

Article

Does Google Analytics Improve the Prediction of Tourism Demand Recovery?

Ilsé Botha ^{1,*}  and Andrea Saayman ² 

¹ School of Accounting, Department of Accountancy, College of Business and Economics, University of Johannesburg, Aucklandpark Campus, Johannesburg 2006, South Africa

² School of Economic Sciences and Tourism Research in Economics, Environs and Society (TREES), North West University, Potchefstroom Campus, Potchefstroom 2531, South Africa; andrea.saayman@nwu.ac.za

* Correspondence: ilseb@uj.ac.za

Abstract: Research shows that Google Trend indices can improve tourism-demand forecasts. Given the impact of the recent pandemic, this may prove to be an important predictor of tourism recovery in countries that are still struggling to recover, including South Africa. The purpose of this paper is firstly, to build on previous research that indicates that Google Trends improves tourism-demand forecasting by testing this within the context of tourism recovery. Secondly, this paper extends previous research by not only including Google Trends in time-series forecasting models but also typical tourism-demand covariates in an econometric specification. Finally, we test the performance of Google Trends in forecasting over a longer time period, because the destination country is a long-haul destination where more lead time may be required in decision-making. Additionally, this research contributes to the body of knowledge by including lower frequency data (quarterly) instead of the higher frequency data commonly used in current research, while also focusing on an important destination country in Africa. Due to the differing data frequencies, the MIDAS modelling approach is used. The MIDAS models are compared to typical time-series and naïve benchmarks. The findings show that monthly Google Trends improve forecasts on lower frequency data. Furthermore, forecasts that include Google Trends are more effective in forecasting one to two quarters ahead, pre-COVID. This trend changed after COVID, when Google Trends led to improved recovery forecasts even over a longer term.

Keywords: Google analytics; tourism forecasting; MIDAS; South Africa; recovery; long haul destinations



Citation: Botha, I.; Saayman, A. Does Google Analytics Improve the Prediction of Tourism Demand Recovery? *Forecasting* **2024**, *6*, 908–924. <https://doi.org/10.3390/forecast6040045>

Academic Editors: Konstantinos Nikolopoulos and Vasileios Bougioukos

Received: 17 September 2024

Revised: 11 October 2024

Accepted: 16 October 2024

Published: 18 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recent years have seen an increase in the use of Big Data in forecasting studies, and tourism forecasting is no exception. According to [1], Big Data can be defined using the 4V concept—namely, big data is characterised by volume, variety, velocity and value. The rationale for using Big Data in tourism forecasting studies is based on the proposition that it reflects the preferences and decision-making processes of tourists [2]. According to [3], four types of Big Data can be identified in the tourism-demand literature, namely, web search data, online textual data, online photo data and social media data.

The most predominant type of Big Data found in tourism forecasting studies is web search data. Ref. [4] report that web search data from Google and Baidu search engines are most frequently employed, with [5] paving the way by using Google Trends data to forecast Hong Kong tourism. Since then, data from these search engines have been used to not only forecast tourism demand for countries and cities but also hotel demand [4]. The recent COVID-19 pandemic that disrupted tourism around the world caused a structural break in tourist arrivals to all countries. Given the impact that the recent pandemic had on the tourism industry, Big Data may prove to be an important predictor of tourism recovery in countries that are still struggling to recover. One such country is South Africa, where

tourism demand is still well below its pre-pandemic levels, it having recovered just more than 80% of pre-pandemic levels by the end of 2023.

Arrivals from Africa are South Africa's main tourism source markets, accounting for 75% of all tourist arrivals, with neighbouring countries accounting for the lion's share. In terms of the remaining 25% of overseas arrivals (2019 figures), the country's main tourism source markets are the UK (14.4%), the USA (14.2%), Germany (12.2%), France (6.3%), the Netherlands (5.5%), Australia (4.2%) and India (3.6%). In terms of recovery, arrivals from Africa have recovered quicker than overseas arrivals (84% versus 79%), underscoring the importance of improving recovery forecasting.

The purpose of this paper is, firstly, to build on previous research that indicates that Google Trends has the potential to improve tourism-demand forecasting by testing this within the context of tourism recovery after a shock. Secondly, this paper extends previous research by not only including Google Trends in time-series forecasting models but also including typical tourism-demand covariates in an econometric forecasting specification. Finally, we test the performance of Google Trends in forecasting over a longer time period because the destination country is a long-haul destination where more lead time may be required in decision-making.

The rationale for including Google Trends data in forecasting tourism arrivals can be found in the notion that potential tourists use the internet to plan their travel activities [5]. Therefore, the internet has become an important source of information for potential tourists and, according to [6], it signals tourists' consumption preferences and informs their decision-making. It is therefore not surprising that [1] reports that forecasting using web search data is more prevalent in tourism than in any other subject area.

When considering only tourism demand (and not papers that focus only on hotel demand, which is a subset of tourism demand), web search data has been used to forecast demand for countries, regions/provinces, cities and attractions. In terms of country research, tourism to Spain was investigated by [7,8]. Ref. [9] forecasted tourism arrivals to South Korea, [10] China, the USA, [11] Austria, [12] Germany, and [13] to China, Turkey and the US. Table 1 provides a brief overview of the country, web search engine, data frequency and estimation methods used in these studies.

In terms of regions/provinces/states, tourism to Hong Kong leads the way and has been studied by [4,5,14–17]. Ref. [2] forecasted tourism arrivals to Hainan, China, [18] Macau, China [19], Taiwan, China, and [20] the Caribbean. Table 2 provides a summary of these studies in terms of data frequency, search data and estimation method.

The research focussing on forecasting tourism to cities is substantial, with Beijing, China, the most popular city, studied by [21–25]. While most of these studies used the Baidu Index, refs. [23,25] used both the Baidu and Google Indices in their studies. Ref. [26] forecasted tourism to four Taiwan cities. For European cities, [27] forecasted tourist arrivals to Prague using Google Trends data, [28] tourism to Vienna and [29] forecasted tourism to Vienna and Barcelona. These studies are summarised in Table 3.

Table 1. Forecasting tourism to countries using web search data.

Countries	Author	Web Search Engine	Data Frequency	Estimation Method
Spain	Artola et al. (2015) [7]	Google	Monthly	ARIMAX
	Maximo and Jose (2018) [8]	Google	Monthly	AR-X
South Korea	Park et al. (2016) [9]	Google	Monthly	ARIMAX
China	Lv et al. (2018) [10]	Google and Baidu	Weekly and Monthly	SEAN * regression
	Wang et al. (2020) [13]	Google	Monthly	ANN-based, SARIMA
USA	Lv et al. (2018) [10]	Google and Baidu	Weekly and Monthly	SEAN * regression
	Wang et al. (2020) [13]	Google	Monthly	ANN-based, SARIMA
Turkey	Wang et al. (2020) [13]	Google	Monthly	ANN-based, SARIMA
Austria	Önder (2017) [29]	Google	Monthly	ADLM, Naïve, AR, Holt-Winters
Spain	Önder (2017) [29]	Google	Monthly	ADLM, Naïve, AR, Holt-Winters
Germany	Bokelman and Lessmann (2019) [12]	Google	Monthly	SARIMA, DLM *

* SEAN = stacked autoencoder with echo-state (ANN-based); DLM = distributed lag model.

Table 2. Forecasting tourism to regions/provinces using web search data.

Regions/Provinces	Author	Web Search Engine	Data Frequency	Estimation Method
Hong Kong	Gawlik et al. (2011) [14]	Google	Monthly	Weighted linear regression
	Choi and Varian (2012) [5]	Google	Monthly	AR-X
	Wen et al. (2019) [6]	Baidu	Monthly	ARIMA, ARIMAX, NAR, NARX, Hybrid
	Bai and Hao (2021) [15]	Baidu and Google	Monthly	Random Walk, ARIMAX, SVR, ANN, Two-step DB-ensemble DBN
	Li and Law (2020) [16]	Google	Monthly	AR, Empirical Mode Decomposition ARX
	Wen et al. (2021) [17]	Baidu	Monthly	Naïve, ETS, SARIMA, SARIMAX, MIDAS
	Xie et al. (2021) [4]	Baidu and Google	Monthly	ARIMA, BPNN, SVR, LSSVR, MA-LSSVR
Hainan, China	Yang et al. (2015) [2]	Baidu and Google	Monthly	ARIMAX
Macau, China	Hu and Song (2021) [18]	Google	Monthly	ANN
Taiwan	Huang and Yu (2019) [19]	Google	Monthly	Algorithms
Caribbean	Bangwayo-Skeete and Skeete (2015) [20]	Google	Monthly	AR-MIDAS

Tourism to specific attractions is also forecasted using search data with [30] forecasting tourism to the Miao Village in China, [31] forecasting demand for five London museums, [26,32] to Mount Siguniang, China, [33] to the Forbidden City, Beijing, and [21,34] to Jiuzhaigou, China.

From the tables above, some clear patterns can be identified. Firstly, there is a paucity of studies focussing on forecasting tourism to Africa, with Asia and Europe attracting most of the research interest. Secondly, the studies mainly use monthly data with monthly forecasts, with some utilising weekly data as well. Although not shown in the tables, forecasting attraction demand mainly uses daily and weekly data. The reason for this can

be found in the assertion that web search data may be helpful in predicting the present, in other words, for nowcasting [5]. Consequently, the current body of research tends to focus more on forecasting the near future than forecasting the longer term or using quarterly data.

Table 3. Forecasting tourism to cities using web search data.

Cities	Author	Web Search Engine	Data Frequency	Estimation Method
Beijing	Li et al. (2017) [21]	Baidu	Monthly	ARMA, Dynamic Factor Model
	Li et al. (2018b) [22]	Baidu	Monthly	BPNN
	Sun et al. (2019) [23]	Baidu and Google	Monthly	Extreme Machine Learning (EML), ARIMA, ARIMAX, ANN, SVR, LSSVR
	Li et al. (2021) [24]	Baidu	Monthly	ARIMA, ARIMAX, ML
	Sun et al. (2022) [25]	Google	Monthly	SN, SARIMA, SES, ARDL, SARIMAX, MLP, B-MLP, KELM, B-KELM, and SAKE
	Wu et al. (2023) [35]	Baidu	Monthly	SARIMA-MIDAS, DFM, ETS, SNaive
Taiwan cities	Hu & Wu (2022) [26]	Google	Monthly	Grey models (AI) and combinations
Prague	Havranek and Zeynalov (2019) [27]	Google	Weekly and Monthly	MIDAS
Vienna	Önder and Günter (2016) [28]	Google	Monthly	ADLM, Naïve, AR, Holt-Winters
	Önder (2017) [29]	Google	Monthly	ADLM, Naïve, AR, Holt-Winters
Barcelona	Önder (2017) [29]	Google	Monthly	ADLM, Naïve, AR, Holt-Winters

The tables thirdly provide a brief overview of the various forecasting methods employed. When considering country tourism forecasts, two methods dominate the research, namely, artificial neural networks or (ANN)-based methods and time-series methods, specifically ARIMA-type models and their extension to include web search data (ARX and ARIMAX). The only exceptions are [12,29], who used autoregressive distributed lag and distributed lag models, where various lags of the web search index are included. Noteworthy is the fact that none of the studies include any typical tourism demand or economic variables in the specification.

The same trend is visible in the provincial/regional and city tourism forecasts. The most predominant forecasting methods remain time-series-based methods (SARIMA, AR, ARX, ARIMA and ARIMAX) and ANN methods, with machine learning methods also becoming popular. In addition, mixed data sampling (MIDAS) was used by [17,20,27,35]. Similar to the country studies, none of the research considers other typical tourism demand or economic variables in the forecasting models, except [25], who include price and income data, and [35], who include economic index data.

Only two papers consider web search data in forecasting tourism over the COVID-19 period. One paper, by [36], uses hotel demand as the dependent variable and found that Google Trends does not improve the accuracy of demand recovery forecasts but concludes that it might be more useful for longer-haul destinations. The second paper [35] forecasted Chinese tourism recovery to Hong Kong. They found that web search data significantly enhances the recovery forecasting accuracy.

The current research addresses the gaps identified above by focussing on tourism to an African country, namely, South Africa, and specifically focussed on forecasting tourism recovery after COVID-19. In addition, we use mixed data sampling with monthly Google Trends data combined with quarterly tourist arrivals, price and income data. We therefore forecast over a longer time horizon and assess the efficacy of search data in improving forecasts over the medium-term. Additionally, this research contributes to the body of knowledge by including lower-frequency data (quarterly) instead of the higher-frequency data commonly used in current research, while also focusing on an important destination country in Africa.

2. Materials and Methods

As indicated, the objective of this research is to determine whether Google trend data improves the prediction of tourism-demand recovery in light of the COVID pandemic that started in 2020.

The data used in this research were tourist arrivals for the main source markets to South Africa and GDP and CPI data on a quarterly basis from 2004 Q1 to 2023 Q4. The data were sourced from Statistics South Africa (arrivals data, <https://www.statssa.gov.za/> accessed on 30 March 2024) and the OECD (<https://www.oecd.org/en/data.html> accessed on 30 March 2024) and IFS databases (<https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b> accessed 30 March 2024) for the GDP and CPI data, respectively. The focus was on arrivals from the UK, the USA, Germany, France, the Netherlands, Australia and India. The Google Trends data (web search, travel category) for these source markets are available on a monthly basis from January 2004 to May 2024. In addition, another set of Google Trends data was added to determine whether news searches (all categories) on South Africa from source markets improved the forecasts. These data are available from January 2008 to May 2024.

The search query for the Google Trends data for web searches was South Africa under the travel category for each source market. The related queries that fall under the travel category for each source market as well as for the worldwide search are shown in the word clouds below.

From Figure 1, apart from the general travel searches (flights, accommodation, travel, costs, holiday), it is evident that the South African source markets were interested in certain tourist attractions, such as South African safaris, Cape Town, Johannesburg, game reserves and the Kruger National Park. The searches for the Indian source market were mainly in terms of cricket.

The scaled and normalised Google Trends data were used. Figure 2 plots the total overseas tourist arrivals and the Google web search data (monthly) to the destination country, South Africa. These figures show that there may be an association between the monthly search data and the monthly tourist arrivals, with the Google Trends series a leading indicator of tourist arrivals.

In terms of methods, the naïve (without drift), ARIMA (1,1) optimized to convergence with the Berndt–Hall–Hall–Hausman and ARDL models, were used as the benchmark models. These models were expanded into an AR-MIDAS-web search model, an AR-MIDAS-web and news search model, an ARDL-MIDAS-web search model and an ARDL-MIDAS-web and news search model. The MIDAS approach was appropriate because it is able to incorporate mixed frequencies [20,28,37]; in this case, monthly Google Trends data, quarterly arrivals and price and income data.

The United States



The United Kingdom



The Netherlands



Germany



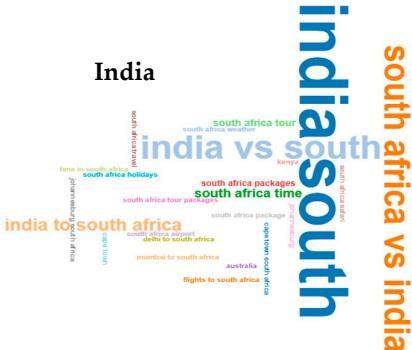
France



Australia



India



Worldwide



Figure 1. Related search queries per source market.

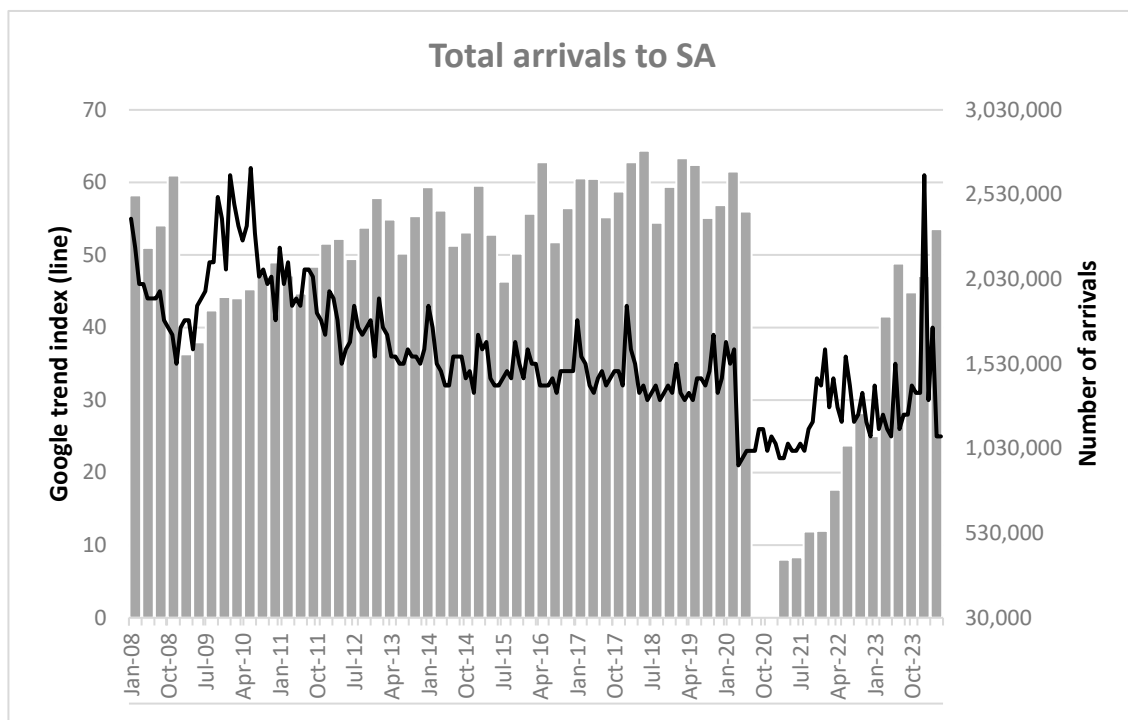


Figure 2. Total arrivals vs. the Google Trend index.

The Augmented Dickey–Fuller test was used to determine the integration order of the time series. The breakpoint unit root test (ADF) was also used to account for the structural break in the data. The results rejected the null hypothesis of a unit root, and therefore, all the series were $I(0)$.

The specification of the AR-MIDAS model for each source market is [20,28,37]:

$$TA_t = \alpha + \beta TA_{t-1} + \gamma \sum_{k=1}^m \omega(k, \theta) L_{HF}^k GT_t + \varepsilon_t$$

In the model, TA_t represents tourist arrivals, the low frequency dependent variable (quarterly basis), and TA_{t-1} represents the lag of tourist arrivals; these variables are transformed in logarithmic form. GT_t represents the high-frequency independent variable on a monthly basis. L_{HF}^k denotes the lag operator for monthly lags k of GT_t and m represents the maximum lag order. This was set to $m = 4$, which means that four months of search queries have an impact on current tourist arrivals. $\omega(k, \theta)$ represents the weighing function that can have several functional forms (e.g., Almon exponential, Almon polynomial, beta) that determine the weight of the temporal aggregation of the high-frequency observations. α, β, γ are the parameters to be estimated, and ε_t is the random error term.

As indicated above, two AR-MIDAS models were estimated for each source market, one including the Google Trends index (web search) under the travel category (MIDAS_W) and another including both the web and news search Google Trends indices (MIDAS_WN). The optimal specification for each country was determined using the Schwarz Criterion.

The alternative econometric model for tourism demand is the ARDL model:

$$TA_t = \alpha_0 + \sum_{i=1}^p \alpha_1 TA_{t-i} + \sum_{j=0}^q \alpha_2 Y_{t-j} + \sum_{j=0}^q \alpha_3 P_{t-j} + \varepsilon_t$$

where TA_t is arrivals, Y_t is income of the origin country, P_t is relative prices and ε_t is an i.i.d. white-noise error term. because inflation measures the cost of living in the destination country, the variable P_t is determined as follows:

$$P_t = (CPI_{jt}) / (CPI_{it})$$

where CPI_j is the consumer price index of the origin country, and CPI_i is the consumer price index of the destination country.

The ARDL model is expanded to include Google Trends data, and therefore, two ARDL-MIDAS models are estimated for each country—one with the monthly web search Google Trends index (ARDL_M_W) and another with both the web and news search Google Trends indices (ARDL_M_WN) [20,28,37]:

$$TA_t = \alpha + \beta TA_{t-1} + \sum_{j=0}^q \alpha_2 Y_{t-j} + \sum_{j=0}^q \alpha_3 P_{t-j} + \gamma \sum_{k=1}^m \omega(k, \theta) L_{HF}^k GT_t + \varepsilon_t$$

In the model, TA_t represents tourist arrivals, the low frequency dependent variable (quarterly basis), TA_{t-1} represents the lag of tourist arrivals, Y_t is income of the origin country and P_t is relative prices. GT_t represents the high-frequency independent variable/s on a monthly basis. L_{HF}^k denotes the lag operator for monthly lags k of GT_t , and m represents the maximum lag order.

Given the focus on how the inclusion of Google searches improves the predictive power during a recovery period, the forecasts were evaluated pre- and post-COVID. Rolling forecasts were used due to the structural break in the data. A rolling window of 30 observations was used. The pre-COVID forecast evaluation was from 2014Q4 to 2019Q4, and the post-COVID forecast evaluation was from 2022Q1 to 2023Q4, due to most travel restrictions having been lifted in late 2021 or early 2022. The MAPE and RMSE methods were used to evaluate the forecasts on $h = 1, 2, 4, 8$ and 12 , with four quarters being one year.

3. Results

The forecast evaluations between the benchmark models and the models including Google Trends indices (i.e., MIDAS_W, MIDAS_WN, ARDL_M_W, ARDL_M_WN) are shown in Tables 4 and 5 below.

For the pre-COVID forecast evaluations in Table 4, the results are as follows:

The addition of Google trends as an additional explanatory variable outperformed on three forecast horizons for the US source market ($h = 2, 8, 12$). For forecasts over a longer forecasting horizon ($h = 8, 12$), the addition of the news search query as an additional explanatory variable improved the forecasts. This may be an indication that prior to COVID, news from South Africa influenced longer-term travel decisions (i.e., 8- to 12-quarters ahead) through changes in perceptions of the country.

For the UK, the MIDAS_ARDL models outperformed on $h = 1, 2$ and 4 , and for $h = 1, 2$, the addition of the news search improved the forecast. Because the UK is traditionally South Africa's most important international source market, Google searches may be more related to logistical travel concerns (see Figure 1) than perceptions about the country. A similar trend is visible for tourists from the Netherlands, which is also traditionally an important source market for South Africa and was a previous coloniser. The MIDAS_ARDL models outperformed on all five forecast horizons, with the addition of the news search variable improving $h = 1, 2$.

For the German source market, the MIDAS_ARDL with the additional news search variable outperformed on the shorter forecast horizons, $h = 1, 2$, which is more in line with other research that shows the significance of Google Trends data in shorter-term forecasts. For the Indian source market, the MIDAS with the web search variable outperformed on $h = 8$, which is more aligned with the results from the US source market and may be an indication that Google was used in the decision whether or not to travel to South Africa pre-COVID.

Table 4. Forecast evaluation pre-COVID.

	US		UK		NED		GER		FR		AUS		IND	
Model (h)	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
ARIMA (1)	9340.14	7.59	34,115.51	34.38	12,110.48	35.69	36,647.78	40.43	9507.99	19.60	2741.76	7.59	5725.67	20.87
MIDAS_W (1)	10,766.81	9.14	24,671.46	26.06	12,188.95	33.10	32,639.61	35.37	11,797.02	22.52	3624.73	10.92	5920.53	25.04
MIDAS_WN (1)	12,570.69	10.96	23,483.75	23.08	12,578.82	32.57	34,803.48	38.16	15,109.35	29.65	3842.60	11.93	5736.33	22.98
NAïVE (1)	11,312.74	11.72	45,592.31	44.69	14,677.95	42.07	45,848.42	59.32	12,342.69	28.88	4230.02	13.77	7741.25	31.09
ARDL (1)	9676.32	9.12	19,307.34	18.92	8109.22	19.00	18,601.41	21.30	5891.18	10.13	3757.25	11.78	5078.68	20.10
ARDL_M_W (1)	9339.28	7.64	16,136.01	16.16	7585.71	19.30	53,268.64	52.03	8634.47	16.25	3737.74	11.08	7047.20	27.69
ARDL_M_WN (1)	10,015.28	7.93	15,435.09	16.15	7709.16	16.71	17,344.43	13.56	22,164.22	28.09	3812.20	11.08	6036.20	21.98
ARIMA (2)	13,666.86	11.42	29,986.97	33.32	11,076.54	31.73	35,086.80	35.55	13,481.69	23.58	3211.15	9.53	5555.57	20.04
MIDAS_W (2)	12,900.84	11.23	23,974.93	25.88	11,484.43	30.81	30,017.28	31.95	13,326.65	24.73	3982.69	11.92	5645.99	23.86
MIDAS_WN (2)	14,087.51	12.68	22,081.21	23.01	12,545.68	32.63	32,485.73	34.13	17,467.34	35.62	3981.28	12.23	5552.10	21.50
NAïVE (2)	15,332.95	15.37	56,297.86	64.21	17,667.42	50.52	63,392.45	93.90	17,701.98	44.00	5905.77	18.25	6767.45	26.49
ARDL (2)	17,798.92	16.40	16,086.84	16.80	8370.16	19.20	17,839.68	21.59	9279.06	17.43	4972.26	15.62	7165.35	28.58
ARDL_M_W (2)	16,626.00	13.55	15,385.83	16.06	7422.50	19.16	53,241.51	51.73	13,045.63	25.77	4692.26	14.43	8064.61	32.55
ARDL_M_WN (2)	15,815.72	15.12	15,063.49	15.82	7361.22	18.42	15,956.93	14.66	71,419.60	77.89	4653.94	13.86	8174.69	30.60
ARIMA (4)	15,041.06	13.22	31,730.99	35.79	11,271.36	31.74	36,487.34	36.62	13,912.48	24.49	3187.85	9.13	5364.08	19.89
MIDAS_W (4)	14,072.53	12.54	22,350.26	23.86	11,498.17	29.50	30,526.86	32.19	13,545.01	24.30	3994.05	12.00	4948.78	20.97
MIDAS_WN (4)	14,656.01	12.44	20,857.73	21.32	12,374.06	30.26	34,779.25	36.60	19,197.83	37.88	4077.77	12.67	5106.68	18.99
NAïVE (4)	8654.72	8.26	14,617.32	15.96	6322.82	16.35	14,178.86	15.25	7528.58	14.18	3092.59	9.35	7111.56	28.08
ARDL (4)	23,459.99	21.95	16,149.71	17.43	8590.86	19.68	14,896.67	16.72	8759.78	17.50	4200.52	13.15	17,574.33	64.98
ARDL_M_W (4)	15,986.33	14.87	14,554.14	15.33	6162.75	16.35	52,555.51	52.22	11,153.65	21.93	4231.38	13.18	22,020.73	64.61
ARDL_M_WN (4)	16,091.54	15.94	14,535.31	16.01	7709.26	20.90	16,781.78	16.58	78,005.33	89.63	3880.18	11.86	16,796.86	53.97
ARIMA (8)	16,944.37	15.53	31,409.02	35.99	11,352.43	31.76	37,427.31	36.33	14,684.05	25.65	3343.95	9.50	5411.06	21.03
MIDAS_W (8)	15,112.25	13.20	24,396.88	26.68	11,995.02	30.46	28,868.17	31.35	14,274.77	24.73	3917.88	11.52	4833.94	19.50
MIDAS_WN (8)	13,391.50	10.63	21,959.86	21.39	13,155.43	31.97	31,306.56	31.53	22,547.10	43.17	3954.93	12.00	5345.51	22.11
NAïVE (8)	15,111.98	15.02	18,064.61	17.32	6974.64	16.14	17,305.35	18.74	11,841.61	21.16	5904.45	16.99	10,265.24	41.23

Table 4. Cont.

	US		UK		NED		GER		FR		AUS		IND	
Model (h)	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
ARDL (8)	43,885.71	42.17	23,986.18	22.03	8396.99	19.11	21,572.60	24.62	11,368.28	23.34	5191.58	15.65	215,099.40	449.20
ARDL_M_W (8)	40,544.77	38.24	21,176.89	22.66	5582.41	14.15	52,950.23	51.36	14,080.18	26.53	5071.77	15.89	500,417.80	694.51
ARDL_M_WN (8)	29,829.86	29.60	80,196.38	41.48	11,433.78	31.00	26,149.52	28.32	49,880.15	71.85	4690.12	15.13	140,245.80	297.99
ARIMA (12)	18,675.02	17.81	31,834.41	37.40	11,526.17	31.99	38,082.59	36.03	15,714.93	27.33	3645.85	10.09	4548.83	16.55
MIDAS_W (12)	16,041.41	14.24	25,010.19	27.77	12,295.17	31.63	30,165.89	32.40	15,616.38	27.07	3964.81	12.42	5531.31	23.40
MIDAS_WN (12)	13,208.65	11.35	23,090.43	21.47	13,363.16	36.34	33,954.87	32.96	349,789.80	354.50	3909.07	11.93	5127.93	22.78
NAïVE (12)	17,332.90	17.63	20,790.71	21.15	8653.09	19.32	17,842.00	19.45	10,559.46	19.32	7311.32	21.62	13,147.40	48.62
ARDL (12)	51,931.92	50.10	29,942.77	28.08	11,499.01	22.48	34,454.61	29.65	11,704.10	22.03	5579.41	17.81	11,866,709.	14,434.61
ARDL_M_W (12)	73,970.88	56.49	26,409.89	27.95	5518.42	13.76	53,219.89	52.96	12,182.12	23.13	5510.79	16.93	122,000,000	86,532.82
ARDL_M_WN (12)	53,446.39	51.02	1,063,032.00	237.46	20,931.14	63.09	41,463.31	39.27	34,217.98	62.11	13,027.68	27.84	2,936,087.0	3535.15

The MAPE in bold indicate the best performing model in each forecast horizon.

Table 5. Forecast evaluation post-COVID.

	US		UK		NED		GER		FR		AUS		IND	
Model (h)	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
ARIMA (1)	18,538.82	20.62	25,257.54	30.48	9285.18	27.93	16,942.56	31.81	4066.95	14.39	5135.67	23.67	4390.72	17.07
MIDAS_W (1)	11,535.91	11.64	28,700.64	24.16	8642.96	28.27	30,763.67	36.79	4831.81	14.03	4552.72	25.02	4633.43	23.60
MIDAS_WN (1)	17,347.02	23.59	27,083.50	27.27	6168.09	21.24	28,086.56	39.69	5778.08	20.66	4664.22	29.51	4772.80	23.70
NAïVE (1)	14,248.33	14.35	32,447.43	36.04	9714.54	29.97	33,201.44	57.71	4662.24	17.60	3431.84	22.84	4976.74	20.20
ARDL (1)	17,286.30	22.97	37,650.28	45.98	11,015.34	37.70	20,297.54	32.43	7086.91	30.94	7741.15	48.58	5755.94	30.99
ARDL_M_W (1)	19,602.71	25.18	35,133.23	43.31	11,373.61	39.75	24,889.00	39.66	5054.96	19.98	6544.16	34.27	10,351.96	63.02
ARDL_M_WN (1)	17,842.67	24.93	21,843.20	22.67	13,344.15	44.39	12,531.25	25.91	4277.10	17.32	6823.18	35.53	8206.87	48.55
ARIMA (2)	21,276.87	24.13	35,781.85	34.95	11,132.86	35.89	30,100.94	55.71	6874.01	23.81	5457.94	35.09	5005.01	24.20
MIDAS_W (2)	14,122.50	16.65	29,197.14	28.83	10,963.10	37.15	36,803.52	50.74	6118.87	24.14	6618.71	40.48	2923.34	13.94
MIDAS_WN (2)	15,048.74	23.26	33,101.99	35.55	7568.91	26.69	35,778.43	54.13	8633.62	31.22	6845.54	48.24	4329.37	19.71
NAïVE (2)	22,079.80	22.05	45,426.83	54.88	11,859.61	38.30	46,051.58	84.48	7383.02	32.26	6046.95	40.21	5890.47	28.37
ARDL (2)	21,822.64	30.50	79,331.40	85.36	12,857.40	44.59	23,287.12	42.19	10252.61	46.90	14,178.16	76.19	31,868.72	176.04
ARDL_M_W (2)	27,307.05	38.57	31,795.89	42.97	11,521.74	42.98	24,931.05	44.61	7875.97	37.74	16,884.15	85.78	21,875.07	125.68
ARDL_M_WN (2)	22,050.08	34.20	20,849.30	24.11	12,982.47	42.18	18,945.96	34.52	8515.58	38.99	16,094.28	80.42	12,472.70	66.53
ARIMA (4)	28,522.66	38.59	32,692.53	34.25	12,577.83	41.85	23,028.54	32.91	8626.11	35.66	9649.47	56.58	7186.19	39.05
MIDAS_W (4)	17,721.52	25.67	30,664.56	30.91	9429.09	28.50	38,509.38	57.66	7736.59	32.08	10,102.54	63.14	3770.90	23.31
MIDAS_WN (4)	14,886.65	22.85	37,064.55	43.60	7793.72	23.54	35,582.21	56.56	12,201.69	55.13	10,061.26	69.49	5460.50	33.29
NAïVE (4)	34,732.12	45.48	44,552.87	50.46	13,109.98	46.77	26,443.24	47.36	11,517.27	48.62	11,196.49	64.89	7846.23	45.23
ARDL (4)	27,717.13	30.90	125,580.70	144.46	12,669.79	44.80	21,675.13	42.56	10,836.47	49.45	15,308.15	81.38	18,505.67	92.58
ARDL_M_W (4)	36,063.88	48.96	41,712.73	55.36	14,793.45	44.61	31,793.02	55.43	11,457.86	55.46	59,729.03	218.19	44,004.82	181.03
ARDL_M_WN (4)	34,395.43	47.23	37,826.01	48.98	14,522.65	45.69	15,076.94	25.17	12,280.52	58.85	37,890.29	119.88	13,672.11	69.91
ARIMA (8)	45,673.55	58.75	37,879.24	44.11	16,081.81	55.97	30,751.06	67.20	17,186.57	79.05	18,088.86	124.01	10,740.38	65.39
MIDAS_W (8)	30,622.84	39.05	34,423.48	36.15	13,233.33	47.67	36,236.18	50.96	19,876.99	92.93	18,391.15	126.96	8409.17	53.83
MIDAS_WN (8)	38,869.97	51.58	36,189.51	42.68	9624.35	42.59	23,400.78	40.20	14,268.06	56.27	14,889.71	106.04	11,123.09	74.99
NAïVE (8)	62,978.51	79.88	71,861.94	85.61	24,560.51	85.95	51,642.60	90.16	22,768.90	97.60	17,855.14	106.35	14,375.42	81.64
ARDL (8)	56,106.12	74.97	976,554.90	750.16	14,403.45	49.52	30,499.09	44.98	82,333.98	296.21	23,248.57	177.85	10,194.18	46.83
ARDL_M_W (8)	64,799.95	88.28	79,718.41	97.37	14,572.83	49.03	30,205.95	50.41	22,243.91	82.70	14,896.69	103.56	15,079.56	66.73

Table 5. Cont.

	US		UK		NED		GER		FR		AUS		IND	
Model (h)	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
ARDL_M_WN (8)	61,708.79	87.75	75,923.78	88.08	25,822.05	69.04	30,624.56	43.49	21,880.37	85.01	10,673.86	86.08	12,037.69	72.40
ARIMA (12)	33,607.25	36.59	31,500.39	41.25	9756.03	39.84	30,277.04	74.71	14,704.83	68.84	13,764.52	98.84	8175.77	52.15
MIDAS_W (12)	17,458.13	22.60	45,837.80	52.36	9425.45	37.73	18,291.95	23.74	17,184.68	79.76	16,302.95	123.72	8762.34	57.27
MIDAS_WN (12)	18,240.52	26.93	44,778.36	57.18	7768.55	33.53	12,134.32	23.99	6258.72	31.65	16,191.22	134.78	10,780.02	75.08
NAÏVE (12)	59,634.82	72.17	57,772.93	65.55	21,855.22	65.11	46,476.41	77.85	28,116.80	122.90	16,719.66	107.60	14,301.56	74.07
ARDL (12)	55,295.05	67.53	139,857.70	169.99	11,441.40	40.85	58,948.40	101.32	60,801.60	203.31	14,938.24	99.69	6796.39	38.64
ARDL_M_W (12)	56,460.36	70.45	917,644.60	382.17	10,569.65	41.57	28,924.18	37.66	19,873.47	73.92	9811.49	76.42	8343.48	49.32
ARDL_M_WN (12)	58,416.17	78.46	238,076.60	222.72	11,482.70	38.81	32,620.85	68.24	19,714.61	79.65	11,137.16	94.44	60,629.92	341.51

The MAPE in bold indicate the best performing model in each forecast horizon.

Contrary to the other source markets, the benchmark models outperformed the MIDAS models on all forecast horizons for the French and Australian source markets. This shows that prior to COVID, the underlying trends in the data were sufficient in forecasting future tourism from these two source markets.

Table 5 shows the forecast evaluations for the post-COVID period after the structural break due to the COVID pandemic. The results show the following:

For the US source market, the MIDAS web search models outperformed on all forecast horizons, and the addition of the news search improved $h = 4$. Post-COVID, web search data therefore became a more important source of information on which this market based their travel decisions in both the short- and long-term.

Similar to the results for the US, for the UK source market, the MIDAS models outperformed on all forecast horizons except for $h = 12$. The MIDAS_ARDL with the web and news search variables outperformed on $h = 1, 2$, and for the longer horizons ($h = 4, 8$), the MIDAS web search models outperformed. Compared to pre-COVID, web search data became important also in longer-term travel forecasts, which may be an indication of the increased uncertainty created by the COVID pandemic.

The MIDAS web and news model outperformed on all forecast horizons for the Netherlands source market, which is similar to the pre-COVID results for this market. Although, post-COVID, the inclusion of economic covariates did not improve the forecasting performance for tourist arrivals from the Netherlands.

For Germany, the MIDAS models outperformed on all forecast horizons; MIDAS_ARDL with web and news search outperformed on the shorter horizons ($h = 1, 2, 4$) and the MIDAS_WN for $h = 8$ and MIDAS_W for $h = 12$. This result may again indicate that, similar to the UK, web search data also became more important in the longer term due to the increased uncertainty caused by the pandemic. However, for India, MIDAS with web search outperformed on $h = 2, 4$, showing an improvement in short-term forecasts rather than the pre-COVID longer-term forecasts ($h = 8$).

In contrast to the pre-COVID forecast evaluation, the addition of Google Trends data did improve the forecasts for the French and Australian source markets. For the French source market, the MIDAS models outperformed on all forecast horizons except $h = 2$, and the addition of the news search variable improved the longer forecast horizons $h = 8, 12$. For Australia, the MIDAS_ARDL outperformed on $h = 8, 12$; the addition of the news search variable improved $h = 8$. This again highlights the increased importance of web search data for information about South Africa after the pandemic.

Figures 3 and 4 compare the forecasting performance of the various models over the different forecasting horizons. Figure 3 shows that pre-COVID, with $h = 1, 2$, the MIDAS_ARDL with the Google Trends web and news search indices was the best model in three of the seven instances, with the ARIMA model and ARDL model in second and third places. Over the longer term, the naïve model was difficult to outperform, and only the MIDAS_ARDL model that included the web search Google index appeared as the best model in two instances over $h = 4$ and once when $h = 8, 12$. Figures 3 and 4 compare the forecasting performance of the various models over the different forecasting horizons. Figure 3 shows that pre-COVID, with $h = 1, 2$, the MIDAS_ARDL with the Google Trends web and news search indices was the best model in three of the seven instances, with the ARIMA model and ARDL model in second and third places. Over the longer term, the naïve model was difficult to outperform, and only the MIDAS_ARDL model that included the web search Google index appeared as the best model in two instances over $h = 4$ and once when $h = 8, 12$.

Figure 4 shows the same figure but after COVID-19, and it is evident that the models that include Google search data outperformed the other models consistently. For example, over $h = 1, 2, 8, 12$, five out of the seven winning forecasts included Google Trends data. For $h = 4$, this improved to six out of the seven forecasts.

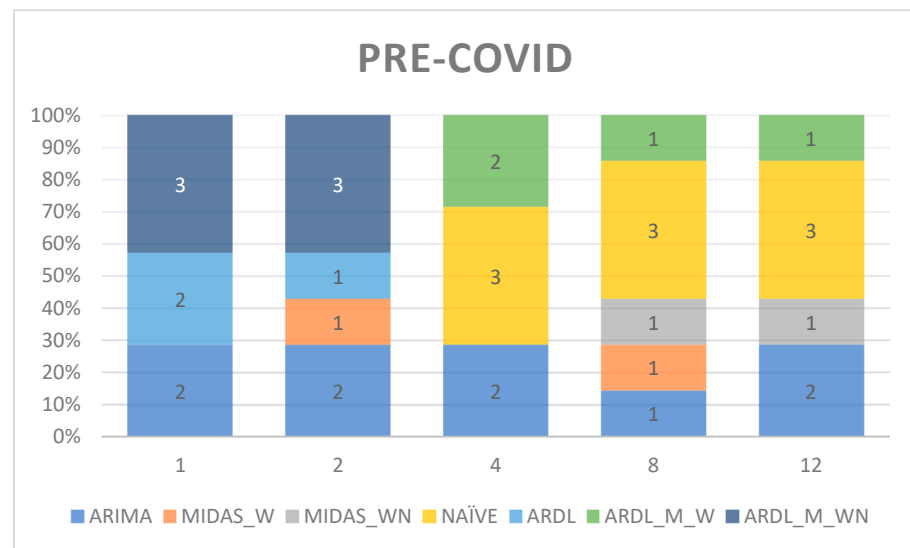


Figure 3. Best forecasting models over different forecasting horizons pre-COVID.

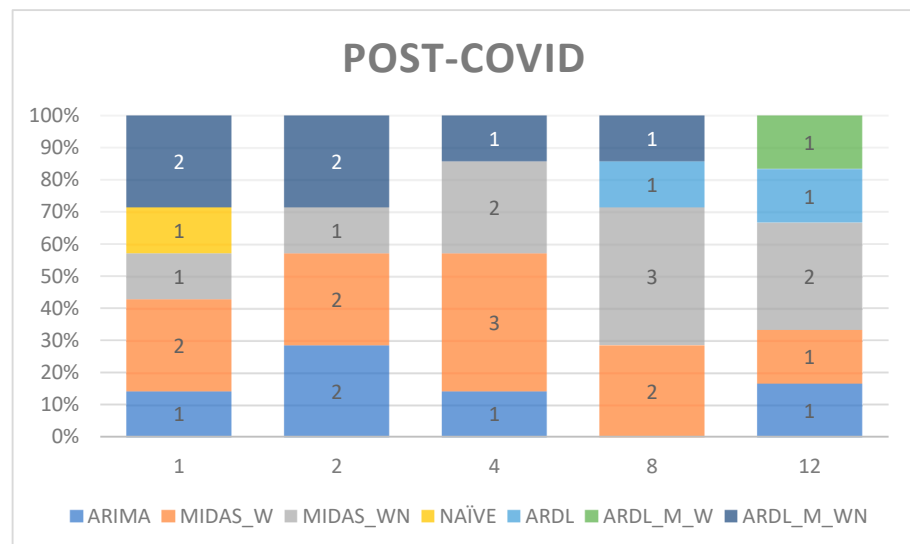


Figure 4. Best forecasting models over different forecasting horizons, post-COVID.

When one compares the models that included economic variables (ARDL-based models) with the pure time-series models pre-COVID, the ARDL models as a group fared better in forecasting tourism to South Africa in the shorter term, i.e., one- and two-quarters. Over one to three years, the time series models outperformed. Post-COVID, this trend continued, but instead of the naïve model faring well for the time-series methods, the MIDAS models that included web search data were the clear winners.

4. Discussion and Conclusions

This research set out to answer three questions, and the results are consequently considered within the context of the questions.

Firstly, does Google Trends have the potential to improve tourism-demand forecasting during recovery after a shock? Overall, the results show that for the pre-COVID period, the MIDAS models that included Google Trends data outperformed the benchmarks in 40% (14/35) of the pre-COVID forecast models, and post-COVID, this percentage increased to 77% (27/35). This shows that adding Google Trends data as an explanatory variable does improve forecast during recovery. These results are in line with the findings of previous research that show that search engine data improves tourism-demand forecasts.

Specifically, it confirms research by [35] that web search data improves tourism recovery forecasts, contradicting results by [36].

Secondly, does the inclusion of typical tourism-demand covariates in an econometric forecasting specification together with Google Trends data lead to improved forecasts? To test this, MIDAS-ARDL models were estimated, and the results show that in the majority of forecasts over all time horizons pre- and post-COVID, the inclusion of Google Trends data improved the forecasting accuracy of the ARDL model.

Compared to time-series forecasts, the ARDL-based models forecast better over the shorter term, with one and two quarters ahead forecasting results, both pre- and post-COVID, and in more instances than not being the best forecasts. It is noteworthy that the ARDL-based models were never the best forecasting model for the Indian and Australian source markets pre-COVID. Post-COVID, this situation changed for these source markets, and these models delivered the most accurate forecasts, especially over the longer term. In contrast, the MIDAS-ARDL models were always the best in forecasting tourist arrivals from the Netherlands in normal (non-recovery) times.

Finally, can Google Trends improve forecasting over a longer time period? Most research on Google Trends and tourism demand is done on monthly or higher-frequency data [24]. This research considers quarterly forecasts and forecasts for up to three years (12 quarters). The rationale for this can be found in the reality that South Africa is not a well-known destination, and the country is a long-haul destination where more lead time may be required in decision-making.

This research shows that monthly Google Trends also improves forecasts on lower-frequency data (quarterly) using a mixed data-sampling framework. The results show that, pre-COVID, forecasts that included Google Trends were more effective in forecasting one to two quarters ahead than further into the future. This trend changed after COVID, where Google Trends led to improved recovery forecasts even over the longer term (4–12 quarters ahead). Additionally, this research incorporates web and/or news search indices as suggested by [29]. The addition of the news search variable improved the forecasts with the web search data by 43% (6/14) pre-COVID and by 59% (16/27) post-COVID.

In conclusion, this paper makes a significant contribution to current research by showing the significance of Google Trends data in (i) improving tourism-forecasting recovery after a shock to the tourism industry, (ii) forecasting tourism using low-frequency data and (iii) forecasting over a longer time horizon. It is also the first time that Google Trends data have been used in forecasting tourism to an African destination, which opens up future research avenues to explore in terms of tourism forecasting in African destinations.

Our research is, however, not without limitations. This research focuses only on one destination, namely, South Africa, and only models intercontinental arrivals, even though neighbouring countries are South Africa's main tourism source markets. This is due to the limited availability of Google Trends and also, in some instances, economic data from these countries. In fact, even Google Trends search data are quite limited, limiting what could be done with them, which indicates that South Africa is a lesser-known destination that attracts fewer international internet searches. However, even with these limitations, our research shows that the inclusion of Big Data such as Google Trends can enhance tourism forecasting. Therefore, even for lesser-known destinations, it makes a difference. Future research could therefore also expand the current research by including other Big Data sources that would allow the use of, among others, sentiment analysis. Further research is needed to explain why post-COVID Google forecasts work better than pre-COVID. Additionally, it would be interesting to also verify whether forecasts that include Google indices improve long-term forecast horizons rather than short-term forecast horizons, as well as for other long-haul destinations.

Author Contributions: Conceptualization, A.S. and I.B.; methodology, I.B.; validation, A.S., formal analysis, I.B.; investigation, A.S.; writing—original draft preparation, A.S.; writing—review and editing, A.S. and I.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Secondary data were used, and the data are publicly available. The data were sourced from Statistics South Africa (arrivals data, <https://www.statssa.gov.za/> accessed on 30 March 2024) and the OECD (<https://www.oecd.org/en/data.html> accessed on 30 March 2024) and IFS databases (<https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b> accessed on 30 March 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. *Tour. Manag.* **2018**, *68*, 301–323. [\[CrossRef\]](#)
- Yang, X.; Pan, B.; Evans, J.A.; Lv, B. Forecasting Chinese tourist volume with search engine data. *Tour. Manag.* **2015**, *46*, 386–397. [\[CrossRef\]](#)
- Wu, D.C.; Zhong, S.; Qiu, R.T.; Wu, J. Are customer reviews just reviews? Hotel forecasting using sentiment analysis. *Tour. Econ.* **2021**, *28*, 795–816. [\[CrossRef\]](#)
- Xie, G.; Li, X.; Qian, Y.; Wang, S. Forecasting tourism demand with KPCA-based web search indexes. *Tour. Econ.* **2021**, *27*, 721–743. [\[CrossRef\]](#)
- Choi, H.; Varian, H. Predicting the present with Google Trends. *Econ. Rec.* **2012**, *88*, 2–9. [\[CrossRef\]](#)
- Wen, L.; Liu, C.; Song, H. Forecasting tourism demand using search query data: A hybrid modelling approach. *Tour. Econ.* **2019**, *25*, 309–329. [\[CrossRef\]](#)
- Artola, C.; Pinto, F.; de Pedraza García, P. Can internet searches forecast tourism inflows? *Int. J. Manpow.* **2015**, *36*, 103–116. [\[CrossRef\]](#)
- Camacho, M.; Pacce, M.J. Forecasting travellers in Spain with google search volume indices. *Tour. Econ.* **2018**, *24*, 434–448. [\[CrossRef\]](#)
- Park, S.; Lee, J.; Song, W. Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data. *J. Travel Tour. Mark.* **2016**, *34*, 357–368. [\[CrossRef\]](#)
- Lv, S.-X.; Peng, L.; Wang, L. Stacked autoencoder with echo-state regression for tourism demand forecasting using search query data. *Appl. Soft Comput.* **2018**, *73*, 119–133. [\[CrossRef\]](#)
- Önder, I.; Gunter, U.; Scharl, A. Forecasting Tourist Arrivals with the Help of Web Sentiment: A Mixed-frequency Modeling Approach for Big Data. *Tour. Anal.* **2019**, *24*, 437–452. [\[CrossRef\]](#)
- Bokelmann, B.; Lessmann, S. Spurious patterns in Google Trends data—An analysis of the effects on tourism demand forecasting in Germany. *Tour. Manag.* **2019**, *75*, 1–12. [\[CrossRef\]](#)
- Wang, L.; Wu, B.; Zhu, Q.; Zeng, Y.R. Forecasting Monthly Tourism Demand Using Enhanced Backpropagation Neural Network. *Neural Process. Lett.* **2020**, *52*, 2607–2636. [\[CrossRef\]](#)
- Gawlik, E.; Kabaria, H.; Kaur, S. Predicting tourism trends with Google Insights. *Accessed Dec.* **2011**, *1*, 2012.
- Bai, H.; Hao, H. A novel two-step procedure for tourism demand forecasting. *Curr. Issues Tour.* **2021**, *24*, 1199–1210.
- Li, X.; Law, R. Forecasting Tourism Demand with Decomposed Search Cycles. *J. Travel Res.* **2019**, *59*, 52–68. [\[CrossRef\]](#)
- Wen, L.; Liu, C.; Song, H.; Liu, H. Forecasting Tourism Demand with an Improved Mixed Data Sampling Model. *J. Travel Res.* **2021**, *60*, 336–353. [\[CrossRef\]](#)
- Hu, M.; Song, H. Data source combination for tourism demand forecasting. *Tour. Econ.* **2019**, *26*, 1248–1265. [\[CrossRef\]](#)
- Huang, K.-H.; Yu, T.H.-K. Application of Google trends to forecast tourism demand. *J. Internet Technol.* **2019**, *20*, 1273–1280.
- Bangwayo-Skeete, P.F.; Skeete, R.W. Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tour. Manag.* **2015**, *46*, 454–464. [\[CrossRef\]](#)
- Li, X.; Pan, B.; Law, R.; Huang, X. Forecasting tourism demand with composite search index. *Tour. Manag.* **2017**, *59*, 57–66. [\[CrossRef\]](#)
- Li, S.; Chen, T.; Wang, L.; Ming, C. Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. *Tour. Manag.* **2018**, *68*, 116–126. [\[CrossRef\]](#)
- Sun, S.; Wei, Y.; Tsui, K.L.; Wang, S. Forecasting tourist arrivals with machine learning and internet search index. *Tour. Econ.* **2019**, *70*, 1–10. [\[CrossRef\]](#)
- Li, X.; Law, R.; Xie, G.; Wang, S. Review of tourism forecasting research with internet data. *Tour. Manag.* **2021**, *83*, 104245. [\[CrossRef\]](#)
- Sun, S.; Li, Y.; Guo, J.E.; Wang, S. Tourism demand forecasting: An ensemble deep learning approach. *Tour. Econ.* **2022**, *28*, 2021–2049. [\[CrossRef\]](#)
- Hu, Y.-C.; Wu, G. The impact of Google Trends index and encompassing tests on forecast combinations in tourism. *Tour. Rev.* **2022**, *77*, 1276–1298. [\[CrossRef\]](#)
- Havranek, T.; Zeynalov, A. Forecasting tourist arrivals: Google Trends meets mixed-frequency data. *Tour. Econ.* **2021**, *27*, 129–148. [\[CrossRef\]](#)
- Önder, I.; Gunter, U. Forecasting Tourism Demand with Google Trends For a Major European City Destination. *Tour. Anal.* **2016**, *21*, 203–220. [\[CrossRef\]](#)

29. Önder, I. Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. *Int. J. Tour. Res.* **2017**, *19*, 648–660. [[CrossRef](#)]
30. Liu, Y.-Y.; Tseng, F.-M.; Tseng, Y.-H. Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model. *Technol. Forecast. Soc. Chang.* **2018**, *130*, 123–134. [[CrossRef](#)]
31. Volchek, K.; Liu, A.; Song, H.; Buhalis, D. Forecasting tourist arrivals at attractions: Search engine empowered methodologies. *Tour. Econ.* **2018**, *25*, 425–447. [[CrossRef](#)]
32. Li, X.; Li, H.; Pan, B.; Law, R. Machine Learning in Internet Search Query Selection for Tourism Forecasting. *J. Travel Res.* **2020**, *60*, 1213–1231. [[CrossRef](#)]
33. Huang, X.; Zhang, L.; Ding, Y. The Baidu Index: Uses in predicting tourism flows—A case study of the Forbidden City. *Tour. Manag.* **2017**, *58*, 301–306. [[CrossRef](#)]
34. Peng, L.; Wang, L.; Ai, X.Y.; Zeng, Y.R. Forecasting Tourist Arrivals via Random Forest and Long Short-term Memory. *Cogn. Comput.* **2020**, *13*, 125–138. [[CrossRef](#)]
35. Wu, J.; Li, M.; Zhao, E.; Sun, S.; Wang, S. Can multi-source heterogeneous data improve the forecasting performance of tourist arrivals amid COVID-19? Mixed-data sampling approach. *Tour. Manag.* **2023**, *98*, 104759. [[CrossRef](#)]
36. Yang, Y.; Fan, Y.; Jiang, L.; Liu, X. Search query and tourism forecasting during the pandemic: When and where can digital footprints be helpful as predictors? *Ann. Tour. Res.* **2022**, *93*, 103365. [[CrossRef](#)]
37. Li, H.; Hu, M.; Li, G. Forecasting tourism demand with multisource big data. *Ann. Tour. Res.* **2020**, *83*, 102912. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.