Vol-11, Issue-1, Year-2024
This Article is under Formatting, the PDF's ready file will be replaced soon.

Unveiling Library Usage Patterns: A Python-based Descriptive Analysis of Big Data in Librarianship

Neeraj[1]*(Research Scholar),* ***Prof. Nirmal Kumar Swain[2]*****(Professor)*
Department of Library and Information Science,
Maharshi Dayanand University, Rohtak, Haryana, India

Abstract

This study analyses circulation data from Vivekananda Library, Maharshi Dayanand University, Rohtak, Haryana, India, for the academic year 2022-23. The analysis is conducted using descriptive statistics and Python programming to understand library resource utilisation patterns, temporal trends, and user behaviour. The findings reveal distinct usage patterns across the four quarters, with variations in the number and types of transactions. Temporal trends show fluctuations in library usage based on academic calendars and semester schedules. Popular materials are identified, providing insights into user preferences and needs. The study demonstrates the value of data analytics in informing decision-making processes in librarianship and highlights the practical application of data-driven approaches in improving library services. The findings have implications for resource allocation, collection development, and service provision, emphasising the importance of library usage analytics in enhancing the user experience and demonstrating the value of libraries in academic settings.

Keywords: *Library usage analytics, Circulation data analysis, Academic libraries, Descriptive statistics, Python programming, User behaviour, Resource allocation, Collection development, Service provision, User experience*

***E-mail:** ntanwar91@gmail.com, **Mob.:** +91-8053500255

1. INTRODUCTION

1.1 Background of the Study

Libraries are crucial hubs of knowledge dissemination and academic support in educational institutions. In today's digital age, big data has become a significant phenomenon, greatly influencing people worldwide (Hussain & Shahid, 2022). Due to the abundance of data

available in diverse formats, companies and organisations across nearly every sector are concentrating on leveraging data to gain a competitive edge (Hamad et al., 2022). Similarly, like other institutions, academic libraries are seizing the opportunity to use big data to enhance their operations and deliver relevant content to their users more effectively (Tella & Kadri, 2021). With the advent of big data analytics, libraries can now harness the power of data to gain insights into user behaviour and optimise their services. Understanding how patrons interact with library resources can improve resource allocation, collection development, and user satisfaction.

However, despite the potential benefits, there is a gap in the literature regarding the application of big data analytics in librarianship, particularly in the context of circulation data analysis. This study seeks to bridge this gap by employing descriptive statistics and Python programming to analyse circulation data from the academic year 2022-23.

This study aims to provide librarians and information professionals with valuable insights into user behaviour and preferences by exploring library usage analytics. The findings of this study can inform decision-making processes in library management, ultimately leading to more efficient and user-centric library services.

1.2 Importance of Library Usage Analytics in Academic Settings

Library usage analytics play a crucial role in academic settings by providing valuable insights into how students, faculty, and staff utilise library resources. By analysing circulation data, library usage patterns, and user behaviour, libraries can optimise their services and resources to better meet their patrons' needs. This data-driven approach allows libraries to make informed decisions about collection development, resource allocation, and service provision, ultimately enhancing the overall user experience. Additionally, library usage analytics can help libraries demonstrate their value and impact within the academic community, supporting strategic planning and decision-making processes. Library usage analytics are essential for improving library services, increasing user satisfaction, and demonstrating the value of libraries in academic settings.

1.3 Purpose of the Study

This study aims to explore library usage analytics in the context of academic librarianship, explicitly focusing on the analysis of circulation data using descriptive statistics and Python programming. The study aims to achieve the following objectives:

i. To analyse circulation data from the academic year 2022-23 to understand patterns and trends in library resource utilisation.
ii. To apply descriptive statistics techniques to the circulation data, including count, unique, top, and frequency measures, to gain insights into user behaviour.
iii. To divide the circulation data into four quarters (first quarter, second quarter, third quarter, and fourth quarter) and analyse each quarter separately to identify any temporal trends.
iv. To demonstrate the utility of Python programming for analysing large-scale circulation data sets in a library setting.
v. To provide librarians and information professionals with practical insights into how they can use data analytics to enhance library services and improve user satisfaction.

1.4 Research Questions

1. What are the patterns and trends in library resource utilisation based on circulation data from the academic year 2022-23?
2. How do circulation statistics vary across different quarters (First, Second, Third, Fourth) of the academic year?
3. What are the most frequently borrowed items (books, journals, etc.) during the academic year?
4. What are the most common transaction types (checkouts, renewals, returns, etc.) recorded in the circulation data?
5. How can descriptive statistics and Python programming be used to analyse and interpret circulation data in a library setting?
6. What insights can be gained from analysing circulation data to improve library services and user satisfaction?
7. How does the analysis of circulation data contribute to the broader field of librarianship and information science?

2. LITERATURE REVIEW

2.1 Overview of Big Data in Librarianship

Big data has emerged as a valuable library resource, offering new opportunities for enhancing services, improving operations, and gaining insights into user behaviour. In the context of librarianship, big data refers to the vast amounts of data generated by library systems, including circulation records, catalogue searches, and user interactions with digital resources.

This data can be analysed to uncover patterns, trends, and correlations that can inform decision-making and improve service delivery.

One of the key benefits of big data in librarianship is its ability to provide librarians and information professionals with a deeper understanding of user needs and preferences. By analysing how patrons interact with library resources, libraries can tailor their services to meet their users' needs better. For example, analysis of circulation data can help libraries identify popular materials and allocate resources accordingly, ensuring that high-demand items are readily available.

Big data analytics also offer opportunities for libraries to enhance their digital services and collections. By analysing data on how users access and use digital resources, libraries can optimise their digital collections, improve search capabilities, and enhance the overall user experience. Additionally, big data can be used to personalise services for users, providing them with recommendations based on their past behaviour and preferences.

2.2 Previous Studies on Library Usage Analytics

(Oladokun et al., 2023) in their article titled *"Global challenge and opportunities for libraries and big data"* explore the intersection of big data and libraries, focusing on the challenges and opportunities this presents to the global library community. The authors begin by highlighting the increasing importance of big data in the digital age and its potential to revolutionise how libraries operate. They argue that libraries can harness the power of big data to improve their services, enhance user experience, and demonstrate their value to stakeholders. The authors discuss issues such as data privacy, security, and ethics, emphasising the need for libraries to consider these factors when implementing big data projects carefully. The authors also highlight the opportunities that big data presents to libraries, including the ability to analyse large amounts of data to gain insights into user behaviour and preferences. They argue that by leveraging big data, libraries can better tailor their services to meet the needs of their users, ultimately enhancing their relevance and impact.

(Zhang et al., 2023) in their article "The Impact of Big Data on Research Methods in Information Science" explore the transformative effects of big data on research methodologies within the field of information science. The authors contend that the proliferation of big data has necessitated reevaluating and adapting traditional research methods to leverage the vast amounts of data now available effectively. The authors highlight the shift from hypothesis-driven research to data-driven research, where researchers analyse large datasets to uncover

patterns and insights that may not have been apparent through traditional methods. This approach has led to new research opportunities and the ability to address complex research questions that were previously challenging or impossible to answer.

(Showers, 2015) in his book *"Library Analytics and Metrics: Using Data to Drive Decisions and Services"*, provides an overview of the use of analytics in libraries, including discussions on data-driven decision-making, collection management, and user engagement. The book explores how libraries can harness the power of data to make informed decisions, improve services, and demonstrate value to stakeholders. Showers argues that in an increasingly digital and data-driven world, libraries must adapt by adopting a more analytical approach to their operations. Showers provides case studies from real libraries, demonstrating how they have successfully implemented analytics projects to improve user experience, collection management, and resource allocation.

(Orcutt, 2009) their book *"Library Data: Empowering Practice and Persuasion"* provides a timely and insightful exploration of the role of data in modern librarianship. The authors argue that data has the power to transform library practice, enabling librarians to make more informed decisions and advocate for the value of their services. One of the book's key strengths is its practical approach, providing librarians with actionable strategies for collecting, analysing, and presenting data. Orcutt drew on their experiences as librarians to offer real-world examples of how data can improve services, assess impact, and demonstrate value to stakeholders.

2.3 Theoretical Framework

The theoretical framework for this study is grounded in the concept of information behaviour, which is the interdisciplinary study of how people seek, evaluate, and use information. Information behaviour theory provides a lens through which to understand how library users interact with library resources and services, and how these interactions can be influenced by factors such as information needs, information-seeking strategies, and contextual variables.

Within the framework of information behaviour, this study also draws on the concept of user-centred design, which emphasises the importance of designing services and systems based on the needs and preferences of users. By analysing circulation data and understanding patterns of library resource utilisation, libraries can better tailor their services to meet the needs of their users, ultimately enhancing the user experience and improving outcomes.

Additionally, this study incorporates elements of data-driven decision-making, which involves using data analysis to inform organisational decisions and improve performance. By applying descriptive statistics and Python programming to analyse circulation data, libraries can gain valuable insights that can guide decision-making processes and lead to more effective resource allocation and service delivery.

3. METHODOLOGY

This study utilises a quantitative research approach to analyse circulation data from Vivekananda Library, Maharshi Dayanand University, Rohtak, Haryana, India, for the academic year 2022-23. The research design involves the use of descriptive statistics and Python programming for data analysis and is as follows:

- Data Collection: Circulation data for the academic year 2022-23 was obtained from Vivekananda Library, including information on dates, transaction types, initials, and locations.
- Data Preprocessing: The collected data was cleaned and transformed to ensure consistency and compatibility with the analysis techniques.
- Data Analysis: Descriptive statistics, including count, unique, top, and frequency, were used to analyse the circulation data for each quarter. Python libraries such as Pandas, NumPy, and Matplotlib were employed for data manipulation and visualisation.
- Findings Interpretation: The findings were interpreted to understand patterns, trends, and correlations in library resource utilisation, temporal trends, and user behaviour.
- Implications and Recommendations: The study discusses the implications of the findings for library management, including resource allocation, collection development, and service provision. Recommendations for future research are also provided.

3.1 Data Collection

The circulation data for the academic year 2022-23 was collected from Vivekananda Library, Maharshi Dayanand University, Rohtak, Haryana, India. The data was obtained from the library management system (LMS) of the institution, which records information about the borrowing and returning of library materials. The circulation data included the following variables:

- ate: The date on which the transaction (borrowing or returning of a library item) took place.
- Transaction Type: The type of transaction, such as check-out, renewal, or return.
- Initial: The initial of the staff member who processed the transaction.
- Location: The location within the library where the transaction took place.

The data was collected for the entire academic year, from 01st April 2022 to 31st March 2023. The data was anonymised to ensure the privacy of library users and aggregated to provide a comprehensive view of library resource utilisation throughout the academic year. The circulation data collected from Vivekananda Library forms the basis for the analysis conducted in this research, providing insights into how library resources were used by students, faculty, and staff during the academic year 2022-23.

3.2 Data Preprocessing

Data preprocessing is an essential step in preparing the circulation data for analysis. It involves cleaning and transforming the raw data to make it suitable for analysis. The following steps were performed as part of the data preprocessing for this study:

- **Handling Missing Values:** Any missing values in the data, such as incomplete transaction records, were identified and either removed or imputed based on the nature of the missing data.
- **Data Transformation:** The data may have been transformed to ensure consistency and compatibility with the analysis techniques used. For example, date formats may have been standardized, and categorical variables may have been encoded for analysis.
- **Removing Duplicates:** Duplicate entries in the data, if any, were identified and removed to ensure that each transaction is represented only once in the dataset.
- **Aggregation:** The data may have been aggregated to a suitable level for analysis. For example, daily transaction counts may have been aggregated to monthly or quarterly totals to analyse trends over time.
- **Normalization or Standardization:** Depending on the analysis techniques used, numerical variables in the data may have been normalized or standardized to ensure that they are on a similar scale.
- **Data Splitting:** The data may have been split into training and testing datasets for model validation purposes, especially if machine learning techniques were used for analysis.

By performing these preprocessing steps, the circulation data was cleaned and transformed into a format that was ready for analysis using descriptive statistics and Python programming.

3.3 Descriptive Statistics Used

The descriptive statistics provide a basic overview of the circulation data and can help identify patterns, trends, and anomalies in the data. They are useful for summarising and interpreting the data, providing valuable insights into library resource utilisation and user behaviour. The following descriptive statistics were used for the study:

- **Count:** Count is a basic descriptive statistic that simply calculates the number of observations in a dataset. In the context of circulation data analysis, the count can be used to determine the total number of transactions or occurrences of a particular event, such as the number of books borrowed or returned during a specific period.
- **Unique:** Unique calculates the number of distinct values in a dataset. For example, in the context of circulation data, the number of unique books borrowed or the number of unique users who borrowed books can provide insights into the diversity of library usage.
- **Top:** Top identifies the most frequent value in a dataset. In the context of circulation data, the top transaction type or the top book borrowed can indicate the most common or popular items or activities in the library.
- **Frequency:** Frequency calculates the number of times each value occurs in a dataset. In circulation data analysis, the frequency can be used to determine how often certain books are borrowed, how often certain transaction types occur, or how often certain users visit the library.

3.4 Python Libraries and Tools Used for Analysis

By using these Python libraries and tools, researchers can efficiently analyse circulation data, perform descriptive statistics, visualise the results, and gain valuable insights into library resource utilisation and user behaviour.

- **Pandas:** Pandas is a powerful data manipulation library in Python that provides data structures like DataFrames, which are particularly useful for working with tabular data such as circulation data. Pandas allow for data cleaning, filtering, aggregation, and analysis.

- **NumPy:** NumPy is a fundamental package for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays. NumPy is often used in conjunction with Pandas for data manipulation and analysis.
- **Matplotlib:** Matplotlib is a popular plotting library in Python that provides a variety of functions for creating static, animated, and interactive visualisations in Python. Matplotlib can be used to create plots such as histograms, line plots, and scatter plots to visualise circulation data and analysis results.
- **Seaborn:** Seaborn is a statistical data visualisation library built on top of Matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics. Seaborn is particularly useful for creating complex visualisations and is often used in conjunction with Pandas for data analysis.
- **Scikit-learn:** Scikit-learn is a machine learning library in Python that provides simple and efficient tools for data mining and data analysis. While machine learning may not be directly applicable to circulation data analysis, Scikit-learn provides useful functions for data preprocessing and clustering that can complement descriptive statistics analysis.

4. RESULTS

These basic statistics provide an overview of library activity, helping to identify high or low-traffic periods and average user behaviour. Quarterly Descriptive Statistics are as follows:

Table 1: *Descriptive Statistics: First Quarter*

	Date	**Transaction Type**	**Initial**	**Location**
Count	25727	25727	25727	25727
Unique	88	2	26	1
Top	15/07/2022	Dchrgng	Dropbox (Drop)	Miscellaneous
Frequency	788	15231	5837	25727

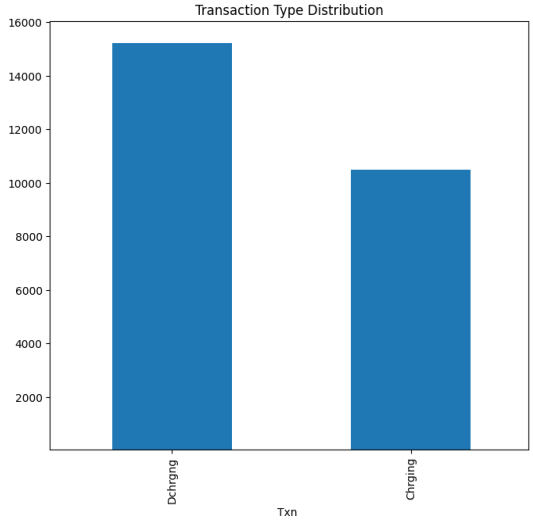

Figure 1: *Distribution of Transaction Types in First Quarter*

In the "Date" column, there are 25,727 recorded dates during the first quarter. There are 88 unique recorded entries for the date column. The maximum number of transactions happened on 15/07/2022 which are 788. There are two Transaction types i.e., Charging and Discharging named as "Dchrgng" which counts to 15,231 and "Chrging" counting to10,496. Total number of charging entries for the top date are 166 and total number of discharging entries are 622. The "Initial" column represents the employee whose ID was used for the transactions. In the first quarter there are 26 unique IDs recorded in the data. But the maximum number of times students used Drop Box in the library for return of books. Out of the 25,727 records drop box was used by students 5837 times. Which shows that students prefer technology provided by the library for easy access.

Table 2: *Descriptive Statistics: Second Quarter*

	Date	**Transaction Type**	**Initial**	**Location**
Count	46386	46386	46386	46386
Unique	88	2	27	1
Top	22/12/2022	Dchrgng	Dropbox (Drop)	Miscellaneous
Frequency	1489	25398	11392	46386

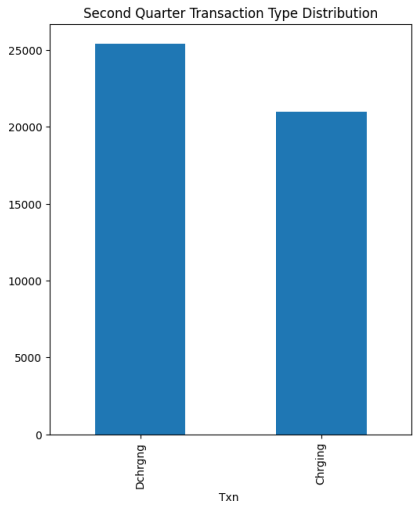

Figure 2: *Distribution of Transaction Types in Second Quarter*

In the "Date" column, there are 46,386 recorded dates during the second quarter. There are 88 unique recorded entries for the date column. The maximum number of transactions happened on 22/12/2022 which are 1489. There are two Transaction types i.e., Charging and Discharging named as "Dchrgng" which counts to 25,398 and "Chrging" counting to 20,988. Total number of charging entries for the top date are 620 and total number of discharging entries are 869. The "Initial" column represents the employee whose ID was used for the transactions. In the second quarter there are 27 unique IDs recorded in the data. In the second quarter also the maximum number of times students used Drop Box in the library for return of books. Out of the 46,386 records drop box was used by students 11392 times.

Table 3: *Descriptive Statistics: Third Quarter*

	Date	**Transaction Type**	**Initial**	**Location**
Count	39966	39966	39966	39966
Unique	86	2	23	1
Top	13/03/2023	Dchrgng	BHUPENDER TFS (BHUPE)	Miscellaneous

Frequency	949	22419	10036	39966

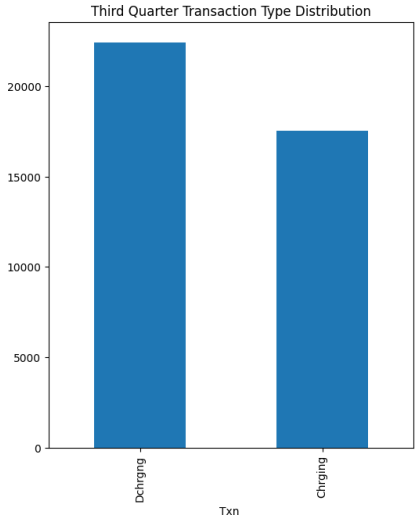

Figure 3: *Distribution of Transaction Types in Third Quarter*

In the "Date" column, there are 39,966 recorded dates during the third quarter. There are 86 unique recorded entries for the date column. The maximum number of transactions happened on 13/03/2023 are 949. Charging and Discharging named as "Dchrgng" which counts to 22,419 and "Chrging" counting to 17,547. Total number of charging entries for this date are 325 and total number of discharging entries are 624. The "Initial" column represents the employee whose ID was used for the transactions. In the third quarter 23 unique IDs are recorded in the data. In the third quarter employee ID "BHUPE" contributed most to the circulation section with 10036 transactions out of the total 39,966 with about 25 percent of overall transactions in the quarter.

Table 4: *Descriptive Statistics: Fourth Quarter*

	Date	**Transaction Type**	**Initial**	**Location**
Count	32998	32998	32998	32998
Unique	89	2	27	1
Top	05/04/2023	Dchrgng	RAJINDER (RAJIND)	Miscellaneous

Frequency	832	17952	7481	32998

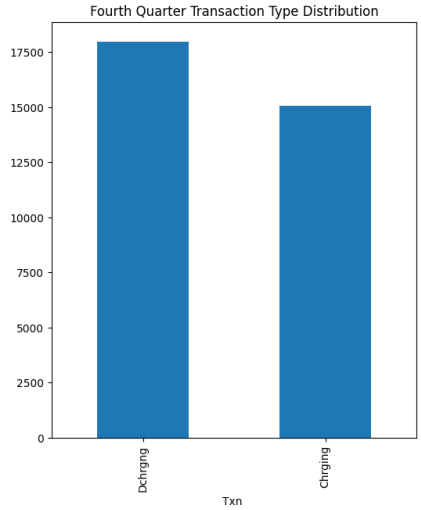

Figure 4: *Distribution of Transaction Types in Fourth Quarter*

In the "Date" column, there are 32,998 recorded dates during the fourth quarter. There are 89 unique recorded entries for the date column. The maximum number of transactions happened on 13/03/2023 which are 832. Charging and Discharging named as "Dchrgng" which counts to 17,952 and "Chrging" counting to 15,046. The day on which maximum number of transactions happened, total number of charging entries are 357 and total number of discharging entries are 475. The "Initial" column represents the employee whose ID was used for the transactions. In the fourth quarter 27 unique IDs are recorded in the data. Employee ID "RAJIND" contributed most to the circulation section with 7481 transactions out of the total 32998 with about 23 percent of overall transactions in the quarter.

5. DISCUSSION

Based on the results for each quarter of the academic year 2022-23 at Vivekananda Library, the following key observations can be made:

- Date Analysis: The number of recorded dates and unique entries for each quarter indicates the frequency and variety of transactions occurring in the library. The maximum number of transactions on a single day highlights peak usage periods, which can inform staffing and resource allocation.

- Transaction Types: The presence of two transaction types, “Charging” and “Discharging,” shows the flow of library materials, indicating borrowing and returning activities. The distribution of these transaction types can provide insights into user behaviour and preferences.
- Employee ID Analysis: The analysis of employee IDs provides an understanding of staff involvement in library transactions. Identification of top contributors like “BHUPE" and “RAJIND” sheds light on the individuals responsible for a significant portion of transactions, which can inform recognition and reward systems.
- Drop Box Usage: The frequent use of the Drop Box by students for book returns suggests a preference for convenient and technology-enabled services. This highlights the importance of offering user-friendly services to enhance user experience.

5.1 Implications of the Findings for Librarianship:

The analysis of circulation data at Vivekananda Library, Maharshi Dayanand University, Rohtak, Haryana, India, has several implications for librarianship:

- Resource Allocation: The findings can help librarians allocate resources more effectively by identifying high-demand materials and popular services. This can lead to improved user satisfaction and more efficient use of library resources.
- Collection Development: The analysis can inform collection development decisions by identifying popular materials and trends in resource utilisation. Librarians can use this information to ensure that the library's collection remains relevant and meets the needs of its users.
- Service Improvement: The findings can be used to improve library services, such as by adjusting opening hours or reorganising library spaces based on usage patterns. This can lead to a better user experience and increased engagement with library services.

5.2 Limitations of the Study:

- Single Library Focus: The study focused on circulation data from Vivekananda Library, which may limit the generalizability of the findings to other libraries.
- Data Quality: The analysis is dependent on the quality of the data collected. If the circulation data contains errors or inconsistencies, it may impact the accuracy and reliability of the findings.

- Scope of Analysis: The analysis focused on descriptive statistics and Python programming. While these methods provide valuable insights, they may not capture the full range of factors that influence user behaviour and library services.

5.3 Suggestions for Future Research:

- Longitudinal Studies: Conducting longitudinal studies to analyse circulation data over multiple academic years can provide a more comprehensive understanding of library resource utilisation patterns and trends.
- Comparative Studies: Comparing circulation data across different libraries or academic institutions can help identify best practices and factors that influence library resource utilisation.
- Qualitative Studies: Supplementing quantitative analysis with qualitative studies, such as user surveys or interviews, can provide deeper insights into user behaviour and preferences.

6. CONCLUSION

Summary of Key Findings: The study analysed circulation data from Vivekananda Library, Maharshi Dayanand University, Rohtak, Haryana, India, for the academic year 2022-23. The key findings include patterns in library resource utilisation, temporal trends in usage, popular materials, and insights into user behaviour. These findings provide valuable insights into how library services can be optimised to better meet the needs of users.

Contribution to the Field of Librarianship: This study contributes to the field of librarianship by demonstrating the value of data analytics in understanding user behaviour and improving library services. The use of descriptive statistics and Python programming highlights the practical application of data-driven approaches in librarianship. The study also contributes to the growing body of literature on library usage analytics, providing a foundation for future research in this area.

Practical Implications of the Study: The findings of this study have several practical implications for librarianship. They can inform decisions related to resource allocation, collection development, and service provision. For example, librarians can use the insights gained from the analysis to ensure that popular materials are readily available to users and to tailor services to better meet user needs. Overall, the study demonstrates how data analytics can be used to improve library services and enhance the user experience.

REFERENCES

1. Hamad, F., Fakhuri, H., & Abdel Jabbar, S. (2022). Big Data Opportunities and Challenges for Analytics Strategies in Jordanian Academic Libraries. *New Review of Academic Librarianship*, *28*(1), 37–60. https://doi.org/10.1080/13614533.2020.1764071
2. Hussain, A., & Shahid, R. (2022). Impact of big data on library services: prospect and challenges. *Library Hi Tech News*, *July*. https://doi.org/10.1108/LHTN-02-2022-0022
3. Oladokun, B. D., Aboyade, M. A., & Aboyade, W. A. (2023). Global challenge and opportunities for libraries and big data. *Library Hi Tech News*, *ahead-of-p*(ahead-of-print). https://doi.org/10.1108/LHTN-12-2022-0138
4. Orcutt, D. (2009). *Library Data: Empowering Practice and Persuasion*. Bloomsbury Publishing. https://books.google.co.in/books?id=SnfDEAAAQBAJ
5. Showers, B. (2015). *Library Analytics and Metrics: Using data to drive decisions and services*. Facet Publishing. https://books.google.co.in/books?id=dgoUDgAAQBAJ
6. Tella, A., & Kadri, K. K. (2021). Big data and academic libraries: is it big for something or big for nothing? *Library Hi Tech News*, *ahead-of-p*(ahead-of-print). https://doi.org/10.1108/LHTN-07-2020-0069
7. Zhang, J., Wolfram, D., & Ma, F. (2023). The impact of big data on research methods in information science. *Data and Information Management*. https://doi.org/10.1016/j.dim.2023.100038
8. Enakrire, R.T., 2015. The use of information and communication technologies for the support of knowledge management in selected academic libraries in Nigeria and South Africa (Doctoral dissertation, University of Zululand).
4 (2017): 12-18.

9. Enakrire, Rexwhite Tega. "The use of information and communication technologies for the support of knowledge management in selected academic libraries in Nigeria and South Africa." PhD diss., University of Zululand, 2015.

10. Chen, Hsin-liang, and Sarah Burns Gilchrist. "Online access to higher education on YouTubeEDU." New Library World 114, no. 3/4 (2013): 99-109.