Comparative analysis of selected data visualization methods

Damian Węcławski*, Radosław Tomczyk, Paweł Powroźnik

Department of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract

This article compares five data visualization tools: Tableau, Power BI, Python libraries, Apache Superset, and Metabase. The paper comprises two studies. The first study analyzes general data visualization capabilities, comparing commercial solutions (Tableau, Power BI) and Python libraries. The second study focuses on the use of Tableau, Apache Superset, and Metabase in the context of big data. In both studies, user satisfaction was measured through surveys conducted after participants completed tasks designed for specific analyses. The results provide insights into the strengths and limitations of each tool in terms of functionality and usability.

Keywords: Big Data; Data Visualization; Tableau; Power BI; Python; Apache Superset; Metabase

*Corresponding author

Email address: **damianweclawski@gmail.com** (D. Węcławski)

1. Introduction

In today's data driven world, the ability to transform complex data sets into clear, compelling visual narratives is essential for effective decision making. Whether you are a business analyst, data analyst or manager, choosing the right data visualization tool can make a significant difference in clarity, accessibility and actionability.

The main purpose of this article is to provide a comparative analysis by evaluating several commonly used data visualization methods, focusing on their efficiency and usability. A selection of popular tools will be evaluated, including Tableau, Power BI, and Python libraries such as Seaborn, Plotly and Matplotlib. The comparison will include an evaluation of each tool based on a number of criteria.

2. Related works

Liu et al. conducted an extensive survey of the various resources available for visualization. The authors have identified and classified a wide range of resources to support audiences from beginners to advanced practitioners. Their study focuses on collections of resources of different types, such as literature-based, web-based, developer focused and special topics, providing a reader-centric resource classification [1].

In their study, Skender and Manevska analyze various data visualization tools, highlighting their importance in transforming complex data into comprehensible visual formats. They point out that effective data visualization makes it easier for the human brain to process and interpret large data sets, thus supporting informed decision-making. As a result, they listed the advantages and disadvantages of each tool. They state that data visualization has become an essential technique for companies and in general for the daily needs of visual data analysis in various aspects of life [2].

Balaji et al. carried out a case study focusing on data analysis and visualization techniques, particularly in handling machine-generated log data. They indicated that textual data can be difficult to analyze quickly, and the use of visualization techniques can help better understand the data at a glance. The study examined the use of Splunk and Tableau to process and visualize data, transforming it into comprehensible visual representations. The authors stated that effective data visualization not only helps analyze data quickly, but also improves decision-making processes, transforming huge amounts of data into the most relevant information [3].

Akhtar et al. demonstrated how Tableau features such as data blending and real-time collaboration can be used to analyze and visualize COVID-19 data, increasing the transparency and accessibility of information. They summarize that with Tableau, you can quickly access reliable information and even analyze data yourself, whether you are a scientist, researcher or ordinary person [4].

Kadam and Akhade study provides an insightful analysis of several well-known data visualization tools, highlighting their features, strengths and weaknesses. The survey brings attention to the evolution of data visualization from static charts to dynamic, interactive representations, emphasizing the importance of choosing tools that are suited for specific data types, target audiences and intended purposes. Among the tools evaluated are Tableau, Power BI and Python libraries, Matplotlib and Seaborn. The authors advise a careful assessment of needs when choosing a visualization tool, taking into account factors such as data complexity and user proficiency. They also highlight emerging trends in data visualization, including the integration of augmented reality, virtual reality, machine learning and artificial intelligence, which are expected to increase the power, availability and usability of these tools in the future [5].

Lavanya et al. look into different data visualization tools, focusing on their features, strengths and weaknesses. The authors cover popular tools like Tableau, Power BI, Google Data Studio, D3.js and Python libraries such as Matplotlib, Seaborn and Plotly. They point out that while Tableau and Power BI offer user-friendly interfaces and robust features, tools such as D3.js and Python libraries provide greater customization, albeit with a steeper learning curve. Future trends that the study identified include the integration of augmented and

virtual reality, as well as the integration of machine learning and artificial intelligence into data visualization tools [6].

Dogadina and Voronin evaluates several prominent data visualization tools, focusing on their functionalities, strengths, and weaknesses. The tools that the evaluation was conducted on include Tableau, Power BI, D3.js and Matplotlib. The analysis shows that each tool has unique strengths suited to different user needs, making it important to choose the right tool based on our requirements, experience and the type of data involved [7].

Zion and Tripathy emphasize the need to select appropriate visualization tools to effectively manage and interpret large, dynamic and heterogeneous datasets. They discuss the evolution from traditional systems that were limited in their ability to handle diverse and voluminous data, to modern tools capable of handling the complexity of large data sets [8].

Chavan provides analysis of various libraries and business intelligence software showcasing their capabilities in the field of data visualization. He identifies challenges that arise when using visualization tools, this includes, ensuring data quality and data integration. The study foresees the possibility of a number of improvements to these tools, such as automated visualization creation using AI and ML capabilities, cross-platform integration and making them more accessible to users with disabilities [9].

Santos, Sério, Abrantes, Sá, Loureiro, Wanzeller, and Martins explore the capabilities of open-source BI tools, focusing specifically on Metabase and Redash. Their study compares these platforms in terms of overall performance, key features, and suitability for organizational needs, emphasizing that the choice of a BI tool should align with a company's specific requirements to gain actionable insights [10].

Michele, Fallucchi, and De Luca present a case study within the Italian government's Data Analytics Framework (DAF), demonstrating the use of Apache Superset to create dashboards and data stories from public administration datasets. Their work highlights how Superset can support interoperability, facilitate large-scale data analysis, and provide clear and accessible visual representations, using tourism data from Sardinia as a real-world example [11].

Purich, Srinivasan, Correll, Battle, Setlur, and Crisan conduct a large-scale, systematic analysis of dashboards in the wild using Tableau Public. They introduce a schematic representation of dashboards as node-link graphs to capture spatial and interactive structures, applying it to a dataset of 25,620 dashboards. This study identifies common building blocks and design patterns, contributing a publicly available corpus to guide future research and dashboard development [12].

3. Materials and methods

3.1. Datasets

The datasets used in this analysis contains detailed information on the course and impact of the COVID-19 pandemic. It consists of seven files, each of which offers detailed insights into various aspects of the epidemic. These files contain data on confirmed cases, deaths and recoveries. Depending on the file, the data is grouped by various dimensions, such as country, specific dates, and in some cases more detailed territorial divisions such as states, provinces or regions [13, 14].

The second study used a public dataset of sales transactions from the UC Irvine Machine Learning Repository, containing approximately 50,000 records with information on products, categories, purchase time, location, and transaction value. The dataset was processed and loaded into a MySQL relational database, with which all three tools were integrated [15].

3.2. Tools

3.2.1. Tableau

Tableau is a leading data visualization and business analysis tool known for its user-friendly drag and drop interface and powerful visual analysis capabilities. It provides users with the ability to create interactive dashboards and dynamic reports without requiring extensive technical skills. Tableau supports a wide range of data sources. Its strength lies in its visually intuitive design, ease of sharing visualizations and strong support for storytelling with data.

3.2.2. Power BI

Power BI, developed by Microsoft, is a complex business analytics platform that integrates with the Microsoft ecosystem. Like Tableau, it offers an intuitive interface for creating interactive dashboards and reports, but with a particular focus on integration with existing enterprise tools. Power BI is particularly suited to organizations already using Microsoft products and offers solid capabilities for real-time data integration, data modeling and collaborative reporting.

3.2.3. Python libraries

Python's visualization libraries offer a powerful and flexible approach to data visualization, designed mainly for users with programming knowledge.

Matplotlib is the primary library for static 2D charts in Python. It provides wide control over chart elements, making it ideal for highly customized visualizations, though it generally requires more coding.

Seaborn, built on Matplotlib, simplify the process of creating statistical graphics. It includes built-in themes and functions for complex visualizations such as heatmaps and pair plots, all with minimal code.

Plotly is a modern, interactive visualization library that supports both 2D and 3D plots. Unlike Matplotlib and Seaborn, Plotly allows to create interactive dashboards and web visualizations with zoom, pan and hover features.

3.2.4. Apache Superset

Apache Superset is an open-source data visualization and business intelligence platform developed under the Apache Software Foundation. It offers more than 40 chart types and includes its own SQL query engine,

allowing advanced analytics directly within the interface. Superset provides strong options for role man-age-ment, user permissions, and large-scale data integra-tion. However, the platform requires more technical knowledge from users, particularly during installation, configuration, and when building complex dashboards. Its flexibility makes it attractive for experienced ana-lysts and organizations looking for scalable, customiza-ble BI solutions.

3.2.5. Metabase

Metabase is an open-source BI tool focused on simplic-ity and accessibility. Its interface allows users to create queries without SQL knowledge, making it a popular choice in non-technical teams. Metabase supports quick creation of dashboards and reports through a user-friendly graphical interface. Its main advantages are ease of use, fast deployment, and low entry thresh-old. On the other hand, Metabase provides fewer chart types and lim-ited customization options compared to Tableau or Su-perset, which may restrict its use in more advanced ana-lytical projects.

3.3. Scenarios

The user of the study begins by watching a prepared video showing the basics of using the tool. The user then receives a set of tasks including a description and an ex-ample of the end result. After completing a set of tasks, the user evaluates the tool by filling out a survey.

The aim of the second study was to empirically ana-lyze three Big Data visualization tools: Tableau, Apache Superset, and Metabase, from the perspective of user us-ability and efficiency. The study was comparative in na-ture and based on participants' performance on tasks us-ing a single dataset and a survey assessing their experi-ence working in each environment.

3.3.1. Tableau scenario

Using the worldometer_data.csv file, create a symbol map chart showing all cases per 1 million population. The size and color of each marker on the map should depend on the value of the parameter. Adjust the size of the mark-ers so that smaller values are visible.

Using the country_wise_latest.csv file, create a hori-zontal bar chart showing recovered per 100 cases by country. Use a filter to remove records with a value of 100 and limit the number of records displayed to the 20 countries with the highest parameter value. Each bar of the chart should have a label with the parameter value. Sort the values in descending order.

Use WHO-COVID-19-global-daily-data.csv to create two area charts showing the number of cumulative cases and cumulative deaths by date. Change the X-axis labels to show the month and year number. Add trend lines to both charts.

3.3.2. Power BI scenario

Using the worldometer_data.csv file, create a filled map graph showing all cases per1 million population. The

color of each country on the map should depend on the value of the parameter.

Using the country_wise_latest.csv file, create a col-umn chart showing Recovered per 100 cases by country. Each bar of the chart should have a label with the param-eter value. Ensure that the parameter value is displayed after the decimal point. Sort the values in ascending or-der. Use filters to remove records with a value of 100 and limit the number of records displayed to the 20 countries with the highest parameter value. Make sure that 20 val-ues are displayed after applying the second filter. Adjust the Y-axis range so that the lower range starts with the minimum parameter value.

Use WHO-COVID-19-global-daily-data.csv to create a double area chart showing the number of new cases and new deaths by date. Change the X-axis labels to show the month name and year. Add a median line to both data series. Change the chart title to "New Deaths and New Cases by Date".

3.3.3. Second study scenario for Tableau, Apache Su-perset, and Metabase

First, create a bar chart displaying the number of transac-tions per product category, showing the total count of transactions on each bar and sorting them in descending order. Next, create a line chart to visualize revenue trends over time, formatting the X-axis to display months and years and adding a trend line to highlight revenue changes. Then, calculate two key performance indicators: the total transaction value using SUM(amount) and the total number of unique transactions using COUNT_DIS-TINCT(transaction_id). Finally, assemble an interactive dashboard with at least three visualizations, including the bar chart, line chart, and KPIs, and enable global filters so users can dynamically explore the data.

4. Experiments

4.1. Participants

The first study group consisted mainly of students be-tween the ages of 23 and 25. This core group consists mostly of those in the final stages of undergraduate or early postgraduate studies. In addition, the study also in-cluded a smaller number of older participants, providing a broader perspective and introducing some diversity to the results. The entire group consisted of 20 people.

The second study group of tools evaluation included 9 participants with technical experience (data analysts, programmers) and 11 without technical experience (mar-keting staff, non-technical students, purchasing depart-ment employees), enabling differentiated assessment based on user technical level.

4.2. Evaluation criteria

One of the criteria examined was ease of use. The evalu-ation was influenced by whether the tool's interface is in-tuitive and whether an inexperienced user can quickly learn to use the tool.

Other criteria tested were visualization capabilities, such as the number and variety of available chart types,

interactivity and the ability to control the appearance of visualization elements.

It was also checked whether the tools have an active user community or forums to provide users with additional support.

4.3. Results and discussion

4.3.1. Tableau and Power BI survey

A significant number of respondents rated highly the intuitiveness of the interface of both tools, with a slight advantage for Tableau. Despite a few low ratings, it is important to remember that for the majority of respondents this was their first contact with the tool. Detailed results are shown in Figure 1.
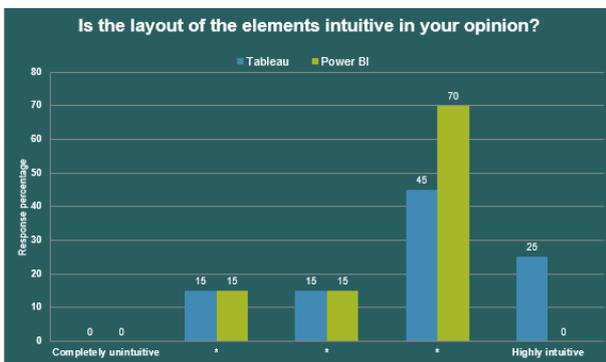
Figure 1: Results of the first question of the survey.

When it came to finding the needed elements, there was no significant difference in results. The fact that almost 50 percent of respondents created the required charts without additional help is a good result, knowing that for most of them this was their first contact with the tool. Detailed results are shown in Figure 2.
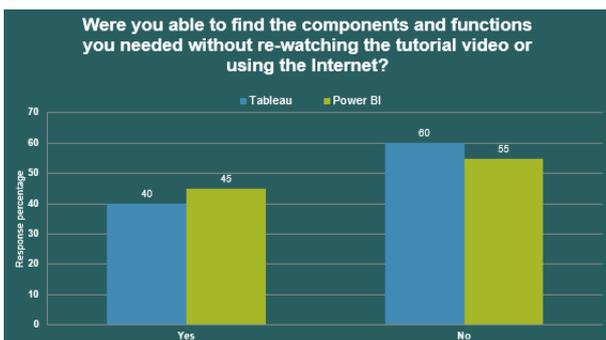
Figure 2: Results of the second question of the survey.

For both tools, most respondents had not used the tool before. Detailed results are shown in Figure 3.
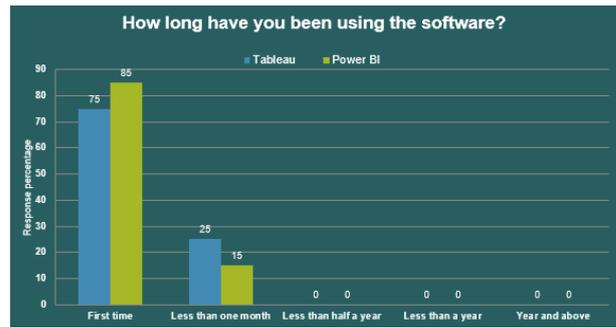

Figure 3: Results of the third question of the survey.

For experience with other data visualization tools, the level of respondents was more divided. The declarations of those testing Power BI were almost evenly split between no or very little experience and medium experience. For those testing Tableau, most respondents had basic familiarity with other tools, and there were a few who were proficient with them. Detailed results are shown in Figure 4.
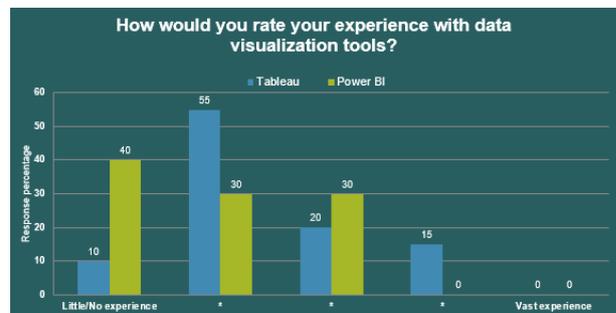

Figure 4: Results of the fourth question of the survey.

Some of the respondents testing Power BI were neutral when it came to the variety of visualization types. For both tools, the vast majority of respondents were satisfied. It is worth noting that respondents evaluated only the visualization types available by default. Detailed results are shown in Figure 5.
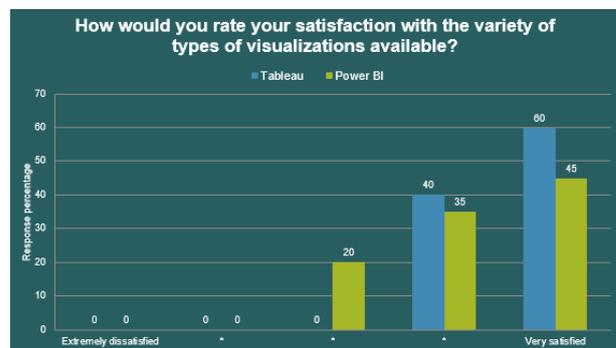

Figure 5: Results of the fifth question of the survey.

Respondents were satisfied with the Responsiveness of the interactive elements for both tools. In the case of Power BI, some respondents' evaluations were neutral. Detailed results are shown in Figure 6.
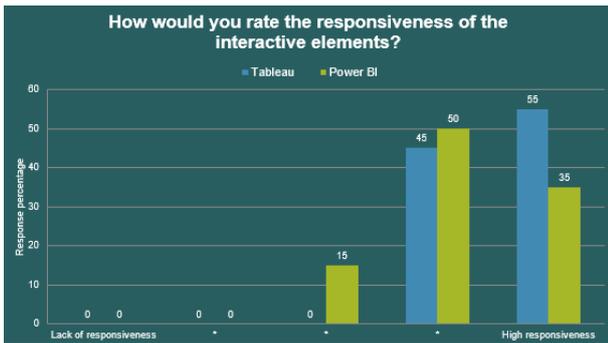
Figure 6: Results of the sixth question of the survey.

In terms of satisfaction with customizing visual elements, Power BI achieved the greatest advantage over Tableau here. Both tools were rated mostly positively. Detailed results are shown in Figure 7.
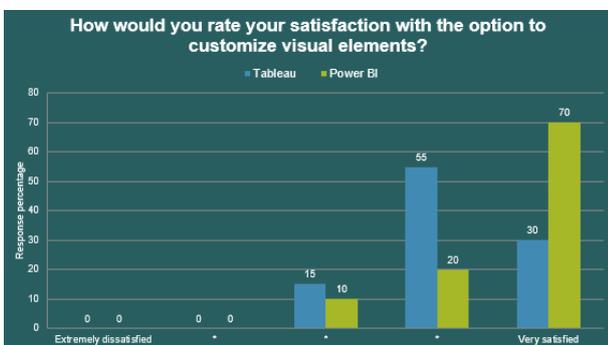

Figure 7: Results of the seventh question of the survey.

4.3.2. Charts using Python libraries

The first graph shows new cases and new deaths by month. The data is presented as a double area chart with additional trend lines. Additionally, the charts have been formatted to improve readability. To create this graph, the matplotlib library was used. The resulting graph is shown in Figure 8.
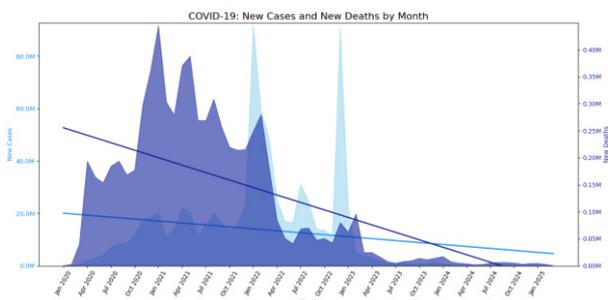

Figure 8: Matplotlib area charts.

The second chart shows the 20 countries with the highest number of recoveries per 100 cases in bar chart form. In addition, the values of each bar have been added as labels, and the y-axis range has been formatted. The seaborn and matplotlib libraries were used to create the chart. Seaborn allowed for easy, low-code plot creation, while matplotlib provided the ability to customize it. The resulting plot is shown in Figure 9.
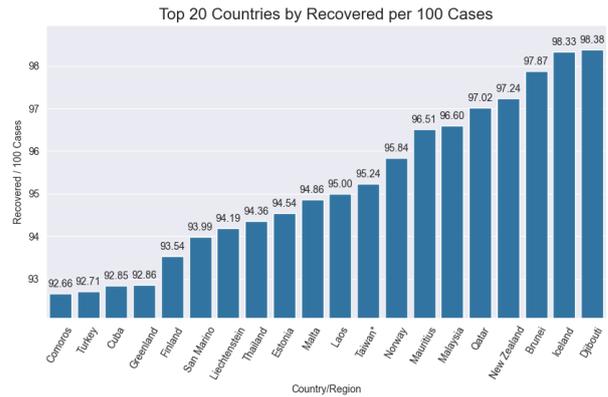

Figure 9: Seaborn bar chart.

The last graph created shows the total number of cases per 1 million population in a country in the form of a symbol map chart. The plotly library was used to creation an interactive graph that allows the hovering over each symbol to display detailed data and to move the resulting map as desired. Additional formatting was applied to change used colors and show country outlines. The resulting plot is shown in Figure 10.

Figure 10: Plotly scatter map.

4.3.3. Results from complementary study (Tableau, Superset, Metabase)

Participants in the complementary study completed five analytical and visualization tasks using Tableau, Apache Superset, and Metabase. The group consisted of 20 participants with varying levels of technical experience. Each participant filled out a survey evaluating ease of use, customization of visualizations, and overall satisfaction.

Regarding ease of use, Metabase received the highest ratings, indicating it was the most intuitive tool for users without technical experience, which aligns with Santos et al. [10], who emphasized Metabase's user-friendly interface and low learning curve. Tableau was rated slightly lower, consistent with Purich et al. [12], who highlighted Tableau's powerful but more complex dashboard capabilities. Superset received the lowest scores due to its more complex interface, reflecting Michele et al. [11], who noted that Superset requires greater technical knowledge for configuration and advanced analytics. Detailed results are shown in Figure 11.
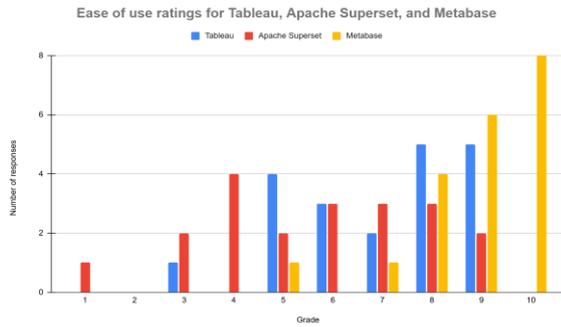
Figure 11: Distribution of ease of use ratings for Tableau, Apache Superset, and Metabase.

For customization of visualizations, Tableau achieved the highest scores, reflecting its strong flexibility and advanced visualization capabilities, in agreement with Lavanya et al. [6], who emphasized Tableau's versatility and rich chart types. Superset received moderate ratings, consistent with Michele et al. [11], which highlighted Superset's flexibility for experienced users but limited out-of-the-box customization. Metabase scored lowest due to its limited customization options [10]. Detailed results are shown in Figure 12.
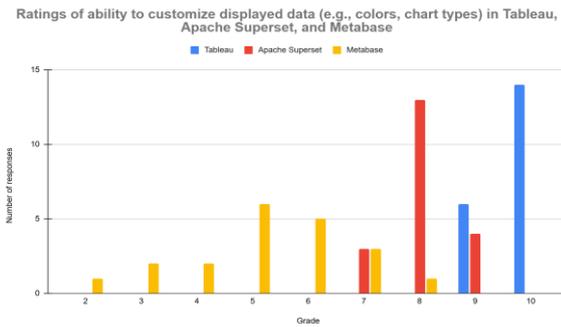

Figure 12: Distribution of customization ratings for Tableau, Apache Superset, and Metabase.

Regarding overall satisfaction, Metabase scored highest due to its simplicity and ease of use [10], followed by Tableau [6, 12]. Superset's overall satisfaction was lower, reflecting the challenges participants faced with dashboard configuration [11]. Detailed results are shown in Figure 13.
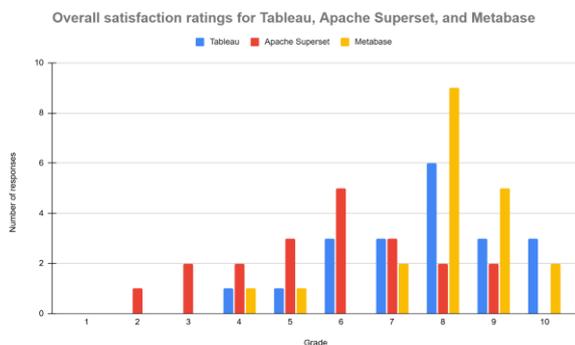

Figure 13: Distribution of overall satisfaction ratings for Tableau, Apache Superset, and Metabase.

Based on the survey results presented in Figure 14, Tableau appears to be the preferred data visualization tool, with 85% of respondents choosing the current version and 65% interested in using it in the future, confirming Purich et al. [12] findings on Tableau's popularity and adoption. Apache Superset shows lower adoption at 25%, consistent with Michele et al. [11] regarding its steeper learning curve.
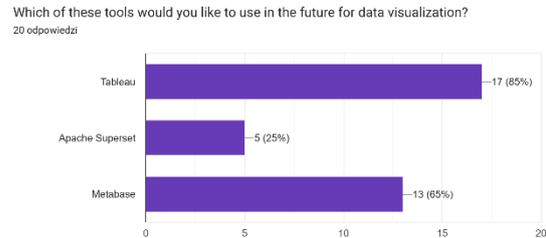

Figure 14: Results of preferred tools ratings.

5. Conclusions

Based on survey responses, Tableau and Power BI both offer user-friendly interfaces, high interactivity and extensive visualization capabilities, with Tableau slightly outperforming Power BI in terms of responsive interactive elements and Power BI slightly better in terms of customizable visual elements.

In terms of learning support, both Tableau, Power BI and Python libraries are supported by extensive documentation and active user communities. Python libraries offer greater flexibility and programmability, especially valued by data analysts and developers. Plotly in particular supports interactive web visualizations, balancing usability and customization.

The complementary study comparing Tableau, Apache Superset, and Metabase revealed clear differences in usability, customization, and overall user satisfaction. Metabase proved to be the most intuitive tool for participants without technical experience, enabling quick and easy creation of dashboards and reports. Tableau demonstrated the highest flexibility and visualization capabilities, making it suitable for users seeking advanced features and polished visual presentations. Superset, while offering high flexibility and scalability, required greater technical expertise, which affected ease of use and overall satisfaction for less experienced participants.

Participants consistently favored tools that balanced ease of use with visual appeal. Tableau emerged as the most versatile option, providing both a user-friendly interface and a wide range of customization features. Metabase was appreciated for its simplicity and rapid deployment, making it ideal for beginners or teams with limited technical knowledge. Superset was valued for its advanced functionality and scalability, particularly for users with programming skills or those managing large datasets, but its complexity presented a higher learning curve.

These findings indicate that the selection of a data visualization tool should be guided by the intended user group, technical proficiency, and the complexity of data

analysis tasks. Tools like Metabase are ideal for quick insights and low-barrier adoption, while Tableau and Superset are better suited for detailed, professional, or large-scale analytical work. Organizations can optimize outcomes by aligning the choice of tool with user experience and project requirements.

The results reveal that participants preferred tools that are either easy to use or visually appealing: Tableau was recognized as the most versatile and aesthetically pleasing, Metabase was appreciated for its simplicity and intuitiveness, while Superset was valued for its flexibility but required greater technical expertise. Ultimately, the choice of visualization tool depends on the user's priorities. Business users seeking ease of use and professional results may prefer Tableau or Power BI, whereas Python libraries offer greater control and flexibility, making them more suitable for users with coding experience and complex visualization needs.

References

[1] X. Liu, M. S. Alharbi, J. Chen, A. Diehl, D. Rees, E. E. Firat, Q. Wang, R. S. Laramee, Visualization Resources: A Survey, Information Visualization 22 (2023) 3-30, https://doi.org/10.1177/14738716221126992.

[2] F. Skender, V. Manevska, Data Visualization Tools - Preview and Comparison, Journal of Emerging Computer Technologies 2 (2022) 30-35.

[3] N. Balaji, B. H. K. Pai, B. Bhat, B. Praveen, Data Visualization in Splunk and Tableau: A Case Study Demonstration, Journal of Physics: Conference Series 1767 (2020) 012008, https://doi.org/10.1088/1742-6596/1767/1/012008.

[4] N. Akhtar, N. Tabassum, A. Perwej, Y. Perwej, Data analytics and visualization using Tableau utilitarian for COVID-19, Global Journal of Engineering and Technology Advances 3 (2020) 28-50, https://doi.org/10.30574/gjeta.2020.3.2.0029.

[5] A. J. Kadam, K. Akhade, A Review on Comparative Study of Popular Data Visualization Tools, Alochana Journal 13 (2024) 532-538.

[6] A. Lavanya, S. Sindhuja, L. Gaurav, W. Ali, A Comprehensive Review of Data Visualization Tools: Features, Strengths, and Weaknesses, International Journal of Computer Engineering in Research Trends 10 (2023) 10-20, https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0102.

[7] V. V. Dogadina, A. V. Voronin, Comparative Analysis of Data Visualization Tools, International Journal of Computing, Programming and Database Management 5 (2024) 49-51, https://doi.org/10.33545/27076636.2024.v5.i1a.93.

[8] G. D. Zion, B. K. Tripathy, Comparative Analysis of Tools for Big Data Visualization and Challenges, In: S. Anouncia, H. Gohel, S. Vairamuthu (eds), Data Visualization (2020) 33-52, http://dx.doi.org/10.1007/978-981-15-2282-6_3.

[9] A. Chavan, Data Visualization Tools Used for Decision Making: A Review, International Research Journal of Modernization in Engineering Technology and Science 5 (2023) 1338-1347, https://www.doi.org/10.56726/IRJMETS33640.

[10] B. Santos, F. Sério, S. Abrantes, F. Sá, J. Loureiro, C. Wanzeller, P. Martins, Open Source Business Intelligence Tools: Metabase and Redash, Proceedings of the 11th International Joint Conference on Knowledge Discovery. Knowledge Engineering and Knowledge Management (IC3K 2019) – KDIR, SciTePress (2019) 467-474, https://doi.org/10.5220/0008351704670474.

[11] P. Michele, F. Fallucchi, E. W. De Luca, Create Dashboards and Data Story with the Data & Analytics Frameworks, In: E. Garoufallou, F. Fallucchi, E. W. De Luca, (eds) Metadata and Semantic Research. Communications in Computer and Information Science, Springer (2019) 272-283, https://doi.org/10.1007/978-3-030-36599-8_24.

[12] J. Purich, A. Srinivasan, M. Correll, L. Battle, V. Setlur, A. Crisan, Toward a Scalable Census of Dashboard Designs in the Wild: A Case Study with Tableau Public, arXiv preprint arXiv:2306.16513 (2023), https://doi.org/10.48550/arXiv.2306.16513.

[13] Covid-19 Worldometer Dataset, https://www.kaggle.com/datasets/pvaibhav1995/covid19-worldometer-dataset/data, [25.03.2025].

[14] COVID-19 Dataset, https://www.kaggle.com/datasets/imdevskp/corona-virus-report, [25.03.2025].

[15] Online Retail Dataset, https://archive.ics.uci.edu/dataset/352/online+retail, [31.06.2025].