriate transformations - such as differencing or logarithmic scaling - are implemented when necessary to stabilize the series for subsequent modeling.

To synthesize the keyword data, a $T \times N$ matrix X is constructed, where T represents the number of time periods and N the number of keywords. PCA is applied to extract the top k components that explain the majority of variance in the data. This process involves two key steps. First, the eigenvalue decomposition of XX' is performed to identify the top k eigenvectors corresponding to the largest eigenvalues, which yield the factor matrix F, representing the latent common factors shared across keyword trends. Second, the factor loading matrix β is estimated using the model $X = F\beta + \varepsilon$, where ε denotes the residual matrix. The resulting diffusion indices, denoted $F1$, $F2$,…, Fn, are interpreted as latent indicators of collective public interest in recreational and tourist destinations and are subsequently used as explanatory variables in the VAR model.

In the VAR framework, the diffusion indices and the series for inbound tourist arrivals are modeled as a vector of endogenous variables. Let Z_t represent the vector at time t, which includes inbound tourist arrivals Y_t and the three diffusion indices $F1_t$, $F2_t$, and $F3_t$. The VAR model is formally expressed as:

$$Z_t = \alpha + \mathbf{B} Z_{t-1} + \varepsilon_t, \tag{1}$$

where α is a vector of intercept terms, **B** is the matrix of autoregressive coefficients $\hat{\beta}_1$ to $\hat{\beta}_{80}$, and ε_t is the vector of error terms. In this specification, each variable is regressed on its own lagged values and those of the other variables, allowing for a comprehensive representation of their interdependencies[5].

[5] According to the principle of selecting the model with smaller AIC, BIC, and HQ values, we choose a five-lag vector autoregression model (VAR), denoted as VAR(5). The results of the VAR model are presented with its four equations labeled sequentially as Model 1, Model 2, Model 3, and Model 4. The following are the specifications for Models 1 through 4:

Model 1: $Y_t = \alpha_1 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \beta_5 Y_{t-5} + \beta_6 F1_{t-1} + \beta_7 F1_{t-2} + \beta_8 F1_{t-3} + \beta_9 F1_{t-4} + \beta_{10} F1_{t-5} + \beta_{11} F2_{t-1} + \beta_{12} F2_{t-2} + \beta_{13} F2_{t-3} + \beta_{14} F2_{t-4} + \beta_{15} F2_{t-5} + \beta_{16} F3_{t-1} + \beta_{17} F3_{t-2} + \beta_{18} F3_{t-3} + \beta_{19} F3_{t-4} + \beta_{20} F3_{t-5}$

Model 2: $F1_t = \alpha_2 + \beta_{21} Y_{t-1} + \beta_{22} Y_{t-2} + \beta_{23} Y_{t-3} + \beta_{24} Y_{t-4} + \beta_{25} Y_{t-5} + \beta_{26} F1_{t-1} + \beta_{27} F1_{t-2} + \beta_{28} F1_{t-3} + \beta_{29} F1_{t-4} + \beta_{30} F1_{t-5} + \beta_{31} F2_{t-1} + \beta_{32} F2_{t-2} + \beta_{33} F2_{t-3} + \beta_{34} F2_{t-4} + \beta_{35} F2_{t-5} + \beta_{36} F3_{t-1} + \beta_{37} F3_{t-2} + \beta_{38} F3_{t-3} + \beta_{39} F3_{t-4} + \beta_{40} F3_{t-5}$

Model 3: $F2_t = \alpha_3 + \beta_{41} Y_{t-1} + \beta_{42} Y_{t-2} + \beta_{43} Y_{t-3} + \beta_{44} Y_{t-4} + \beta_{45} Y_{t-5} + \beta_{46} F1_{t-1} + \beta_{47} F1_{t-2} + \beta_{48} F1_{t-3} + \beta_{49} F1_{t-4} + \beta_{50} F1_{t-5} + \beta_{51} F2_{t-1} + \beta_{52} F2_{t-2} + \beta_{53} F2_{t-3} + \beta_{54} F2_{t-4} + \beta_{55} F2_{t-5} + \beta_{56} F3_{t-1} + \beta_{57} F3_{t-2} + \beta_{58} F3_{t-3} + \beta_{59} F3_{t-4} + \beta_{60} F3_{t-5}$

Model 4: $F3_t = \alpha_4 + \beta_{61} Y_{t-1} + \beta_{62} Y_{t-2} + \beta_{63} Y_{t-3} + \beta_{64} Y_{t-4} + \beta_{65} Y_{t-5} + \beta_{66} F1_{t-1} + \beta_{67} F1_{t-2} + \beta_{68} F1_{t-3} + \beta_{69} F1_{t-4} + \beta_{70} F1_{t-5} + \beta_{71} F2_{t-1} + \beta_{72} F2_{t-2} + \beta_{73} F2_{t-3} + \beta_{74} F2_{t-4} + \beta_{75} F2_{t-5} + \beta_{76} F3_{t-1} + \beta_{77} F3_{t-2} + \beta_{78} F3_{t-3} + \beta_{79} F3_{t-4} + \beta_{80} F3_{t-5}$

$$\begin{bmatrix} Y_t \\ F1_t \\ F2_t \\ F3_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \boldsymbol{B} \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ Y_{t-3} \\ Y_{t-4} \\ Y_{t-5} \\ F1_{t-1} \\ F1_{t-2} \\ F1_{t-3} \\ F1_{t-4} \\ F1_{t-5} \\ F2_{t-1} \\ F2_{t-2} \\ F2_{t-3} \\ F2_{t-4} \\ F2_{t-5} \\ F3_{t-1} \\ F3_{t-2} \\ F3_{t-3} \\ F3_{t-4} \\ F3_{t-5} \end{bmatrix} + \begin{bmatrix} \varepsilon_{Yt} \\ \varepsilon_{F1t} \\ \varepsilon_{F2t} \\ \varepsilon_{F3t} \end{bmatrix} \tag{2}$$

To examine the temporal relationships and potential causality among the variables, two statistical tests are employed. The Granger Causality test is used to evaluate whether past values of the diffusion indices provide significant predictive power for inbound tourist arrivals. If significant, this would suggest that internet search activity precedes changes in tourism demand, implying a leading indicator role for online behavior. Complementing this, the Wald test is applied to assess the joint significance of the estimated coefficients within each equation of the VAR model, thereby validating the structural soundness of the model and the explanatory power of the diffusion indices.

Overall, this study offers a novel methodological contribution to tourism forecasting by demonstrating the feasibility of using high-frequency internet search data as predictive inputs for low-frequency tourism indicators. By transforming keyword trends into latent diffusion indices and embedding them in a dynamic multivariate framework, the model provides a new lens for understanding the informational role of online buzz in shaping tourism patterns. In practical terms, the findings present valuable implications for policymakers, tourism authorities, and marketing professionals. Real-time monitoring of search behavior can serve as an early warning system for shifts in tourist inflows, enabling more agile and data-driven decision-making in tourism planning, promotional strategies, and resource allocation.

4. Data and Empirical Findings

4.1 Data Description

This study utilizes high-frequency keyword search data related to popular recreational and tourism destinations in Taiwan to forecast the lower-frequency data of inbound tourist arrivals[6]. Details regarding the data and their sources are presented in Table 1. As the Google Trends database begins from January 2004, the study period spans from 2004 to 2024. The keyword data were obtained from Google Trends,

[6] The original keyword data in this study is on a weekly basis, while the number of inbound tourists to Taiwan is published annually by government statistics. We converted the weekly keyword data into annual data to conduct a VAR test alongside the annual tourist arrival data. As Google Trends data begin in January 2004, the study period spans from 2004 to 2024.

while data on the number of inbound tourists were collected from the Tourism Statistics Database of the MOTC.

Table 1: Data Sources and Descriptions

	Category	Name	Data Source	Data Period	Frequency
Tourism and recreational site-related keywords	National Parks	Yangmingshan, Taroko, Kenting, Shei-Pa,Yushan	Google trends	2004-2024	Annual
	National Scenic Areas	Yehliu, Baishawan, Alishan, Sun Moon Lake, Penghu			
	Forest Recreation Areas	Xitou, Taipingshan, Lala Mountain, Guanwu, Wuling			
	Leisure Farms and Agricultural Zones	Qingjing, Green World, Flying Cow Ranch, Janfusun, Fushoushan			
	Tourist Regions	Chihkan Tower, Jiufen, Anping Old Street, Taipei 101, Confucius Temple			
	Museums	National Museum of Marine, Biology, Chiang Kai-shek Memorial Hall, Hakka Cultural Park,National Palace Museum, Chimei Museum			
	Religious Sites	Longshan Temple, Dajia Jenn Lann Temple, Lukang Tianhou Temple, Beigang Chaotian Temple, Nankunshen Temple			
	Theme Parks	Lihpao Land, Leofoo Village, Formosan Aboriginal Culture, Village, Farglory Ocean Park, E-Da Theme Park			
	Fishing Ports	Kezailiao, Zhuwai, Yong'an,Wuqi, Hsinchu			
	Natural Landscapes	Shuinandong, Gaomei Wetlands, Shifen Waterfall,Thermal Valley, Qingshui Cliffs			
	Historic Streets	Qishan, Daxi, Tamsui, Sanxia, Lukang			
	Cultural & Creative Sites	Huashan 1914 Creative Park, Songshan Cultural Park, Pier-2 Art Center, National Kaohsiung Center for the Arts, Blueprint Cultural and Creative Park			
	Shopping Areas	Hayashi Department Store, Miramar Entertainment Park, Yizhong Street, Ximending, Ruifeng Night Market			
	Night Markets	Shilin, Ningxia, Fengjia, Kenting Street, Raohe			

4.2 Empirical Results

Part I: Factor Loadings of the Diffusion Index

Table 2 presents the factor loadings derived from Principal Component Analysis (PCA). Following the Kaiser criterion (eigenvalue greater than one), three diffusion indices were extracted. Among these, the first component ($F1$) from the tourism and recreational site-related keyword dataset exhibited the strongest explanatory power. The six keywords with the highest loadings on $F1$ are "Lungshan Temple," "Jiufen," "Miramar Entertainment Park," "Fushoushan Farm," "Yangmingshan National Park," and "Raohe Street Night Market." These findings indicate that search interest in these destinations serves as a leading indicator of increases in inbound tourism, reflecting their significance as major points of interest.

Table 2: The factor loadings of diffusion indices

Tourism and recreational site-related keywords	**Diffusion indices**		
	F1	F2	F3
Components of Diffusion index (keywords)	Ranking of components (by weight)		
Yangmingshan	**5**	–	–
Taroko		–	–
Kenting		–	–
Shei-Pa		–	–
Yushan		–	–
Yehliu		–	–
Baishawan		–	–
Alishan		–	–
Sun Moon Lake		–	–
Penghu		–	–
Xitou		–	–
Taipingshan		–	–
Lala Mountain		–	–
Guanwu		–	–
Wuling		–	–
Qingjing		–	–
Green World		–	–
Flying Cow Ranch		–	–
Janfusun		–	–
Fushoushan	**4**	–	–
Chihkan Tower		–	–
Jiufen	**2**	–	–
Anping Old Street		–	–
Taipei 101		–	–
Confucius Temple		–	–
National Museum of Marine Biology		–	–
Chiang Kai-shek Memorial Hall		–	–
Hakka Cultural Park		–	–
National Palace Museum		–	–
Chimei Museum		–	–
Longshan Temple	**1**	–	–
Dajia Jenn Lann Temple		–	–
Lukang Tianhou Temple		–	–
Beigang Chaotian Temple		–	–

Nankunshen Temple		–	–
Lihpao Land		–	–
Leofoo Village		–	–
Formosan Aboriginal Culture Village		–	–
Farglory Ocean Park		–	–
E-Da Theme Park		–	–
Kezailiao		–	–
Zhuwai		–	–
Yong'an		–	–
Wuqi		–	–
Hsinchu		–	–
Shuinandong		–	–
Gaomei Wetlands		–	–
Shifen Waterfall		–	–
Thermal Valley		–	–
Qingshui Cliffs		–	–
Qishan		–	–
Daxi,		–	–
Tamsui		–	–
Sanxia		–	–
Lukang		–	–
Huashan 1914 Creative Park		–	–
Songshan Cultural Park		–	–
Pier-2 Art Center		–	–
National Kaohsiung Center for the Arts		–	–
Blueprint Cultural and Creative Park		–	–
Hayashi Department Store		–	–
Miramar Entertainment Park	**3**	–	–
Yizhong Street		–	–
Ximending		–	–
Ruifeng Night Market		–	–
Shilin		–	–
Ningxia		–	–
Fengjia		–	–
Kenting Street		–	–
Raohe	**5**	–	–
First reigenvalues of the correlation matrix	47.363	5.306	4.185
Variability explained	0.812		

Source: The authors. 1. In table 2, we choose the diffusion index having significant effects on dependent variables in VAR tests in Table 3, that's F1. 2. Based on the ranking of components, we conclude the higher-weight variables in bold italics, which are "Longshan Temple, Jiufen, Miramar Entertainment Park, Fushoushan Farm, Yangmingshan National Park, and Raohe Street Night Market."

Part II: Results from the Vector Autoregressive (VAR) Model

Table 3 reports the estimation results from the VAR (5) model. In Model 1, the first diffusion index ($F1$) is found to have a statistically significant effect. Specifically, the number of inbound tourists in the previous period (Y_{t-1}) significantly influences the current period (Y_t), and the third lag of $F1$ ($F1_{t-3}$) also has a significant impact on Y_t.

Table 3: VAR results

Model 1		Model 2		Model 3		Model 4	
Coefficient estimate							
$\widehat{\alpha}_1$	**_4.01*** (1.16)_**	$\widehat{\alpha}_2$	0.05(0.05)	$\widehat{\alpha}_3$	0.01(0.07)	$\widehat{\alpha}_4$	0.09(0.08)
$\widehat{\beta}_1$	**_0.86***(0.15)_**	$\widehat{\beta}_{21}$	-0.02(0.01)	$\widehat{\beta}_{41}$	-0.02(0.01)	$\widehat{\beta}_{61}$	0.01(0.01)
$\widehat{\beta}_2$	-0.15(0.21)	$\widehat{\beta}_{22}$	0.03(0.01)	$\widehat{\beta}_{42}$	0.004(0.01)	$\widehat{\beta}_{62}$	-0.03(0.01)
$\widehat{\beta}_3$	-0.10(0.21)	$\widehat{\beta}_{23}$	-0.002(0.01)	$\widehat{\beta}_{43}$	0.01(0.01)	$\widehat{\beta}_{63}$	0.005(0.01)
$\widehat{\beta}_4$	0.13(0.21)	$\widehat{\beta}_{24}$	-0.03(0.01)	$\widehat{\beta}_{44}$	-0.0004(0.01)	$\widehat{\beta}_{64}$	0.02(0.01)
$\widehat{\beta}_5$	-0.26(0.16)	$\widehat{\beta}_{25}$	-0.001(0.01)	$\widehat{\beta}_{45}$	-0.04(0.01)	$\widehat{\beta}_{65}$	-0.01(0.01)
$\widehat{\beta}_6$	-6.77(4.21)	$\widehat{\beta}_{26}$	**_0.58***(0.17)_**	$\widehat{\beta}_{46}$	**_0.67*** (0.25)_**	$\widehat{\beta}_{66}$	0.32(0.28)
$\widehat{\beta}_7$	-1.08(4.97)	$\widehat{\beta}_{27}$	**_0.76*** (0.20)_**	$\widehat{\beta}_{47}$	-0.07(0.29)	$\widehat{\beta}_{67}$	-0.63(0.33)
$\widehat{\beta}_8$	**_10.73*(5.84)_**	$\widehat{\beta}_{28}$	-0.59(0.24)	$\widehat{\beta}_{48}$	-1.33(0.34)	$\widehat{\beta}_{68}$	0.26(0.38)
$\widehat{\beta}_9$	0.64(6.75)	$\widehat{\beta}_{29}$	0.10(0.27)	$\widehat{\beta}_{49}$	**_0.75*(0.40)_**	$\widehat{\beta}_{69}$	0.16(0.44)
$\widehat{\beta}_{10}$	-3.04(5.17)	$\widehat{\beta}_{30}$	-0.03(0.21)	$\widehat{\beta}_{50}$	0.15(0.30)	$\widehat{\beta}_{70}$	-0.28(0.34)
$\widehat{\beta}_{11}$	3.72(2.71)	$\widehat{\beta}_{31}$	0.18(0.11)	$\widehat{\beta}_{51}$	0.08(0.16)	$\widehat{\beta}_{71}$	0.04(0.18)
$\widehat{\beta}_{12}$	-0.38(2.29)	$\widehat{\beta}_{32}$	-0.02(0.09)	$\widehat{\beta}_{52}$	0.15(0.13)	$\widehat{\beta}_{72}$	0.01(0.15)
$\widehat{\beta}_{13}$	1.55(1.95)	$\widehat{\beta}_{33}$	-0.09(0.08)	$\widehat{\beta}_{53}$	-0.28(0.11)	$\widehat{\beta}_{73}$	-0.02(0.13)
$\widehat{\beta}_{14}$	-1.33(1.98)	$\widehat{\beta}_{34}$	-0.01(0.08)	$\widehat{\beta}_{54}$	-0.07(0.12)	$\widehat{\beta}_{74}$	0.18(0.13)
$\widehat{\beta}_{15}$	-1.76(1.97)	$\widehat{\beta}_{35}$	-0.02(0.08)	$\widehat{\beta}_{55}$	-0.12(0.11)	$\widehat{\beta}_{75}$	0.13(0.13)
$\widehat{\beta}_{16}$	1.29(2.45)	$\widehat{\beta}_{36}$	0.04(0.10)	$\widehat{\beta}_{56}$	-0.22(0.14)	$\widehat{\beta}_{76}$	-0.19(0.16)
$\widehat{\beta}_{17}$	1.73(2.51)	$\widehat{\beta}_{37}$	-0.08(0.10)	$\widehat{\beta}_{57}$	0.10(0.15)	$\widehat{\beta}_{77}$	-0.09(0.16)
$\widehat{\beta}_{18}$	-0.80(2.06)	$\widehat{\beta}_{38}$	0.04(0.08)	$\widehat{\beta}_{58}$	-0.34(0.12)	$\widehat{\beta}_{78}$	-0.31(0.13)
$\widehat{\beta}_{19}$	0.04(2.31)	$\widehat{\beta}_{39}$	0.06(0.09)	$\widehat{\beta}_{59}$	-0.26(0.14)	$\widehat{\beta}_{79}$	0.18(0.13)
$\widehat{\beta}_{20}$	3.17(2.48)	$\widehat{\beta}_{40}$	-0.01(0.10)	$\widehat{\beta}_{60}$	-0.02(0.15)	$\widehat{\beta}_{80}$	0.07(0.16)
det(SSE)	0.00002						
AIC	-8.357						
BIC	-5.787						
HQ	-7.336						

1. Here, the results of the VAR model are presented with its four equations labeled sequentially as Model 1, Model 2, Model 3, and Model 4. The following are the specifications for Models 1 through 4:

Model 1: $Y_t = \alpha_1 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \beta_5 Y_{t-5} + \beta_6 F1_{t-1} + \beta_7 F1_{t-2} + \beta_8 F1_{t-3} + \beta_9 F1_{t-4} + \beta_{10} F1_{t-5} + \beta_{11} F2_{t-1} + \beta_{12} F2_{t-2} + \beta_{13} F2_{t-3} + \beta_{14} F2_{t-4} + \beta_{15} F2_{t-5} + \beta_{16} F3_{t-1} + \beta_{17} F3_{t-2} + \beta_{18} F3_{t-3} + \beta_{19} F3_{t-4} + \beta_{20} F3_{t-5}$

Model 2: $F1_t = \alpha_2 + \beta_{21} Y_{t-1} + \beta_{22} Y_{t-2} + \beta_{23} Y_{t-3} + \beta_{24} Y_{t-4} + \beta_{25} Y_{t-5} + \beta_{26} F1_{t-1} + \beta_{27} F1_{t-2} + \beta_{28} F1_{t-3} + \beta_{29} F1_{t-4} + \beta_{30} F1_{t-5} + \beta_{31} F2_{t-1} + \beta_{32} F2_{t-2} + \beta_{33} F2_{t-3} + \beta_{34} F2_{t-4} + \beta_{35} F2_{t-5} + \beta_{36} F3_{t-1} + \beta_{37} F3_{t-2} + \beta_{38} F3_{t-3} + \beta_{39} F3_{t-4} + \beta_{40} F3_{t-5}$

Model 3: $F2_t = \alpha_3 + \beta_{41} Y_{t-1} + \beta_{42} Y_{t-2} + \beta_{43} Y_{t-3} + \beta_{44} Y_{t-4} + \beta_{45} Y_{t-5} + \beta_{46} F1_{t-1} + \beta_{47} F1_{t-2} + \beta_{48} F1_{t-3} + \beta_{49} F1_{t-4} + \beta_{50} F1_{t-5} + \beta_{51} F2_{t-1} + \beta_{52} F2_{t-2} + \beta_{53} F2_{t-3} + \beta_{54} F2_{t-4} + \beta_{55} F2_{t-5} + \beta_{56} F3_{t-1} + \beta_{57} F3_{t-2} + \beta_{58} F3_{t-3} + \beta_{59} F3_{t-4} + \beta_{60} F3_{t-5}$

Model 4: $F3_t = \alpha_4 + \beta_{61} Y_{t-1} + \beta_{62} Y_{t-2} + \beta_{63} Y_{t-3} + \beta_{64} Y_{t-4} + \beta_{65} Y_{t-5} + \beta_{66} F1_{t-1} + \beta_{67} F1_{t-2} + \beta_{68} F1_{t-3} + \beta_{69} F1_{t-4} + \beta_{70} F1_{t-5} + \beta_{71} F2_{t-1} + \beta_{72} F2_{t-2} + \beta_{73} F2_{t-3} + \beta_{74} F2_{t-4} + \beta_{75} F2_{t-5} + \beta_{76} F3_{t-1} + \beta_{77} F3_{t-2} + \beta_{78} F3_{t-3} + \beta_{79} F3_{t-4} + \beta_{80} F3_{t-5}$

2. The values in parentheses represent standard deviations. Statistical significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

These results highlight the predictive power of $F1$, implying a causal relationship whereby keyword search interest - particularly that captured by $F1$ - serves as an effective predictor of future inbound tourism demand. The six most influential keywords within $F1$ further reinforce the role of these destinations - Lungshan Temple, Jiufen, Miramar Entertainment Park, Fushoushan Farm, Yangmingshan National Park, and Raohe Street Night Market - as critical drivers of tourist activity.

Although no bidirectional causality was found between tourist arrivals and the diffusion index, the significant influence of lagged keyword indicators suggests that public interest in recreational activities, as reflected in online search behavior, contributes meaningfully to the decision-making process of potential visitors.

In 2024, Taiwan's tourism industry continued its post-pandemic recovery. According to data from the National Statistics, R.O.C. (Taiwan), the total number of inbound visitors to Taiwan reached 7.85 million in the period from January to December 2024, marking an increase from 2023 and indicating sustained international appeal. Taiwan's diverse natural landscapes, rich cultural heritage, and vibrant local cuisine make it an attractive destination. Whether historical landmarks, night markets, or its convenient transportation network, these features enhance the country's tourism competitiveness. A large proportion of tourists came from Japan, South Korea, and Singapore, underscoring Taiwan's prominence in the Asian tourism market. The industry is thus experiencing a steady rebound, supported by efforts to improve tourism quality and diversify target markets.

The empirical results of this study reveal that specific destinations - Lungshan Temple, Jiufen, Miramar Entertainment Park, Fushoushan Farm, Yangmingshan National Park, and Raohe Street Night Market - exert significant influence on inbound tourist numbers. These destinations represent six distinct tourism categories: tourist regions (Jiufen), religious sites (Lungshan Temple), shopping areas (Miramar), leisure farms and agricultural zones (Fushoushan Farm), national parks (Yangmingshan), and night markets (Raohe Street). A common characteristic among these sites is their widespread recognition and popularity, both domestically and internationally. Their influence may be attributed not only to their inherent touristic appeal but also to the Taiwanese government's focused promotion strategies, including digital marketing and media exposure which have increased their search visibility and online presence. For instance:

Lungshan Temple, with its deep historical and religious roots, serves as a key cultural landmark that attracts both domestic and international visitors. **Jiufen**, famed for its gold mining past, Japanese colonial architecture, and nostalgic mountain-town ambiance, has become globally renowned through film and animation portrayals. **Miramar Entertainment Park**, featuring the iconic Ferris wheel launched in 2004, has emerged as a popular urban attraction in Taipei. **Fushoushan Farm** offers seasonal landscapes and immersive farming experiences, making it a favored destination for nature-based tourism. **Yangmingshan National Park**, recognized in 2020 as the world's first "Urban Quiet Park," boasts hot springs and ecological diversity. It was also listed among *Time Magazine*'s 100 Greatest Places in 2021. **Raohe Street Night Market** leverages a unique culinary culture and effective international marketing to attract visitors seeking authentic local experiences.

5. Conclusion

This study identifies a set of key tourism-related keywords—"Lungshan Temple," "Jiufen," "Miramar Entertainment Park," "Fushoushan Farm," "Yangmingshan National Park," and "Raohe Street Night Market"—which represent six major tourism categories and significantly influence the number of inbound tourists to Taiwan.

In the digital era, online search behavior plays a pivotal role in shaping travel decisions. Data derived from search engines reveals a strong positive correlation between the search volume of tourism-related keywords and actual visitor numbers. This study applies digital trace data to examine the relationship between online interest in recreational destinations and inbound tourism, finding that both digital visibility and government policies contribute meaningfully to attracting visitors.

The findings highlight three main contributing factors: (1) the appeal of Taiwan's natural and cultural attractions, supported by government efforts to promote tourism rooted in local identity and the arts; (2) a well-connected transportation network that enhances accessibility; and (3) policy instruments such as travel subsidies and promotional campaigns, which effectively boost tourist arrivals.

In summary, Taiwan's flourishing tourism sector is underpinned by strategic policy support, robust digital marketing, and continuous improvements in infrastructure. When rising online interest is successfully converted into physical visits and bolstered by supportive policy interventions, a virtuous cycle can emerge—stimulating the growth of the tourism industry and contributing to a sustained increase in inbound tourist arrivals.

As noted in the literature, converting unstructured social-media and multi-source data into forecastable time series requires further advances in text mining, sentiment analysis, and social network analytics. Future research should prioritize multi-source data integration; with ongoing progress in digital technologies and analytics, scholars are encouraged to incorporate social-media corpora and sentiment analysis to quantify review content, discussion intensity, and affective valence on online platforms, and to embed these measures in predictive models to further enhance the accuracy of tourism forecasts.

References

Askitas, N. and K. F. Zimmermann (2009). Google Econometrics and Unemployment Forecasting. Discussion Paper No. 4201.

Banbura, M. (2011) Nowcasting with daily data. European Central Bank, Working Paper.

Barhoumi, K., O. Darné and L. Ferrara (2010). Are Disaggregate Data Useful for Factor Analysis in Forecasting French GDP? Journal of Forecasting, 29(1-2), 132–144.

Boivin, J. and S. Ng (2005). Understanding and Comparing Factor-Based Forecasts. International Journal of Central Banking, 1(3), 117-151.

Cebrián, E. and J. Domenech (2022). Is Google Trends a quality data source? Applied Economics Letters, 30(6), 811–815.

Choi, H., and H. Varian (2012). Predicting the present with google trends. Economic Record,88(s1):2-9.

Clark, M., E.J. Wilkins, D.T. Dagan, R. Powell, R.L. Sharp and V. Hillis (2019). Bringing forecasting into the future: using Google to predict visitation in U.S. national parks. Journal of Environmental Management, 243, 88-94.

D' Amuri, F., and J. Marcucci (2010). "Googleit!" Forecasting the US Unemployment Rate with a Google Job Search index. Fondazione Eni Enrico Mattei Working Papers.

De Luca, G. and M. Rosciano (2024). Google Trends data and transfer function models to predict tourism demand in Italy. Journal of Tourism Futures 2024; https://doi.org/10.1108/JTF-01-2023-0018

Evans, Martin D. D. (2005). Where Are We Now? Real-Time Estimates of the Macroeconomy. International Journal of Central Banking, 1(2), September.

Giannone, D., L. Reichlin and L. Sala (2005). Monetary Policy in Real Time. NBER Chapters, in: NBER Macroeconomics Annual 2004, 19, 161-224

Höpken, W., T. Eberle, M. Fuchs and M. Lexhagen (2020). Improving Tourist Arrival Prediction: A Big Data and Artificial Neural Network Approach. Journal of Travel Research, 60(5), 998-1017.

Kholodilin, Konstantin A., M. Podstawski, and B. Siliverstovs. (2010). Do Google searches help in nowcasting private consumption? A real-time evidence for the US. KOF Swiss Economic Institute working Paper256.

Li, X., R. Law, G. Xie and S. Wang (2021). Review of tourism forecasting research with internet data. Tourism Management, 83, doi: https://doi.org/10.1016/j.tourman.2020.104245.

Liebermann, J. (2012) Short-term forecasting of quarterly gross domestic product growth. Central Bank of Ireland, Working Paper, 74-84.

Luciani, M. and L. Ricci (2013). Nowcasting Norway. ECARES, Working Papers.

Marcellino, M., J. Stock and M. Watson (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. European Economic Review, 47(1), 1-18

Matsumoto, A., K. Matsumura and N. Shiraki (2013). Potential of Search Data in Assessment of Current Economic Conditions. No. 2013-04-18. Bank of Japan.

National Statistics, R.O.C. (Taiwan)(2024). Number of Inbound Visitors to Taiwan. 2024 survey of "Key Economic and Social Indicators" from statistics database of the National Statistics, R.O.C. (Taiwan). https://www.stat.gov.tw/Point.aspx?sid=t.12&n=3591&sms=11480

Notini, H. H., J. V. Issler, C. Rodrigues, S. Matos and R. Bonelli (2012). Nowcasting Brazil monthly GDP: A State-Space approach. Economics, Working Papers.

Takeda, F. and T. Wakao (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. Pacific-Basin Finance Journal, 27, 1-18.

Taiwan's Ministry of Transportation and Communications (MOTC) (2023). Cumulative Statistics of Visitors to Major Tourist and Recreational Sites by Type. 2023 survey from Tourism statistics database of the tourism administration (MOTC).

Vosen, S. and T. Schmidt (2011). Forecasting private consumption: survey-based indicators vs. Google trends. Journal of Forecasting,30.6: 565-578.

Vosen, S. and T. Schmidt (2012). A monthly consumption indicator for Germany based on Internet search query data. Applied Economics Letters19.7: 683-687.