iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52836

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Data Analysis Made Easy with UNkNOT

Vivek Ghildiyal[1], Priyanshu Singh[2], Aditee Mattoo[3], Rahul Gupta[4]
[1, 2, 3, 4]*Department of Computer Science Engineering,M.Tech Int., Noida Institute of Engineering and Technology, U.P 201306, India*

Abstract: ***Analysts need to be able to identify insights into the data as it grows and becomes more complex. In the business world, it's common for organizations, even small ones, to be overwhelmed by data. They have a lot of spreadsheets, databases, and other documents that need to be looked at to help them make decisions. [5]Unfortunately, this procedure takes a lot of time and a lot of manual labor.***
Additionally, it costs a lot, especially if the user implements it. The systematic application of statistical [3] and logical techniques to describe and illustrate, summarize, and evaluate data is known as data analysis [22-23]. One of the fast-growing techniques for identifying data trends is data analysis [13]. Because speed and accuracy are the foundations of this system, they are why it is so well-known.
The entire procedure by which this tool could take the place of the current method is detailed in this paper. A device that brings those fields together is integrating the various tools used by individual users. The UNkNOT tool's flexibility for integration into existing security systems and frameworks is designed to guarantee data integrity and confidentiality.
Keywords: Data Analysis Tool, Data Analytics, Data Viz, Data Visualization, Unknot, ExploratoryData Analysis

I. INTRODUCTION

Unknot Data Analysis Tool [23] is a highly customizable program that helps users unknot their databases. The tool provides users with a straightforward approach to exploring and researching their data, which can help them source new information or discover patterns within their databases.

The tool will allow its users to create their parameters for exploration and analysis. When it comes to performance monitoring there are several tools on the market, but sadly many of these tools do not allow monitoring databases in a very simple way. By making it possible to efficiently handle and manipulate large amounts of data, automate tasks, and train and deploy AI models, this tool can help with AI-based data analysis. This tool can help data scientists and AI engineers focus on building better models and analyzing the results rather than getting bogged down in data management and manipulation by offering effective methods for handling large amounts of data.

The tool will provide an easy way to access, process, and share centralized information on various stages of your workflow, as well as help with managing security policies. It also allows users to personalize their tool by altering its functionality and user interface. The Unknot tool is the ultimate data validation and extraction tool, designed to provide users with the most robust and thorough results. It provides a holistic solution in 4 phases: 1st phase focuses on working with existing data on a local device; 2nd phase focuses on taking data from the user (can be local or global) given that the device is local; 3rd phase focuses on securing the data integrity and confidentiality. And finally, after 3rd phase, the tool can be hosted on a global level/platform.

II. FRAMEWORK USED

There are several approaches and techniques are commonly used in data analysis, including statistical analysis, machine learning, and data visualization [15-16]. It is particularly useful for understanding patterns and trends in data and can be applied to various fields, including social science, medicine, and engineering. Machine learning is a rapidly growing field involving algorithms to learn from data and make predictions or decisions. It can be applied to a wide range of problems, including image recognition, natural language processing, and predictive modeling. Python [4-10-11-12] is a powerful and versatile programming language widely used for data analysis. It offers a wide range of libraries and frameworks that make it easy to work with data, perform complex calculations, and visualize results. One of the most popular libraries used for data analysis in Python is NumPy [8-9-21]. It is an open-source library that provides support for large and multi-dimensional arrays and matrices of numerical data, as well as a collection of mathematical functions to operate on these. The below table displays the used libraries along with their merits and demerits.

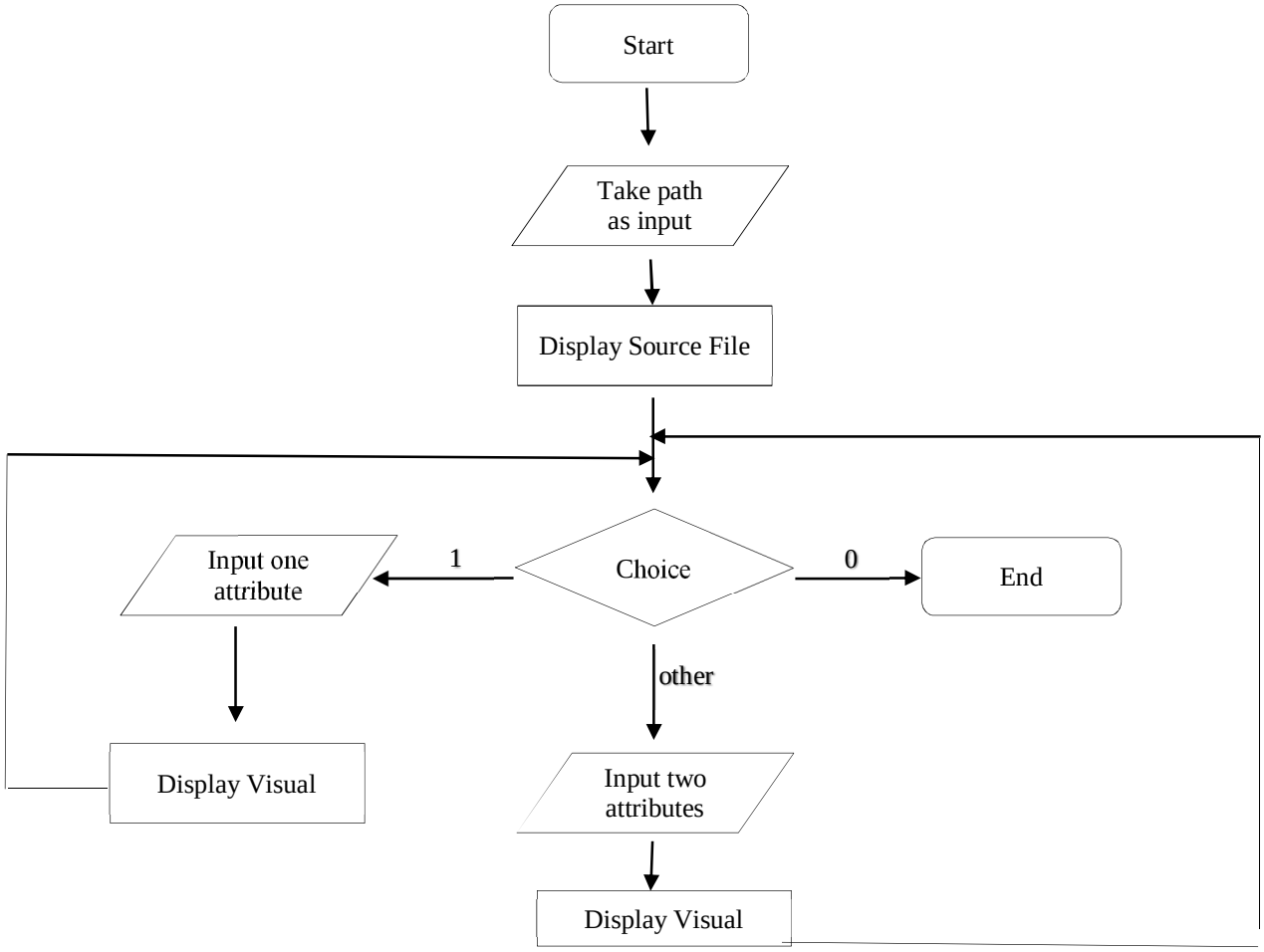
©IJRASET: All Rights are Reserved

Table 2.1: Python Libraries

S.NO.	LIBRARY USED	ADVANTAGES	DISADVANTAGES
1.	PANDAS	Pandas [21-24] provide powerful data manipulation functions, such as group by, join, and reshape, which allow for flexible and efficient data processing. Pandas also provide powerful IO tools for loading and saving data in various formats, such as CSV, Excel, JSON, and SQL.	Difficult syntax, Bad documentation.
2.	OPENPYXL	Openpyxl is a powerful Python library for reading and writing Excel files. The library supports all features of the latest version of Excel including, but not limited to, adding and removing sheets, rows, columns, and cells, as well as updating cell values and formatting.	It is currently not possible to maintain links between files.
3.	DASH	Dash [14-17] is a Python framework for building web-based data visualization and analysis tools. One of the key features of Dash is its ability to connect to a wide range of data sources.	Complex, Integration issues
4.	PLOTLY	Plotly [14-25] is a powerful data visualization library for Python. It allows developers to create interactive, web- based plots and graphics, such as scatter plots, line plots, bar plots, and more.	Confusing initial setup to use Plotly without an online account, and lots of code to write.

III. PROPOSED METHODOLOGY

UNkNOT has a unique methodology for data analysis [18-19]. Users can use UNkNOT's analytics [5] tools to review their own data, or they can use UNkNOT's charts and graphs to gain an understanding of the overall trends in their dataset. UNkNOT will work on a single data set in its initial phases. A data set will be asked from the user irrespective of the contents (can be any data). For analytics, unknot has provided several charts and graphs for the user to understand their data easily. Providing the users to also work on a specific part of the result and the ability to download it displayed in Fig. 3.2.



ig 3.2: Phase 1 Layout

IV. COMPARATIVE ANALYSIS

Unknot is a pioneer in its working field, as the tools available in the market do not serve the level of functionality expected by an expert. The benefit here is it will be the first-of-a-kind tool to be available in the market for users. Table 4.1 lays out the comparison.

Table 4.1: Comparison between Traditional and Proposed Approach

S.NO.	GOOGLE ANALYTICS [4]	TABLEAU	UNkNOT
1.	It is a free web analytics service provided by Google that tracks and reports website traffic.	It is a paid data visualization and business intelligence software used for data analysis and visualizations.	It is a free service for performing Data Analysis.
2.	It provides insights into website user behavior such as page views, sessions, bounce rate, and conversion rate.	It provides the ability to connect to a wide range of data sources and allows for the creation of interactive dashboards and reports.	It provides the ability to work with complex datasets.
3.	It is mainly used for tracking and analyzing website performance.	It is mainly used for data visualization and analysis, but also provides advanced analytics capabilities such as predictive modeling and statistical analysis.	It is mainly used for data tracking, visualization, and analysis.
4.	It provides real-time data and allows for the customization of reports through custom dimensions and metrics.	It provides a visual interface for creating reports and dashboards and allows for real-time data exploration.	It allows working with real-time data for creating reports and dashboards.

V. PHASE 1 FINDINGS

During the first part of the process, the user must enter the path of the file present in his local server/system/storage. Users will be given a list of choices from which the user can operate based on their preference for visuals shown in below figures.

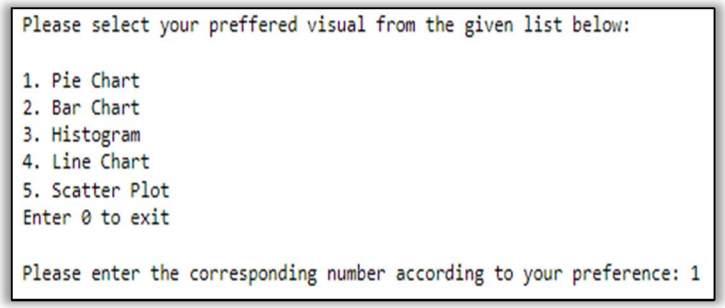


Fig 5.1: Visuals to Choose From

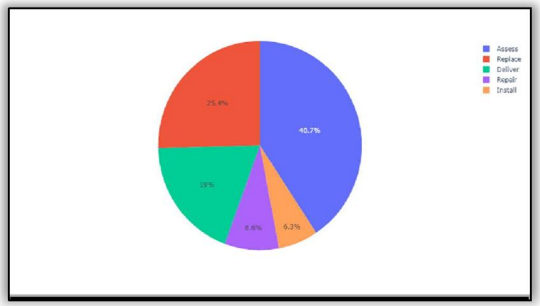


Fig 5.2: Pie Chart Visual

User will enter their preferred choice and will be asked for the respective attribute/attributes for analysis. After entering the path, the system will display data for verification. To prevent re-running of the program, the User will get the choice to choose exit or the next preferred chart/visual for analysis. In charts, users will get features like downloading the whole chart, or the preferred region only, taking visuals for a specific attribute or an area only. Users can dismiss the dashboard at any point by entering '0'.

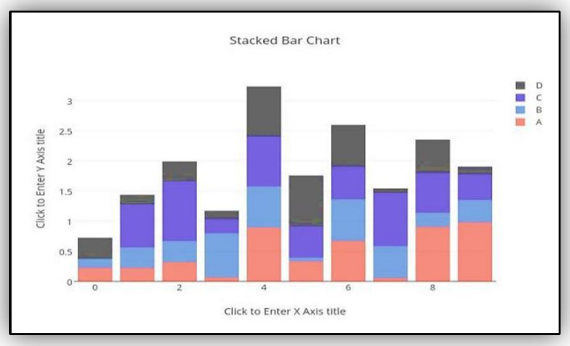


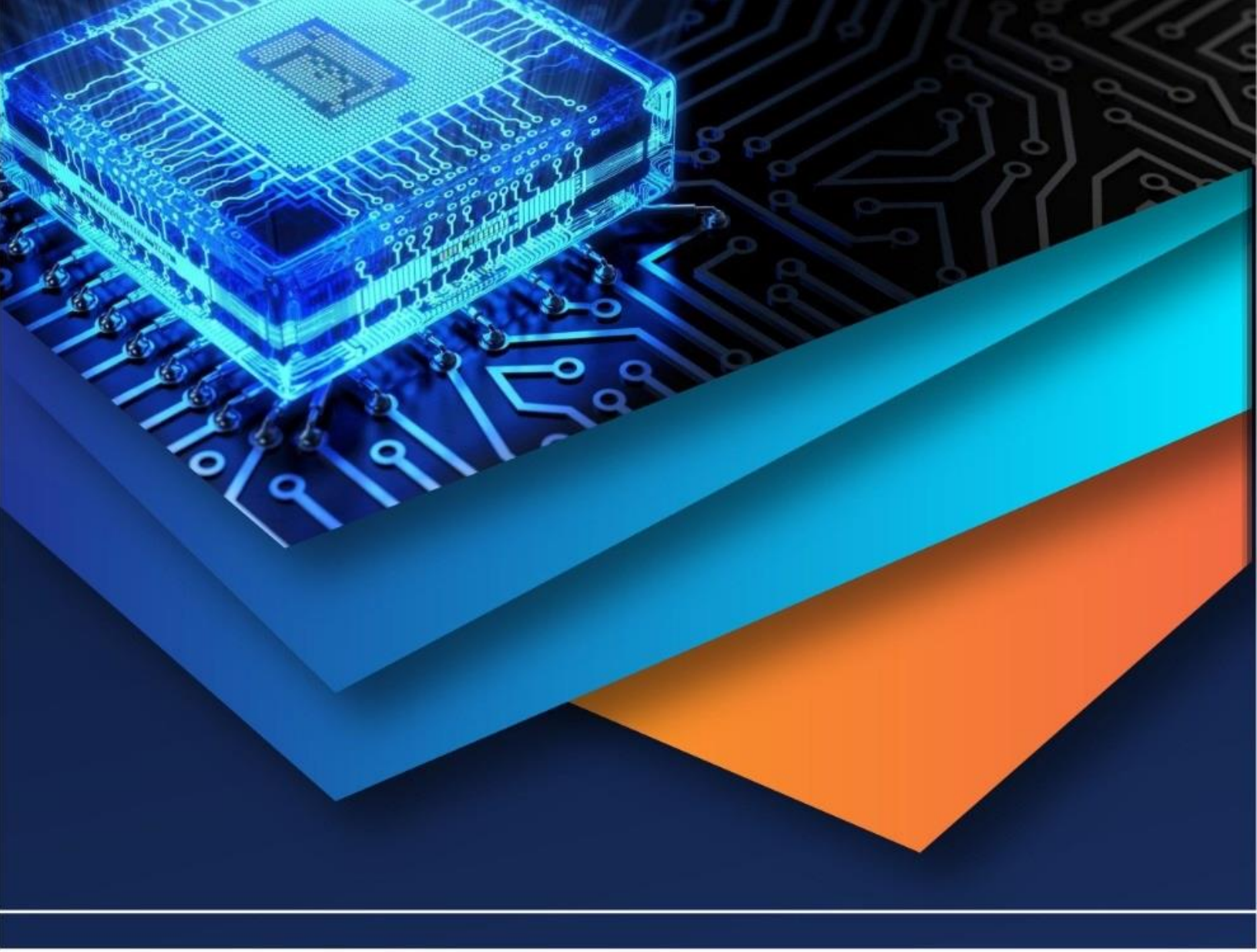
Fig 5.3: Bar Chart Visual

VI. CONCLUSION & FUTURE SCOPE

Unknot is a data analysis [1-2-7] tool that helps you under- stand your data easily and intuitively. We've completed our current phase with the help of libraries like pandas [8-21], plotly[20], and openpyxl. The dashboard [6] is a great way to display the data and provide an overall view. Updates in the upcoming version of Unknot- Analysis/Visualization of a specific number of rows or a particular set of data, more interactive environment, predictions based on data, analysis of more than attributes, and multiple dataset analysis[8]. The current version of UNKNOT is in its initial state, taking data from the end user and doing visualization on it. With the aim of making UNkNOT available to users with no prior knowledge of Data Analysis [16- 17], it is planned that it will go full global—or more precisely, be made available around the world. UNkNOT's functionality is currently at the end phase of 1 and the beginning of phase 2, but many features from phase 3 have also been included. The security focus will remain in the new version, which will also offer more functionality. Phase 1, in its simplicity, only presents visual representations of the data it inputs. Phase 2 attempts to make progress on the project by working with user-provided data. For now, this data is stored locally—in the future, it will be possible for users in other locations to provide global analysis. The Visualizer [24] will undergo several upgrades during its development. This is when a new user with no prior knowledge of data analysis will be able to work on UNkNOT.

REFERENCES

[1] K. Johnson, B. Lee, and J. Smith. (2020). Data analysis methods for large datasets. Journal of Big Data, 7(2), 23-38.
[2] S. Chen, X. Zhang, and Y. Liu. (2021). Machine learning approaches for predictive analytics. Data Mining and Knowledge Discovery, 35(1), 73-8
[3] W. McKinney, "pandas: a foundational Python library for data analysis and statistics", Python for High Performance and Scientific Computing, vol. 14, no. 9, 2011
[4] X. Cai, H. Langtangen and H. Moe, "On the Performance of the Python Programming Language for Serial and Parallel Scientific Computations", Scientific Programming, vol. 13, no. 1, pp. 31-56, 200
[5] J. Van Der Donckt, J. Van der Donckt, E. Deprost and S. Van Hoecke, "Plotly-Resampler: Effective Visual Analytics for Large Time Series," 2022 IEEE Visualization and Visual Analytics (VIS), Oklahoma City, OK, USA, 2022, pp. 21-25, doi: 10.1109/VIS54862.2022.00013
[6] G. Iyer, S. DuttaDuwarah and A. Sharma, "DataScope: Interactive visual exploratory dashboards for large multidimensional data," 2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC), Phoenix, AZ, USA, 2017, pp. 17-23, doi: 10.1109/VAHC.2017.8387496
[7] Kabita Sahoo, Abhaya Kumar Samal, Jitendra Pramanik, and Subhendu Kumar Pani. Exploratory data analysis using python. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019
[8] Wes McKinney. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. OReilly Media, Inc., 2012
[9] Fabio Nelli. Python data analytics: Data analysis and science using PANDAs, Matplotlib and the Python Programming Language. Apress, 2015.
[10] Dr Ossama Embarak, Embarak, and Karkal. Data analysis and visualization using python. Springer, 2018.
[11] Pramanik, Jitendra & Samal, Abhaya Kumar & Sahoo, Kabita & Pani, Dr. Subhendu. (2019). Exploratory Data Analysis using Python. International Journal of Innovative Technology and Exploring Engineering. 8. 4727-4735
[12] Kiranbala Nongthombam , Deepika Sharma, 2021, Data Analysis using Python, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 07 (July 2021
[13] Wes McKinney and the Pandas Development Team,pandas: powerful Python data analysi
[14] Stancin, Igor and Alan Jović. “An overview and comparison of free Python libraries for data mining and big data analysis.” 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (2019): 977-982
[15] Harshal S. Kudale, Mihir V. Phadnis, Pooja J. Chittar, Kalpesh P. Zarkar,DATA ANALYSIS AND VISUALIZATION OF OLYMPICS USING PYSPARK AND DASH-PLOTLY,202
[16] Pritchard, L., White, J. A., Birch, P. R. J., Toth, I. K. GenomeDiagram: a python package for the visualization of large-scale genomic data. Bioinformatics, Volume 22, Issue 5, 1 March 2006, Pages 616–617. DOI: 10.1093/bioinformatics/btk021
[17] Shammamah Hossain,Visualization of Bioinformatics Data with Dash Bio,201
[18] Nagpal, Abhinav & Gabrani, Goldie. (2019). Python for Data Analytics, Scientific and Technical Applications. 140-145. 10.1109/AICAI.2019.8701341
[19] Wes McKinney, Python for Data Analysis(BookZZ.org),201
[20] Carson Sievert,Interactive web-based data visualization with R, plotly, and shiny(CRC press),202
[21] Nelli, Fabio. (2018). Python Data Analytics: With Pandas, NumPy, and Matplotlib. 10.1007/978-1-4842-3913-1
[22] "Data Wrangling with Python" by Jacqueline Kazil and Katharine Jarmul (2017) - O'Reilly Media, ISBN: 978-1491948811
[23] "Data Analysis with Pandas and Python" by Fabio Nelli (2017) - Packt Publishing, ISBN: 978-1787125933
[24] "Hands-On Data Analysis with Pandas" by Kevin Markham (2019) - Packt Publishing, ISBN: 978-1801092913
[25] "Python for Data Analysis and Visualization: A Hands-On Guide to Pandas, Matplotlib, Seaborn and Plotly" by Hadelin de Ponteves (2021) - Udemy, ISBN: 978-1801249073



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)