

ORIGINAL

Bibliometric analysis of scientific output on retail trade in OpenAlex

Análisis bibliométrico de la producción científica sobre el comercio minorista en OpenAlex

Yanet de la Concepción Avila¹  , Damian Valdés Santiago²  , Danay Milagros Bulnes Mann¹  

¹Universidad de La Habana, Facultad de Turismo. La Habana, Cuba.

²Universidad de La Habana, Facultad de Matemática y Computación. La Habana, Cuba.

Cite as: Avila Y de la C, Santiago DV, Bulnes Mann DM. Bibliometric analysis of scientific output on retail trade in OpenAlex. Data and Metadata. 2025; 4:827. <https://doi.org/10.56294/dm2025827>

Submitted: 14-03-2025

Revised: 20-06-2025

Accepted: 14-10-2025

Published: 15-10-2025

Editor: Dr. Adrián Alejandro Vitón Castillo 

Corresponding Author: Yanet de la Concepción Avila 

ABSTRACT

Objective: to conduct a bibliometric analysis of scientific output published in journals included in the OpenAlex bibliographic catalog between 2020 and 2024 on about the relations between retail and tourism.

Method: a three-stage, five-step procedure was applied for the bibliographic review and bibliometric analysis, using different methods, tools, and programs such as the biblioshiny web application of the R programming language.

Results: it was created, the classification and analysis of the scientific output included in OpenAlex, based on various indicators such as productivity per year, journals, authors, countries, and institutions, grouping of journals according to Bradford's Law, impact of journals according to the H-index, citation analysis by authors, conceptual structure of the research field, as well as its social structure by authors and countries.

Conclusions: there was an initial increase in publications during the pandemic, followed by a gradual decline. Non-commercial literature emerged as a relevant source of knowledge. Colombia positioned itself as a leader in academic production, although with limited international cooperation. Research focused predominantly on economic and health aspects, leaving significant gaps in the theoretical conceptualization of the topic and in the analysis of contemporary trends such as sustainability and digital transformation of the sector.

Keywords: Bibliometric Analysis; Retail; Tourism.

RESUMEN

Objetivo: realizar un análisis bibliométrico de la producción científica publicada en las revistas incluidas en el catálogo bibliográfico OpenAlex entre los años 2020 y 2024, sobre la relación entre el comercio minorista y el turismo.

Método: se aplicó un procedimiento el cual incluye tres etapas y cinco pasos para la revisión bibliográfica y el análisis bibliométrico, en el cual se utilizaron diferentes métodos, herramientas y programas como la aplicación web biblioshiny del lenguaje de programación R.

Resultados: se logró la clasificación y análisis de la producción científica incluida en OpenAlex, a partir de diversos indicadores como productividad por año, revistas, autores, países e instituciones, agrupamiento de las revistas según la Ley de Bradford, impacto de las revistas según el índice H, análisis de citación por autores, estructura conceptual del campo de investigación, así como su estructura social por autores y países.

Conclusiones: se evidenció un incremento inicial de publicaciones durante la pandemia, seguido de una gradual disminución. La literatura no comercial emergió como fuente relevante de conocimiento. Colombia se posicionó como líder en producción académica, aunque con limitada cooperación internacional. La

investigación se centró predominantemente en aspectos económicos y sanitarios, dejando vacíos significativos en la conceptualización teórica del tema y en el análisis de tendencias contemporáneas como la sostenibilidad y transformación digital del sector.

Palabras clave: Análisis Bibliométrico; Comercio Minorista; Turismo.

INTRODUCTION

Retail trade is the activity of selling domestically produced or imported goods at retail or wholesale. In the 1950s and 1960s, stores began to specialize, and it was from there that the retail business evolved on a large scale. Various chains joined together under the concept of franchising, resulting in companies not only selling but also managing their logistics activities.⁽¹⁾

According to Webster, tourism and commerce are two activities that have a very important impact on the development of society. Speaking specifically about the link between tourism and retail, Cárdeno states that the former drives the growth and development of the latter by increasing the influx of potential customers to commercial areas, while commerce provides another asset of great importance to the tourism sector.⁽³⁾

To the authors' knowledge, research on retail trade has been addressed by a small number of academics, resulting in limited previous bibliometric production and hindering its systematic analysis. Given the predominance of qualitative and subjective approaches in the existing literature on retail, the application of quantitative methodologies in scientific production will contribute to a more rigorous examination of the phenomenon, allowing for the identification of historiographical trends, the evolution of theoretical approaches, and correlations between variables.

In this sense, bibliometric analyses have established themselves as a fundamental tool for evaluating the impact and structure of scientific knowledge,⁽⁴⁾ allowing the identification of publication patterns and emerging areas. In the field of tourism, this methodology has proven useful for synthesizing large volumes of literature.^(5,6)

Through a bibliometric study, it is possible to determine the main research topics, the gaps and similarities between the work of scientists and leading researchers, the most influential journals, and the most prolific countries in a given discipline or area of knowledge. This method is extremely useful in the formulation of scientific policies.⁽⁷⁾

This study is justified for several reasons: the need to systematically evaluate scientific production in a key area for tourism management, the importance of identifying research trends in a rapidly evolving sector, and the lack of specific bibliometric reviews on retail trade and its relationship with tourism, despite its strategic relevance.

To this end, this research used OpenAlex as a source of information, which indexes more than 200 million academic documents, including articles, theses, preprints, and patents.^(11,12) It is more inclusive than Scopus and Web of Science, as it covers open-access journals, institutional repositories (such as arXiv, PubMed Central), and gray literature. Unlike Scopus and Web of Science, which require a subscription, OpenAlex is completely free. Its open-data model allows for mass downloads and analysis without license restrictions.⁽¹³⁾

OpenAlex provides structured data in different formats, facilitating the automation and replicability of bibliometric studies. It also integrates with R (bibliometrix), VOSviewer, and other bibliometric analysis software.

With the help of bibliometrics, this study aims to analyze the evolution and behavior of scientific research related to the relationship between retail and tourism published in journals indexed in the OpenAlex database between 2020 and 2024. The study applies a bibliometric approach that combines the analysis of co-authorship, citation, and productivity by country, among others. The analyses were performed using the biblioshiny web application.

METHOD

The qualitative and quantitative assessment was obtained from the procedure developed based on the study of the methodologies proposed by several authors.^(7,16,17,18,19,20,21,22) This procedure consists of three stages and five steps, which are shown in figure 1.

Stage 1: Research design

Step 1.1. Research objective: to analyze the behavior and evolution of scientific publications on the relationship between retail trade and tourism between 2020 and 2024.

Step 1.2. Selection of information sources: to perform the systemic and quantitative analysis of scientific literature published in indexed journals, the OpenAlex database was selected.⁽¹²⁾ OpenAlex indexes more than

200 million academic documents, including articles, theses, preprints, and patents. It covers open access journals, institutional repositories (such as arXiv, PubMed Central), and gray literature.

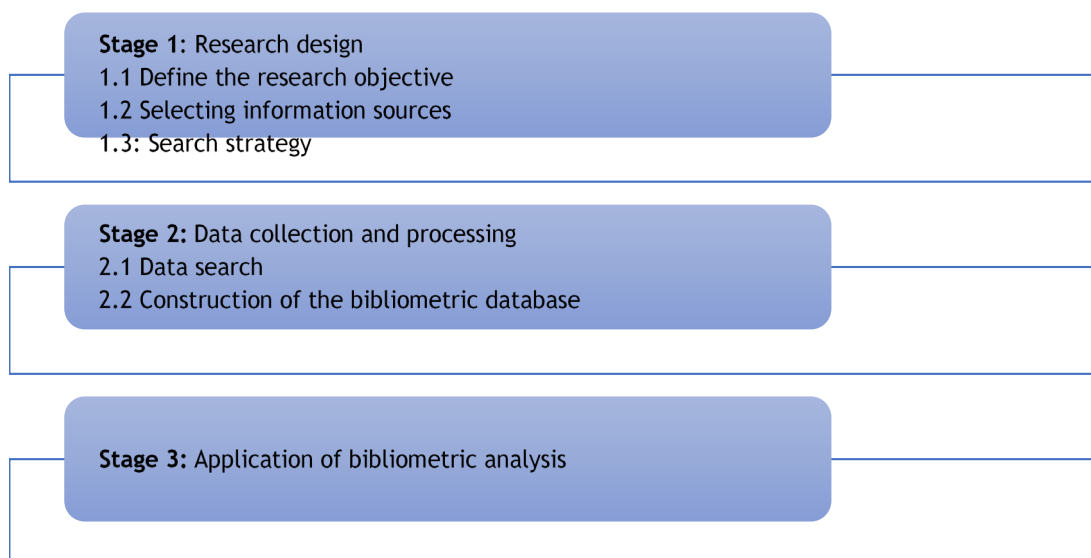


Figure 1. Procedure for bibliometric analysis

Grey (or non-conventional) literature comprises documents of academic, technical, or institutional value that are not commercially distributed, such as theses, reports, preprints, conference proceedings, and government documents. Although access to them is limited and they are not always peer-reviewed, they are relevant sources for research as they contain up-to-date and specialized information. Locating them requires the use of academic repositories, specialized databases, or institutional archives, which highlights their importance in complementing traditional scientific literature.

Step 1.3. Search strategy: the research covers the period from 2020 to 2024, in order to show the evolution and behavior of publications during this period. Literature from several countries and in several languages was examined using the query “retail OR tourism OR merchandising OR marketing OR products.” Scientific articles, theses, books, and technical reports were included as document types. The results were filtered into the subject areas of Business, Innovation, and Economics; Various Aspects of Tourism Research; Urban Planning; and Tourism Studies.

Stage 2: Data processing

2.1. Data search: the bibliometric dataset covers the period 2020-2024, with 2146 documents published in 650 different sources. The types of documents include 1584 articles, 105 books, 252 theses, and 205 reports.

2.2 Construction and processing of the bibliometric database: a CSV file was downloaded from the OpenAlex website and imported into the biblioshiny web application, based on the bibliometrix package of the R programming language. To do this, the R software, its RStudio development environment and the bibliometrix package, along with its dependencies were installed on a personal computer

A database was created with the 2146 publications, including the following information: title of the research, author or authors, year, source or journals, number of citations, and keywords.

Stage 3: Application of bibliometric analysis

After the database was created and processed, bibliometric analysis was applied using the biblioshiny web application, and the following aspects were analyzed and graphed: productivity by year, journals, authors, countries, and institutions; grouping of journals according to Bradford’s Law; impact of journals according to the H-index; citation analysis by authors; and social structure by authors and countries.

Bradford’s law states that if journals are ordered in descending order according to the number of articles they contain on the subject, successive zones of periodicals containing the same number of articles on the subject form a simple geometric series.

Bradford called the first zone the core of journals specifically dedicated to the topic in question. Bradford’s Law can be used to identify the fundamental or core journals of a discipline and focus the analysis on the documents in the fundamental zone.

The H-index (Hirsch index) is the number of articles published by a journal (H) each of which has been cited in other articles at least H times.⁽²⁴⁾

RESULTS AND DISCUSSION

Productivity per year

As shown in figure 2, the largest number of publications are concentrated in 2020, followed by a decline in the number of papers published on tourism product portfolios. This is reflected in the annual growth rate, which was -21,79 %. The documents have an average age of 3,4 years and receive an average of 0,97 citations per document, with a total of 9193 references.

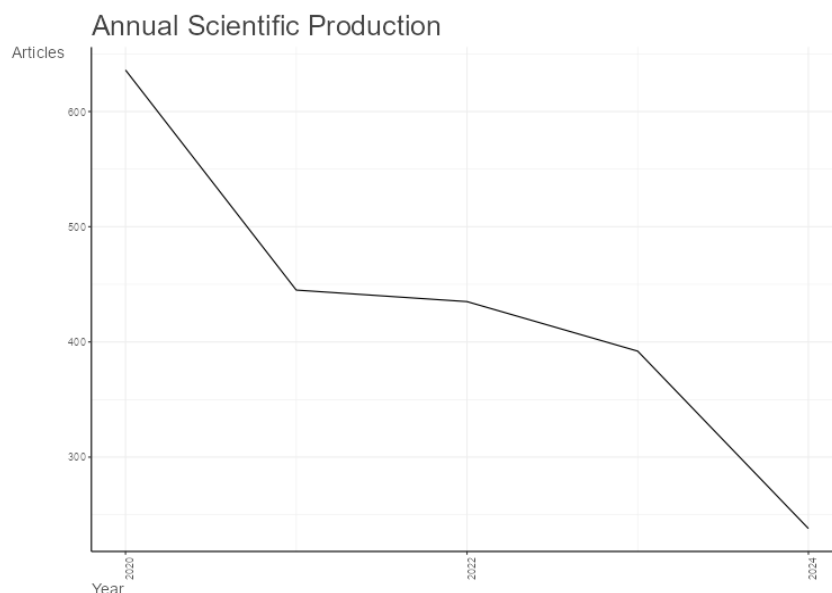


Figure 2. Number of publications per year

Productivity by journal

The scientific output obtained, recorded in OpenAlex, was published in a total of 649 academic journals. Figure 3 shows the journals with the highest contributions. In this section, the journal 593 Digital Publisher CEIT stands out with a total of 64 materials, which represents 2,98 % of the total manuscripts published on the subject. It is followed in descending order by number of publications: Los Angeles Catholic University of Chimbote, Ciencia Latina Multidisciplinary Scientific Journal, Deleted Journal, Vinculatégica Efan, Knowledge Pole: Scientific-Professional Journal, Science Proficiency, Contemporary Dilemmas Political Education and Values, and INNOVA Research Journal. These academic journals account for 9,23 % of the total manuscripts analyzed. There are 614 publications belonging to gray literature, representing 28,61 % of the total materials.

Productivity by authors

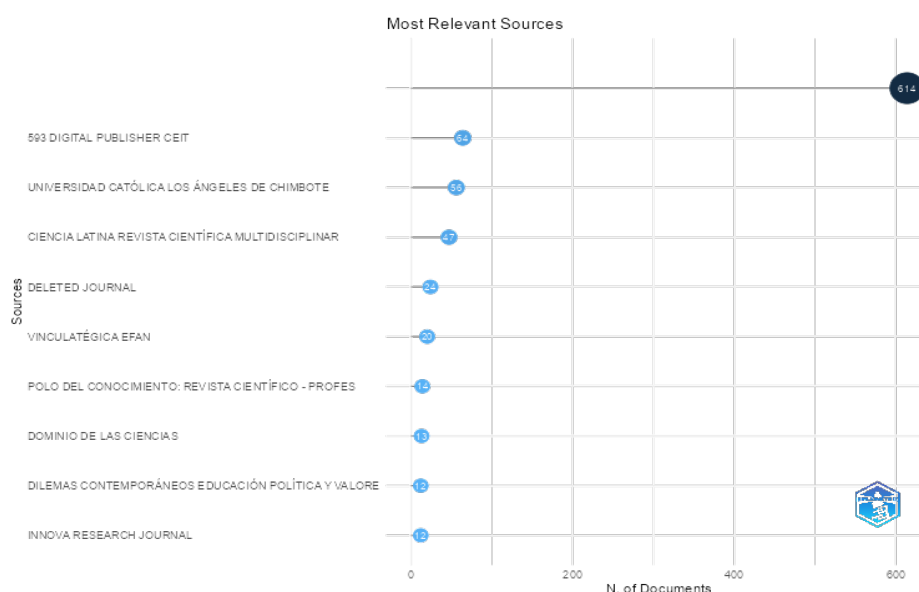


Figure 3. Journals with the highest number of publications

A total of 4708 authors participated, of whom 450 wrote documents alone (500 in total). There are 2,5 co-authors per document and 9,32 % international co-authorships. The authors who stand out most for their productivity among the scientific community in terms of the relationship between trade and tourism in OpenAlex are shown in figure 4. At the top of this list are Dora Alicia Mora-Pérez with 15 materials, Dora Alicia Mora with 14 materials, Kathia Michalczewsky with 9 publications, the Ministry of Foreign Trade and Tourism (MINCETUR) with 8 contributions, Lady Andrea León Serrano, Paolo Robuffo Giordano, and Sandra Paola González Cerón, all three authors with 7 materials each; and Edison Becerra Molina, Francisco Javier Campines Barría, and Francisco Vladimir Benítez Luzuriaga, all with 5 publications. These authors account for 3,82 % of all publications.

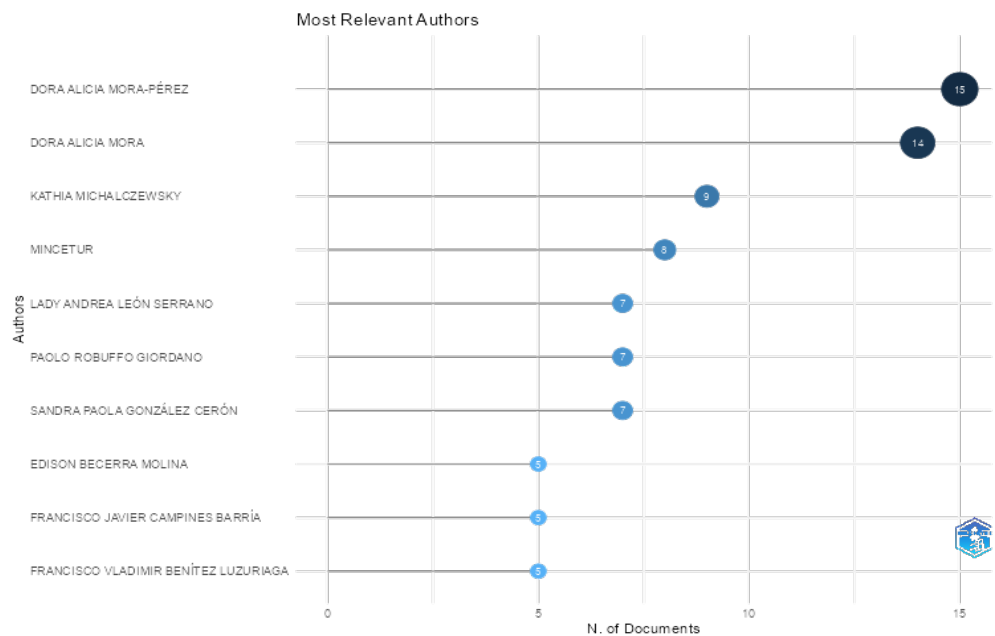


Figure 4. Authors with the highest number of publications

Productivity by institution

The authors come from 1079 institutions. The institutions that stand out most for their productivity among the scientific community in terms of tourism product portfolios in *OpenAlex* are shown in figure 5. At the top of this list are the Technical University of Machala, the Catholic University of Cuenca, the Regional Autonomous University of Los Andes, the State University of Southern Manabí, the Technical University of Manabí, the Bank of the Republic, the University of Guayaquil, and the State Technical University of Quevedo. 2,88 % of publications did not report an institution (NOTREPORTED). The most productive institutions account for 20,32 % of all publications.

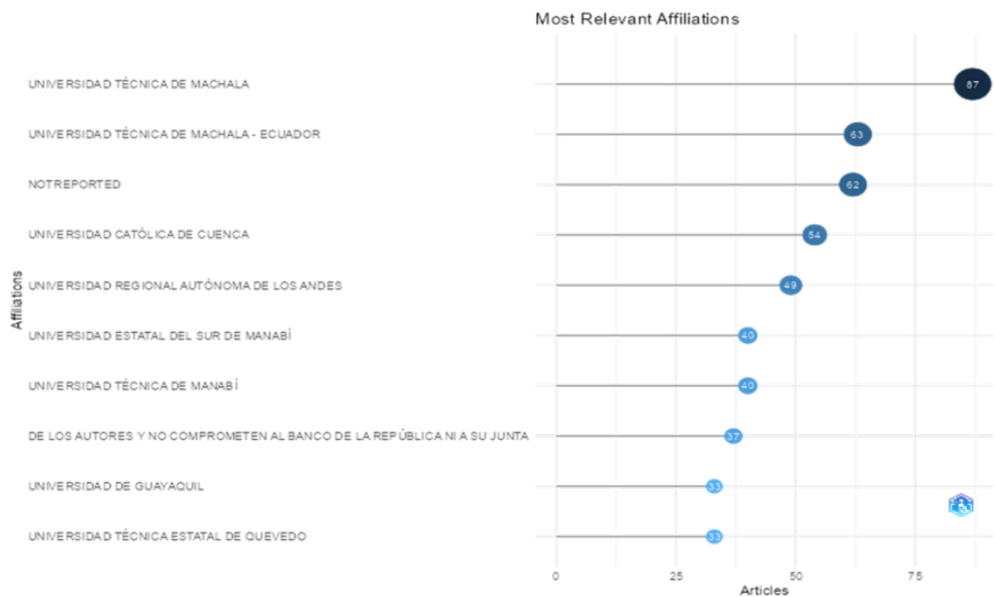


Figure 5. Institutions with the highest number of publications

Grouping of journals according to Bradford's Law

Figure 6 shows the results of Bradford's law in the bibliometric database used, where it can be seen that gray literature, the journal 593 Digital Publisher CEIT, and the Catholic University of Los Angeles in Chimbote are in the core zone, where 32,22 % (734) of the materials are grouped. These three sources represent 0,46 % of the total sources (649).

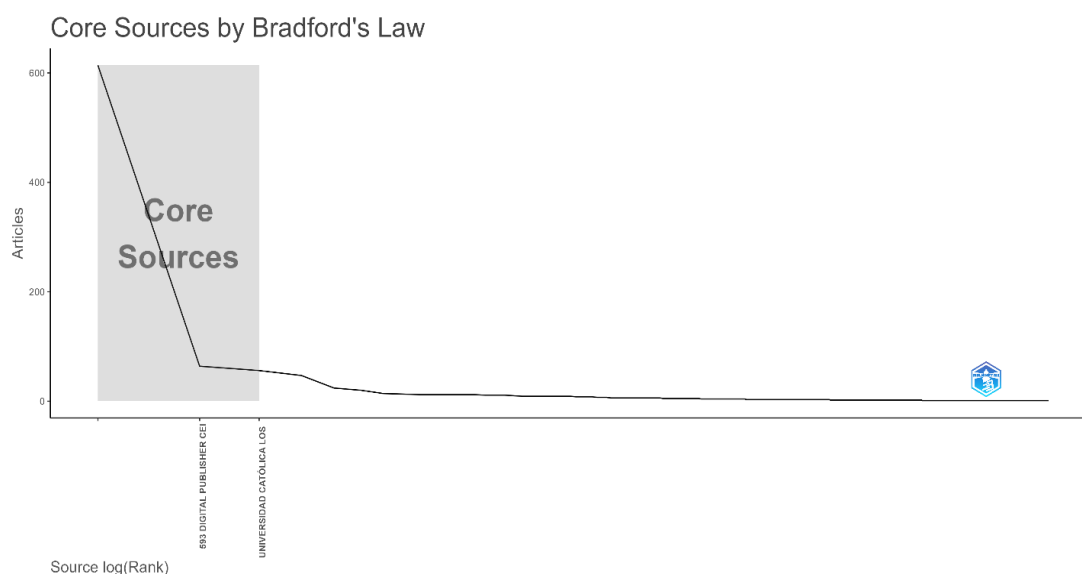


Figure 6. Journals in the core zone according to Bradford's law

This suggests that there are a large number of sources with a smaller contribution, while a very small core concentrates a significant part of academic production. Grey literature remains a key source, which could be due to its accessibility or because it includes reports, theses, and unconventional documents with high relevance in the field studied. The two remaining publishing institutions in the core area emerge as central players in the field of trade and tourism.

Impact of journals according to the H-index

Figure 7 shows the H-index for the main publication sources in the bibliometric database used in this research. It can be seen that grey literature has the highest H-index (6), while the sources 593 Digital Publisher CEIT, INNOVA Research Journal, Social Science Journal

and Venezuelan Management Magazine report an H-index = 4. On the other hand, the sources Ciencia Latina Multidisciplinary Scientific Journal, Contemporary Dilemmas Political Education and Values, Mastering the Sciences, E-IDEA Journal of Business Sciences, and ECA Sinergia show an H-index = 3.

Thus, gray literature has the greatest impact among the sources analyzed, as it has the highest H-index (6), suggesting that its publications are the most cited within this specific set of sources. Journals with H-index values of 4 and 3, although less influential, still contribute significantly.

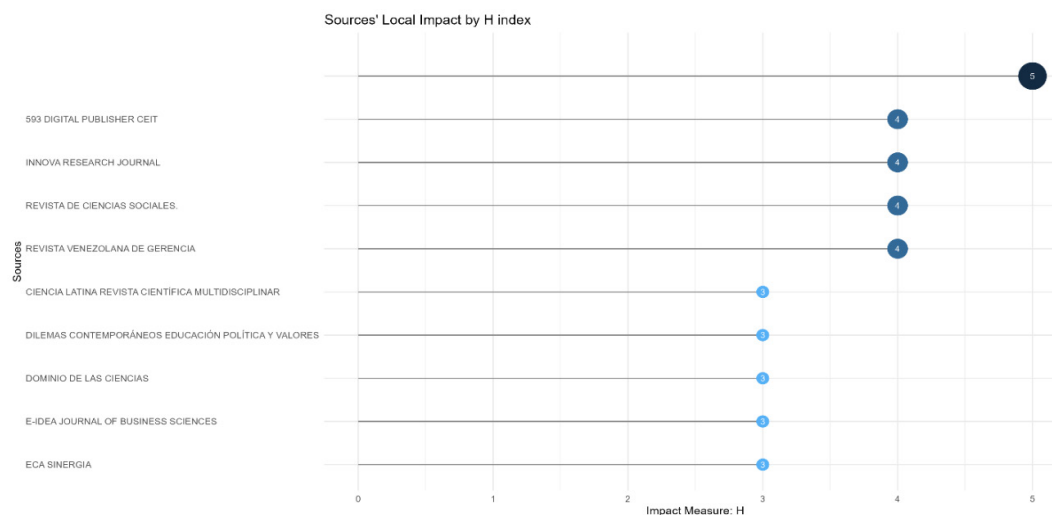


Figure 7. H-index of the main journals

Citation analysis by authors

Citation analyses are highly valued in bibliometric analyses, as they allow the identification of authors with the highest number of citations or leaders within the topics studied.

Figure 8 shows the citation analysis by authors, according to the H-index. The most cited authors are: Andrés Pacheco-Molina (H = 3, 11 citations), Mayiya González Illescas (H = 3, 58 citations), Norman Mora-Sánchez (H = 3, 16 citations), Alberto Magno Cutipa Limache (H = 2, 11 citations), Annherys Paz Marcano (H = 2, 11 citations), Campo Elías López Rodríguez (H = 2, 10 citations), Carlos Arango (H = 2, 6 citations), Edison Becerra Molina (H = 2, 8 citations), Francisco Vladimir Benítez-Luzuriaga (H = 2, 9 citations), and Freddy Lenin Villarreal Satama (H = 2, 9 citations).

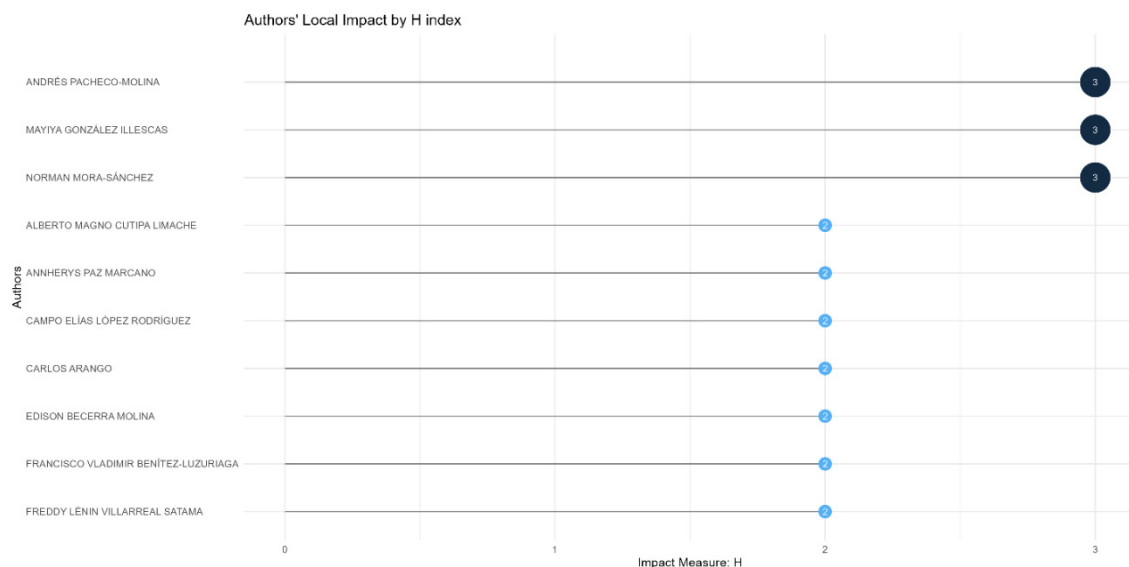


Figure 8. Most cited authors

Productivity by country

Figure 9 shows the geographical distribution of authors by correspondence of the materials in the database constructed. Twenty-four countries are included, and authors who do not declare their country are identified, possibly associated with the grey literature included.

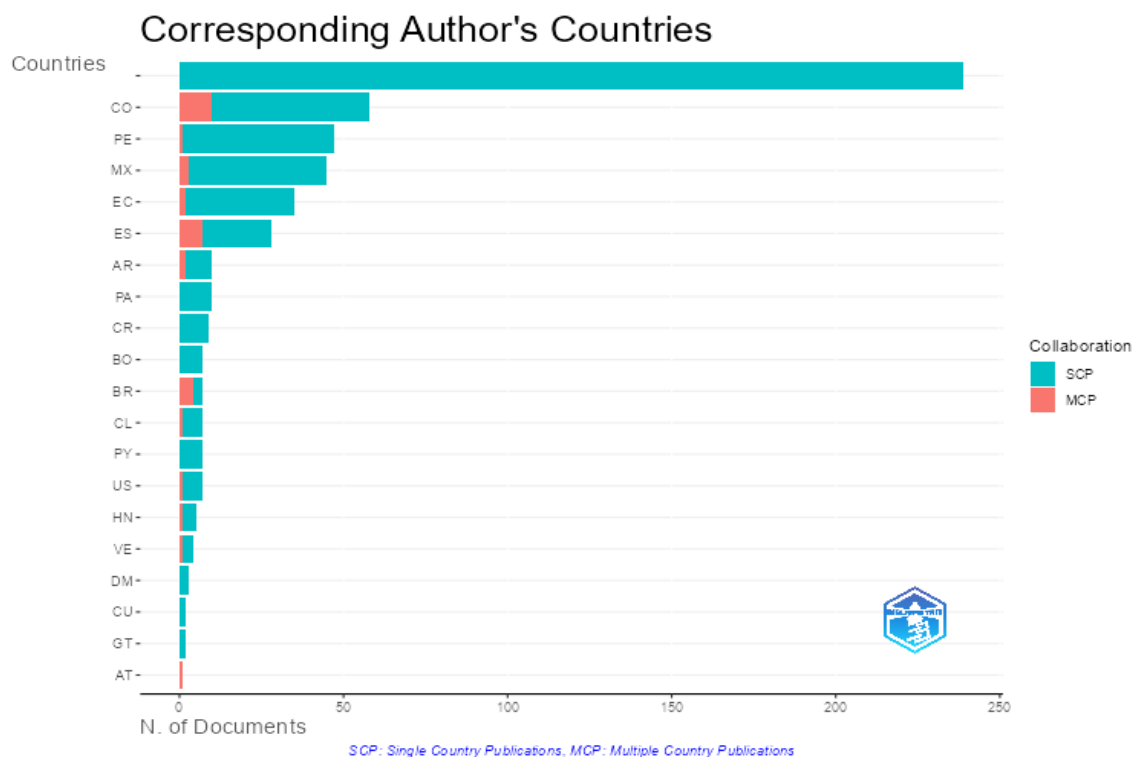


Figure 9. Productivity by country

Colombia (CO) leads in total production and citations (150 citations). Dominica (DM) has the highest average number of citations per article (2,00), although it only has 6 citations, indicating that its few articles are very influential. The United States (US) also stands out with 1,70 citations per article, despite a total of 12 citations, reflecting quality rather than quantity.

Among the countries with moderate volume but uneven impact are Mexico (MX) with a total of 66 citations and an average of 1,50 citations per article, indicating a balance between production and citations; Ecuador (EC) with 46 citations and an average of 1,30 citations per article; and Spain (ES) with a total of 33 citations and an average of 1,20 citations per article.

Peru (PE) has low production and low impact, with a total of 30 citations and a low average of 0,60 citations per article, suggesting that many publications have few citations. as well as Bolivia (BO), Costa Rica (CR), and Brazil (BR), which have a total of fewer than 3 citations and averages of fewer than 0,40 citations per article.

The MCP (orange in figure 9) indicates, for each country, the number of documents in which there is at least one co-author from another country. In this sense, the MCP measures the intensity of a country's international collaboration. The countries with the highest international collaboration were Colombia (10), Spain (7), and Brazil (4).

In summary, Colombia leads in total citations, but countries with less production (such as Dominica or the US) have more influential articles. The Latin American region dominates production, although with notable differences in impact per article.

Social structure of the field of research by country authors

Figure 10 shows the social structure by authors, where seven groups of collaborating authors are reported. For example, the group of Pablo Robuffo Giordano, Kathia Michalczewsky, and Ricardo Rozemberg, or the duo Silvia Lilianza Castano and Claudia Elena Fukau-Quintero, suggests frequent collaborations on the topic. Authors such as Dora Alicia Mora appear to be central, possibly leading projects, while others, such as Luis Eduardo Benavides Pupiales, appear more isolated. The diversity of surnames indicates a potentially international network. Overall, the network shows typical patterns of scientific collaboration, with stable co-authorship clusters.



Figure 10. Most frequent keywords given by authors

Country Collaboration Map

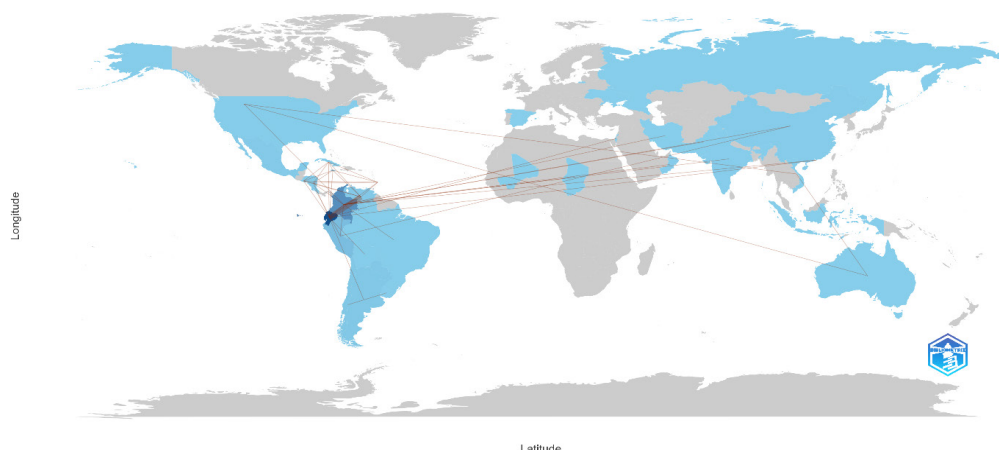


Figure 11. Map of collaboration between authors

When analyzing the social structure by country, figure 11 is obtained, which suggests the existence of globally distributed nodes of international collaboration. It can be seen that collaboration between South American countries is strong, being most prominent between Colombia, Venezuela, and Ecuador. In addition, these countries maintain strong collaboration with Asian countries.

CONCLUSIONS

The bibliometric analysis of scientific output on the relationship between retail trade and tourism in *OpenAlex* (2020-2024) reveals significant trends in research, collaboration, and thematic approaches in this field.

In summary, this study highlights a dynamic but fragmented field of research, with opportunities to consolidate theoretical frameworks and strengthen global collaboration networks. The adaptability of the tourism and commercial sectors to global crises, such as the pandemic, remains a critical area for future research.

BIBLIOGRAPHIC REFERENCES

1. Akehurst G, Alexander N. The internationalisation of retailing. Routledge; 1995. <https://www.routledge.com/The-Internationalisation-of-Retailing/Akehurst-Alexander/p/book/9780714641744>
2. Telfer DJ, Sharpley R. Tourism and Development in the Developing World. 2nd ed. Routledge; 2016. <https://www.routledge.com/Tourism-and-Development-in-the-Developing-World/Telfer-Sharpley/p/book/9781138921740?srsltid=AfmBOoqeTGs-kAUx4xPUR-hddh0TLBZRGeJXsPHlwLQcNEJr-6-caVck>
3. Cárdeno C. Turismo y comercio: España y su potencial como destino turístico de compras. *Distrib y Consum* 2015;25(137):16-21. <https://dialnet.unirioja.es/servlet/articulo?codigo=5233401>
4. Zupic I, Čater T. Bibliometric Methods in Management and Organization. *Organ Res Methods* 2015;18(3):429-72. <https://doi.org/10.1177/1094428114562629>
5. Ramos-Rodríguez AR, Ruíz-Navarro J. Changes in the intellectual structure of strategic management research: A bibliometric study of the Strategic Management Journal, 1980-2000. *Strateg Manag J* 2004;25(10):981-1004. <https://doi.org/10.1002/smj.397>
6. Mulet-Forteza C, Genovart-Balaguer J, Mauleon-Mendez E, Merigó JM. A bibliometric research in the tourism, leisure and hospitality fields. *J Bus Res* 2018;101:819-27. <https://doi.org/10.1016/j.jbusres.2018.12.002>
7. Rodríguez Jiménez G, Martínez Martínez CC, Alfonso Serafín Ad., Campos Cardoso LM. Análisis bibliométrico de la producción científica en gestión de destinos turísticos en SCOPUS. *Mikarimin Rev Científica Multidiscip* 2023;IX(1):151-70. <http://https://dialnet.unirioja.es%3Earticulo>
8. Pike S, Page SJ. Destination marketing organizations and destination marketing: A narrative analysis of the literature. *Tour Manag* 2014;41:202-27. <https://doi-org/10.1016/j.tourman.2013.09.009>
9. Gössling S, Scott D, Hall CM. Pandemics, tourism and global change: A rapid assessment of COVID-19. *J Sustain Tour* 2021;29(1):1-20. <https://doi.org/10.1080/09669582.2020.1758708>
10. López-Sanz JM, Penelas-Leguía A, Gutiérrez-Rodríguez P, Cuesta-Valiño P. Sustainable development and consumer behavior in rural tourism - Are they related? *Sustainability* 2021;13(5):2596. <https://doi.org/10.3390/su13052596>
11. Piwowar H, Priem J, Orr R. The open bibliometrics revolution: How OpenAlex enables transparent and inclusive research evaluation. *PLOS Biol* 2023;21(4):e3002032. <https://doi.org/10.1371/journal.pbio.3002032>
12. Priem J, Piwowar H, Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv* 2022. <https://arxiv.org/abs/2205.01833>
13. Visser M, van Eck NJ, Waltman L. OpenAlex as a credible alternative to commercial bibliometric databases: A cross-disciplinary analysis. *J Informetr* 2023;17(3):101428. <https://doi.org/10.1016/j.joi.2023.101428>
14. Huang W, Wu Y. A comparative study of OpenAlex, Scopus, and Web of Science for bibliometric analysis. *Scientometrics* 2022;127(12):6163-85. <https://doi.org/10.1007/s11192-022-04575-4>

15. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. J Informetr 2017;11(4):959-75. <https://doi.org/10.1016/j.joi.2017.08.007>
16. Duque P, Cervantes Cervantes LS. Responsabilidad Social Universitaria: una revisión sistemática y análisis bibliométrico. Estud Gerenciales 2019;35(153):451-64. <https://doi.org/10.18046/j.estger.2019.153.3389>
17. Aguiar Barbosa A d., Fumi Chim-Miki A. El surgimiento de la perspectiva de la competitividad turística social. Mapeo y clusterización de Redes Bibliométricas. Estud y Perspect en Tur 2020;29:390-405. <https://dialnet.unirioja.es/descarga/articulo/7316507.pdf>
18. Cardoso de Santana J, Maracajá KF, de Araújo Machado P. Turismo cultural y sostenibilidad turística: mapeo del desempeño científico desde Web of Science. Tur y Soc 2021;28:95-113. <https://doi.org/10.18601/01207555.n28.05>
19. Serrano Leyva B, Díaz Pompa F, Feria Velázquez FF, Richards G. Estudio bibliométrico sobre turismo creativo en Scopus, tendencias y futuras líneas de investigación. Biblios 2021;(82). <https://doi.org/10.5195/biblios.2021.979>
20. Pérez Anzardo LE, González Ortega R, Ortiz Pérez OL, Pérez Campdesuñer R. Análisis de la producción científica de Agroturismo desde sus inicios hasta la actualidad. Bibl An Investig 2022;18(3):1-9. <https://dialnet.unirioja.es/descarga/articulo/8742707.pdf>
21. Díaz-Pompa F, Serrano-Leyva B, Feria-Velázquez FF, Cruz-Aguilera N. Estudio bibliométrico sobre turismo cultural en SCOPUS. Líneas para futuras investigaciones. Bibl An Investig 2022;18(3):4-25. <https://doi.org/10.5195/biblios.2021.979>
22. de la Cruz Piña JM, Ferrero Ronda R, Rivas Nuila JR, Cruz Aquilera N. Turismo de bienestar como segmento en crecimiento: una mirada desde los estudios bibliométricos. ReHuSo 2023;8(1):98-112. <https://doi.org/10.33936/rehuso.v8il.5080>
23. Okoye K, Hosseini S. Introduction to R Programming and RStudio Integrated Development Environment (IDE). In: R Programming: Statistical Data Analysis in Research. Springer; 2024. page 3-24. https://doi.org/10.1007/978-981-97-3385-9_1
24. Bornmann L, Marx W. How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations. Scientometrics 2014;98(1):487-509. <http://link.springer.com/10.1007/s11192-013-1161-y>

FINANCING

The authors did not receive funding for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Yanet de la Concepción Avila, Damian Valdés Santiago, Danay Milagros Bulnes Mann.

Data curation: Yanet de la Concepción Avila, Damian Valdés Santiago.

Formal analysis: Damian Valdés Santiago, Danay Milagros Bulnes Mann.

Research: Yanet de la Concepción Avila, Damian Valdés Santiago, Danay Milagros Bulnes Mann.

Methodology: Damian Valdés Santiago.

Software: Damian Valdés Santiago.

Supervision: Danay Milagros Bulnes Mann.

Validation: Damian Valdés Santiago, Danay Milagros Bulnes Mann.

Visualization: Yanet de la Concepción Avila, Damian Valdés Santiago.

Writing - original draft: Yanet de la Concepción Avila, Damian Valdés Santiago.

Writing - revision and editing: Yanet de la Concepción Avila, Damian Valdés Santiago, Danay Milagros Bulnes Mann.