TECHNICAL NOTE Nº IDB-TN-2882

Language Models and Google Trends

An Application to Tourism in the Andean Countries

Cristhian Larrahondo
Emily Díaz
Diego A. Guerrero

Inter-American Development Bank
Country Department Andean Group

February 2024

Language Models and Google Trends

An Application to Tourism in the Andean Countries

Cristhian Larrahondo*
Emily Díaz*
Diego A. Guerrero**

* Inter-American Development Bank
** The University of Tennessee

Inter-American Development Bank
Country Department Andean Group

February 2024

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Language models and Google Trends: an application to tourism in the Andean Countries / Cristhian Larrahondo, Emily Díaz, Diego Guerrero.
p. cm. — (IDB Technical Note ; 2882)
Includes bibliographical references.
1. Tourism-Data processing-Andean Region. 2. Internet searching-Andean Region. 3. Information storage and retrieval systems-Artificial intelligence. I. Larrahondo, Cristhian. II. Diaz, Emily. III. Guerrero, Diego. IV. Inter-American Development Bank. Country Department Andean Group. V. Title. VI. Series.
IDB-TN-2882

http://www.iadb.org

Language Models and Google Trends: An Application to Tourism in the Andean Countries

Authors: Cristhian Larrahondo, Emily Díaz, and Diego A. Guerrero

February 2024

JEL Codes: C82, L83, O10, O20, O54, Z32, Z38

Keywords: Tourism, Google Trends, Language models, ChatGPT, Bard, Andean Region

Acknowledgments: This technical note is part of a regional study on tourism published by the Andean Countries Department of the Inter-American Development Bank, where the current state and potential of tourist resources in Peru, Ecuador, Colombia, Bolivia, and Venezuela are examined. The authors thank Laura Giles Álvarez and Wladimir Zanoni for their comments, reviews, edits, and suggestions. They also thank research assistants Luis Serrudo and Miguel Alzamora for their guidance in data acquisition.

Executive Summary

Using Google Trends data, this study leverages high-frequency unstructured data to characterize tourism demand in the Andean countries. The paper explores real-time data to monitor trends in the tourism industry, leveraging language models to identify suitable search terms. The document presents a methodology based on keywords related to tourism products in Bolivia, Colombia, Ecuador, Peru, and Venezuela between 2010 and 2023. Results show decline and recovery trends in the interest in traveling to these destinations. Additionally, trends for local tourist destinations and specific products are analyzed. The real-time data-driven results can guide industrial policies in the region, emphasizing the importance of understanding and adapting to changing dynamics of tourism demand in the Andean Region.

1. Introduction

All five countries comprising the Andean Region (Bolivia, Colombia, Ecuador, Peru, and Venezuela) have received ten million tourists annually since 2010 (World Bank, 2023a). This number represents 100 million tourists over a decade or 5% of those received in Latin America and the Caribbean. On average, tourism accounts for 7% of total exports in Bolivia, 9% in Colombia, 6% in Ecuador, 7% in Peru, and 1% in Venezuela (World Bank, 2023b). Therefore, tourism stands out as a key industry for economic development, characterized by generating spillovers to other industries and localities within a country that may not be inherently tourist-oriented (Faber & Gauber, 2019).

Similarly, it is a key sector for productive transformation, as tourism represents an opportunity for the Andean Region. First, tourism is a major stimuli for service exports and has a great potential attracting income (Abuelafia et al, 2023).[1] Additionally, as emphasized by Fayissa et al. (2011), tourism positively contributes to per capita GDP growth of countries in Latin America and can enhance short-term economic growth. Along the same lines, negative shocks in the tourism industry significantly affects economic development. For example, Mulder (2020) warns that the impact of the COVID-19 pandemic on the tourism industry may have decreased employment and GDP by 1% in Latin America.

Given that tourism substantially contributes to economic growth in the Andean Region, it is necessary to have indicators that allow monitoring the sector's

[1] Tourism records levels above the world average within service exports. In the Andean countries, receptor tourist spending accounts for up to 2.5% of the GDP between 2015 and 2019.

performance. However, existing statistics have limitations in terms of availability, quality, and comparability between units of analysis. The quintessential tourism indicator is international tourist arrivals, which tracks the number of non-resident visitors arriving in a country over a period.[2] Others include the number of hotel guests, total spending on travel and transportation, the number of employees in the industry, and macroeconomic variables reflecting the economic activity of tourism –for example, the contribution of tourism to GDP (World Tourism Organization, 2021). However, these indicators often differ in their measurement methodology from country to country or are not universally published, limiting comparative analysis and regional sector diagnosis. A significant milestone in the literature has been the use of high-frequency unstructured data from Google Trends to analyze and forecast tourism demand (Bokelmann & Lessmann, 2019; Park et al., 2017; Bangwayo & Skeete, 2015). In general, macro-econometric studies incorporating internet searches can improve their predictive power by up to 30% (Cevik, 2020).

In this context, the question arises: Can high-frequency unstructured data be used to monitor tourism demand in the Andean Region? This study answers this question developing a methodology for monitoring and evaluating the industry. High-frequency unstructured data is constructed to approximate Internet users' interest in tourism (later referred to as *potential demand*).[3] In this way, the aim is to have useful, timely, and frequent time series describing tourism demand patterns in the Andean Region countries. In addition, this study leverages language models to generate search terms that refine the approximation of tourism demand with Google Trends.[4] Thus, a tool is outlined that uses Google Trends data as its main input and can be generalized to analyze not only the tourism sector's situation but other sectors.

In summary, with this real-time indicator, the goal is to characterize recent trends in the potential tourism demand in the Andean countries and extend the analysis to identify specific local destinations with tourism potential. The methodology used is relevant because it allows evaluating fluctuations in potential tourism demand by identifying possible changes in user interests in destinations and tourism products in real-time, which may vary because of, for example, the pandemic or other exogenous shocks.

Specifically, as an approximation of tourism demand, time series data are extracted from Google searches between 2010 and 2023, representing the interest of potential tourists in destinations within the Andean countries. Google Trends website provides a time series with a normalized index that captures weekly interest from search

[2] This document focuses on a one-month period for this indicator.
[3] This refers to an approximation of the actual demand for tourism.
[4] Large Language Models (LLM) are Artificial Intelligence (AI) algorithms using machine learning techniques designed to understand and generate human-like text coherently, such as OpenAI's ChatGPT and Google's Bard.

queries relative to a historical point, reflecting the quantity of Google searches made by users. Google is commonly used for searching, buying, and comparing prices, and booking accommodation and tickets using words like "hotels" or "flights." Therefore, in line with previous studies forecasting tourism demand (Bokelmann & Lessmann, 2019; Park et al., 2017; Bangwayo & Skeete, 2015), a time series is constructed with a set of frequently used keywords in the search engine by users who potentially will travel to a tourist destination. Additionally, this paper evaluates the evolution of specific tourist products, defining a set of categories or types of tourism that define specialized tourist products using specific keywords. The categorization of tourist products captures trends interest in each country, including their cultural, historical, recreational, commercial, and ecological "products" that interest travelers.

This paper has two key takeaways regarding the characterization of tourism demand in the Andean Region. First, the document defines a potential demand for tourism in each country, showing promise in capturing user interest in tourism-related activities in the region. It also compares the real-time potential demand to traditional indicators and show these two are closely associated with over 0.6 correlation in most Andrean countries. Second, the potential demand is broken down by types of tourism, providing a deeper understanding of the preferences and needs of potential tourists. Together, the results highlight opportunities in the development and promotion of the tourism industry in the Andean Region. These two outcomes suggest that the methodology addresses limitations and shortcomings from traditional tourism sector data, as it allows for an early and homogeneous approximation of tourism demand.

The analysis of tourism demand using Google Trends is highly relevant to authorities, government institutions, and stakeholders involved in promoting and developing tourism. The indicators enable timely information on the industry and the study of possible gaps and opportunities in the available tourism supply in each country or region.[5] The findings provide a foundation for strategic decision-making, as they identify opportunities based on the specific preferences of potential tourists. This approach, in turn, can guide the design of public policies and promotion strategies that maximize the tourism potential of the Andean Region. Furthermore, the information is valuable for investors and stakeholders in the tourism industry, offering solid insights to support decision-making. In summary, this analysis has the potential to catalyze sustainable and beneficial growth for the economy and communities in the Andean Region, serving as a valuable tool for decision-makers and stakeholders in the industry.

Furthermore, the approximation to tourism demand according to its type holds fundamental importance, as it provides a solid foundation for the sustainable

[5] Characterizing the demand would allow to strengthen or make up for deficiencies in the current supply.

development of the tourism industry in the region. By understanding and categorizing different tourist segments, authorities, entrepreneurs, and stakeholders can tailor specific strategies to meet the needs and expectations of visitors, promote lesser-known destinations, and diversify the tourism offering. Additionally, this identification allows for optimizing natural and cultural resource management, preserving historical and ecological heritage, and generating economic opportunities for local communities. Ultimately, understanding the types of tourism in the Andean Region is essential to ensure responsible and beneficial tourism for visitors and local populations, contributing to the sustainable growth of this industry in the region.

This study contributes to the literature in three ways. First, it adds to a niche of economic literature that has evaluated the use of real-time series for predicting tourism demand (Bokelmann & Lessmann, 2019; Park et al., 2017; Bangwayo & Skeete, 2015). Instead of the macroeconomic approach in previous studies, this paper defines a methodology for characterizing tourist products and a list of words associated with types of tourism that are relevant for characterizing the industry at the local level. The product-focused study allows for developing new terms of interest and evaluate locally relevant products that can expand models in the style of Wu et al. (2023).[6] Second, the paper expands the scope of literature using Google Trends, documenting the relevance of this indicator in five countries that have not been previously assessed with this data. Focusing on new countries allows for expanding evidence and characterizing local trends for developing countries. Finally, the paper adds to the evidence on the relevance of this data by comparing it with visitor number data published by each country.

This technical note is divided into five sections. The second section presents recent tourism trends in the Andean Region with some stylized facts as context. The third section is the proposed methodological section for using the Google Trends tool and language models to approximate tourism demand and its characterization. The fourth section presents the methodology's application and discusses its results as an analysis of tourism trends and types. Finally, the fifth section contains the main conclusions of this technical note.

2. Context

The tourism sector is still a developing industry in the Andean Region. Between 2011 and 2019, approximately 89 million people arrived in the region's countries, showing a

[6] This study shows how the heterogeneity of the data, the level of disaggregation, and different sources can be used to improve the predictive capacity of certain models.

year-on-year increase in arrivals (UNWTO, 2023). Figure 1 illustrates the total number of international tourists arriving in five countries and the regional total during this period. Peru and Colombia received the highest flow of international arrivals: in 2019 alone, Peru and Colombia recorded the entry of 4.2 million and 4.4 million people, respectively. Ecuador and Bolivia also experienced progressive increases in international arrivals between 2011 and 2019. However, to a lesser extent than Colombia and Peru, and in some years, there were decreases in the number of entries to these countries. On the contrary, Venezuela has shown a gradual decline in the number of international arrivals during this period, so that by 2019, only 260,000 people arrived (UNWTO, 2023).

The COVID-19 pandemic altered the patterns and outcomes of the tourism sector worldwide and for the countries of the Andean Region. According to available data from the UNWTO (2023), international arrivals worldwide dropped by 72% in 2020 compared to 2019.[7] In the Andean Region, a sharp decline in the international arrivals indicator has been observed since 2020, as shown in Figure 1, driven by mobility restrictions during the pandemic. These figures gradually improved over the past two years, although pre-pandemic levels have not yet been fully recovered for most countries; as of 2022, only Colombia has recovered to 2019 levels.

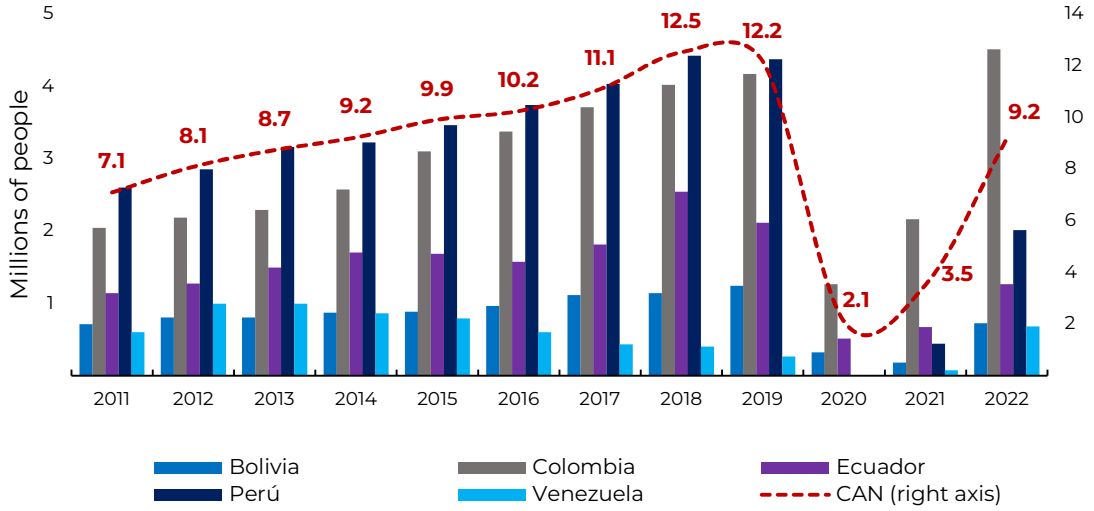
Other economic indicators depict the region's tourism situation before and after the pandemic. Figure 2 uses data from the UNWTO (2023) and national sources to illustrates the relative importance of the tourism sector in the Gross Domestic Product (GDP) of Colombia, Ecuador, and Peru. The tourism sector's contribution to GDP between 2011 and 2019 remained relatively stable from year to year.[8] In Peru, the tourism sector represented, on average, 3.5% of the GDP, while in Ecuador and Colombia, it was around 2.0% for each country. These figures show a decline starting in 2020, with tourism contributions decreasing to 1.5% in Colombia and 1.1% in Ecuador (see Figure 2). Revenues from tourism-related activities also exhibit a similar trend in the periods before and after the pandemic. For all countries in the region, tourism revenues increased between 2011 and 2019, even in the case of Venezuela, which showed these increases to a lesser extent until 2016.[9]

As shown in Figure 3, on average, income as a percentage of total exports increased between 2011 and 2019 and, in 2020, experienced a decline of 65.8% (from 10.0% in 2019 to 3.4% in 2020). The figures reflect the outcomes of sectors directly related to tourism, such as accommodation services, travel, and the sale of tourist products. However, the

[7] World Tourism Organization (2023), Dashboard UNWTO Tourism Recovery Tracker. https://www.unwto.org/es/unwto-tourism-recovery-tracker.
[8] No data is available for Peru. However, due to global trends and what has been observed for each country in the region, a decline similar to that of Colombia and Ecuador can be expected.
[9] Data for Venezuela is only available up to 2016.

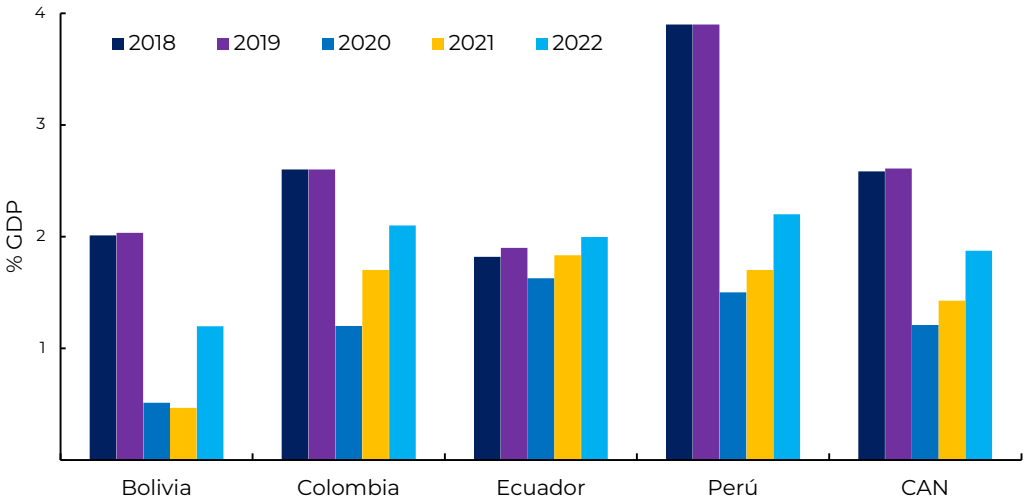
tourism industry contributes to the economy through other activities, for instance, in the consumption of goods and services, which can foster the growth of productive sectors (Richardson, 2010; Rogerson, 2012; Scarlett, 2021).

Figure 1. International tourists arrivals



Note: For 2020, UNWTO does not have international arrivals data for Peru and Venezuela.
Source: Tourism data from the World Tourism Organization (UNWTO).

Figure 2. Tourism GDP

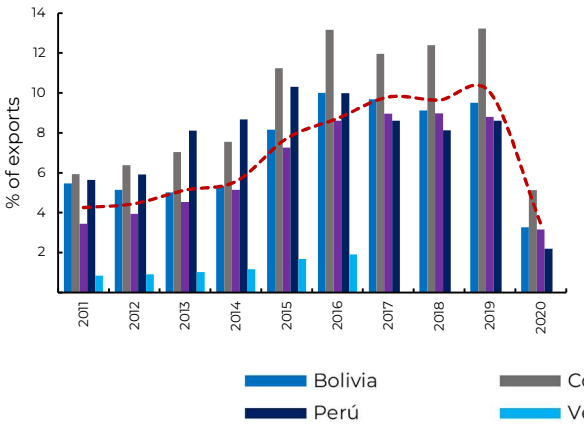


Note: The figure shows tourism GDP as a proportion of total GDP, with tourism GDP being the contributions of tourism-related industries in each country. Data available through 2019 for Peru, and through 2021 for Ecuador and Colombia. No data are available for Bolivia and Venezuela. The Andean Region (CAN) is the simple average of the countries without considering Venezuela, due to data availability.
Source: UNWTO, Instituto Nacional de Estadística (Bolivia), Departamento Administrativo Nacional de Estadística (Colombia), Banco Central del Ecuador (Ecuador), and Ministerio de Comercio Exterior y Turismo (Perú).

Finally, despite the trends shown so far, average spending of international tourists arriving in the Andean countries remained stable before, during, and after the pandemic. Figure 4 shows that tourist spending in the region was USD 950 between 2011 and 2016 but dropped to USD 717 between 2017 and 2019. It is noted that, during the analyzed period, tourists arriving in Colombia incur higher tourism expenses, exceeding USD 1200; this figure has been consistent between 2011 and 2022. On the other hand, tourist spending in Ecuador, Peru, and Bolivia increased from 2020 relative to the pre-pandemic period (2011 to 2019). Meanwhile, expenditure per tourist in Venezuela declined after 2020.

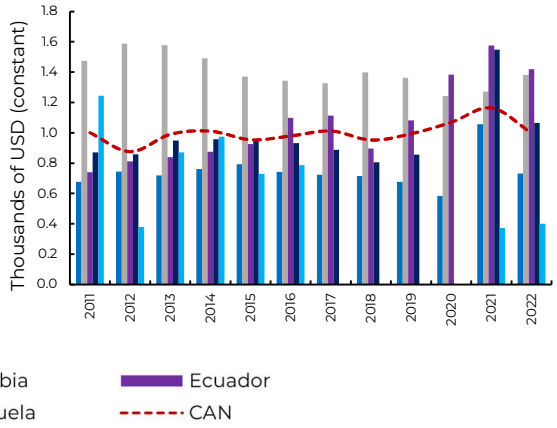
In summary, data shows that the tourism sector in the region has grown over the last decade in terms of international arrivals and income as exports, making significant contributions to the regional economy. The notable increase in international arrivals and the significant contribution of tourism to GDP are highlighted, demonstrating the strategic potential of the sector for the Andean economies. However, the COVID-19 pandemic negatively affected the performance of these indicators, manifested in the sharp decline in international arrivals and the reduction in the contribution of tourism to the GDP. These stylized facts highlight the vulnerability of the tourism industry in times of crisis. Understanding the situation is essential for addressing the reactivation and sustainable growth of the sector in the Andean Region, as well as anticipating and adapting to future challenges. To this end, a viable approach to consider is the dynamics of potential tourism demand, which is the proposal of this study.

Figure 3. Income from tourism



Note: According to the metadata, the indicator refers to a country's income from expenditures by international visitors due to tourism, weighted by income due to total exports of goods and services. Data are available up to 2020 for all countries except Venezuela; for Venezuela data are reported up to 2016.
Source: World Development Indicators, World Bank.

Figure 4. Average expenditure per international tourist (constant USD 2012)



Note: The World Tourism Organization's Dashboard of tourism data does not clarify the methodology to ensure comparability of the average expenditure per tourist. It states that the unit is "U.S. Dollars". Between 2017 and 2020 no data was reported for Venezuela; in 2020 no data is available for Peru.
Source: Tourism data of the WTO.

3. Methodology

This section details the methodology used to calculate the indicator capturing potential tourism demand. It also emphasizes the use of Google Trends data, its characteristics, and its applications. Likewise, the section describes the strategy used to determine potential tourism demand and the types of tourism based on Google Trends data, as well as to select keywords and search terms. The indicator allows for analyzing tourism demand trends and different types of tourism in the region.

3.1. Google Trends

Google Trends is a tool for monitoring and analyzing the popularity of search queries over time. The tool allows for identifying relevant trends, comparing between different topics, and examining how interest in a topic varies based on geographical location and demographic characteristics. Thus, this tool enables the analysis of temporal variation in tourism when data is linked to terms related to tourism. The result is a data series that allow an approximation to the analysis of the "potential" demand for tourism in each country.

The resulting indicators are considered an approximation to potential demand rather than actual or effective demand for four reasons. First, the index measures demand interest and not the demand, acting as a proxy for travel intentions rather than capturing travel itself. Second, the index is based on a sample of people using the search engine, which could exclude population segments that demand tourism but do not use the tool. Third, the index is sensitive to keywords used to generate searches. Finally, the data may exhibit a time lag between the search time and actual tourism.

The indicator is a normalized or relative index ranging from 0 to 100, capturing the search volume in Google. The highest value represents the peak popularity of a keyword or term in a specific period and location. This index is normalized with respect to an unknown baseline that depends on the periodization and the number of searches. Since it is an index, it does not reflect absolute values, and the normalization process used by the company is not publicly known. Additionally, Google Trends data is based on a selective sample of searches on Google.

To extract data from Google Trends, the researcher enters a series of parameters on the website. First, the keywords that capture search trends. Second, a time range of interest between 2004 and the present date. Third, the country or geographic region where searches originate, which may also determine the language of the keywords. Fourth, the researcher may select one of 67 categories determined by Google's algorithm (e.g., arts and entertainment, business, finance, or travel). Finally, the type of search engine in Google's network: web searches, images, news, shopping, or YouTube.

Three features related to these parameters determine the validity and reliability of the analysis. First, the precise selection of relevant keywords used by the target population in their searches. Various studies have suggested a series of keywords that are relevant to the analysis of tourism demand. Choi and Varian (2012) show that searches for "Hong Kong" categorized as tourism from nine countries have an average correlation of 73% with reported monthly visits to the region. Since then, more sophisticated studies have extended the list of words to include "hotels," "flights," or "travel" for predicting tourism demand (Bokelmann & Lessmann, 2019; Park et al., 2017; Bangwayo & Skeete, 2015; Feng et al., 2019). Other studies suggest that the average tourist searches for climate, shopping, and destination-related attractions (Wu et al., 2023).

Second, it is important to consider the idiosyncrasies of geographies and time intervals. Each country has idiosyncratic differences in its use of technological tools, where regulation, cultural patterns, and linguistic differences imply that other search engines enjoy more popularity than Google. For example, geographical selection can also affect the language of keywords for searches, as pointed out by Dergiades et al. (2018). Similarly, Yang et al. (2015) uses data from multiple search engines and conclude that the accuracy of forecast models depends on the popularity of search engines in particular countries. Additionally, it is challenging to choose a representative sample of tourists' origin countries for the Andean Region. This study uses the parameter of global searches, simplifying language difficulties and identifying general tourism-demanding countries.

Finally, it is imperative to recognize the inherent limitations of Google Trends data, as these are not exempt from inaccuracies. The search interest index is a construction made with reference to a baseline that fluctuates depending on the search parameters. For example, searching the keyword "flights to Ecuador" may have a specific value of "84" in the first week of 2022 when the time range begins in 2018, but the value in the same date is "67" when studying the 2004-2023 period. The difference stems from changes in the index baseline, but fluctuations and trends over time remain unchanged.

The changing baseline motivates that indices are not comparable unless they result from the same search and, therefore, have the same reference point. Some studies have addressed this issue by applying transformations to series values, as in Siliverstovs and Wochner (2018), which compare different search queries by transforming them into ratios and applying logarithms so that the resulting series reflects the same baseline in their index.

In summary, this study uses averages to analyze potential tourism trends based on the Google Trends index . It uses the monthly average to transform the frequency for comparison with other indicators. Additionally, this paper compares the search of the simple name of a tourist destination ("Ecuador") with average interest for the tourist

destination in conjunction with tourism-related keywords suggested by the literature (the average searches for "Ecuador flights" and "Ecuador hotels," for example).

3.2. Tourism Demand

The strategy for identifying keywords was divided in two stages. First, potential demand is characterized at an aggregate level by selecting a set of keywords or phrases related to tourism. Selected words are commonly used in searches related to tourism or travel intentions and are combined with the specific tourist destination of interest. ChatGPT and Google Bard language models were used to list these terms. Searches were conducted in both Spanish and English to cover a broader spectrum of users.[10] This methodology allows for an approximation of tourism demand using search terms highly related to the theme.

In the second stage, demand for tourist products in each country is characterized. The process begins by defining new keywords and specific tourist destinations or products of interest in each country.[11] To determine touristic destinations in each country, the study further utilizes information from online tourism sites, official national sources from each country, and feedback from local tourism experts. Preliminary tests were conducted in both stages to filter out words that best reflect the expected dynamics of tourism demand.[12]

Google Trends' series are used in two ways. First, time series are constructed to understand the general tourism dynamics by country. Second, searches are focused on specific tourist products within each country in the region, showing specific components of tourism demand. In both phases, search terms were constructed by combining a key phrase or concept (hereafter, *keyword*) and a tourist destination (hereafter, *destination*). Additionally, the country of origin is left undefined in both stages, meaning that results reflect global interest or "demand" for the selected destinations. The process of identifying search terms for each stage and the results of the exercise are detailed below.

[10] To improve the accuracy of the index, searches were performed with all possible combinations of words, considering only lowercase letters, without special characters, and variations of plural and singular forms of words. For example, "travel" is the keyword, and the resulting search is "travel colombia" or " trips colombia". Preliminary tests were used to refine the keywords that best reflected expected tourism demand dynamics, such as drops in tourism searches during the COVID-19 pandemic.

[11] Language models were used, as in the previous stage. Some examples of searches were keywords describing "How to get to [destination]" or "Things to do in [destination]" or simply "Hotels in [destination]" or "Tourism in [destination]"; for example: "things to do in bogota", "cuzco tourism" or "how to get to the salar de uyuni".

[12] That is, with series that show consistent patterns. For example, it was expected that the series would show drops in tourism searches during the pandemic.

3.3. Types of tourism

This paper generates novel series of tourism products capturing specific types of tourism in each country. The attractions or tourist products of countries are characterized into six types: leisure (beach vacations, resorts, shopping), adventure (mountaineering, camping, extreme sports), culture (museums, monuments, festivals), ecotourism (natural reserves, national parks), medical (cosmetic surgeries, medical procedures), and religious (pilgrimages, churches, religious sites). A complete list of associated words for each category is available in Table A2 of the Annexes.

An interest index is generated using the average interest related to keywords in each category. Series are adjusted in order to change the baseline, as Google Trends data is non-comparable due to their reference point (since, as previously explained, interest data is constructed as an index based on an implicit reference point). To generate this conversion factor, the methodology leverages the fact that Google Trends allows searches for up to five words at the same time, all sharing the same reference point. Therefore, one common word of reference is introduced in all searches to generate a conversion factor (in this case, the word "museums" is the baseline).

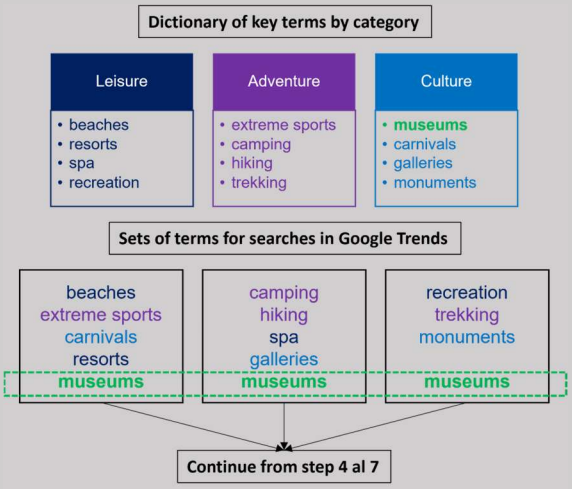
The values for tourism categories are developed in several steps. First, searches are conducted for five words, including the selected reference word. Second, the search process is iterated with a set of four additional words and the selected reference word. Third, the average of the reference word series in each search is computed, generating a conversion factor $\left(\frac{r_1}{r_2}\right)$, where r is the mean of the reference series in the first and second searches, respectively. The result is that the nine keywords can be expressed approximately with respect to the same base index (r_1). Finally, the potential demand for a category is calculated as the mean of all keywords in that category.

To construct the conversion factor, consider three time series corresponding to different keywords k_{tni} where t is the time index, n is an index representing the order of the search, and an index for each keyword i. Each time serie is an index based on the total underlying searches $\tilde{k}_{ti}$ and a normalization factor r_n such that $k_{tni} = \tilde{k}_{ti} r_n$. The normalization factor depends on the periodization and the subset of words selected: two keywords estimated in the same search n have the same normalization factor. To simplify, suppose searches are made in pairs of words (Google Trends allows quintuples). Three search combinations are possible: $\{[k_{tA1}, k_{tA2}],\ [k_{tB2}, k_{tB3}], [k_{tC1}, k_{tC3}]\}$. Consider any two searches: $A = [k_{tA1}, k_{tA2}], B = [k_{tB2}, k_{tB3}]$. Although all three times series can be generated with these two subsets, the data generation process differs for each search as they have different normalization factors. Note that in the two subsets, there is a common element k_{t2} whose conversion factor differs in subset A (r_A) and subset B (r_B), but the underlying value $\tilde{k}_{t2}$ is the same in both searches. This common element is the chosen reference series.

Box 1. Step-by-Step Guide to Generate Comparable Series Using Google Trends (allows calculating demand by type of tourism)

1. Define the components of the category of interest. In this case, the types of tourism: leisure, adventure, cultural, ecotourism, medical, and religious.
2. Create a dictionary of key terms related to each component of the category. Each component will have its respective list of key terms.
   - ✓ **Example:** Key terms related to leisure tourism include beach vacations, resorts, shopping, spa, etc.
3. Define a term from the dictionary in the previous step to function as an anchor. This term serves as a linking element in the next step.
   - ✓ **Note:** It should be a term that is not so popular as to diminish the importance of the other terms but is also not of low interest, preferably in the average interest range.
4. Generate simultaneous queries [to make the series comparable[13]] for five terms[14], always including the anchor term from the previous step. That is, there will always be 4 key terms + the anchor term. This generates a time series for each term.
   - ✓ **Note:** This step should be repeated until series are generated for all key terms in the dictionary from step 2. Each set of searches should not follow a particular order; it is recommended that the key terms (of the 4 to be searched) be random to ensure variability in the interest Identified by Google Trends.
5. Calculate the conversion factor.
6. Transform each series from step 4 using the conversion factor.
   - ✓ **Note:** With this, a set of comparable terms among all terms would be obtained.
7. Group the terms of each category and calculate the average value.
   - ✓ **Note:** The average series for each category would be the approximation to the demand for tourism according to each type.

Figure 5. Step-by-step



Source: Prepared by the authors.

[13] As Google performs index normalization, they are only comparable when searches for key terms are made simultaneously.
[14] This is the limit allowed by Google Trends.

To transform the time series so that subsets A and B have the same normalization factor (to be comparable), a conversion factor is estimated as follows: $v = \frac{\mathbb{E}[k_{ti,n}]}{\mathbb{E}[k_{ti,-n}]}$.[15] In summary, for each query A and B, the following is estimated: $v = \frac{\mathbb{E}[k_{t2A}]}{\mathbb{E}[k_{t2B}]} = \frac{r_A}{r_B}$. Finally, the series in subset B is transformed using the factor: $B \times v = [k_{tB}\ v, k_{tB}\ v] = [k_{tA}\ , k_{tA3}]$.

4. Tourism trends

This section presents the results derived from Google Trends data and the aforementioned strategy. The results are divided in two sections. The former focuses on the potential demand for tourism. The latter section characterizes the potential demand according to different types of tourism. In both cases, an analysis is conducted at both the regional and disaggregated levels for each country in the region.

4.1. Tourism demand

4.1.1. Regional trends

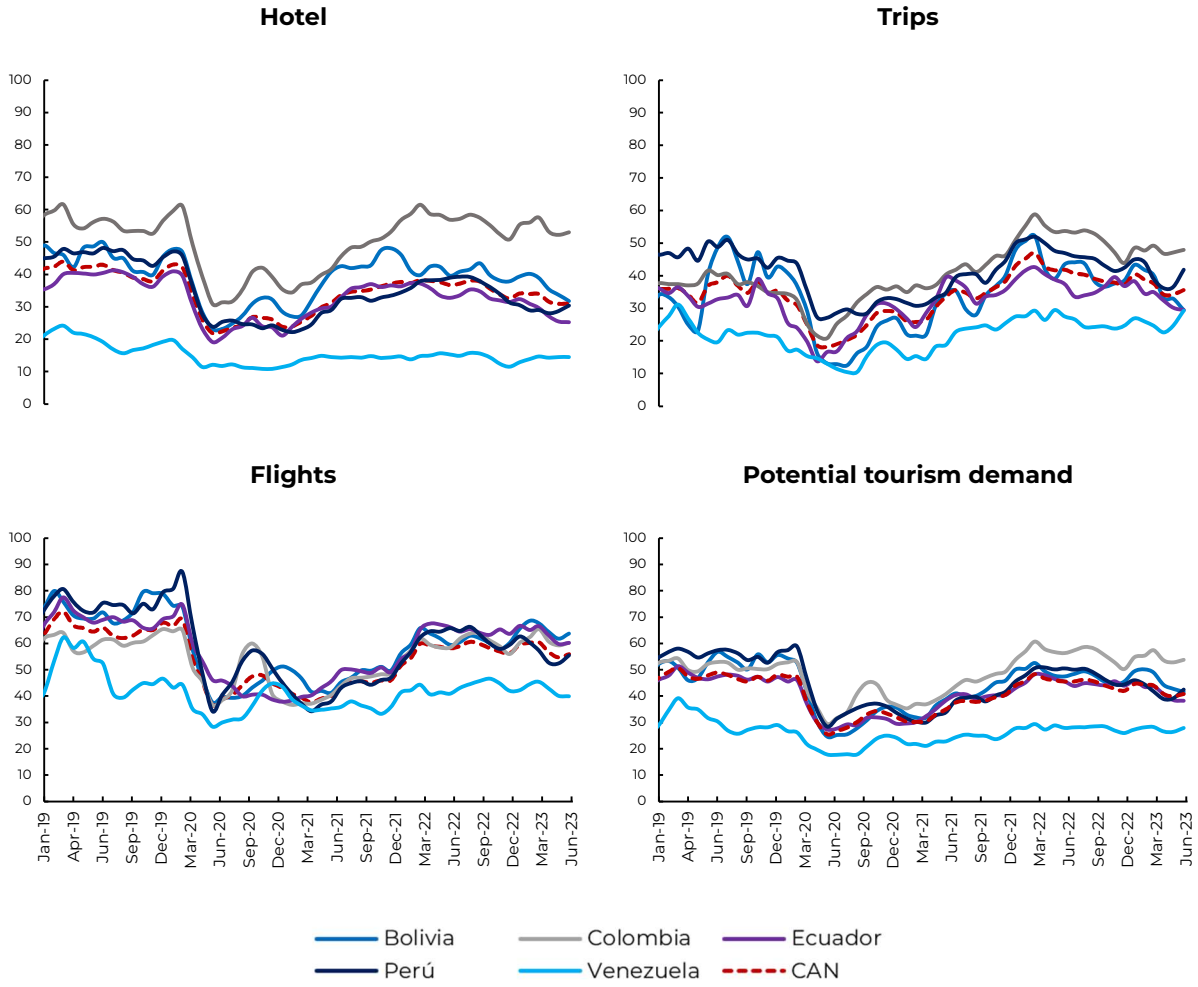
The themes related to tourism demand that best capture the dynamics of the sector are those related to queries for hotels, travel, and flights, as they could represent potential tourism demand.[16] Figure 6 presents the trends in these searches for each of the countries and the Andean Region[17], including the trend or "tourism demand."[18] Of the three themes, hotel and flight searches best exhibit the impact of the COVID-19 pandemic in the first quarter of 2020. As will be seen later, the validity of using data series extracted from Google Trends is reaffirmed by comparing the series with actual demand, expressed in the number of international arrivals. In general, it is observed that these have above 0.6 correlation across most geographies (see Box 2).

The most substantial annual declines in tourism demand were recorded between the end of the second quarter and the beginning of the third quarter of 2020.[19] Notably, in June 2020, the drop in the average of the Andean Region was 46.6%, 53.3% in Bolivia, 49.4% in Peru, 43.7% in Venezuela, 43.3% in Colombia, and 41.1% in Ecuador. A year later,

[15] Although the normalization factor should not vary over time, in practice there are differences at the decimal level. Therefore, the mean of the conversion factor reduces this variation.
[16] Studies such as Bangwayo-Skeete and Skeete (2015), Cevik (2020), and Dergiades et al. (2018) have shown this.
[17] This was constructed as the simple average of the indices of the five countries in the region.
[18] Defined as the average of the three themes in each country. For the region's case, once the tourism demand in each country is obtained, that of the region is calculated, again, as the simple average of the five countries.
[19] It refers to the percentage change between June 2019 and June 2020.

tourism demand began to show positive annual growth. However, tourism demand has not recovered homogeneously among the Andean Region.

Figure 6. Trends related to tourism in each country



Source: Prepared by authors.

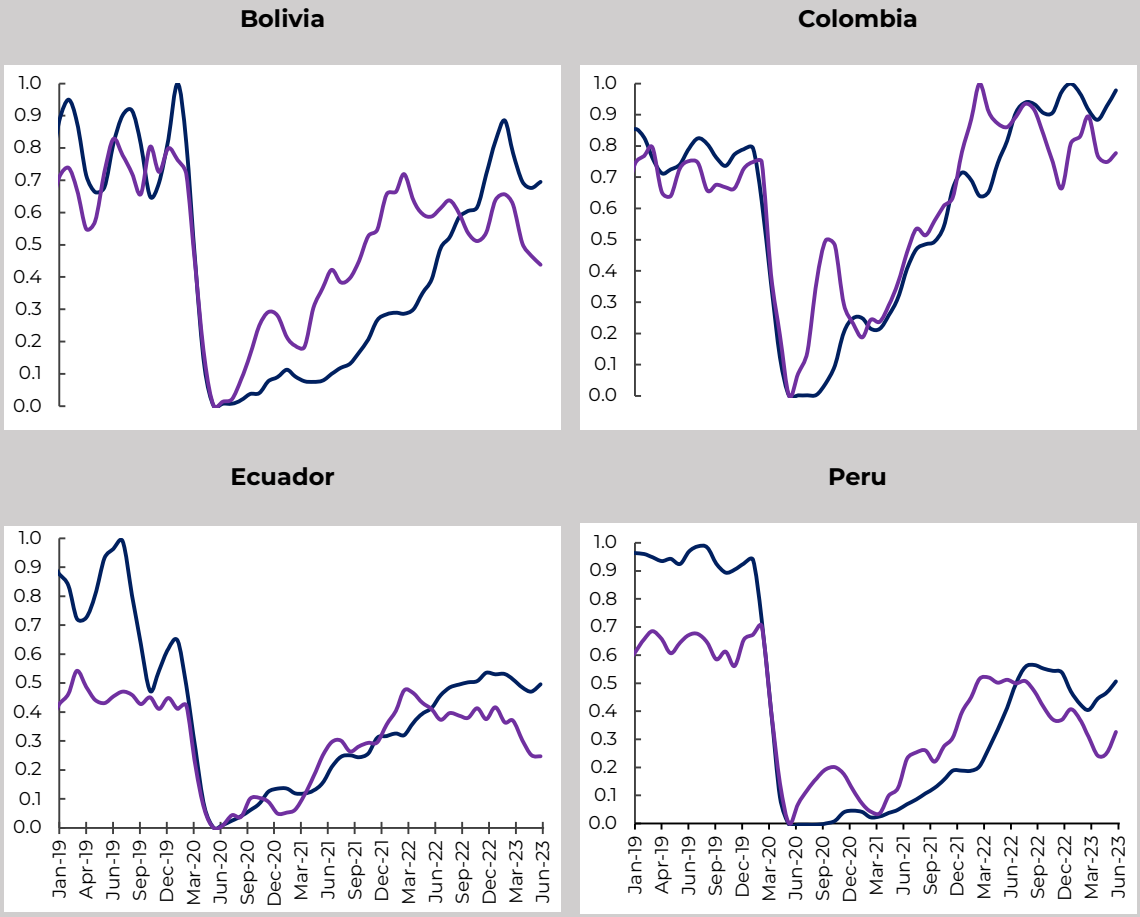
Box 2. Validity and Utility of the Proposed Measure

A comparison between potential and actual tourism demand was conducted to validate the accuracy of the potential tourism data. International arrivals data for each country was used to evaluate effective tourism in the region.[20] The potential tourism measure derived from

[20] Taken from official sources of each country: Bolivia ("Estadísticas de Viajeros Internacionales", Instituto Nacional de Estadística; database only considers entry of tourists residing abroad); Colombia (Aerocivil; database 'Origen-Destino' only considering passengers that will enter from abroad); Ecuador ("Entradas y Salidas de Visitantes al Ecuador", Ministerio de Turismo; database only considers entries of passengers residing abroad), and Peru ("Llegada de Turistas Internacionales", Ministerio de Comercio Exterior y Turismo; database only considers entries of international tourists,

Google Trends instead captures real-time data, serving as a lead indicator of tourism trends. The data is normalized using the Min-Max method to allow for comparisons.[21] The results highlight a strong correlation between these two data series (see Figures 7 and 8), emphasizing the utility of the measure based on Google Trends as an effective tool for anticipating and understanding tourism trends in the Andean Region. This correlation provides decision-makers with the ability to implement strategies based on up-to-date data, thereby enhancing the management and promotion of tourism in the region.

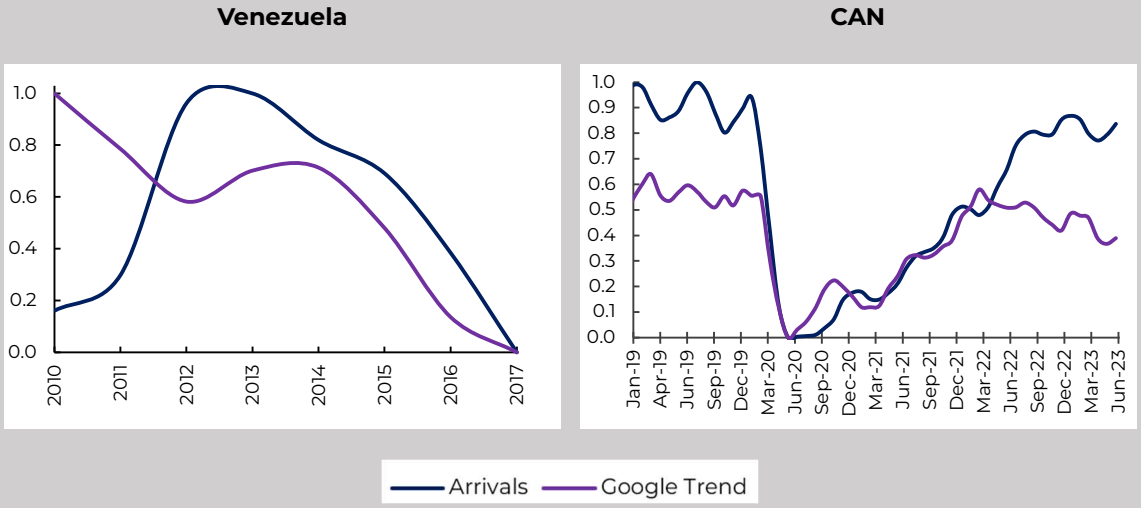
Figure 7. International arrivals vs. potential demand[22]



which includes foreigners and Peruvian tourists residing abroad). In the case of Venezuela, because it was not possible to have monthly data on international arrivals, the exercise is done with annual data taken from the World Bank available here (unfortunately only available until 2017).

[21] This normalization method transforms both series to a common scale, which facilitates comparison between patterns and trends in the data. Normalized series are rescaled in a range between 0 and 1. This eliminates differences in the units of measurement and the original magnitudes of the series, enabling a meaningful and accurate comparison of their relative behaviors. In addition, Min-Max normalization preserves the proportional relationships between the values of the series, which is essential for assessing their correlation and obtaining relevant insights in data and trend analysis.

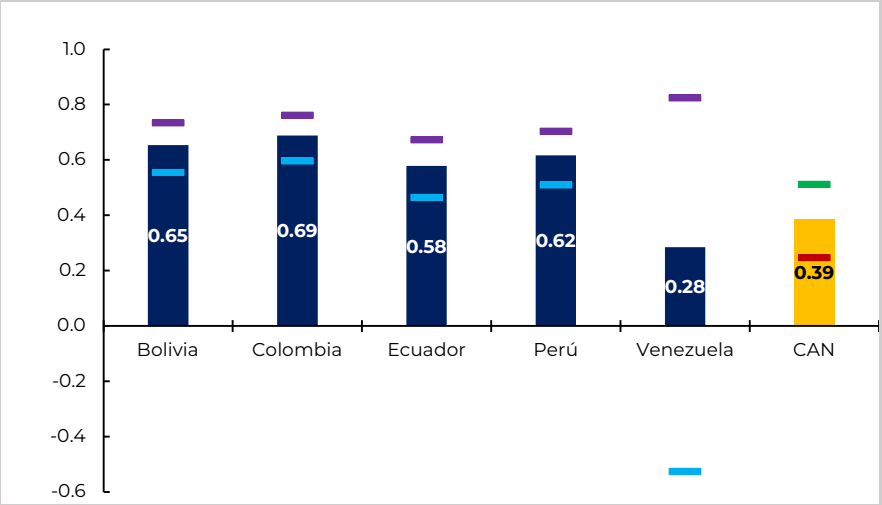
[22] This refers to the index resulting from searches related to tourism built from Google Trends.



Source: Prepared by authors.

As observed in Figures 6 and 7, among the five countries in the Andean Region, the tool performs with better accuracy for Colombia and Bolivia. The two series (arrivals and Google Trends) have correlations of 0.69 and 0.65, respectively, followed by Peru and Ecuador (0.62 and 0.58). Particularly, the series exhibit its best alignment during the COVID-19 crisis period, providing the potential tourism demand measure with high value in the absence of data during such episodes. However, for Venezuela, the measure does not seem to perform well. This may be associated with idiosyncratic issues in the data from and available for the country. This validation result highlights the relevance of using search trends in policymaking, especially due to their high-frequency and timely publication.

Figure 8. Correlation between Arrivals and Potential Tourism Demand



Note: Lines indicate the 95% confidence Interval of the correlation coefficient.
Source: Prepared by the authors.

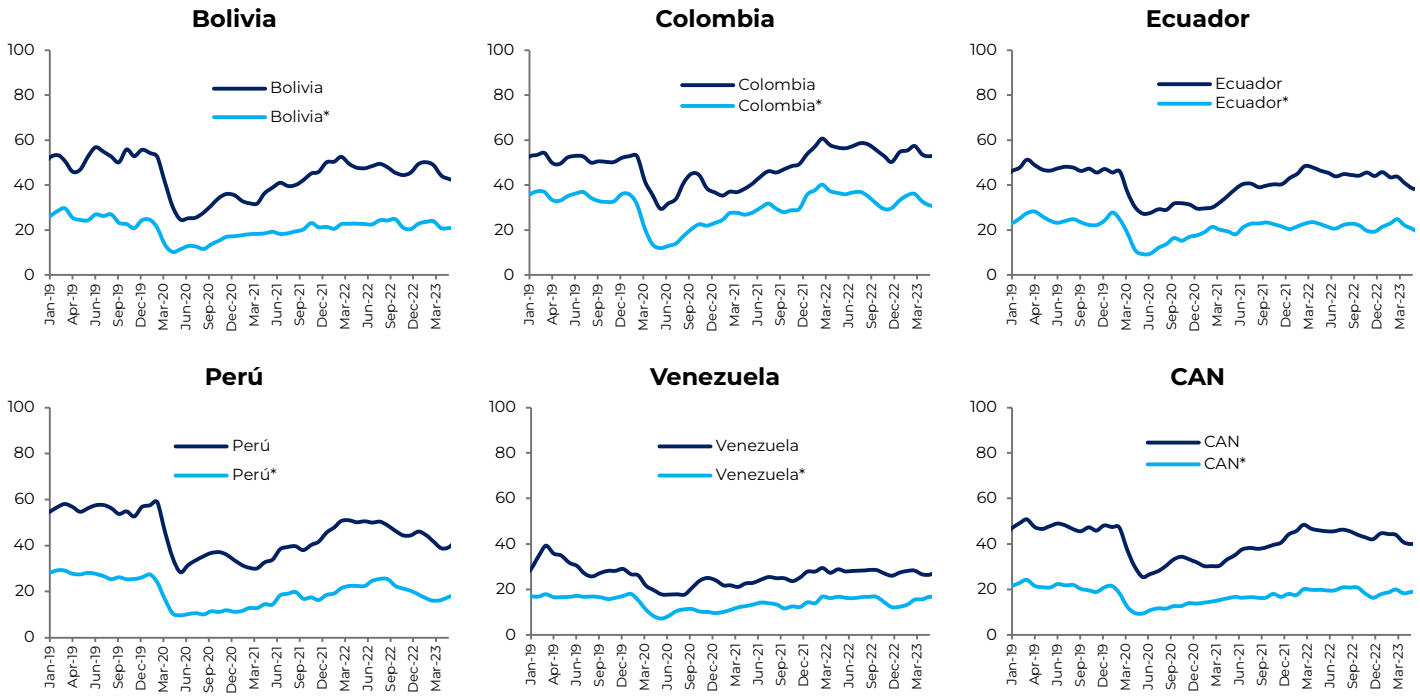
Until June 2023, Colombia and Venezuela had recovered their demand levels and even surpassed pre-pandemic records. Since January 2022, they recorded more than ten consecutive months with average growth rates of 8.2% and 7.5% compared to March

2020, respectively[23]. Between January 2021 and June 2023, on average, Venezuela is 6.2% above the pre-pandemic level, and Colombia is 5.4%, while Ecuador is 4.2% below, Bolivia 9.5%, and Peru 21.3%.

4.1.2. National trends

The aggregated demand series at the country level and the demand for tourist products are compared in this section, both for each country and for the Andean Region as a whole. In Figure 9, both series follow similar trends but differ in the levels of search intensity, where the demand by country shows more search intensity than the product series. This difference may be related to the keywords used. The country series is constructed with terms intended to reflect the generic potential demand per country. On the other hand, the product series uses terms that capture specific touristic destinations and this analysis do not exhaustively represent the existing supply of destinations per country.

Figure 9. Tourism demand by country and in the CAN region



Note: Dark blue lines are series obtained after searching the country's name as the keyword. *Sky blue lines are series obtained from the aggregation of results for all tourism products in each country.
Source: Prepared by authors.

[23] March 2020 is used as the reference month for the pre-pandemic level.

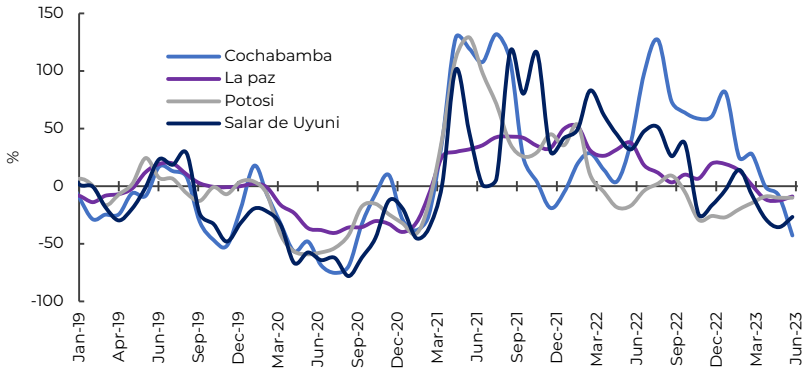
The impact of the pandemic is clearly observed in both series for Bolivia, Colombia, Ecuador, and Peru, showing declines from March and April 2020 and recovering from May and June 2021 onwards. Venezuela does not exhibit a notable decrease in tourism demand during the pandemic, although the series shows a slight decline. The pattern observed for Venezuela may be attributed to a low search intensity in pre-pandemic years, such that the pandemic had a limited impact in the already low tourism demand which has also been unable to recover.

a. Bolivia

Between 2010 and 2019, Bolivia experienced relatively stable tourism demand, but this trend shifted after COVID-19. From March 2020, tourism began to register substantial declines, reaching a year-to-year decline of 55.6% in July 2020. Despite the gradual economic recovery, tourism demand is still lagging. In the first quarter of 2022, demand levels similar to the pre-pandemic period were reached in the potential demand indicator. However, demand decreased again and has yet to rebound.[24] In fact, in the second quarter of 2023, demand started to decline again at an average annual rate of 10%.

Figures 10 and A1 depict the annual change in demand and tourism demand for four of the country's most well-known tourist destinations, including La Paz, Cochabamba, Salar de Uyuni, and Potosí. These destinations experienced an average annual decline of 50.6% in the first six months following the COVID-19 pandemic in March 2020. It was not until April 2021 that the first growth in tourism demand was recorded in these destinations.

Figure 10. Tourism demand – main destination in Bolivia (% annual variation)



Source: Prepared by authors.

[24] This may also represent a limitation of this paper. Alternatively, it may capture a structural change in search patterns for these keywords since the pandemic.

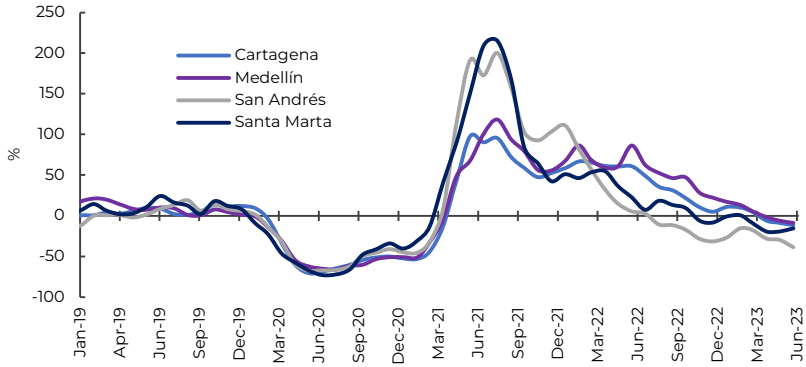
Following the recovery of the tourism sector, the four destinations managed to regain pre-pandemic levels. However, this recovery has not been equal among the destinations. Since the beginning of 2022, tourism demand in La Paz has recorded over 17 consecutive months of positive growth rates compared to March 2020, even surpassing pre-pandemic levels. On the other hand, Cochabamba exhibits a declining trend before the impact of COVID-19, but touristic demand stabilized after. The same applies to Salar de Uyuni and Potosí.

b. Colombia

Between 2010 and 2019, Colombia saw relatively stable tourism demand. The trend changed after the COVID-19 pandemic. From March 2020, tourism demand faced frequent annual declines, with a 43.6% drop in June 2020. The sector was part of a reactivation and economic recovery plan which managed to reach pre-pandemic levels in January 2022. The recovery has been such that potential tourism surpassed pre-pandemic levels. Nevertheless, in 2023, the tourism sector also reflected the slowdown experienced by the rest of the Colombian economy, with an average annual decline of 11% in the first quarter.

Figures 11 and A2 present the annual change in demand and tourism demand for four of the country's most well-known tourist destinations: Cartagena, Medellín, San Andrés, and Santa Marta. All destinations faced the similar challenges during the pandemic; between the second and third quarters of 2020, they experienced annual decreases above 70% in tourism demand. Although the recovery in the four destinations is parallel to the overall country's demand recovery, it is not symmetrical among the destinations.

Figure 11. Tourism demand – main destination in Colombia (% annual variation)



Source: Prepared by authors.

Cartagena and San Andrés have not yet recovered pre-pandemic demand levels, while Medellín and Santa Marta have. Cartagena's slow recovery may be related to two factors: citizen security and the change in tourism type post-COVID-19, resulting in

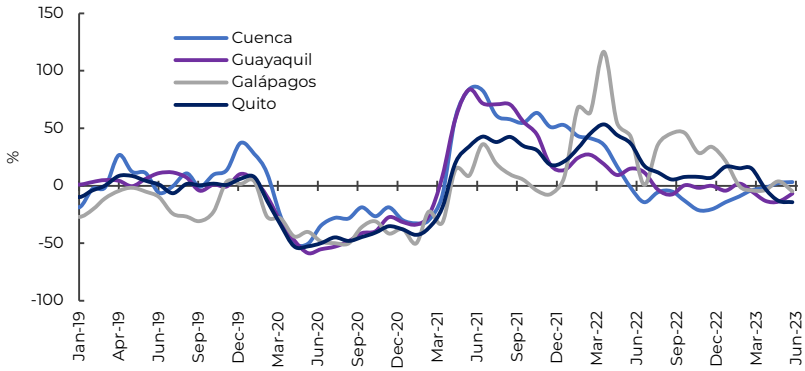
tourist flows towards different destinations. This explanation also applies to other traditional destinations in the country. In addition, San Andrés experienced Hurricane Iota in November 2020, which left damages on the island that have not yet been repaired.

c. Ecuador

Between 2010 and 2019, Ecuador had generally stable tourism demand. In 2013, however, the country shows an average annual increase of 10.3% compared to the period's overall average of -3.1%. With the impact of the COVID-19 pandemic, demand significantly dropped from March 2020, and around May 2021, demand began to recover. During the pandemic, the largest drop in tourism demand occurred in July 2020, with a -41.6% variation compared to the same month in 2019. Despite the marked recovery in the second quarter of 2021, demand gradually declined to reach an average decrease of -16% in the second quarter of 2023, possibly motivated by the country's political and social climate. Nevertheless, post-pandemic recovery has returned to pre-pandemic levels, as illustrated in Figures 12 and A3.

Figures 12 and A3 also show the annual variation in tourism demand for four representative destinations in Ecuador: Cuenca, Guayaquil, Quito, and the Galápagos Islands. Demand for these destinations exhibits similar patterns during the COVID-19 pandemic period, with a significant drop from March 2020 and a strong recovery from March 2021, consistent with country-level demand patterns. The most significant drop in demand for these destinations in 2020 compared to 2021 was nearly -50%, reported in June 2020.

Figure 12. Tourism demand – main destination in Ecuador (% annual variation)



Source: Prepared by authors.

The recovery followed relatively similar patterns for Quito, Guayaquil, and Cuenca. Meanwhile, potential tourism in the Galápagos Islands slowly recovered from March 2021, and only in early 2022 did demand for this destination rebound. In April 2022, the

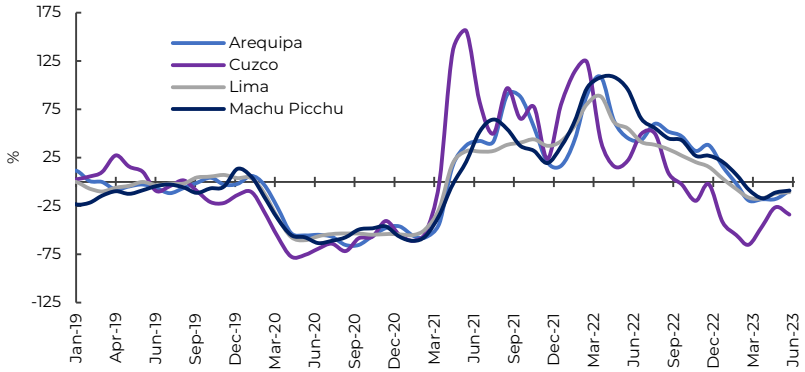
demand for Galápagos Islands showed an increase of 116.5% compared to the same month in 2021.

Even when these destinations surpassed pre-pandemic levels since the second quarter of 2021, the increase has slowly declined in recent months to the same levels as in 2019. It is worth mentioning that the tourist product that showed the best results during the second half of 2022 was the Galápagos Islands; however, it has experienced a significant drop in demand in 2023, averaging 2.1% compared to the same months from January to June in 2022, which saw a 58.7% increase.

d. Peru

Between 2010 and 2019, Peru had a stable potential tourism demand. During the COVID-19 pandemic, demand declined by -19.1% in April 2020 compared to 2021. Subsequently, tourism demand continued to fall at an average year-to-year change of -38.0% from April 2020 to April 2021. Since June 2021, there has been an important recovery in demand, with an annual increase of 19.3% compared to 2020. The recovery was sustained until the end of 2022, and from 2023 onwards, tourism demand declined again.

Figure 13. Tourism demand – main destination in Perú (% annual variation)



Source: Prepared by authors.

Figures 13 and A4 show the annual variation in tourism demand for four destinations in Peru: Arequipa, Cuzco, Lima, and Machu Picchu. Demand for these destinations exhibits parallel patterns, behaving similarly to country-level demand. As observed in Figures 13 and A4, the pandemic negatively impacted demand for these destinations, marked from April 2020 and lasting until May 2021. Cuzco experienced the most pronounced decline during the early months of the pandemic, reaching an annual change of -77.3% compared to 2019 by May 2020. Additionally, as shown in Figure A4, search intensity for this destination is relatively lower compared to Arequipa, Lima, and Machu Picchu.

The demand for these destinations clearly reflects post-pandemic recovery. From May 2021, demand for these tourist products grew significantly, with Cuzco experiencing the strongest recovery. Demand recovery remained stable until the end of 2022 and began to decline in 2023. In the second quarter of 2023, the average variation in demand for Arequipa, Lima, and Machu Picchu was around -13.50%, while for Cuzco, it was -35.2%, both figures compared to the same period in 2022.

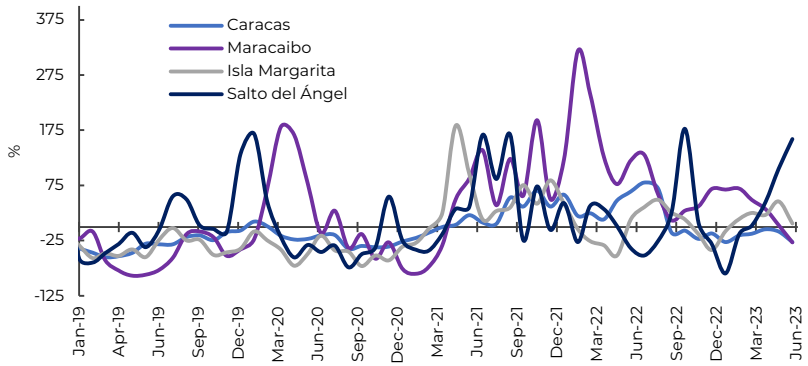
e. Venezuela

Between 2010 and 2019, the potential tourism demand for Venezuela remained relatively stable, albeit with a gradual decline over the period. For 2020, during the pandemic, there was a slight decrease in search intensity for this country. The most pronounced drop in demand occurred in June 2020, reaching -43.8% with respect to the same month in 2019. On average, during 2020, tourism demand fell by -28.2% compared to 2019. Recovery began in May 2021, and the levels have remained close to pre-pandemic demand from mid-2021 to mid-2023. However, search intensity for tourism in Venezuela was comparatively lower than in the rest of the region, even before the pandemic. This fact is evident in Figure 6, where the decline in demand during the pandemic months is less severe for Venezuela compared to the rest of the countries in the region, and search intensity in 2019 was also lower.

To understand national tourism demand, four representative products for Venezuela were considered: Caracas, Maracaibo, Isla Margarita, and Salto Ángel. As seen in Figures 14 and A5, the annual variation in demand during the pandemic months shows a very slight or no decline. In fact, for Caracas, Maracaibo, and Salto Ángel, the variation in demand in March 2020, the first month of the pandemic, compared to 2021 was positive: 1.8% (Caracas), 66.7% (Maracaibo), and 47.9% (Salto Ángel); only for Isla Margarita was a decrease in demand observed, which, however, was evident since mid-2019. Additionally, contrary to expectations, search intensity for Maracaibo seems to have continued increasing during the pandemic period and dropped significantly from September 2020.

Nevertheless, there has been a recovery in demand for all these products since May 2021, which currently equals pre-pandemic levels. The behavior of the series is consistent with what is observed in the overall demand for Venezuela, where the intensity of tourist searches for Venezuela was previously low. Consequently, the series had comparatively smaller declines during the pandemic, relative to the rest of the countries in the region. For the second quarter of 2023, compared to the same months in 2022, the demand for these products was -27.1% (Caracas), -27.9% (Maracaibo), 6.0% (Isla Margarita), 158.9% (Salto Ángel).

Figure 14. Tourism demand – main destination in Venezuela (% annual variation)



Source: Prepared by authors.

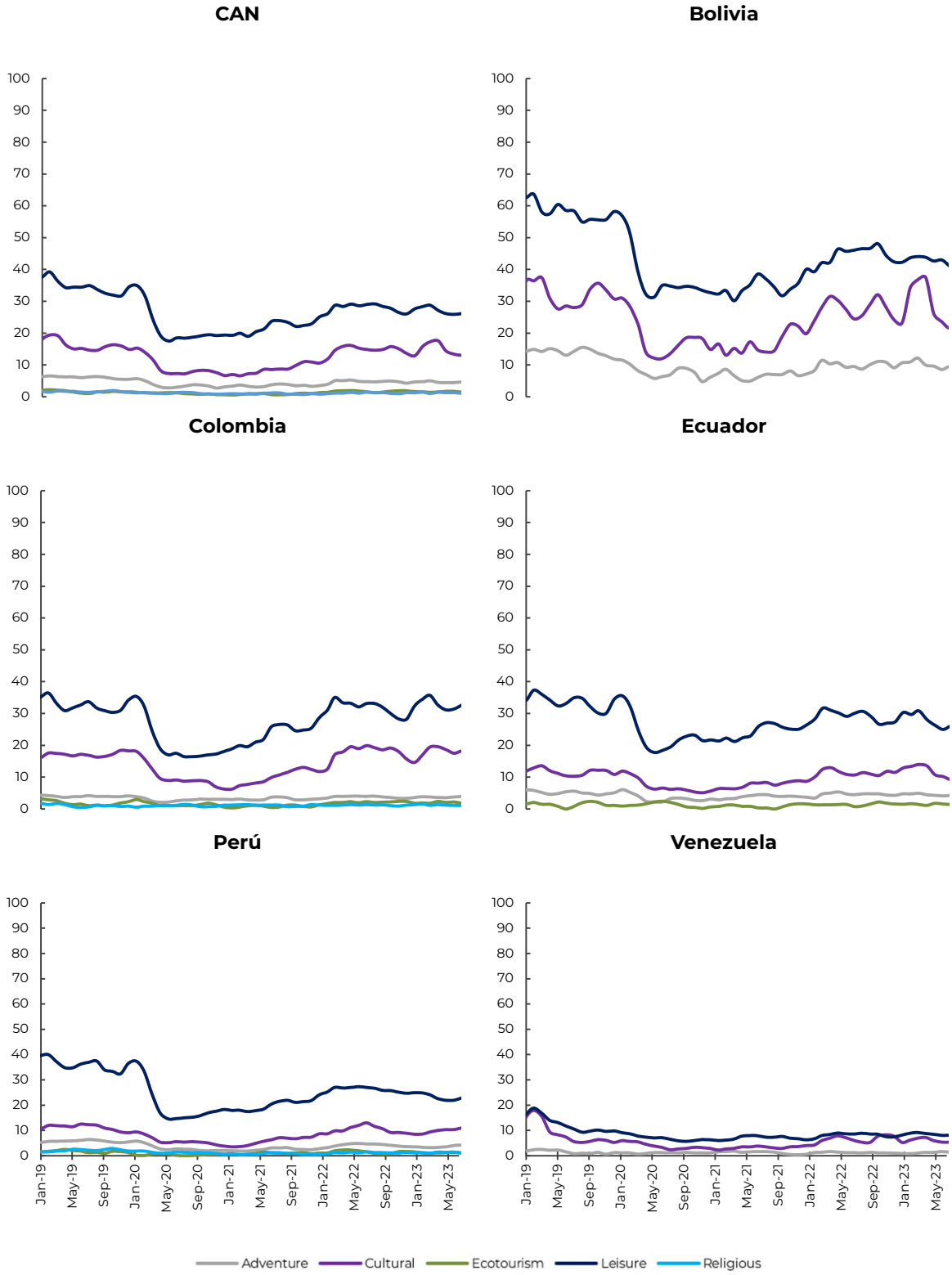
4.2. Type of Tourism

In the rich and diverse Andean Region, it is possible to find a wide range of touristic attractions that thrive due to the geographic, historical, and cultural features of the area. The variability of the Andean landscapes, from towering mountains to dense tropical jungles, allows for the development of nature tourism (or adventure tourism), where tourists can explore hiking trails, observe native fauna, and enjoy exceptional landscapes. Likewise, the region hosts an impressive cultural wealth, giving rise to cultural tourism with archaeological sites, festivals, and unique cuisine attracting travelers interested in the traditions and history of the region. For similar reasons, there are opportunities for religious tourism, ecotourism, and leisure tourism.[25] The latter also responds to the fact that the region offers an advantage in local currency spending compared to foreign currencies, making leisure tourism relatively affordable for foreigners.

According to Figure 15, over the last decade, leisure tourism and cultural tourism have been the most popular types of tourism in the Andean Region. Beaches, hotels, and resorts have become the main tourist attractions in the region, along with historical, cultural, and heritage sites. With the COVID-19 pandemic, both types of tourism saw a reduction in levels starting from March 2020. Despite this decline, in line with the recovery of the region's tourism sector, demand for these two types of tourism is also experiencing a slow but steady recovery, although not reaching pre-pandemic levels.

[25] Eco-tourism is a relatively new pattern responding to changes in the dynamics of tourism, and idiosyncratic changes in the meaning of terms. Eco-tourism is closely associated with nature and adventure tourism.

Figure 15. Demand by tourism type



Source: Prepared by authors.

It is worth noting that leisure and cultural tourism are predominant in all series. Only in Colombia and Peru there is potential demand for all five categories. Specifically, in Bolivia and Venezuela, there is no evidence of demand for religious and ecotourism, while in Ecuador, there is no demand for religious tourism. At the national level, these tourism segments are still in recovery post-pandemic. In the case of Colombia, cultural and leisure tourism are approaching the levels recorded in 2019, while in Bolivia, Ecuador, Peru, and Venezuela, these categories are stagnant. On the other hand, adventure tourism, which is not the most relevant within the region, does exhibit a particular dynamic: it remained at constant levels over time and across countries, demonstrating resilience to the COVID-19 pandemic.

The results of this analysis reveal an interesting insight into tourism demand trends by type in the Andean Region. Leisure tourism and cultural tourism have been the most popular types of tourism in the last decade, reflecting the region's popularity for its beaches, hotels, resorts, and historical and cultural richness. However, these categories dropped during the COVID-19 pandemic, suggesting vulnerability to unexpected shocks across the region.

Despite the initial decline, there is a slow but steady recovery in the demand for these types of tourism. Additionally, it has been highlighted that adventure tourism shows notable resilience, maintaining consistent levels over time and across countries, even during the pandemic. At the national level, the results vary, with Colombia approaching pre-pandemic levels in cultural and leisure tourism while the other countries are still in the process of recovery. These findings emphasize the importance of diversifying the supply of tourism in the Andean Region and the need to develop sustainable and crisis-resistant recovery strategies while providing valuable insights into tourism demand trends in the region.

5. Conclusion

This study focused on using Google Trends data to approximate tourism demand in the Andean Region. The analysis shows the temporal evolution of Internet searches related to tourism on Google, which can be interpreted as an approximation to users' interest in visiting destinations in this region. General and product-specific trends have demonstrated parallel behavioral patterns, especially aligned with global trends during the analysis period. The trends are also remarkably correlated with traditional indicators for this sector. These results suggest that selected keywords in the analysis are suitable for trend analysis at both the regional and national levels, approximating tourism demand.

The methodological approach provides consistent and useful results. Among the most relevant conclusions regarding the approximation of tourism demand is that the

demand recovered in a timely and heterogeneous manner after experiencing a regional year-to-year decline exceeding 40% in June 2020. A year after, a gradual increase in demand was observed but the recovery has been slow and differs across countries. Even after registering positive growth rates for two years since mid-2021, some countries have failed to reach pre-pandemic levels. The case of Venezuela is particularly noteworthy. Venezuela shows a relatively lower potential demand compared to the rest of the region and greater resilience to declines during the pandemic, suggesting the existence of additional factors influencing the industry.

The comparative result between the potential demand indicator and international arrivals data reveals that real-time data is valuable to the decision-making process of the tourism industry. While international arrivals provide a historical reflection of tourism activity, Google Trends data offers an up-to-date leading indicator and a dynamic glimpse to emerging trends. The strong correlation between both data sources underscores their complementarity and suggests that their combined use provides a competitive advantage in the management and promotion of tourist destinations. These insights allow decision-makers to anticipate changes, adapt strategies, and maximize responsiveness to challenges, thus contributing to the sustainable growth of the tourism industry in the Andean Region.

Finally, the use of Google Trends as a tool to approximate tourism demand in the Andean Region provides authorities, the different stakeholders, and the Inter-American Development Bank (IDB) with real-time data on tourism-related search trends. The approach allows for identifying areas of opportunity and tourist preferences, as well as the adaptation of strategies to promote lesser-known destinations and diversify the tourism offering. It is recommended that authorities monitor the evolution of demand, anticipate changes in tourist preferences, and design public policies that maximize the tourism potential of the region. It is crucial to promote collaboration with the private sector and other stakeholders in the tourism industry. The IDB Group can support initiatives that champion usage of Google Trends data, including training and funding research projects. This tool is powerful for strategic decision-making and the promotion of sustainable growth in the tourism industry in the Andean Region.

References

Abuelafia, E., Andrian, L. G., Beverinotti, J., Castilleja Vargas, L., Diaz, L. M., Gutiérrez Juárez, P., Manzano, O., & Moreno, K. (2023). *Nuevos horizontes de transformación productiva en la Región Andina*. Banco Interamericano de Desarrollo. https://doi.org/10.18235/0004691

Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, *46*, 454-464. https://doi.org/10.1016/j.tourman.2014.07.014

Cevik, S. (2020). Where Should We Go? Internet Searches and Tourist Arrivals. *IMF Working Paper,* (2020/022), 1-15.

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *The Economic Record, 88*, 2-9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Dergiades, T., Mavragani, E., & Pan, B. (2018). Google Trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management*, *66*, 108-120. https://doi.org/10.1016/j.tourman.2017.10.014

Faber, B., & Gaubert, C. (2019). Tourism and economic development: Evidence from Mexico's coastline. *American Economic Review*, *109*(6), 2245-2293.

Fayissa, B., Nsiah, C., & Tadesse, B. (2011). Research Note: Tourism and Economic Growth in Latin American Countries – Further Empirical Evidence. *Tourism Economics, 17*(6), 1365-1373. https://doi.org/10.5367/te.2011.0095

Feng, Y., Li, G., Sun, X., & Li, J. (2019). Forecasting the number of inbound tourists with Google Trends. *Procedia Computer Science*, *162*, 628-633.

Mulder, N. (2020). *The impact of the COVID-19 pandemic on the tourism sector in Latin America and the Caribbean, and options for a sustainable and resilient recovery*. ECLAC Series.

Richardson, R. B. (2010). *The contribution of tourism to economic growth and food security (No. 1093-2016-88021).* http://dx.doi.org/10.22004/ag.econ.97140

Rogerson, C. M. (2012). Strengthening agriculture-tourism linkages in the developing world: Opportunities, barriers and current initiatives. *African Journal of Agricultural Research*, 7(4), 616-623. https://doi.org/10.5897/AJARX11.046

Scarlett, H. G. (2021). *Tourism recovery and the economic impact: A panel assessment. Research in Globalization*, 3, 100044. https://doi.org/10.1016/j.resglo.2021.100044

Siliverstovs, B., & Wochner, D. S. (2018). Google Trends and reality: Do the proportions match?: Appraising the informational value of online search behavior: Evidence from Swiss tourism regions. J*ournal of Economic Behavior & Organization*, *145*, 1-23.

World Bank. (2023a). "*International tourism, number of arrivals*." World Development Indicators, The World Bank Group, data.worldbank.org/indicator/ ST.INT.ARVL.

World Bank. (2023b). "*International tourism, receipts (% of total exports).*" World Development Indicators, The World Bank Group, data.worldbank.org/indicator/ST.INT.RCPT.XP.ZS.

World Tourism Organization (2021). *Yearbook of Tourism Statistics, Data 2015 – 2019*, 2021 Edition, UNWTO, Madrid, DOI: https://doi.org/10.18111/9789284422487

Wu, J., Li, M., Zhao, E., Sun, S., & Wang, S. (2023). Can multi-source heterogeneous data improve the forecasting performance of tourist arrivals amid COVID-19? Mixed-data sampling approach. *Tourism Management, 98*, 2-17. https://doi.org/10.1016/j.tourman.2023.104759

Xiao, Y., Tian, X., & Xiao, M. (2020). Tourism Traffic Demand Prediction Using Google Trends Based on EEMD-DBN. *Engineering, 12*(3), 194-215. https://doi.org/10.4236/eng.2020.123016

Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism management, 46*, 386-397.

Annexes

Table A1. Prompts used for searches in ChatGPT and Bard.

Stage 1: Search of regional trends	
- "List keywords related to travel" - "List keywords related to tourism"	
Stage 2: Search of national trends	
Keywords	*Destinations*
"Which are the most common phrases used to search a tourist destination in order to travel there?"	"List the main tourist destinations in [COUNTRY]"

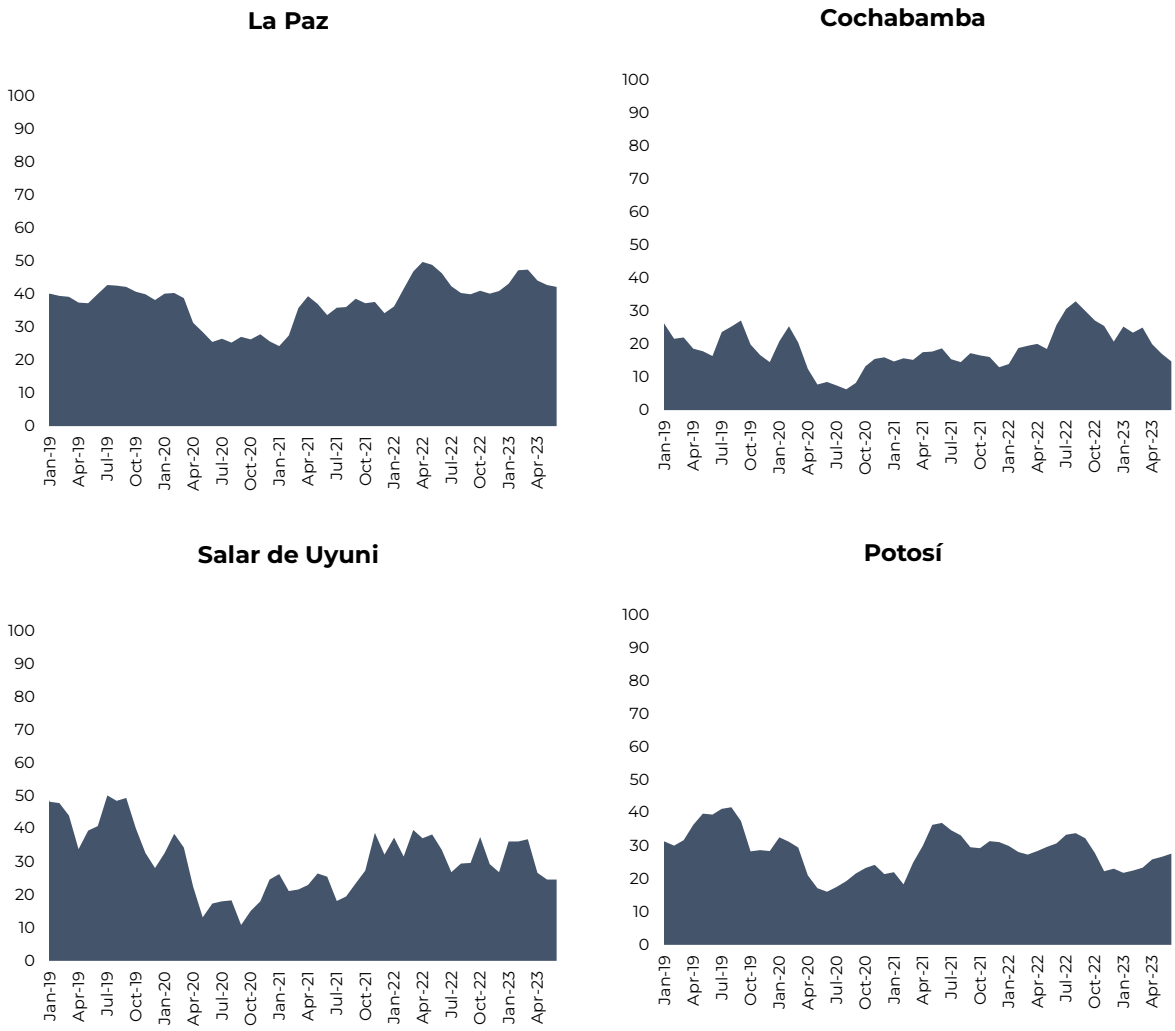
Source: Prepared by the authors.

Table A2. Search terms to identify the type of tourism

Leisure Tourism	**Adventure Tourism**	**Cultural Tourism**	**Ecotourism**	**Religious Tourism**
• seaside vacation	• extreme sports	• historical places	• birdwatching	• pilgrimages
• recreational tourism	• camping	• heritage places	• national parks	• churches
• beaches	• mountain tours	• traditional places	• ecolodges	• spiritual centers
• resorts	• hiking	• cultural festivals	• ecological reserves	• spiritual journeys
• amusement parks	• trekking	• museums	• wildlife tours	• meditation retreats
• pleasure travel	• rock climbing	• festivals	• green tourism	• holy sites
• shopping vacation	• bungee jumping	• carnivals	• nature hikes	• religious festivals
• sunbathing	• zipline	• monuments	• sustainable tourism	• religious site
• sightseeing tours	• rafting	• art galleries	• natural reserves	• sacred site
• relaxing getaways	• paragliding	• traditional dance	• ecofriendly camping	• spiritual workshops

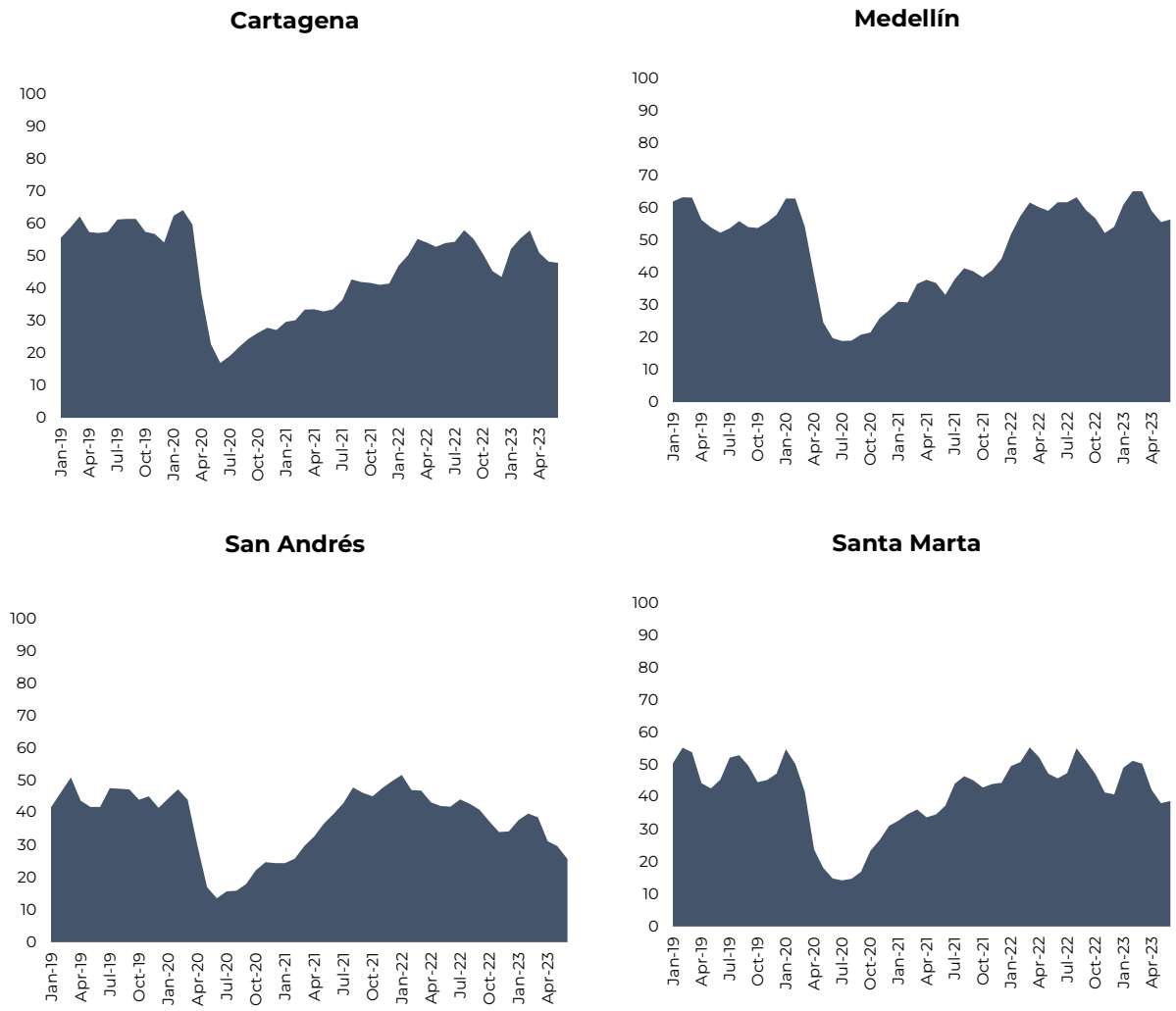
Note: Tourism types were searched in both Spanish and English.
Source: Prepared by the authors using the results of the strategy in Table A1.

Figure A1. Tourism demand in Bolivia's most popular destinations



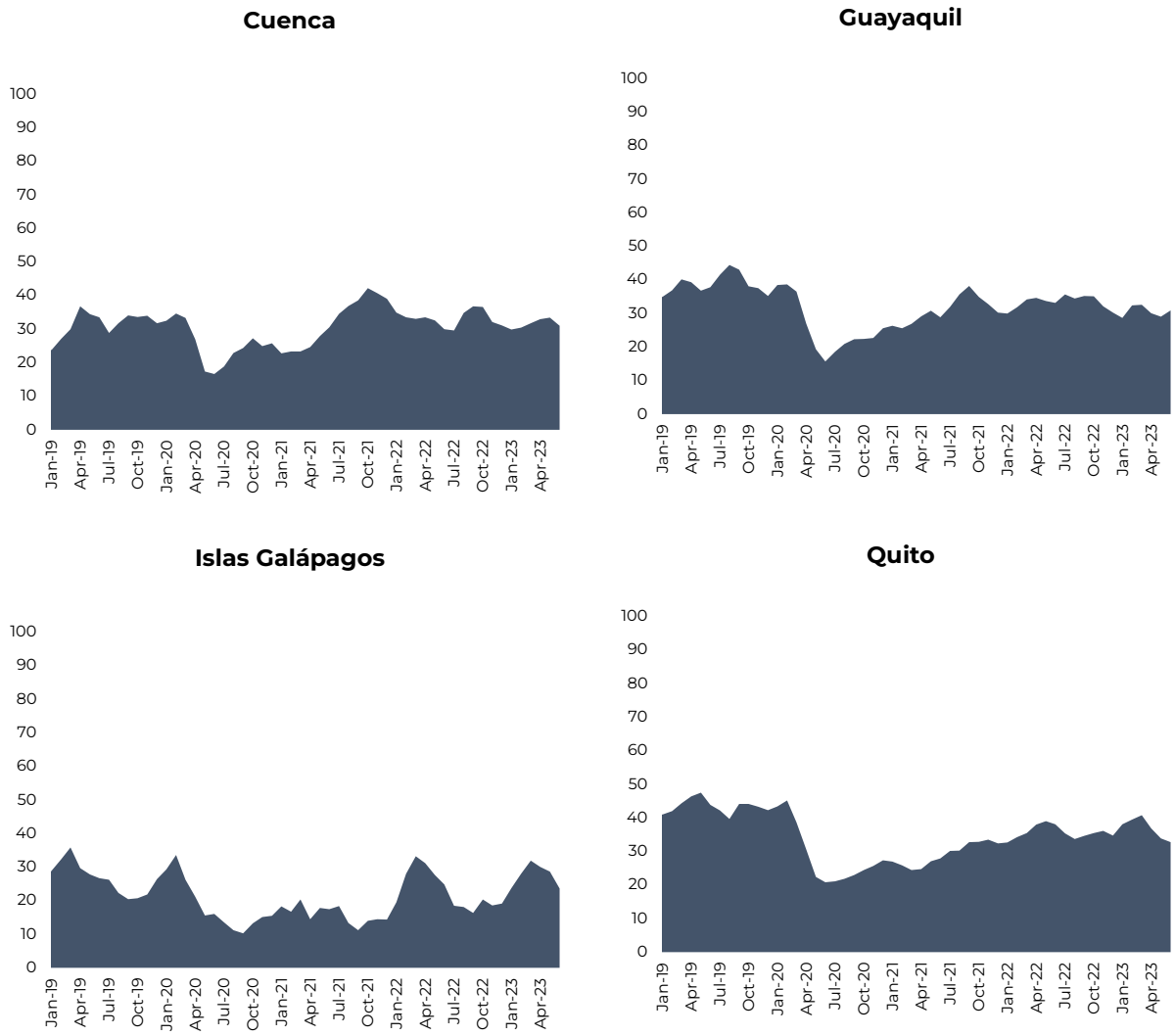
Source: Prepared by the authors

Figure A2. Tourism demand in Colombia's most popular destinations



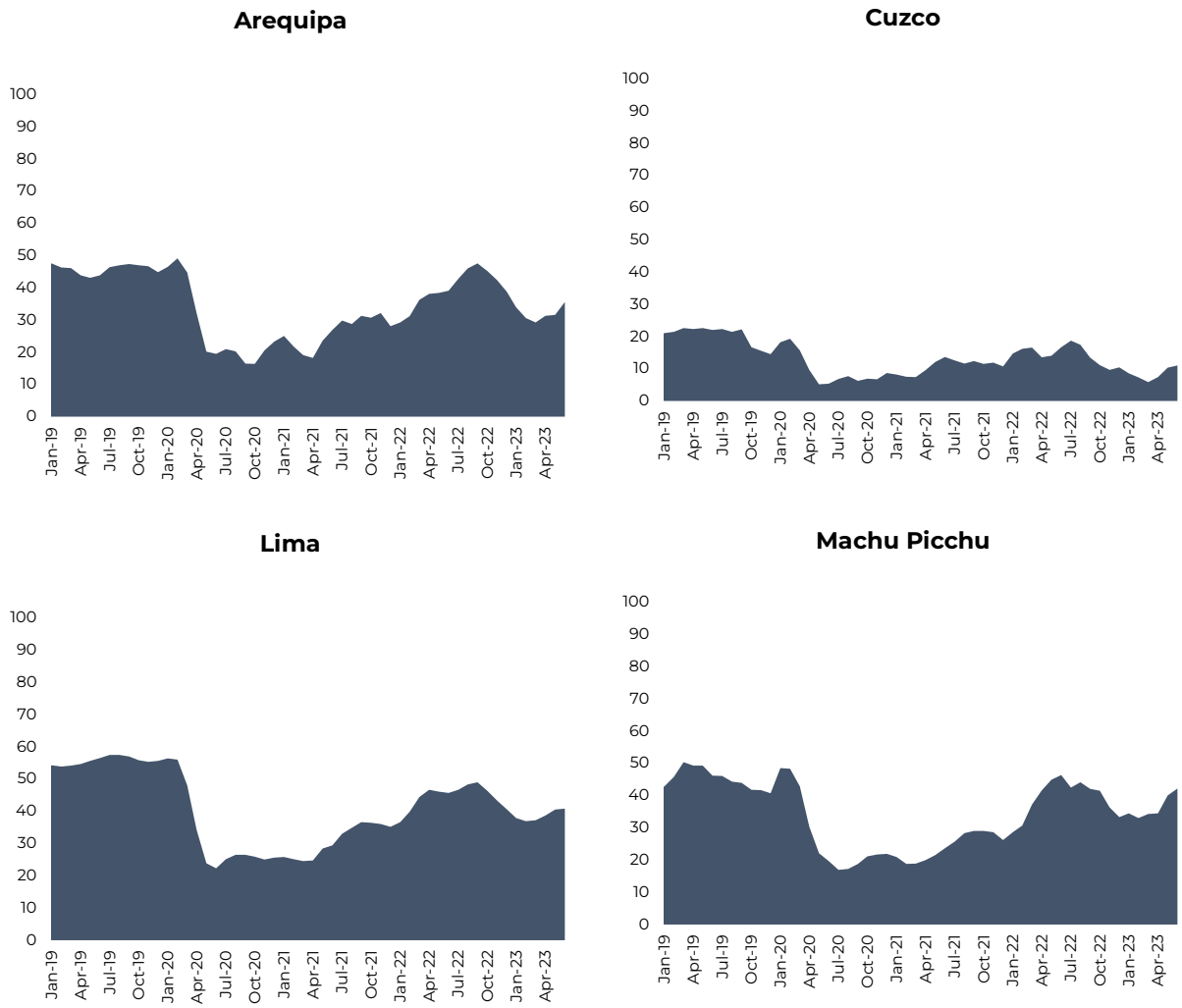
Source: Prepared by the authors

Figure A3. Tourism demand in Ecuador's most popular destinations



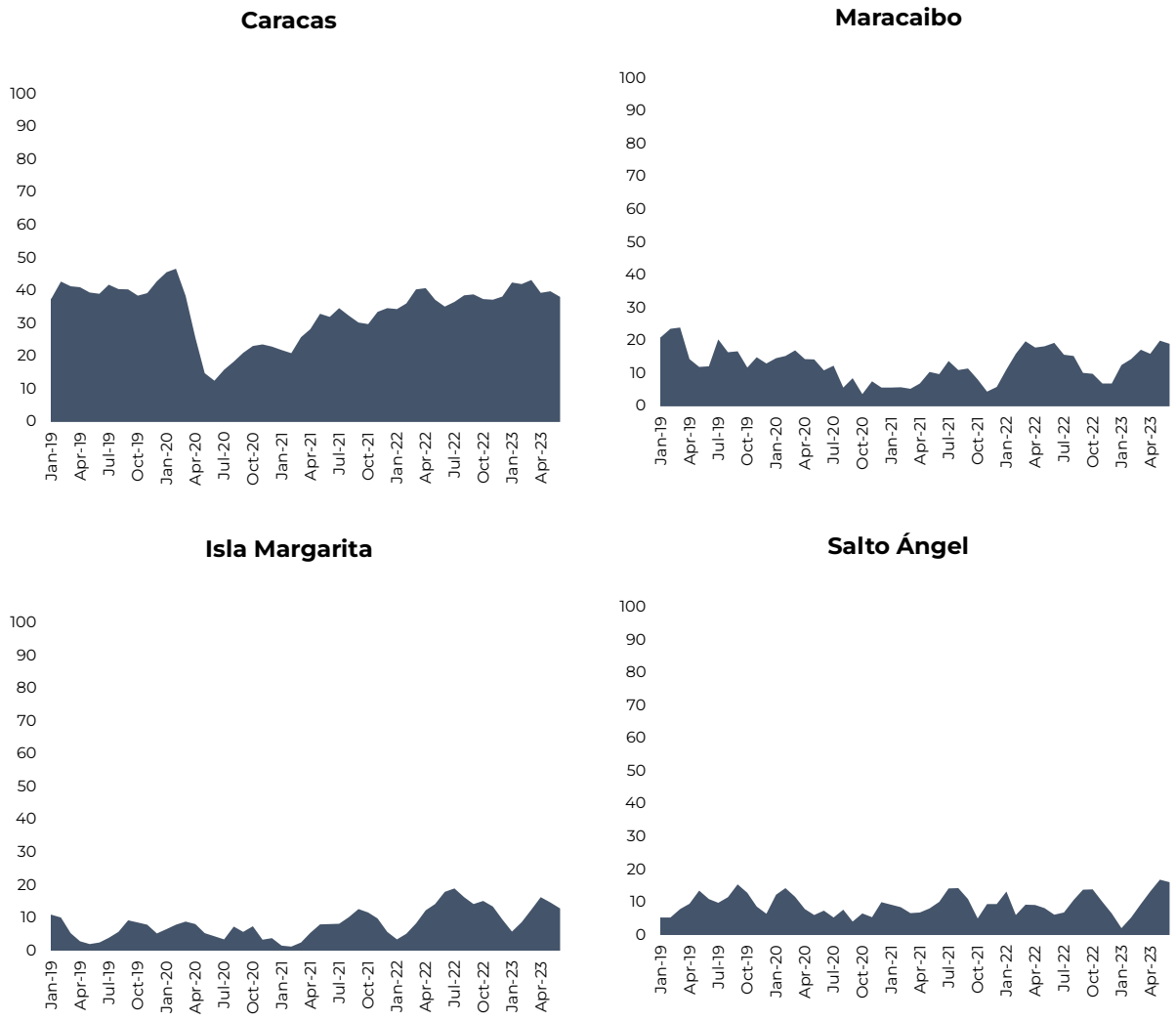
Source: Prepared by the authors

Figure A4. Tourism demand in Peru's most popular destinations



Source: Prepared by the authors

Figure A5. Tourism demand in Venezuela's most popular destinations



Source: Prepared by the authors