The academic performance of students, particularly in challenging core subjects like Mathematics and Reasoning, remains a critical area of concern, as highlighted in various studies. Recent advancements in machine learning and data mining techniques offer promising solutions for analysing educational datasets and predicting student success. [1] Kumar (2021) conducted a survey leveraging the student performance dataset from the UCI Machine Learning Repository, utilising Python and Jupyter Notebook for exploratory data analysis and grade prediction. This work demonstrates the potential of machine learning models to extract meaningful insights from raw data, providing a data-driven approach to understanding and improving academic outcomes. The study underscores the importance of technological intervention in identifying key factors influencing student performance and devising strategies to address learning challenges.

Various studies highlight the potential of data analysis and machine learning in addressing domain-specific challenges. For instance, Kumar (2021) utilized machine learning techniques to predict student performance, demonstrating how exploratory data analysis can uncover meaningful insights from educational datasets. Similarly, Da Silva et al. [16] introduced *Dominoes*, an interactive data exploration tool designed for project managers and developers to analyze software relationships. Unlike traditional approaches that rely on post hoc analyses or complex scripts, *Dominoes* enables users to interactively investigate project data, visualize intermediate results, and save exploration paths for reuse. Through scenario-based evaluations, the tool achieved an 86% success rate, highlighting its effectiveness in answering project-specific questions within constrained timeframes. Together, these works illustrate the growing role of advanced data analysis tools and machine learning in providing actionable insights across diverse domains, including education and software engineering.

Hassan-Montero et al. [17] introduced *SCImago Graphica*, a no-code tool designed to simplify the creation of complex data visualizations through drag-and-drop interactions. By allowing users to bind data variables to various encoding channels and adjust their settings, the tool generates interactive graphical displays that are suitable for both exploratory data analysis and data communication. The study evaluates the tool's expressiveness and ease of use through diverse examples and a visualization catalog, demonstrating its capability to quickly and efficiently produce a wide range of data visualizations. *SCImago Graphica* addresses the long-standing challenge of combining high expressive power with user-friendliness, making it a valuable resource for data exploration and presentation.

Palocsay et al. [18] proposed using Microsoft Excel as an introductory tool for business intelligence (BI) and decision support system (DSS) applications, emphasizing its accessibility and versatility for managers and business students. The study highlights three key capabilities of Excel: manipulating records as a database, creating PivotTables® and PivotCharts® for data analysis, and importing data as an automation container. These foundational skills enable users to leverage Excel's potential for exploratory data analysis and support information discovery in business contexts. By demonstrating how Excel can serve as a gateway to advanced BI tools, the article underscores its importance in fostering data-driven decision-making across organizations.

Mackeprang et al. [19] introduced *Kaleidoscope*, an exploratory data analytics tool designed to assist in the critical and complex task of evaluating and selecting ideas during collaborative ideation. By integrating automatic computational methods with human sensemaking, *Kaleidoscope* allows users to explore and annotate ideas interactively. The tool’s design principles emphasize semantic technologies to support intuitive data exploration. Based on qualitative feedback on a prototype, the study identified potential enhancements and outlined directions for future development. This work highlights the importance of combining computational efficiency with human judgment to ensure valuable ideas are not overlooked in ideation processes.

Furcila [20] proposed *InTool Explorer*, an innovative interactive exploratory data analysis tool designed to address the complexities inherent in studying Alzheimer's disease (AD). AD is characterized by progressive cognitive decline and diverse neuropathological changes that vary across brain regions and between patients, making data analysis challenging. *InTool Explorer* integrates multidisciplinary data, including histological, neuropsychological, and clinical variables, into a unified platform. In a study using data from 11 AD patients, the tool enabled the identification of regional differences in tau and amyloid protein expression, as well as patient-specific characteristics related to disease progression. This interactive tool demonstrates the potential to uncover novel insights beyond conventional statistical analysis, improving neuroscientific research productivity and enhancing understanding of AD.

Pitroda [21] introduced *QuickViz*, an interactive web application tool designed to automate the Exploratory Data Analysis (EDA) phase of the data science lifecycle. Data science has become integral across industries, requiring efficient methods for deriving insights from complex datasets. EDA, a crucial early phase, helps identify irregularities and uncover patterns in data but can be time-intensive when performed manually. *QuickViz* simplifies this process by offering a user-friendly, no-code platform that allows users to upload datasets and explore descriptive and graphical EDA techniques. The tool's web-based dashboard facilitates the visualization of findings, making it accessible to stakeholders without prior programming knowledge. By conserving time and enhancing accessibility, *QuickViz* empowers users to derive insights efficiently and present them effectively.

Otero-Escobar and Velasco-Ramírez [22] emphasized the pivotal role of Exploratory Data Analysis (EDA) in educational data mining, highlighting its application in analyzing academic achievement among high school students in Veracruz, Mexico. By utilizing tools like Google Colaboratory and Python libraries such as Pandas, Numpy, Matplotlib, and Seaborn, the study identified key variables influencing academic performance while ensuring data integrity. EDA techniques, including histograms, provided insights into data dispersion, concentration, and outliers, offering actionable recommendations for adjusting teaching strategies and educational tools. This research underscores how EDA not only enhances understanding of student data but also supports interventions aimed at reducing dropout rates and improving academic outcomes.

Chen et al. [23] introduced WhatsNext, an interactive notebook framework designed to enhance exploratory data analysis (EDA) for users with limited coding expertise. The framework aims to support low-code visual data exploration by integrating a recommendation panel in notebook cells, suggesting potential next-step exploration questions or actions. Additionally, WhatsNext features a dynamic tree visualization, allowing users to easily trace analytic dependencies between notebook cells. This approach addresses the challenge of disorganized data analysis workflows in traditional computational notebooks, offering a structured, insight-driven environment for EDA, which enhances both the accessibility and effectiveness of data exploration.

In an era marked by rapid technological advancements, individuals strive to keep their knowledge and skills up-to-date, often at the expense of their health and well-being. While physical fitness is often considered a priority for maintaining a healthy lifestyle, eating habits—the cornerstone of nourishment—are frequently neglected. According to RamyaSri et al. [9], food provides the essential proteins and minerals necessary to sustain productivity and overall health. Despite this, global obesity rates have risen by 27.5% over the past 33 years, highlighting the widespread disconnect between perceived and actual healthy eating practices. This study investigates generational views on healthy lifestyles by analyzing daily eating habits and individuals’ perceptions of their health, aiming to uncover the gap between self-reported and actual dietary behaviours.

Exploratory Data Analysis (EDA) plays a pivotal role in understanding and preparing data for predictive modeling. Deming et al. [24] elaborate on the advantages of EDA, emphasizing its utility in preprocessing data and setting a foundation for statistical and machine learning models. By illustrating how predictive modeling techniques vary in performance when applied to the same dataset, the study underscores the importance of proper data exploration and validation. Through detailed discussions on data comprehension, preparation, and visualization, this work highlights EDA's significance in ensuring robust analytical outcomes and forming the basis for effective business analytics.

Exploratory Data Analysis (EDA) is a cornerstone of the data preparation and cleaning process, often consuming a significant portion of the entire statistical analysis workflow. Rahmany et al. [25] emphasize the iterative and dynamic nature of EDA, highlighting its role in uncovering patterns, detecting anomalies, testing hypotheses, and validating assumptions through descriptive statistics and graphical tools. The study also underscores the accessibility of EDA, which allows both statisticians and non-statisticians to derive actionable insights from data. Despite its longstanding presence in data analysis, EDA remains a critical area of research, with continuous advancements and applications reinforcing its importance in modern data science practices.

The analysis of resistomes in complex microbial communities, such as the human microbiome, has been significantly advanced by whole metagenomic sequencing. However, despite the development of numerous pipelines for data processing and annotation, the lack of user-friendly tools for visual, statistical, and functional analysis remains a key bottleneck. Dhariwal et al. [26] address this gap with ResistoXplorer, a web-based tool designed to simplify high-throughput analysis of resistome data. ResistoXplorer integrates statistical and visualization advancements with functional annotations to streamline the exploration of antimicrobial resistance data. It features three modules: the *Antimicrobial Resistance Gene Table* for profiling and comparative analysis, the *Integration* module for exploratory analysis of resistome and microbiome profiles, and the *Antimicrobial Resistance Gene List* for network visual analytics to identify associations between resistance genes and microbial hosts. This tool lowers the technical barrier for resistome analysis, facilitating biological insights and driving progress in antimicrobial resistance research.

References

1. Da Silva, J. R., Junior, Campagna, D. P., Clua, E., Sarma, A., & Murta, L. (2020). Dominoes: An Interactive Exploratory Data Analysis Tool for Software Relationships. *IEEE Transactions on Software Engineering*, *48*(2), 377–396. <https://doi.org/10.1109/tse.2020.2988241>
2. Hassan-Montero, Y., De-Moya-Anegón, F., & Guerrero-Bote, V. P. (2022). SCImago Graphica: a new tool for exploring and visually communicating data. *El Profesional De La Informacion*. <https://doi.org/10.3145/epi.2022.sep.02>
3. Palocsay, S. W., Markham, I. S., & Markham, S. E. (2009). Utilizing and teaching data tools in Excel for exploratory analysis. *Journal of Business Research*, *63*(2), 191–206. <https://doi.org/10.1016/j.jbusres.2009.03.008>
4. Mackeprang, M., Strama, J., Schneider, G., Kuhnz, P., Benjamin, J. J., & Müller-Birn, C. (2018). Kaleidoscope. *Adjunct Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, 75–77. <https://doi.org/10.1145/3266037.3266106>
5. Furcila, D. (2017). [P2-426]: INTOOL EXPLORER: A NEW INTERACTIVE EXPLORATORY DATA ANALYSIS TOOL IN ALZHEIMER’s DISEASE. *Alzheimer S & Dementia*, *13*(7S\_Part\_16). <https://doi.org/10.1016/j.jalz.2017.06.1082>
6. Pitroda, H. (2022). A Proposal of an Interactive Web Application Tool QuickViz: To Automate Exploratory Data Analysis. *2022 IEEE 7th International Conference for Convergence in Technology (I2CT)*. <https://doi.org/10.1109/i2ct54291.2022.9824068>
7. Otero-Escobar, A. D., & Velasco-Ramírez, M. L. (2023). Study on Exploratory Data Analysis Applied to Education. *Study on Exploratory Data Analysis Applied to Education*. <https://doi.org/10.1109/icev59168.2023.10329702>
8. Chen, C., Hoffswell, J., Guo, S., Rossi, R., Chan, Y., Du, F., Koh, E., & Liu, Z. (2023). WhatsNext: Guidance-enriched Exploratory Data Analysis with Interactive, Low-Code Notebooks. In *IEEE Symposium on Visual Languages and Human Centric Computing (VL/HCC)* (Vol. 15, pp. 209–214). <https://doi.org/10.1109/vl-hcc57772.2023.00033>
9. Deming, C., Dekkati, S., & Desamsetti, H. (2018). Exploratory Data Analysis and Visualization for Business Analytics. Asian Journal of Applied Science and Engineering, 7(1), 93–100. <https://doi.org/10.18034/ajase.v7i1.53>
10. Rahmany, M., Zin, A. M., & Sundararajan, E. A. (2020). COMPARING TOOLS PROVIDED BY PYTHON AND R FOR EXPLORATORY DATA ANALYSIS. *IJISCS (International Journal of Information System and Computer Science)*, *4*(3), 131. <https://doi.org/10.56327/ijiscs.v4i3.933>
11. Dhariwal, A., Junges, R., Chen, T., & Petersen, F. C. (2021). ResistoXplorer: a web-based tool for visual, statistical and exploratory data analysis of resistome data. *NAR Genomics and Bioinformatics*, *3*(1). <https://doi.org/10.1093/nargab/lqab018>