



وقائع مؤتمرات جامعة سبها
Sebha University Conference Proceedings

Conference Proceeding homepage: <http://www.sebhau.edu.ly/journal/CAS>



Analyze data from scholarly articles on Google Scholar for two colleges within Sebha University using an exploratory data analysis approach

*Roqyia Abdelhamed, Mansour Essgaer

Artificial Intelligence Department, Faculty of Information Technology, Sebha University, Sebha, Libya

Keywords:

Scholarly articles
Exploratory data analysis
EDA
Sebha university
Google Scholar
Citations

ABSTRACT

Applying the exploratory stage to academic publication data allows us to uncover valuable insights from unstructured data, aiding Sebha University administrators in enhancing academic scientific output. One major issue in research production is the lack of citations, which academics are keen to understand. They seek to identify key factors that can help them publish in journals where their work will be cited. Additionally, evaluation committees and efforts to improve the quality of educational institutions require structured knowledge and organized data, rather than raw, hard-to-analyze information. By utilizing the exploratory stage, we were able to identify the most active years of scientific publishing and the journals whose research receives few citations.

تحليل البيانات من المقالات العلمية على الباحث العلمي المتعلقة بكليتين بجامعة سبها باستخدام تحليل البيانات الاستكشافية

*رقية عبد الحميد سالم محمد و منصور الصغير

قسم الذكاء الاصطناعي، كلية تقنية المعلومات، جامعة سبها، ليبيا

الكلمات المفتاحية:

المقالات العلمية.
المرحلة الاستكشافية للبيانات
جامعة سبها
الباحث العلمي
الاستشهادات

المخلص

تطبيق المرحلة الاستكشافية على بيانات النشر الأكاديمي يمكننا من الاستفادة من المعلومات المخفية داخل البيانات غير المنظمة، والمساعدة في الحصول على معارف تمكن الإدارات في جامعة سبها من تحسين الإنتاج العلمي الأكاديمي. حيث تُعدّ من أهم مشكلات الإنتاج البحثي هو عدم الاستشهاد بالأبحاث العلمية. كما يتطلع الأكاديميون لمعرفة أهم العوامل التي تساعدهم في النشر في مجلات يمكن الاستشهاد بأبحاثهم فيها. كذلك، لجان التقييم ورفع جودة المؤسسات التعليمية يحتاجون إلى معرفة وجدول مهيكلة بدلاً من بيانات خام يصعب دراستها وتحليلها. حيث تمكّننا من خلال المرحلة الاستكشافية من الحصول على سنوات النشر العلمي الأكثر نشاطاً وأكثر المجالات العلمية التي لا تتلقى أبحاثها استشهاد.

1. Introduction

With the rise of electronic publishing, research papers are now widely accessible in databases, Google Scholar, scientific journals, and other sources. However, universities and research institutions must sift through these resources to gather insights and details about their research papers. The exploratory phase, an initial step in data analysis, plays a vital role in extracting valuable information from scientific publications on Google Scholar[1].

For academic text data, the exploratory stage prepares the ground for more comprehensive studies. Generally speaking, the exploratory stage of academic textual data prepares the ground for more in-depth analysis. It supports our understanding of the data, enables us to spot important themes and trends, pinpoints areas in need of development, and directs future study. It converts data from records and numbers into forms that provide a certain understanding and meaning, allow us to obtain important and valuable information, and lay the groundwork for more in-depth and thorough analyses [2] [3].

In addition to scientific research information, or metadata, the exploratory stage of academic data can yield valuable information

from the publishers, the scientific department, and the university it is linked with. As a result, we gather data, including the busiest publishing years, from which we may identify the publishing years with the greatest publication rates. The titles of the scholarly journals where the study was published can also be examined. It is also feasible to examine all of the scientific research's citations in order to identify the studies with the highest citation rates and to examine the studies with the lowest citation counts. There is a lot of useful information that can be derived from the exploratory phase.[4],[5].

The objective of this research is to conduct an exploratory analysis of the quantity of research papers produced by the College of Science and College of Medicine at Sebha University. The study seeks to identify the years with significant research output, as well as to examine the scientific journals and publishers where Sebha University publications have not received any citations. apply EDA To reveal patterns of publishers and obtain information that helps useful in solving scientific publishing problems.

*Corresponding author:

E-mail addresses: rayan_alshibani@yahoo.com, (M. Essgaer) man.essgaer@sebhau.edu.ly

Article History : Received 02 June 2024 - Received in revised form 30 September 2024 - Accepted 06 October 2024

2. Literature Reviews

Exploratory Data Analysis (EDA) is a quantitative methodology utilized for comprehending data in the absence of specific statistical hypotheses [6]. Recent progressions in machine learning have opened avenues for automating EDA, ranging from recommendation systems for isolated actions to completely automated procedures leveraging deep reinforcement learning [7]. Within educational settings, EDA has been applied to evaluate performance in game-based virtual assessments and vast student datasets, utilizing computationally intensive techniques to offer perspectives for learning evaluation and performance oversight [8]. These investigations showcase the adaptability of EDA in unveiling concealed patterns and insights across various domains, underscoring its significance in the preliminary phases of knowledge exploration processes[9]. Exploratory Data Analysis (EDA) has found application in diverse sectors, including the realms of education and industry, for extracting valuable insights from extensive datasets. Within educational contexts, EDA has been utilized to examine student placement records, forecast academic achievements, and enhance instructional methodologies. Jumana Nagaria & Senthil Velan S [10]and Otero-Escobar & Velasco-Ramírez [11]have explored these applications. Courtney [12]introduced a structured six-step approach to implement EDA in educational institutions, highlighting its pivotal role in data-informed decision-making and continual enhancement. Utilizing tools like histograms and statistical analyses, EDA methodologies have been instrumental in recognizing trends, flagging anomalies, and unravelling interrelations among variables within educational datasets [13]. Exploratory Data Analysis (EDA) stands as a critical phase in data exploration, encompassing the scrutiny and condensation of data attributes to unveil underlying patterns and tendencies [14]. This process typically involves three core stages: data presentation, feature identification, and feature interpretation [15]. EDA commonly kicks off with data preprocessing tasks like cleansing, transforming, and amalgamating data to enhance its integrity [16]. Diverse techniques, including univariate, bivariate, and multivariate analyses, are wielded in EDA to grasp relationships embedded within datasets. Visual representations assume a pivotal role in EDA, aiding in the visualization of data structures and trends.

exploratory data analysis on scientific research was applied in many studies to analyze the produce colleges and departments. Also, Certain research endeavors have employed exploratory techniques alongside machine learning algorithms to delve into university data concerning students, while others have directed such methods toward scientific research documents to uncover insights and latent patterns within the information.

Ailong [17] and Xiaobin [18] concentrate on assessing scientific research within college departments. Ailong [17] introduces an evaluation model and assessment system, while Xiaobin [18] utilizes the DEA method to measure research efficiency. Palacio [19]conducts a qualitative analysis of scientific output within a specific department, emphasizing the influence of university policies and changes in teaching roles. Hong-xi [20] identifies issues in college students' research activities and proposes remedies such as increasing funding and promoting teacher involvement. Together, these studies offer a comprehensive insight into the challenges and potential remedies for evaluating and enhancing scientific research within college departments.

The preliminary phase was also leveraged to extract patterns and insights from student data that could be advantageous to educational institutions such as schools and universities In making administrative decisions.

This paper [21]visually examines student placement data to draw conclusions through mathematical models. The research demonstrates how Exploratory Data Analysis (EDA) can unveil underlying patterns within extensive datasets. This methodology holds broad applicability and enhances managerial decision-making capabilities. In an educational setting, it enables informed decisions about student placements.

also, The domain of exploratory data analysis concerning citations in scientific publications has experienced notable progress. Heimerl [22] unveiled an interactive visual analysis method called CiteRivers, enabling the examination of citation trends. Building upon this,

Dunaiski [23] refined the approach with the creation of the Concept Cloud browser, utilizing concept lattices and tag clouds to streamline exploratory searches within academic publications and citation datasets.

3. Material and method

Exploratory Data Analysis (EDA) serves primarily as an approach to uncover the inherent expressions within data, diverging from the formal processes of modeling or hypothesis testing. It aids in dissecting datasets to depict their statistical characteristics, encompassing metrics such as central tendency (mean, mode, and median), spread (standard deviation and variance), distribution shape, and outlier identification[24] [25]. Fig. 1 illustrates the fundamental stages of the exploratory data analysis phase, elaborated upon in subsequent sections following the data collection phase:

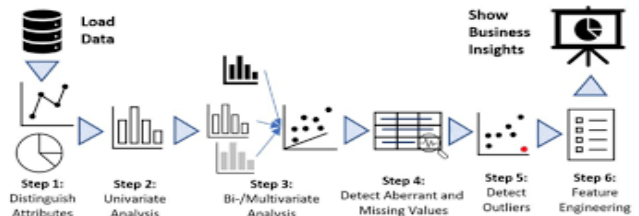


Fig. 1: EDA steps

1. Data collection and description

This research paper centre's around applying exploratory data analysis of publications and scientific papers from Sebha University available on Google Scholar. The data was manually collected and recorded in an Excel file, Due to the lack of time and the difficulty of manual collection, the study was limited to only two colleges, namely the College of Science and the College of Medicine. The data was collected from the first 505 research papers in the Scientific Researcher. The true identity of the publishers was also scrutinized and investigated, and anyone whose data did not contain the name of Sebha University was excluded. The database contains 12 features that were used to implement the exploratory phase. These features are described in the table shown in the TABLE 1.

TABLE 1: dataset features

Articles data	The description
author_google	Publisher name as ranked by Google Scholar
author-one	The name of the publisher who wrote in the author data of the article.
co-author	The name of the co-publisher who wrote in the author data of the article.
depart	The name of the scientific department that the publisher follows at the university.
faculty	The name of the faculty that the publisher follows.
title	the title of the article
abstract	the abstract of the article.
keyword	The keyword of the article.
CITED BY	the number of citations that article got in Google scholar
Publication date	the publication date of the article
Publisher	The entity or institution that published the article
Journal	The scientific journal that the article was published within

After data collection, it's imperative to comprehend and convert the data into a practical format. The exploratory phase can meet this requirement.

2. Exploratory data analysis (EDA)

The EDA approaches can be either non-graphical or graphical. Each of these strategies is either univariate, bivariate, or multivariate:

2.1 Non-graphical analysis:

a. 5 number summaries

For each continuous feature, the statistical description is utilized to provide various summary statistics such as measures of central tendency as mean, and measures of spread including standard deviation and interquartile range. In addition, the maximum, and minimum in each variable. This helps us know the smallest and largest number of citations for scientific papers and research, as well as the smallest or oldest publication date and the largest or most recent publication date.

2.2 Graphical analysis:

a. Univariate analysis tools

It furnishes a statistical overview for every field within the raw dataset or a summary for an individual feature. Examples of such analyses include count plots, pie chart plots, and histograms [26]. The elaboration of several of these tools is presented in more depth below:

- **Histogram:** employed to illustrate the distribution of numerical (continuous) data, dividing the complete value range into intervals. Each interval's frequency distribution is depicted by a rectangle, with the width symbolizing the class interval .
- **Count plot:** tallies the occurrences of observations per category within a categorical feature, displaying the outcomes in the form of a bar chart.
- **Pie chart plot:** illustrates the allocation of a total quantity among different levels of a categorical feature. Each category level is represented by a distinct slice of the circle, with the size of each slice indicating the proportion of the total attributed to that category level .

b. Bivariate analysis tools

Utilized to investigate the connections between two features and search for correlations among them [27]. Further elaboration on some of these tools is provided below:

- **Heat map:** A statistical measure indicating the degree of relationship and association between two features. Positive and negative correlations are the fundamental types of correlation, while a correlation of zero indicates no relationship between the features [28]. It serves to assess hypotheses concerning cause-and-effect correlations among features.
- **Scatter plot:** Data points are positioned both horizontally and vertically to illustrate the influence of one feature on another [29].
- **Side-by-Side Box plots:** A graphical depiction of numerical data presented as quadrants. Commonly employed for identifying outliers within a dataset. The boxplot summarizes sample data using the 25th, 50th, and 75th percentiles, also known as the lower, middle, and upper quartiles [30].

c. Multivariate analysis tools

Methods utilized to uncover relationships among multiple features extend beyond pairwise comparisons. Instances of such approaches include pivot tables and bar plots [31]. Several of these techniques will be explored in detail below:

- **Pivot Table:** This refers to a structured data summary presented in tabular form, incorporating aggregates such as sums, averages, or other statistical measures arranged in a user-friendly layout. The pivot table effectively organizes data, accommodating any applied filters and sorting criteria [32].
- **Bar plot:** The bar chart depicts comparisons between discrete groups. One axis of the graph shows the specific categories being compared, while the other axis represents the corresponding measured values for those categories .

Results and discussions

In this section, we will examine the outcomes of the exploratory phase carried out within the Python environment, summarizing the results as follows:

1. 5-number summaries

After examining TABLE 2, the statistical description helped in forming an overview of the dataset's characteristics, as detailed below:

- Out of the first thousand research papers on the Scientific Researcher of Sebha University, 506 were identified as being authored by publishers from the Faculties of Science and Medicine.
- The scientific papers from the Faculties of Science and Medicine have an average of 26 citations each, with the average year of publication being 2015.
- The standard deviation of citations is 63.8, reflecting the variability or spread of citation counts around the mean. In contrast, the standard deviation for publication years is 5.34,

indicating the variability or dispersion of publication years around the mean.

- The smallest year of publication for both science and medicine during the first thousand publications is 1995, while the smallest citation value is zero.
- The largest citation value was 785, and the most recent publication year was 2023.
- 25% of scientific research received zero citations, which is an unfortunate statistic, while 25% of research was in 2012.
- 50% of the researches received a citation value equal to 6, and half of the scientific researches for both schools were during the year 2015.
- 75% of the research received only 23 citations, which is an unsatisfactory statistic considering the large departments of the two colleges and the number of researchers in them. The statistics also showed that the majority of research for both science and medicine was in the year 2019.

TABLE 2: The statistical description of features

	citations	Publications date
count	506	506
mean	26.2	2015
standard deviation	63.8	5.34
min	0	1995
max	785	2023
25%	1	2012
50%	6	2015
75%	23	2019

2. Comparing the production of scientific research by the faculties of science and medicine compared to the faculties of Sebha University.

From the Fig.2 , we notice that the College of Science has the highest production of scientific research, with a large difference from the rest of the colleges, followed by the College of Medicine. The figure shows that the College of Science has around 400 scientific research papers out of the first 1000, making it the leading contributor. The College of Medicine ranks second with approximately 120 research papers, which is significantly fewer than the College of Science.

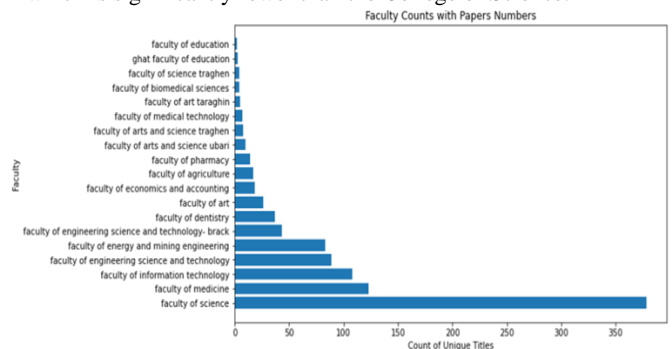


Fig.2: Displaying the Ranking of Sebha University colleges in terms of production of research and scientific papers

3. Investigating the highest and lowest publication years for the faculties of science and medicine.

The College of Science saw its highest production rate in this period in 2012, when scientific output peaked. Compared to the number of professors and researchers in the various departments of the College of Science, the number of research papers published during that year—roughly 40—was modest. It was followed by the publishing years of 2013 and 2014, then gradually decreased until 2016, after which it rose again and decreased again in 2019, and then resumed the rise in 2020, reaching the highest rate during the last five years in 2020 as is. The figure 3 makes this evident.

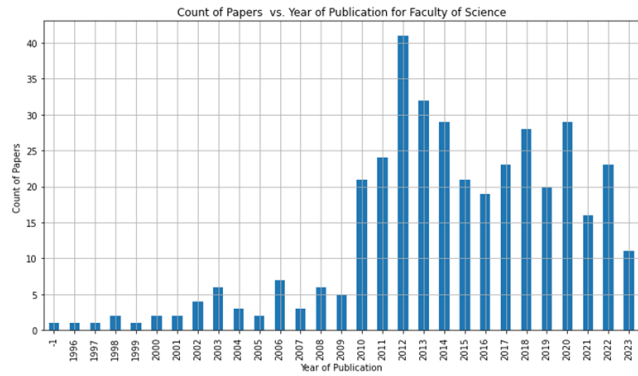


Fig.3:The production of research and scientific papers of faculty of science in their publications years

As is clear in the figure.4, The statistic makes it evident that the College of Medicine produced relatively little research throughout the middle years of publication; 2014 was the highest year, and the total number of publications did not approach 12. The College of Medicine published 20 papers in 2021, the most in its publication history. This is a relatively small amount when considering the size of the college, its departments, and its significance in the health field—after all, it is the only medical college in the city of Sebha. The city's health is negatively impacted by the decline in research projects.



Fig.4: The production of research and scientific papers of faculty of medicine in their publications years

4. Identify attributes associated with scientific research that did not receive citations in the two colleges.

The Fig.5 reveals that approximately 100 scientific papers from the College of Science have no citations, while about 300 papers do have citations. This indicates that roughly a quarter of the research from the College of Science is not cited. Similarly, about a quarter of the research from the Faculty of Medicine is also uncited as in Fig.6.

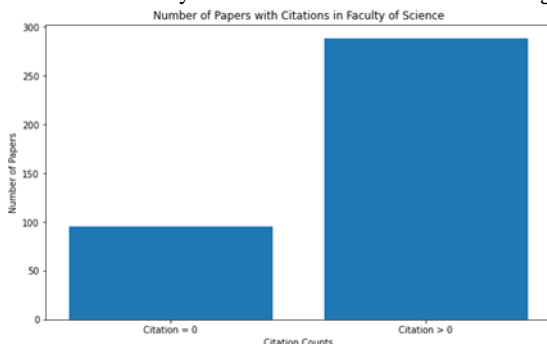


Fig.5 : A comparison between the number of studies with zero citations and those with citations for the Faculty of science

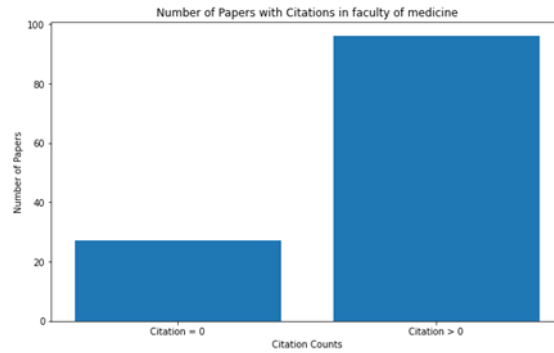


Fig.6 : A comparison between the number of studies with zero citations and those with citations for the Faculty of Medicine

a. the distribution number of research papers .vs. publications date

Figure.7 highlights that the years with the most significant output of scientific research lacking citations for the College of Science are 2018, succeeded by 2022 and 2023. This data suggests that college administrations might consider examining local scientific events and conferences held during these years to understand the underlying factors contributing to scientific output without citations. Notably, recent years have seen the highest productivity in scientific research without corresponding citations.

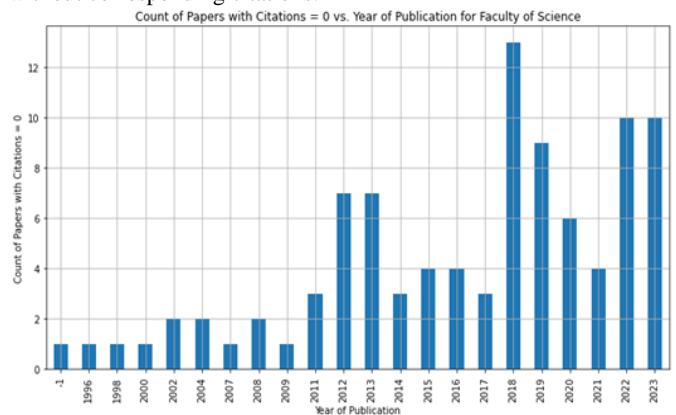


Fig.7: the distribution number of research papers with zero citations .vs. publications date for faculty of science

According to Figure 8, it's evident that in the recent past, the College of Medicine has similarly witnessed a notable amount of scientific research output without citations.

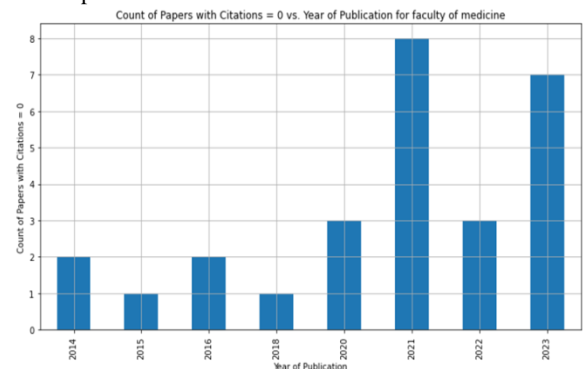


Fig.8: the distribution number of research papers with zero citations .vs. publications date for faculty of medicine

b . determine scientific departments with scientific research that did not receive citations

As shown in the figure. 9, more than half of the uncited scientific research from the College of Science originated from researchers in the Chemistry Department. Additionally, the Departments of Mathematics, Geology, and Physics are among the top scientific departments whose research did not attract citations or attention from the global scientific community.

Distribution of Papers with Citations = 0 by Department within Faculty of Science

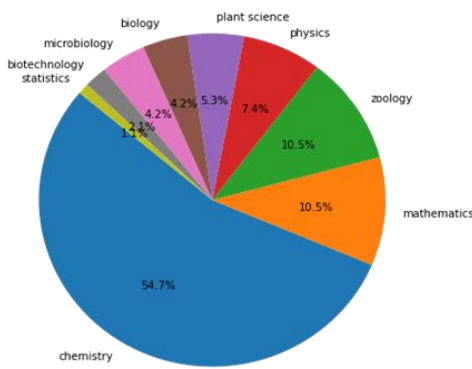


Fig. 9: the distribution number of research papers with zero citations .vs. department for faculty of science.

More than half of the scientific research of the Faculty of Medicine that did not receive any citations was produced by researchers from the Department of Pharmacology, as shown in the figure 10. Also, about a quarter of the zero-cited research production of the Faculty of Medicine was from the Department of Microbiology, followed by the Department of Oral Surgery.

Distribution of Papers with Citations = 0 by Department within faculty of medicine

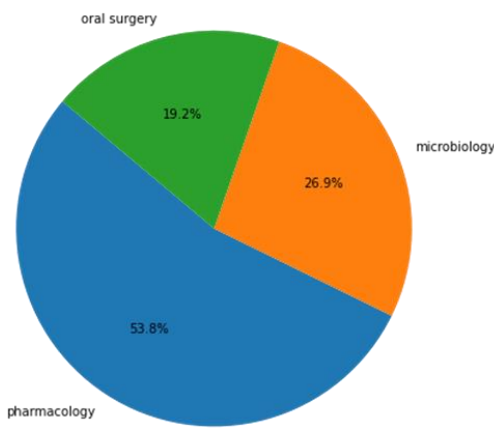


Fig. 10: the distribution number of research papers with zero citations .vs. department for faculty of medicine.

c . Listing scientific journals and publishers that released research without receiving any citations

The figure.11, indicates that the Sebha University Journal has the highest percentage of scientific papers with zero citations. These statistics can be obtained by looking more closely at the research articles published in the university journal and by altering the admission and arbitration policies to stop the publication of more research with zero citations.

Top 15 Journals with zero citations for faculty of science and medicine

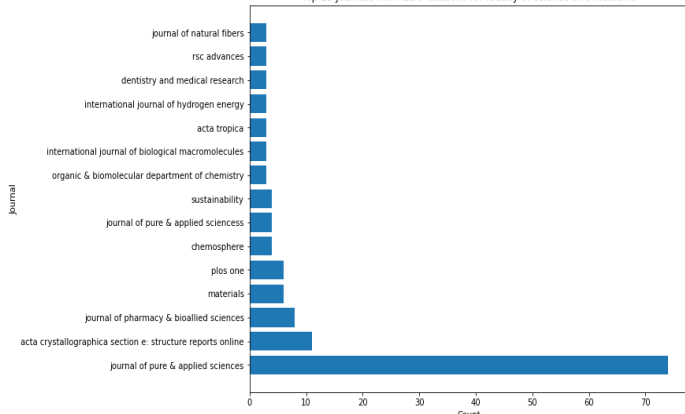


Fig. 11: Most scientific journals in which scientific research is published are devoid of citations

We are unable to conclude that publishing scientific work in journals that are not subject to international classification is the sole cause of the lack of citations. Figure 12 shows that some prominent academic publishers are included in the list, but they are also included in the list of publishers of articles with zero citations. Examples of these publishers include the global publishing firm Al-Safir. The vast majority of publications at Sebha University did not cite the papers.

Top 15 Publisher with zero citations for faculty of science and medicine

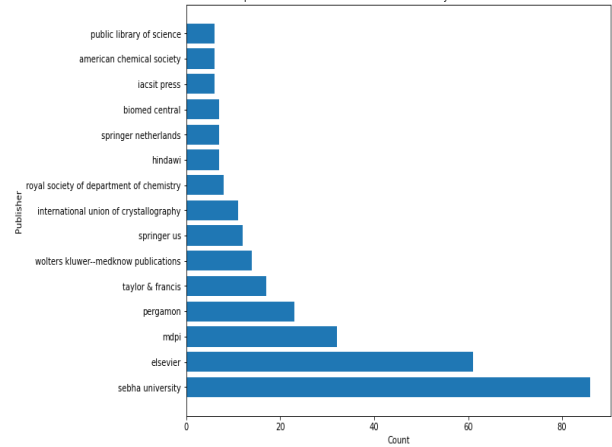


Fig. 12: Most publishers in which scientific research is published are devoid of citations

d. Listing the most frequent words in the titles of zero-cited papers for the College of Science and the College of Medicine

Utilizing the exploratory phase and leveraging natural language processing libraries, we can analyze the titles of scientific research papers with zero citations to identify trending topics by displaying the most frequent words. The figure 13 reveals that the most common words in the titles of zero-citation research within the College of Science are related to topics in the Chemistry Department, which has the largest number of papers.

Top 20 Most Frequent Words in Titles for papers to faculty of science

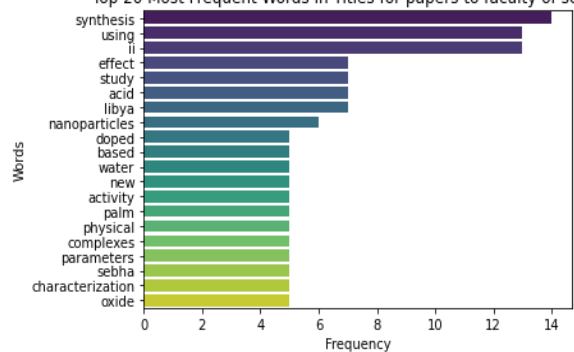


Fig. 13: The most frequent words in the titles of zero-cited papers for the Faculty of Science

The figure 14 also highlights the most frequently used words in the titles from the Faculty of Medicine, and more words can be examined to gain further insights into the topics.

From both figures 13 and 14, we observe the recurrence of the word "Libya," suggesting that scientific research conducted on the local community has not received citations.

Top 20 Most Frequent Words in Titles (Excluding Stopwords and Numbers)

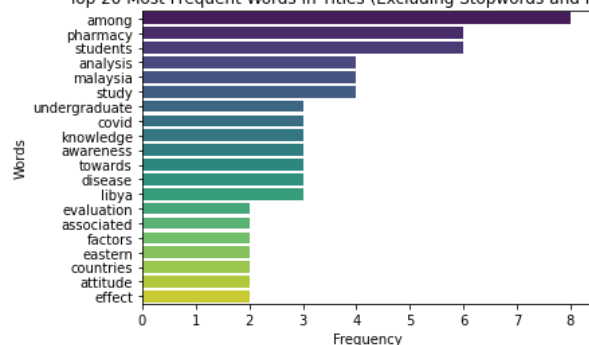


Fig. 14: The most frequent words in the titles of zero-cited papers for the Faculty of Medicine

Conclusion and Recommendations

The focus was placed on scientific journals and publishers that featured research from Sebha University through the Publisher and Journal attributes. This investigation pinpointed the specific scientific journals where Sebha University's research was published and garnered the highest citations. Additionally, we highlighted the publishers and journals associated with Sebha University whose research did not receive any citations. This data holds significance for scholars and university administrators aiming to enhance the caliber of academic output and boost citation frequencies. This study recommends broadening the research scope and incorporating additional analytical features to identify patterns crucial for understanding and predicting Sebha University's future scientific research trajectory. These insights can aid in addressing research challenges such as the prevalence of studies with zero citations, facilitating informed solutions for refining the university's scientific research landscape.

References

- [1]- E. Bell and A. Bryman, "The ethics of management research: an exploratory content analysis," *British journal of management*, vol. 18, no. 1, pp. 63-77, 2007.
- [2]- E. Davidson, R. Edwards, L. Jamieson, and S. Weller, "Big data, qualitative style: a breadth-and-depth method for working with large amounts of secondary qualitative data," *Quality & quantity*, vol. 53, pp. 363-376, 2019.
- [3]- A. Galletta, *Mastering the semi-structured interview and beyond: From research design to analysis and publication*. NYU press, 2013.
- [4]- P. Organisciak, B. M. Schmidt, and J. S. Downie, "Giving shape to large digital libraries through exploratory data analysis," *Journal of the Association for Information Science and Technology*, vol. 73, no. 2, pp. 317-332, 2022.
- [5]- D. S. Siegel, D. Waldman, and A. Link, "Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: an exploratory study," *Research policy*, vol. 32, no. 1, pp. 27-48, 2003.
- [6]- M. Komorowski, D. C. Marshall, J. D. Saliccioli, and Y. Crutain, "Exploratory Data Analysis," 2016.
- [7]- T. Milo and A. Somech, "Automating Exploratory Data Analysis via Machine Learning: An Overview," *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 2020.
- [8]- D. C. Gibson and S. d. Freitas, "Exploratory Analysis in Learning Analytics," *Technology, Knowledge and Learning*, vol. 21, pp. 5 - 19, 2015.
- [9]- R. Yulianto *et al.*, "Exploratory Data Analysis for Building Energy Meters Using Machine Learning," *Journal of Telecommunication Electronics and Control Engineering (JTECE)*, 2023.
- [10]- J. Nagaria and S. V. S, "Utilizing Exploratory Data Analysis for the Prediction of Campus Placement for Educational Institutions," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-7, 2020.
- [11]- A. D. Otero-Escobar and M. L. Velasco-Ramírez, "Study on Exploratory Data Analysis Applied to Education," *2023 IEEE International Conference on Engineering Veracruz (ICEV)*, pp. 1-5, 2023.
- [12]- M. B. Courtney, "Exploratory Data Analysis in Schools: A Logic Model to Guide Implementation," *International Journal of Education Policy and Leadership*, vol. 17, 2021.
- [13]- I. Griha Tofik Isa, Z. Zulkarnaini, L. Novianti, F. Elfaladonna, and S. Agustri, "Exploratory Data Analysis (EDA) dalam Dataset Penerimaan Mahasiswa Baru Universitas XYZ Palembang," *Smart Comp: Jurnalnya Orang Pintar Komputer*, 2023.
- [14]- K. M. A. Kadhar and G. Anand, "Analyzing the Data," *Data Science with Raspberry Pi*, 2021.
- [15]- J. De Mast and A. Trip, "Exploratory Data Analysis in Quality-Improvement Projects," *Journal of Quality Technology*, vol. 39, pp. 301 - 311, 2007.
- [16]- K. Paul, "Data Expedition: Travel Through Data Preprocessing, EDA And PCA," *Educational Administration: Theory and Practice*, 2024.
- [17]- H. Ailong, "ON THE EVALUATION MODEL OF SCIENTIFIC RESEARCH ACHIEVEMENTS AND THEIR GAUGING SYSTEM IN COLLEGES/DEPARTMENTS OF UNIVERSITIES," *Journal of Southwest Agricultural University*, 2007.
- [18]- L. Xiaobin, "Research on Evaluation of the Scientific Research Efficiency of Departments in Universities and Colleges Based on DEA," *Science and Technology Management Research*, 2013.
- [19]- M. I. Palacio and S. G. Martín, "La producción científica en la Facultad de," 2016.
- [20]- J. Hong-xi, "Analysis and Exploration of College Students' Scientific Research Activities," *Journal of Shenyang College of Education*, 2010.
- [21]- J. Nagaria, "Utilizing exploratory data analysis for the prediction of campus placement for educational institutions," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020: IEEE, pp. 1-7.
- [22]- F. Heimerl, Q. Han, S. Koch, and T. Ertl, "CiteRivers: Visual Analytics of Citation Patterns," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, pp. 190-199, 2016.
- [23]- M. Dunaiski, G. J. Greene, and B. Fischer, "Exploratory search of academic publication and citation data using interactive tag cloud visualizations," *Scientometrics*, vol. 110, pp. 1539-1571, 2017.
- [24]- V. Kotu and B. Deshpande, *Data science: concepts and practice*. Morgan Kaufmann, 2018.
- [25]- S. Juggins and R. J. Telford, "Exploratory data analysis and data display," *Tracking environmental change using Lake Sediments: data handling and numerical techniques*, pp. 123-141, 2012.
- [26]- A. Agaal and M. Essgaer, "An Exploratory Data Analysis of Breast Cancer Features in South of Libya," *Journal of Pure & Applied Sciences*, vol. 21, no. 4, pp. 57-64, 2022.
- [27]- A. Ghosh, M. Nashaat, J. Miller, S. Quader, and C. Marston, "A comprehensive review of tools for exploratory analysis of tabular industrial datasets," *Visual Informatics*, vol. 2, no. 4, pp. 235-253, 2018.
- [28]- K. Sahoo, A. K. Samal, J. Pramanik, and S. K. Pani, "Exploratory data analysis using Python," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 4727-4735, 2019.
- [29]- M. Abzalov and M. Abzalov, "Exploratory data analysis," *Applied Mining Geology*, pp. 207-219, 2016.
- [30]- R. K. Pearson, *Exploratory data analysis using R*. Chapman and Hall/CRC, 2018.
- [31]- K. H. Esbensen, D. Guyot, F. Westad, and L. P. Houmoller, *Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design*. Multivariate Data Analysis, 2002.
- [32]- K. Hunter-Thomson, "How Can We Better Move From Data to Meaning?," *Science Scope*, vol. 44, no. 3, pp. 14-19, 2021.