Ornamental Horticulture
ISSN 2447-536X | WWW.ORNAMENTALHORTICULTURE.COM.BR

SBFPO Sociedade Brasileira de Floricultura e Plantas Ornamentais

ARTICLE

Google trends as an early signal in international flower trade

Google Trends como sinal antecipado no comércio internacional de flores

Carlos Fernando Osorio Andrade[1,*], Paula Andrea Lopez Herrera[2], and Diego Alonso Garcia Bonilla[3]

[1]Universidad del Valle, Instituto de Educación Técnica Profesional de Roldanillo, Valle - INTEP, Tuluá, Colombia.
[2]Universidad del Valle, Buga, Colombia.
[3]Universidad del Valle, Cali, Colombia.

Abstract
This study investigates the predictive potential of Google Trends for U.S. flower imports, examining whether fluctuations in online search volumes can anticipate changes in import patterns. Monthly flower import data from 2004 to 2023 were analyzed alongside four Google search indices using a comparative evaluation of time series models, including ARIMA, ARIMAX, and SARIMAX. The analysis reveals clear predictive patterns, emphasizing the importance of incorporating seasonality into forecasting. Results indicate that the SARIMAX model, which explicitly accounts for seasonal components, consistently outperforms simpler models, achieving a mean absolute percentage error (MAPE) of 7.04%. Among the four search indices examined, only the distribution index—comprising terms related to florists and delivery services—demonstrated statistically significant predictive power. This finding suggests that consumer interest in flower accessibility and delivery strongly influences import dynamics. The study contributes original evidence supporting the use of Google Trends as an early indicator within the international flower trade. By integrating online search behavior into forecasting models, stakeholders can gain timely insights into market demand, potentially improving supply chain management, inventory planning, and operational responsiveness. Overall, this research highlights the value of digital search data as a practical, cost-effective tool for anticipating trends in global floral markets and informing strategic decisions.
Keywords: floriculture, imports, online data, predictive power.

Resumo
Este estudo investiga o potencial preditivo do Google Trends para as importações de flores dos EUA, examinando se as flutuações nos volumes de busca online podem antecipar mudanças nos padrões de importação. Dados mensais de importação de flores de 2004 a 2023 foram analisados em conjunto com quatro índices de busca do Google por meio de uma avaliação comparativa de modelos de séries temporais, incluindo ARIMA, ARIMAX e SARIMAX. A análise revela padrões preditivos claros, enfatizando a importância de incorporar a sazonalidade nas previsões.Os resultados indicam que o modelo SARIMAX, que considera explicitamente componentes sazonais, supera consistentemente modelos mais simples, alcançando um erro percentual absoluto médio (MAPE) de 7,04%. Entre os quatro índices de busca examinados, apenas o índice de distribuição — composto por termos relacionados a floriculturas e serviços de entrega — demonstrou poder preditivo estatisticamente significativo. Essa descoberta sugere que o interesse do consumidor na acessibilidade e entrega de flores influencia fortemente a dinâmica das importações.O estudo contribui com evidências originais que apoiam o uso do Google Trends como indicador antecipado no comércio internacional de flores. Ao integrar o comportamento de busca online aos modelos de previsão, os envolvidos podem obter insights oportunos sobre a demanda de mercado, potencialmente melhorando a gestão da cadeia de suprimentos, o planejamento de estoque e a capacidade de resposta operacional.No geral, esta pesquisa destaca o valor dos dados de busca digital como uma ferramenta prática e econômica para antecipar tendências nos mercados globais de flores e embasar decisões estratégicas.
Palavras-chave: floricultura, dados online, importações, poder preditivo.

Introduction

In the contemporary digital economy, the proliferation of big data has radically transformed market analysis methodologies. Consumer behavior metrics in electronic environments are crucial for decision-making and economic research, enabling a proactive strategy that anticipates future trends (De Luca & Rosciano, 2023). Within this paradigm, Google Trends (GT) emerges as a highly relevant information source, as it allows near real-time tracking of user search interest, reflecting the relative popularity of specific terms over time. This ability to capture the pulse of public interest has driven its adoption in academic and business contexts for diverse applications, from sales forecasting to consumer trend analysis (Vosen & Schmidt, cited in Archibugi, 2024).

One application with notable yet under-explored potential lies in using GT data to analyze imports in international trade. It is posited that search patterns, by indicating growing interest in certain products, can serve as early signals of future demand (Yang et al., 2022). These signals, captured before their materialization in formal transactions, offer the possibility of anticipating fluctuations in import volumes, providing a predictive advantage over traditional foreign trade records, which are often published with significant time lags (Bantis et al., 2022). For the floriculture sector - characterized by the high perishability of its products and the need for agile logistics - such anticipation is particularly valuable. It could optimize decisions regarding planting, harvesting, inventory management, and cold chain logistics (Rombach et al., 2021).

Incorporating these metrics into predictive models can enhance the ability to project trade flows, especially in sectors sensitive to consumer preferences. While studies have applied GT to forecast demand in areas such as electric vehicles and tourism (Castellacci and Santoalha, 2025), the flower industry presents a unique opportunity. This sector, marked by pronounced seasonality and ties to festivities, faces the challenge of aligning delicate production with fluctuating and geographically dispersed demand (Sun, 2024). Predictions informed by online interest could lead to substantial improvements in planning, avoiding overproduction or shortages and refining the complex logistics of exportation (Chelliah et al., 2024; Ye & Yue, 2023).

Although specific literature is limited, studies in analogous sectors support the predictive power of these data. Previous research has shown that GT improves forecasts for seafood imports (Chaudhady and Choi, 2023) and automobile sales (Choi and Varian, 2011), and optimizes supply chain operations for agricultural products (Tamasiga et al., 2023). The underlying mechanism suggests that market actors monitor search interest; an increase is interpreted as an indicator of future demand, prompting proactive adjustments to import orders, which are reflected weeks or months later in official statistics (Boone et al., 2018).

Corresponding autor: carlos,fernando.osorio@correounivalle.edu.co | https://doi.org/10.1590/2447-536X.v31.e312959 | Editor: Paulo Rodrigo Ramos Xavier Pereira, Universidade Federal do Piauí, Brasil | Received: May 29, 2025 | Accepted: Oct 07, 2025 | Available online: Nov 17, 2025.

Given the above, this study aims to examine specifically: To what extent can variations in Google Trends search volumes related to flowers predict subsequent changes in floral import volumes? What are the theoretical and empirical mechanisms explaining this relationship? What strategic implications does this dynamic hold for actors in the value chain? To address these questions, the article is structured as follows: After this introduction, the second section reviews the relevant literature. The third section details the methodology employed, with an emphasis on time series modeling through the comparison of ARIMA, ARIMAX, and SARIMAX models. The fourth section presents the results, while the fifth is dedicated to discussing the findings. Finally, the sixth section outlines the general conclusions of the study.

Theoretical Framework

Market Trends in the Flower Industry in the United States

The floral industry in the United States is a consolidated sector that has shown sustained growth in recent years (Niño and Janzen, 2025). According to data from Grand View Research (2025), total consumer spending on products such as flowers, seeds, and potted plants reached the figure of $69 billion in 2024. The specific segment corresponding to floral gifts was valued at $12.18 billion, with projections indicating it could exceed $16 billion by 2030, translating to a compound annual growth rate (CAGR) of over 5% (Arizton Advisory and Intelligence, 2025). Various studies suggest that this market dynamism is driven by the emergence of new trends in consumer purchasing decision patterns. Among these factors, the following can be mentioned: a greater inclination toward self-care practices, increasing interest in incorporating natural elements into home decor, and a marked preference for purchasing personalized gifts with emotional value (Spence et al., 2023).

A majority proportion of cut flowers marketed in the United States originates from international imports, with Latin America being the predominant geographic region, accounting for 85% of the total share. Within this context, Colombia and Ecuador stand out as the main trading partners for the United States in this sector. It is important to note that, at the domestic level, the U.S. floral market has also undergone a transformation in its distribution channels. Although traditional physical channels such as retail florists, supermarkets, and garden centers still manage the majority of sales (approximately 61% of the market in 2023), e-commerce is emerging as the fastest-growing distribution channel (Grand View Research, 2025). This digital expansion is driven by the convenience of 24/7 shopping, a wider selection, various delivery options, and the preferences of younger consumers. Additionally, the specific market for online flower delivery services was estimated to be between $7.3 and $7.7 billion in 2024, with an annual growth projection of approximately 6.8% (Credence Research, 2024).

The Role of Demand Forecasting in Production and Logistics Decisions

The relationship between accurate forecasting and efficient supply chain management is a fundamental factor within the horticultural industry. Authors such as Drechsler and Holzapfel et al. (2022) emphasize that, due to the high volatility of demand and the perishable nature of most horticultural products, one of the main challenges in medium-term planning is to forecast demand as accurately as possible. This statement summarizes the dual challenge faced by the floral sector: uncertainty in demand and the short shelf life of the product, which means that forecasting errors - whether overestimating or underestimating - have significant economic consequences (Rombach et al., 2021).

The literature has shown that predictive analysis is key to strategic decision-making in this context. Demand forecasts are not an end in themselves but rather represent the main input for a series of interconnected operational decisions (Ren et al., 2022). First, an accurate forecast of future demand allows producers and exporters to plan planting and harvesting cycles to align supply with expected demand peaks. Authors such as Kantasa-Ard et al. (2020) have found that demand forecasting models can predict the quantity of agricultural products needed in a future period, helping to plan agricultural production with sufficient lead time. This is crucial for products requiring long planting and harvesting cycles, such as corn, cassava, and pineapple. They also conducted a simulation where the use of forecasted demand reduced storage and transportation costs to less than 2% deviation from actual demand. Additionally, it has been found that anticipating demand allows for more efficient inventory management and procurement of inputs. In this regard, Huang et al. (2023) highlight that the use of historical data enables informed scheduling decisions based on forecasts that balance immediate requirements with long-term operational efficiency. On the other hand, the logistics of cut flowers is a race against time, so accurate prediction of remaining shelf life allows for adjustments in the production and distribution of these products (Tromp et al., 2016). In summary, by combining a forecast of when and where the product will be needed with models of its shelf life, companies can optimize the entire cold chain and distribution process. To achieve this precision, various time-series modeling and machine learning techniques have been explored (Sari et al., 2024)

Impact of Google Trends Searches on Imports

The use of online search data, specifically from Google Trends, has become a consolidated source of leading indicators for consumer behavior and, by extension, market dynamics. Empirical research supports this claim. Silva et al. (2019) demonstrated that purchase intentions emerge during the initial exploration phase through searches, revealing an interest that later translates into potential demand. Specific cases, such as the one examined by Dera et al. (2024), illustrate how Google Trends served as an anticipatory tool for demand in heated tobacco products. Similarly, Diaz-Balteiro et al. (2023) documented the correlation between sustained increases in searches and rises in both demand and price for wild edible mushrooms.

While literature addressing the impact of searches on imports is limited, existing precedents are directly relevant. Chaudhady and Choi (2023) provided significant evidence by demonstrating that integrating Google Trends data improves predictions related to seafood imports. In the export sector, Lessoua et al. (2020) used this tool as an indicator to analyze purchasing trends in the Romanian wine industry. This empirical evidence across various sectors suggests a robust relationship that can be explained through economic and consumer behavior principles.

The mechanism by which search trends translate into commercial actions, such as imports, is supported by well-established theoretical foundations. First, the law of demand (Mankiw, 2012) posits that demand changes based on consumer behavior, preferences, and expectations. In the floral sector, an increase in searches for events like Valentine's Day or Mother's Day directly reflects emerging purchase intentions (Wang et al., 2023). This early rise in digital interest foreshadows future aggregate demand, prompting importers to adjust inventories and place international orders to meet it. Second, the bandwagon effect (Pitchayadol, 2018) explains how the visibility of certain trends accelerates collective purchasing decisions. Companies aware of these trends react proactively by increasing stock, directly impacting import volumes. Finally, contemporary consumer behavior reinforces this dynamic. Before purchasing, individuals conduct searches for inspiration, price comparisons, and quality assessments (Thirumalai and Sinha, 2011). While this process reflects individual intent, when aggregated, it serves as an early and reliable indicator of collective market behavior, enabling companies to more accurately predict demand peaks and optimize planning.

Based on the practical need for better forecasts in the floral supply chain, the empirical evidence of the predictive power of Google Trends, and the theoretical foundations of consumer behavior that explain it, the following hypothesis is established to guide the present research: Variations in the volume of searches related to flowers on Google Trends function as leading indicators to predict changes in imports.

Material and methods

Data

In this study, two distinct types of variables were employed. As the dependent variable, monthly imports of fresh cut flowers into the United States were analyzed, with data obtained from the U.S. International Trade Commission database for the period from January 2004 to December 2023. These data were expressed in CIF terms, which refer to the total cost of bringing the merchandise into the importing country, including the value of the goods, international freight cost, and insurance to the point of entry.

As explanatory variables, monthly Google search indices for flower-related words in the United States during the same period were included.

These data were obtained from the Google Trends website, which provides nearly real-time information on internet search behaviors. It is important to note that these data are obtained in the form of a normalized indicator with positive values ranging from 0 to 100, depending on the search volume for each word. The calculation method involves dividing the total number of searches for a keyword in a specific geographic area by the total number of searches for the same area during a particular time period.

To ensure a rigorous selection of keywords, a systematic three-phase process was followed to accurately capture consumer intent and ensure the robustness of the model. First, in the identification phase, an exhaustive list of relevant terms was generated. This process included brainstorming based on the consumer journey (from inspiration to purchase), analyzing industry reports to identify the most popular flower varieties and sales channels, and using Google Trends' "related queries" feature to discover adjacent search terms. Second, a refinement and filtering phase was carried out. Terms with consistently low search volumes were removed, as their low representativeness could introduce statistical noise into the model. Finally, in the third phase, with the dual objective of building a parsimonious model and creating interpretable explanatory variables, the remaining keywords were grouped thematically. This categorization was not arbitrary but sought to reflect different facets of consumer search intent. The words were grouped into four main categories: terms associated with distribution, which included words such as *florist*, *flowershop*, and *flowerdelivery* (reflecting high purchase intent); terms related to flower varieties, such as *roses*, *lilies*, *tulip*, *orchids*, *sunflowers*, and *chrysanthemum* (indicating interest in specific products); a third group consisting of words associated with seasons, for example, *springflowers*, *summerflowers*, and *winterflowers* (capturing general seasonal demand); and finally, terms associated with special occasions such as *weddingbouquets*, *funeralflowers*, and *sympathyflowers* (linked to specific purchasing events). Below, Table 1 presents the descriptive indicators.

Table 1. Descriptive indicators of the variables flower imports and Google Trends searches.

Variable	Mean	SD	Median
Flower imports	114,477,344	54,371,814	98,031,935
Distribution index	19.95139	6.596912	18.0
Flower varieties index	12.72708	3.476852	11.83333
Seasons index	8.178472	2.224337	8.0
Occasions index	8.343056	1.37805	8.166667

As shown in Table 1, the variable "flower imports" has an average of $114,477.344 million CIF, with a standard deviation of $54,371.814 and a median of $98,031.935. The "distribution index" shows a mean of 19.95 with a standard deviation of 6.59, while the "flower varieties index" reaches an average of 12.73 with a slightly lower median (11.83). Regarding the "seasons index," the mean is 8.18 with a standard deviation of 2.22, while its median is exactly 8, indicating a relatively symmetrical distribution. Finally, the "occasions index" has a mean of 8.34 and a median of 8.17, with a low standard deviation of 1.38.

Stationarity Analysis

Prior to applying time series models, it is essential to test the assumption of stationarity. This involves verifying that the statistical properties of the series, such as the mean and variance, remain constant over time. To analyze this assumption, the time series for each variable were first visualized, a procedure that can be seen in Fig. 1. Additionally, the Elliot, Rothenberg, and Stock unit root test (DF-GLS) was applied, which allows confirming or rejecting the hypothesis that the variables are stationary. The results of this test are shown in Table 2.

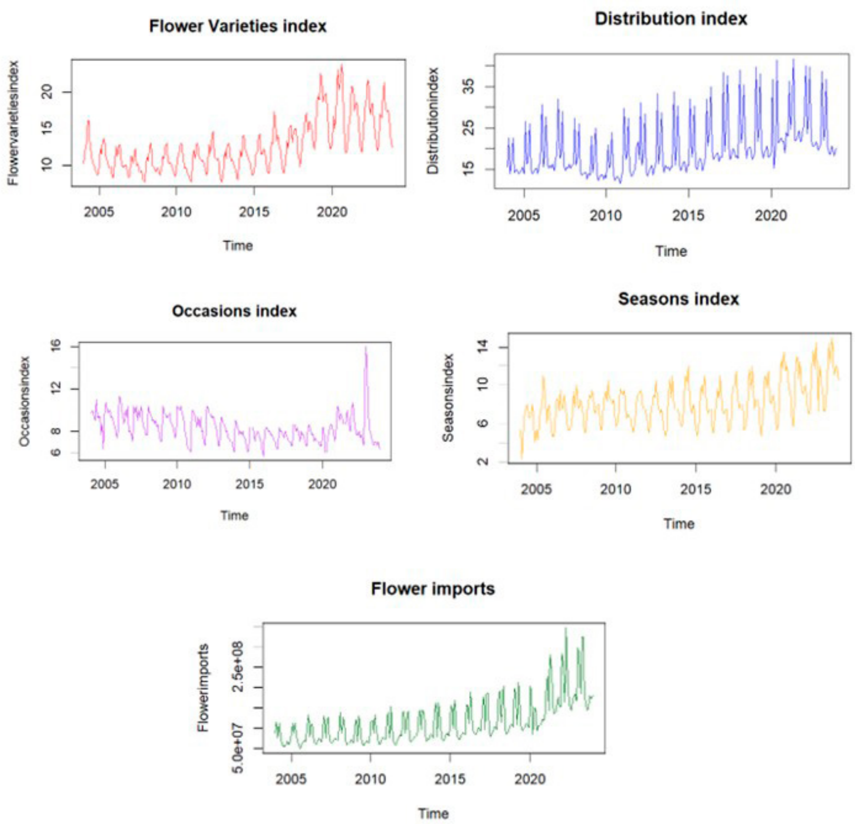

Fig. 1. Time series of flower imports and Google Trends searches.

According to Fig. 1, it can be observed that all variables exhibit seasonal behavior over time, reflecting regular peaks associated with recurring events, possibly special dates with high flower demand. However, some variables also display upward trends, which could suggest a failure to meet the stationarity assumption. Specifically, imports, the distribution index, the seasons index, and the flower varieties index appear to show progressive expansion. To corroborate this finding, it is essential to review the results of the Elliot, Rothenberg, and Stock unit root test (DF-GLS), which are presented below.

Table 2. Results of the stationarity test (DF-GLS) confirming compliance with the assumption in the series.

Variable	Test Statistic	Critical Value (1%)	p-value
Imports	-3.6186	-2.57	< 2.2e-16
Distribution Index	-3.9524		< 2.2e-16
Varieties Index	-3.1344		1.488e-10
Seasons Index	-2.8883		0.0006046
Occasions Index	-3.2198		8.527e-07

Based on the results in Table 2, it can be concluded that all variables are stationary, as the test statistic for each series is lower (or more negative) than the critical value at the 1% level of -2.57. Considering this, it is not necessary to differentiate any of the variables included in the model, and time series estimations can be directly applied using the explanatory variables at their original level.

Time Series Analysis

After verifying the assumptions for applying the models, it was necessary to carry out a process to select the appropriate method. Although machine learning methods have gained popularity in forecasting tasks, this study opted for the family of time series models for two fundamental reasons. First, the objective is not only prediction but also inference and interpretability of the relationship between Google searches and imports. These models offer clear and statistically evaluable coefficients that quantify the impact of each variable, a feature often absent in more complex machine learning models considered “black boxes” (Sullivan, 2020). Second, the size of the dataset (240 monthly observations) is suitable for traditional time series econometrics, whereas many machine learning algorithms require significantly larger data volumes to avoid overfitting (Rajput et al., 2023). Therefore, the econometric approach was considered more appropriate for transparently testing the research hypothesis.

The literature consistently states that there is no single forecasting method that works best in all cases. However, ARIMA-type models and their extensions are often a good choice in this type of analysis where the direction of the effect between the analyzed variables is recognized (Jiao and Chen, 2019). In our case, the research hypothesis posits that the direction of the effect goes from Google Trends searches to imports.

To increase the robustness of our model, we employed the Auto ARIMA method proposed by Hyndman and Khandakar (2008), which consists of an automated approach that seeks to detect the best possible ARIMA model based on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). In addition, we applied three distinct types of models, comparing their AIC and BIC: a simple ARIMA without exogenous variables or seasonal components, using exclusively the autoregressive structure of the imports variable; an ARIMAX that included the exogenous variables, i.e., the four Google Trends indices, but without seasonal components; and finally, a SARIMAX that included both the Google Trends indices and seasonal components.

Each of the applied models can be specified as follows:

The **ARIMA** (p, d, q) model can be specified as follows:

$$\phi(\beta)(1-\beta)^d\ yt = c + \theta\ (\beta)_{\varepsilon_t}$$

Where,

$$\phi(\beta) = 1 - \phi_1(\beta) - \phi_2(\beta)^2 - \cdots - \phi_p(\beta)^p$$

is the autoregressive operator.

$$(1-\beta)^d$$

represents differencing of order

$$\theta(\beta) = 1 + \theta_1(\beta) + \theta_2(\beta)^2 + \cdots + \theta_q\beta^q$$

is the moving average operator.

And finally, ε_t is an error term or white noise.

The ARIMA model serves as our baseline, as it attempts to predict future imports based solely on their historical behavior. In simple terms, the ARIMA model is built on three fundamental components: the autoregressive component (p), which represents the “memory” of the series by assuming that the value of imports in a given month depends on values recorded in previous months; the integration component (d), which acts as a “stabilizer” by adjusting the series if it exhibits trends (e.g., constant growth in imports), making it stationary and more predictable; and finally, the moving average component (q), which functions as an “error corrector” by modeling the impact of errors or surprises from previous forecast periods.

In the case of the **ARIMAX** model, it can be specified as follows,

$$\phi(\beta)(1-\beta)^d yt = \beta X_t + \theta\ (\beta)_{\varepsilon_t}$$

Where X_t represents the matrix of exogenous variables, i.e., distribution index, varieties index, seasons index, and occasions index, and β are their coefficients.

This model is an extension of ARIMA that incorporates “external helpers.” In addition to using the history of imports, it includes the four Google Trends indices as explanatory variables , allowing us to directly measure their predictive power.

For the **SARIMAX** model, it can be specified as follows,

$$\phi(\beta)\Phi(\beta^{12})(1-\beta)^d(1-\beta^{12})^D yt = \beta X_t + \theta\ (\beta)\Theta(\beta^{12})_{\varepsilon_t}$$

Where,

$$\Phi(\beta^{12}) = 1 - \Phi_1(\beta^{12}) - \Phi_2\left(\beta^{2(12)}\right) - \cdots - \Phi_P\beta^{P^{12}}$$

is the seasonal autoregressive operator.

$$(1-\beta^{12})^D$$

represents seasonal differencing of order D

$$\Theta(\beta^{12}) = 1 + \Theta_1(\beta^{12}) + \Theta_2\left(\beta^{2(12)}\right) + \cdots + \Theta_Q\beta^{Q^{12}}$$

is the seasonal moving average operator. The superscript 12 is the seasonal period, in this case, since monthly data is used.

This is the most comprehensive model because it “understands the calendar.” In addition to the ARIMAX components, it adds a seasonal component indicated by P, D, Q. This is crucial for the flower market, as it enables the model to anticipate patterns that repeat annually in the same months (m = 12), such as demand peaks for Valentine’s Day or Mother’s Day.

Model Evaluation

As previously mentioned, the estimation of each model used the method proposed by Hyndman and Khandakar (2008), implemented in the R library *forecast*, which applies the *auto.arima* algorithm to automatically select the values of p, d, q, P, D, and Q. These correspond to the orders of the autoregressive, differencing, and moving average components—both non-seasonal and seasonal, respectively. In other words, the algorithm tests various possible parameter combinations and selects the one that best fits the historical data, eliminating the need for the researcher to manually determine these values through trial and error. Here, the algorithm automatically reviewed the AIC and BIC values to provide the optimal combination for ARIMA, ARIMAX, and SARIMAX. These values are shown in the "Specification" column of Table 3.

To compare the best type of model, in addition to AIC and BIC, we also considered the Mean Absolute Percentage Error (MAPE), the Mean Absolute Scaled Error (MASE), and the autocorrelation of residuals (ACF). These indicators are also presented in Table 3.

Table 3. Comparison of ARIMA, ARIMAX, and SARIMAX models based on fit and predictive accuracy criteria.

Model	Specification	AIC	BIC	MAPE	MASE	ACF
ARIMA	(5,1,1)	8966.01	8993.82	22.55%	2.06	-0.0980809
ARIMAX	(2,1,2) with 4 independent variables	8868.78	8909.54	14.32%	1.50	0.02254698
SARIMAX	(1,0,1) (0,1,1) CE (12) with 4 independent variables	8146.18	8170.15	7.04%	0.72	0.01220668

CE = Seasonal Component (12 months)

As shown in Table 3, the first model, the simple ARIMA without independent variables or seasonal components, exhibited limited performance with a MAPE of 22.55% and a MASE of 1.87, indicating high prediction errors. The second model (ARIMAX), which included dependent variables, notably improved accuracy, achieving a MAPE of 14.32% and a MASE of 1.24. However, it was slightly penalized with higher AIC and BIC values, likely due to increased model complexity. The most significant improvement was observed in the third model (SARIMAX), which reached a MAPE of 7.04% and a MASE of 0.72, outperforming even a naïve model (MASE < 1). Additionally, it achieved the lowest values for both AIC and BIC, as well as the lowest ACF. Based on these results, it is evident that seasonality is key in U.S. flower imports; in other words, import behavior may be dramatically influenced by special dates such as Valentine's Day, Christmas, or Mother's Day.

Model Results

Estimation of the SARIMAX Model

Given that the SARIMAX model presented the best fit to the data, we proceeded to analyze the coefficients of the independent variables and their significance levels, which are presented in Table 4.

Table 4. Estimates of the SARIMAX model coefficients with exogenous variables.

Coefficient	Estimation	p-value
ar1	-7.1015e-01	< 2.2e-16 ***
sma1	-4.7777e-01	6.665e-12 ***
Distribution Index	1.9158e+06	0.0009307 ***
Varieties Index	1.6757e+06	0.1265007
Seasons Index	-1.0317e+05	0.9344244
Occasions Index	1.0507e+06	0.3448635

Model Structure: SARIMAX (1,0,1) (0,1,1) CE = 12, with regressors. ***$p < 0.001$

According to the results obtained from the SARIMAX estimation, both the ar1 component and the Sma1 component are statistically significant. This means that the behavior of imports not only depends on values recorded in immediately preceding periods (autoregressive effect) but is also influenced by repetitive seasonal patterns that systematically occur over time (seasonal moving average effect). In other words, the model confirms the existence of short-term dependencies, as well as regularities associated with seasonality that exert a statistically relevant influence on the dynamics of the series. Regarding the independent variables, only the distribution index shows a statistically significant relationship, with 99% statistical confidence. In simpler terms, based on the applied model, it can only be supported that searches for flowers associated with distribution channels, such as flower delivery or florists, have predictive power over variations in imports.

Model Simplification

Based on the results obtained from the SARIMAX model and with the intention of adhering to the principle of parsimony, we considered simplifying the model by retaining only the distribution index as the exogenous variable. The results of the simplified model are presented below in Table 5.

Table 5. Results of the simplified SARIMAX model with the distribution index as an exogenous variable.

Coefficient	Estimation	p-value
ma1	-7.2361e-01	< 2.2e-16 ***
sma1	-4.8370e-01	3.052e-12 ***
Distribution Index	2.3008e+06	2.190e-05 ***
AIC	8144.28	N/A
BIC	8157.98	N/A
Log-Likelihood	-4068.14	N/A
MAPE	6.856854	N/A
MASE	0.7182014	N/A
ACF	0.0330479	N/A

SARIMAX(0,1,1)(0,1,1), CE = 12. ***$p < 0.001$

With the results of the simplified model presented in Table 5, it can be concluded that reducing the number of exogenous variables substantially improved all metrics. For example, the AIC decreased from 8146.18 in the full model to 8144.28. Similarly, the BIC decreased from 8170.15 to 8157.98. Regarding error metrics, these show excellent values with a MAPE of 6.86% and a MASE of 7.18%, indicating high predictive accuracy. Finally, the correlation of residuals (ACF) obtained a value of 0.033, very close to zero, suggesting that the residuals do not exhibit significant autocorrelation. Regarding the model components, all are statistically significant with a 99% confidence level. Particularly, the exogenous variable distribution index showed a coefficient of 2.30083e+06, indicating that for each unit increase in this indicator, flower imports increased by approximately 2.3 million dollars in CIF.

Cross-Validation

Finally, to verify the predictive capacity of the estimated model, a cross-validation procedure was applied where the data was divided into a training set and a test set. Specifically, data from January 2004 to December 2019 was used for training, while data from January 2020 to December 2022 was used for the test set. Additionally, to establish the robustness of the model, error metrics were calculated for both estimations. The following Table 6 contains the estimation results.

Table 6. Performance metrics of the SARIMAX model in cross-validation.

Indicator	Training	Test
MAPE	6.643922	8.885622
MASE	0.6825996	1.5106706
Theil’s U	N/A	0.4127316
ACF	0.01417434	0.32234744

When analyzing the test statistics from the cross-validation of the SARIMAX model, it is evident that the error metrics (MAPE and MASE) are higher in the test set compared to the training set. However, this is common in this type of evaluation. Moreover, a MAPE below 10% in the test set is still considered very satisfactory for economic predictions (Lewis, 1982). Regarding Theil’s U statistic, it yields a value of 0.41, which is below 1, confirming that the model has superior predictive capacity compared to a naive model, i.e., a model that does not incorporate trends, seasonality, or exogenous variables. Finally, the increase in ACF in the test data may indicate that there is some temporal structure in the residuals not fully captured by the model, possibly due to the COVID-19 pandemic in the test data, which may have influenced this situation. However, this ACF value is not high enough to indicate a misspecification in our model. Overall, the cross-validation results confirm the robustness of the simplified SARIMAX model and its ability to generate predictions for flower imports in the United States.

Discussion

The results partially confirm the initial hypothesis by showing that Google Trends can be useful as a predictive tool in the international flower trade. However, not all search categories have the same explanatory power. The only statistically significant predictor was the distribution index related to terms linked to sales channels, with an estimated effect of $2.3 million in imports for each unit increase. This selectivity is understood from consumer behavior theory, as searches linked to distribution channels reflect more immediate and concrete purchase intentions than those focused on varieties or events. Thus, queries such as “flower delivery” or “florist near me” express a much greater proximity to the purchase decision than merely informational searches, such as those about specific flower types.

At first glance, it may seem counterintuitive that indices related to occasions and seasons lack predictive power, given the clear influence of holidays like Valentine’s Day on demand. However, this apparent contradiction is resolved by analyzing the structure of the SARIMAX model. The seasonal component of the model ((0,1,1)[12]) is precisely designed to capture recurring and predictable patterns throughout the year. Essentially, the model already “knows” that February, May, and December are high-demand months, so the information contained in search spikes for “Valentine’s Day flowers” is largely redundant from a statistical perspective; its explanatory power has already been absorbed by the model’s seasonal component. The distribution index, on the other hand, appears to capture volatility in purchase intent that goes beyond these predictable peaks, thus providing new and valuable information to the model.

The results show that the seasonal component captured by the SARIMAX model reinforces the importance of Google Trends in highly seasonal industries such as floriculture. On key dates like Valentine’s Day or Mother’s Day, its predictive capacity becomes strategic. SARIMAX outperformed simpler models like ARIMA and ARIMAX, demonstrating that ignoring seasonality reduces accuracy. Additionally, the simplified model achieved very low error metrics (MAPE of 6.86% and MASE of 0.718), confirming its robustness and positioning Google Trends as a valuable market intelligence tool in international trade. However, cross-validation revealed a higher error in the test set (2020 - 2023), an increase attributable to the emergence of an unprecedented exogenous shock: the COVID-19 pandemic. This event drastically altered consumption patterns, global supply chains, and online search behavior, introducing volatility that no historically based model could fully anticipate (Hodbod et al., 2021). Therefore, the increased error does not invalidate the model’s relevance but rather highlights its vulnerability to disruptive events and reinforces the need to complement such predictive tools with real-time contextual analysis.

From a theoretical perspective, our results validate the causal mechanism proposed in the theoretical framework: the observed increase in searches related to floral distribution effectively signals a purchase intent that, once aggregated at the population level, allows anticipation of significant movements in the international supply chain. Importers, alerted by these early demand signals, adjust their orders to adequately meet the future market. This finding significantly aligns with previous documentation by Chaudhady and Choi (2023) in the seafood sector, suggesting that the predictive value associated with Google Trends could extend to various consumer products that share similar seasonal patterns.

Conclusions

The present study aimed to evaluate whether variations in Google Trends search volumes function as leading indicators to predict changes in flower imports to the United States. The research partially validates this hypothesis, demonstrating that Google Trends is an effective early prediction tool, but its predictive power is specifically concentrated in searches associated with distribution channels. The central finding is that the "distribution index", which groups terms such as *"florist"* or *"flower delivery"*, was the only statistically significant predictor. This suggests that searches indicating clear and direct purchase intent best anticipate import flows. Conversely, indices related to varieties, seasons, or special occasions did not show significant predictive power, likely because the robust seasonal component of the SARIMAX model already efficiently captures these predictable and recurring demand patterns, such as those for Valentine's Day or Mother's Day. The superiority of this model, with a final MAPE of 6.86%, confirms that seasonality is a dominant factor in this market.

The strategic implications of these results are significant for the floral value chain. They allow a transition from a reactive management model to a proactive one. Producers and importers can use monitoring of the distribution index as an early signal to adjust planting cycles, optimize purchase volumes, and plan logistics further in advance, thereby mitigating risks of overstocking or stockouts. For retailers, this information is key to aligning marketing campaigns and last-mile logistics with peaks in consumer interest, enabling them to capitalize on emerging demand more efficiently. Finally, while this approach demonstrates great potential, it is important to recognize its limitations, particularly its vulnerability to external shocks, such as the COVID-19 pandemic. To address these challenges, future research should focus on three key areas: (1) Developing multimodal models that integrate search data with other real-time sources (sentiment analysis, e-commerce data, weather conditions); (2) Expanding the model geographically to other importing markets to validate its transferability; (3) Corroborating findings with primary industry data to strengthen empirical robustness.

Acknowledgments

The authors declare that there are no acknowledgments to be made.

Author Contribution

CFOA: Conceptualization, Investigation, Methodology, Supervision, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing. **PALH:** Data Curation, Formal Analysis, Methodology, Software, Writing – Original Draft. **DAGB:** Project Administration, Funding Acquisition, Supervision, Validation, Writing – Review & Editing.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

Data will be made available upon request to the authors.

Declaration of generative AI and AI-assisted technologies in the writing process

Generative AI and AI-assisted technologies were used to support the translation of the original text into English with Claude 3.7.

References

ARIZTON ADVISORY AND INTELLIGENCE. **U.S. floral gifting market – Industry outlook & forecast 2025–2030** (Report No. ARZ19657053). United States: Arizton Advisory and Intelligence, 2025. Available at: https://www.arizton.com/market-reports/floral-gifting-market-in-united-states-2025 Accessed on: May 15, 2025.

BANTIS, E.; CLEMENTS, M.; URQUHART, A. Forecasting GDP growth rates in the United States and Brazil using Google Trends. **International Journal of Forecasting**, v.39, n.4, p.1909-1924, 2022. https://doi.org/10.1016/j.ijforecast.2022.10.003

BOONE, T.; GANESHAN, R.; HICKS, R. L.; SANDERS, N. R. Can Google Trends improve your sales forecast? **Production and Operations Management**, v.27, n.10, p.1770-1774, 2018. https://doi.org/10.1111/poms.12839

CASTELLACCI, F.; SANTOALHA, A. Does digitalisation affect the adoption of electric vehicles? New regional-level evidence from Google Trends data. **Regional Studies**, v.59, n.1, p.2358829, 2025. https://doi.org/10.1080/00343404.2024.2358829

CHAUDHADY, A.; CHOI, S. Forecasting fish import using deep learning: a comprehensive analysis of two different fish varieties in South Korea. **Smart Media Journal**, v.12, n.11, p.134-144, 2023. https://doi.org/10.30693/SMJ.2023.12.11.134

CHOI, H.; VARIAN, H. Predicting the present with Google Trends. **Economic Record**, v.88, n.1, p.2-9, 2012. https://doi.org/10.1111/j.1475-4932.2012.00809.x

CREDENCE RESEARCH. **Flower delivery service market by product (rose, tulip, orchid, lily, gerbera, sunflower, carnations, others); by occasion (corporate event, wedding, personal & gifting, sympathy); by type (bouquets, flowers for vases, cut flowers, plants); by application (commercial, household) - Growth, share, opportunities & competitive analysis, 2024-2032**. United States: Credence Research, 2025. Available at: https://www.credenceresearch.com/report/flower-delivery-service-market. Accessed on: 26 April 2025.

DE LUCA, G.; ROSCIANO, M. Google Trends data and transfer function models to predict tourism demand in Italy. **Journal of Tourism Futures**, v.10, n.1, p.1-17, 2024. https://doi.org/10.1108/jtf-01-2023-0018

DERA, P.; KLIMIUK, K.; KALINOWSKA-BESZCZYŃSKA, O.; BALWICKI, Ł. Digital smoke signals: Event-driven online search trends in Heated Tobacco Products in Poland. **Tobacco Prevention & Cessation**, v.10, n.May, p.20-18332, 2024. https://doi.org/10.18332/tpc/187280

DIAZ-BALTEIRO, L.; ALFRANCA, O.; VOCES, R.; SOLIÑO, M. Using Google search patterns to explain the demand for wild edible mushrooms. **Forest Policy and Economics**, v.152, p.102993, 2023. https://doi.org/10.1016/j.forpol.2023.102993

DRECHSLER, M., and HOLZAPFEL, A. Decision support in horticultural supply chains: a planning problem framework for small and medium-sized enterprises. **Agriculture**, v.12, n.11, p.1922, 2022.

GRAND VIEW RESEARCH. **U.S. floriculture market size, share & trends analysis report by product (bedding & garden plants, potted plants), by distribution channel (offline, online), by end use (commercial, industrial), and segment forecasts, 2024–2030 (Report No. GVR-4-68040-395-2)**. United States: Grand View Research, 2025. Available at: https://www.grandviewresearch.com/industry-analysis/us-floriculture-market-report. Accessed on: April 26, 2025.

HUANG, H.; CUAN, X.; CHEN, Z.; ZHANG, L.; CHEN, H. A multiregional agricultural machinery scheduling method based on hybrid particle swarm optimization algorithm. **Agriculture**, v.13, n.5, p.1042, 2023. https://doi.org/10.3390/agriculture13051042

HODBOD, A.; HOMMES, C.; HUBER, S. J.; SALLE, I. The COVID-19 consumption game-changer: Evidence from a large-scale multi-country survey. **European Economic Review**, v.140, p.103953, 2021. https://doi.org/10.1016/j.euroecorev.2021.103953

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Software**, v.27, n.3, p.1–22, 2008. https://doi.org/10.18637/jss.v027.i03

JIAO, E. X.; CHEN, J. L. Tourism forecasting: A review of methodological developments over the last decade. **Tourism Economics**, v.25, n.3, p.469-492, 2019. https://doi.org/10.1177/1354816618812588

KANTASA-ARD, A.; NOUIRI, M.; BEKRAR, A.; AIT EL CADI, A.; SALLEZ, Y. Machine learning for demand forecasting in the physical internet: a case study of agricultural products in Thailand. **International Journal of Production Research,** v.59, n.24, p.7491–7515, 2020. https://doi.org/10.1080/00207543.2020.1844332

LESSOUA, A.; MUTASCU, M.; TURCU, C. Firm performance and exports: evidence from the romanian wine industry. **Journal of Wine Economics**, v.15, n.2, p.207-228, 2020. https://doi.org/10.1017/jwe.2020.28

LEWIS, C.D. **Industrial and Business Forecasting Methods.** Londres: Butterworths, 1982.

MANKIW, N.G. **Principles of Economics**. 6th edition. Mason: Editorial SA Ediciones Paraninfo, 2012. 705p.

NIÑO, G.; JANZEN, J. Valentine's Day and the Gains from Agricultural Trade: Cut Flowers in the US. **Farmdoc Daily**, v.2025, n.2, p.1-8, 2025.

PITCHAYADOL, P. Farmer's adoption intention towards eco-innovation in Thailand. **ASEAN Journal of Management and Innovation**, v.5, n.2, p.62–79, 2018. https://doi.org/10.14456/ajmi.2018.31

RAJPUT, D.; WANG, W.; CHEN, C. Evaluation of a decided sample size in machine learning applications. **BMC Bioinformatics**, v.24, 2023. https://doi.org/10.1186/s12859-023-05156-9

REN, S.; CHAN, H. L.; SIQIN, T. Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. **Annals of Operations Research**, v.291, p.761–777, 2020. https://doi.org/10.1007/s10479-019-03148-8

ROMBACH, M.; DEAN, D. L.; OLYNK WIDMAR, N. J.; BITSCH, V. "Oh, you shouldn't have!" Understanding key factors impacting cut flowers gifting preferences in Germany. **Horticulturae**, v.7, n.10, p.368, 2021. https://doi.org/10.3390/horticulturae7100368

SARI, M.; DURAN, S.; KUTLU, H.; GULOGLU, B.; ATIK, Z. Various optimized machine learning techniques to predict agricultural commodity prices. **Neural Comput & Applic**, v.36, p.11439-11459, 2024. https://doi.org/10.1007/s00521-024-09679-x

SILVA, E.S.; HASSANI, H.; MADSEN, D.Ø.; GEE, L. Googling fashion: forecasting fashion consumer behaviour using Google Trends. **Social Sciences**, v.8, n.4, p.111, 2019. https://doi.org/10.3390/socsci8040111

SPENCE, J.; KOLEVA, B.; BENFORD, S.; DARZENTAS, D.; FLINTHAM, M.; GLOVER, K.; WAGNER, H.; GIBSON, R.; THORN, E. "More than a cliché": Experiencing Hybrid Gifting in the Wild. **ACM Transactions on Computer-Human Interaction**, v.30, p.1-31, 2023. https://doi.org/10.1145/3577015

SULLIVAN, E. Understanding from machine learning models. **The British Journal for the Philosophy of Science**, v.73, p.109-133, 2020. https://doi.org/10.1093/bjps/axz035

SUN, R**.** "You need to listen to the market!": Making decisions through senses in China's largest cut-flower market. **American Behavioral Scientist**, p.00027642241246684, 2024. https://doi.org/10.1177/00027642241246684

TAMASIGA, P.; ONYEAKA, H.; BAKWENA, M.; HAPPONEN, A.; MOLALA, M. Forecasting disruptions in global food value chains to tackle food insecurity: The role of AI and big data analytics – A bibliometric and scientometric analysis. **Journal of Agriculture and Food Research**, v.14, p.100819, 2023.https://doi.org/10.1016/j.jafr.2023.100819

THIRUMALAI, S.; SINHA, K.K. Customization of the online purchase process in electronic retailing and customer satisfaction: An online field study. **Journal of Operations Management**, v.29, n.5, p.477-487, 2011. https://doi.org/10.1016/j.jom.2010.11.009

TROMP, S.O.; VAN DER SMAN, R.G.; VOLLEBREGT, H.M.; WOLTERING, E.J. On the prediction of the remaining vase life of cut roses. **Postharvest Biology and Technology**, v.70, p.42-50, 2012. https://doi.org/10.1016/j.postharvbio.2012.04.003

VOSEN, S.; SCHMIDT, T. Forecasting Private Consumption: Survey Based Indicators vs. Google Trends. **Journal of Forecasting,** v.30, n.6, p.565-578, 2011. https://doi.org/10.1002/for.1213

WANG, S.; CHEN, T.; WANG, C.; LIU, Z.; JIA, L.; ZHAO, X. The effect of customer satisfaction on floral product purchase behavior, evidence from Shanghai, China**. Scientific Reports**, v.13, n.1, p.7945, 2023. https://doi.org/10.1038/s41598-023-35137-0

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*?* **Annals of Tourism Research**, v.93, p.103365, 2022. https://doi.org/10.1016/j.annals.2022.103365

YE, X.; YUE, P. What matters to reshaping consumption patterns in China? Digital inclusion and supply chain. **Finance Research Letters**, v.59, p.104804, 2024. https://doi.org/10.1016/j.frl.2023.104804

CHELLIAH, B. J.; LATCHOUMI, T. P.; SENTHILSELVI, A. Analysis of demand forecasting of agriculture using machine learning algorithm. **Environment, Development and Sustainability**, v.26, p.1731-1747, 2024. https://doi.org/10.1007/s10668-022-02783-9