DATABRICKS VS. TABLEAU: WHICH VISUALIZATION TOOL IS RIGHT FOR YOU?

Krupa Goel,
krupagoel@gmail.com

Abstract

This paper compares DataBricks and Tableau, two commonly used data visualization and analytics tools, with their respective strengths, weaknesses, and recommended applications in business. DataBricks is in the cloud computing domain, and its significant specialties include large-scale data processing, machine learning, and real-time data analytics. It is ideal with prominent data scientists and engineers handling large data projects. This collaborates with key cloud service providers and thereby supports elastic data processing. Tableau is a business intelligence tool aimed at analysts and decision-makers, allowing the creation of unique and interactive data visualizations and dashboards through a drag-and-drop mechanism without necessitating specific programming knowledge. For this reason, it is easy to use and best suited for preparing reports and presenting analytical data to non-technical personnel. The paper also compares and contrasts features like scalability, simplicity, additional analytical tools, and collaboration. DataBricks handles complex data processes, operations, and AI projects, while Tableau is ideal for visualization creation and sharing. Based on the study's analysis, it is evident that using DataBricks or Tableau depends on an organization's analysis needs, technical competence, and intended goals. In many cases, businesses can operate both tools to their advantage. DataBricks serves as the underlying tool for extensive data processing, and Tableau simplifies the findings for the business heads. It also enhances the strengths of both platforms, in which fake decisions based on emotions are eliminated, and better decisions based on data are made.

Keywords; DataBricks, Tableau, Data visualization, Business intelligence, Machine learning, Big data, Real-time analytics, Scalability, Dashboards, Churn rates.

I.   INTRODUCTION

In today's world of globalization, using statistics and visualization is an inevitable part of a manager's and an organization's activity. As data becomes another valuable resource for different types of businesses, the importance of the instruments that would help to manipulate, analyse, and visualize data is high. While many organizations amass large quantities of data, the more significant problem lies not in data analysis but in the ability to present it to users in a way that makes sense to them. Some of the most effective tools used to address these issues are DataBricks and Tableau, which include distinct attributes for various functions. Even in the Data Visualization and Analytics domains, we have fantastic players like DataBricks and Tableau, which serve different user classes. DataBricks was originally designed as the single unified platform for designing analytical workflows and stands out for executing large-scale data processing and

machine learning. It is also much loved by data scientists, engineers, and technical teams dealing with AI projects or handling big data processing. Tableau, in contrast, is a business intelligence software tool conceived for the simplicity and the possibility of creating eye-popping and clickable visualizations. Tableau is favoured by business analysts and decision-makers who do not have a solid technical background since the software does not require any coding but instead relies on the drag-and-drop technique as the primary mechanism for interacting with it.

It is important to note that comparing DataBricks to Tableau is problematic because it depends on several factors. One of the first factors to consider is the type and nature of data you are processing or analysing. This may be most suitable for organizations that handle large amounts of structured or unstructured data, generating real-time analytics and complex machine-learning models. For more technologically sophisticated projects, its extensive data processing capabilities and its compatibility with machine learning frameworks make it a more suitable platform. On the other hand, if the objective is to design interactive dashboards that do not require much coding and are meaningful to non-technical individuals, Tableau could be more suitable. It stands out in its ability to convert simple data to powerful visualizations that can be communicated with cross-functional co-workers for improved decision-making. Tableau's lack of complexity does not diminish the software's functionality; it continues to be a viable software for EDA and visualizing data without much programming involved.

This article aims to compare two popular tools, DataBricks and Tableau, and determine the best fit for a business. In our comparison between the two platforms, we will focus on their main strengths, weaknesses, and suitable applications, trying to outline where each is particularly effective and where it can be problematic. This article will benefit a data scientist looking for a place to perform data analysis and a business analyst who needs an easy way to visualize the information. A superior selection of a proper tool can enhance data value and make decisions even better, keeping the role of data meaningful for the company.

Figure 1: Data Visualization vs. Information Design

II.   OVERVIEW OF DATABRICKS AND TABLEAU
1. Brief Description of DataBricks
DataBricks is a cloud-based platform born from the need to solve sophisticated, significant data challenges. It is an open-source software company founded in 2013 by the same founders of Apache Spark, one of its big data engines. Some features are now Data Engineering, Machine Learning, and Real-Time data analytics, among others, due to the advancement of the platform. DataBricks, an easily scalable tool, has multiple features for handling big data and complex data processing and is widely used in various industries by data scientists, engineers, and analytics teams (Mansoor, 2020). The most significant advantage of DataBricks is the possibility of its

connection to several cloud service providers, IaaS, including Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP). This flexibility enables users to use the resources available in a cloud environment to store and process deploying deployataBrichas integrated notebooks, allowing users to code in Python, SQL, and Scala to meet the technical requirements of various stakeholders due to its highly customizable and complex analysis interface (Chen & Zhang, 2021).

DataBricks is primarily focused on big data processing and Machine Learning. The program is implemented using distributed computing; hence, it is ideal for processing extensive data and offers support for developing machine learning models at scale (Gill 2018). Solutions, for example, in predictive modeling, measuring the degree of data similarity, and recommendation services, can be solved effectively in the DataBricks environment. Moreover, DataBrick also provides connectivity with Apache Spark, which enables the user to do complex data transformations and the capability to do real-time data stream processing (Nath & Sharma, 2021). As such, DataBricks is adopted by organizations, especially those who need highly reliable data processing pipelines, and is most likely to be found in the manufacturing, technology, and e-commerce industries.
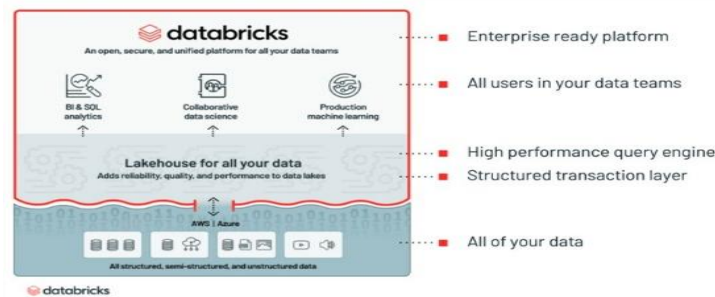

Figure 2: data bricks

2. Brief Description of Tableau

Tableau was founded in 2003 in California and is a popular business intelligence and data visualization software that emphasizes intuitiveness. Its primary purpose is to assist business users, especially those who are not IT professionals, in developing engaging dashboards and captivating reports to analyse different aspects of organizational performance. Tableau has been popular since the user interface's drag-and-drop functionalities enable easy data manipulation and analysis (Davenport & Harris, 2017). Another exciting aspect is the availability of numerous integrated alternatives, mainly bar charts, lines, maps, and scatter. These enable users to display large and complicated data sets in an easily understandable graphic format. Tableau is supposed to be connected to various data sources, such as databases, clouds, or spreadsheets. This extensive compatibility enables users to capture and draw data from various sources without necessarily entwining the data from different origins (Schloss & Smith, 2019).

Tableau is specifically designed to work for business intelligence, which allows the creation of fast and interactive visualizations. It is most prevalent in industries where finance, healthcare, and retail require top-down information analysis derived from big data. Specifically, while DataBricks supports big data query and processing and integration with advanced analytics, including machine learning, Tableau does not. Instead, it emphasizes the possibility of sharing these insights by displaying them within the easy-to-use dashboards (Wills, 2019).

Figure 3: Tableau Business Intelligence Reporting | Tableau Reporting Tool

3.   Importance of Understanding the Differences

When deciding which one is better than the other, DataBricks and Tableau should be considered because each of them is more suitable for different tasks and users. For organizations that are using or looking to use big data and machine learning, DataBricks would be a perfect pick. The platform's capability to ramp up with a cloud resource and compute massive data makes it suitable for data scientists and engineers who need to train a machine learning model, process live data, or handle big data (Chen et al., 2021). Tableau is more appropriate for business analysts and decision-makers who need simple visuals to analyse and share information. Its user-friendly interface and the opportunity to create dynamic dashboards without programming expertise make this product valuable for teams requiring report generation or initial data analysis. Tableau is also more effective when the data is reported to other users who may not need help understanding data analysis, including executives and clients, because of its interactive and appealing visual representations (Schloss & Smith, 2019).

Deciding between DataBricks or Tableau solely relies on the user's ability when it comes to technical competency as well as the complexity of the data being analysed. Havening end-to-end control of data pipelines and the need to implement current machine learning models make DataBricks a better choice for data engineering teams. Tableau is more advantageous for those business personnel who require creating simple layouts and sharing them with others. While both platforms are relevant and present in modern data-driven organizations, sometimes both can coexist, where, for example, DataBricks is more suitable for deep data analysis while the results are presented in Tableau (Nath & Sharma, 2021). Therefore, organizations seeking the best out of their data must fully grasp the critical differences between DataBricks and Tableau. DataBricks is a heavy hitter for extensive data operations, machine learning, and real-time processing. In contrast, Tableau is well-created for easy-to-use visualization and BI. Determining the suitability of the platform is dependent on the organization, its data needs, and the nature of the analysis it wants to achieve.

Figure 4: Databricks vs Snowflake

III.   KEY FEATURES COMPARISON: DATABRICKS VS. TABLEAU

DataBricks, or Apache Spark and Tableau, are handy tools for processing and visualizing big data. However, their workloads, scalability, simplicity, and collaborative features are notably different. This comparison enables users to decide which programs would be most suitable for their needs depending on the size and complexity of their data, the specific analysis they need to complete, and their level of computing sophistication.

1.   Core Strengths

One of the most significant advantages of DataBricks is that it is built for scalability and the capability to include ML and AI facilities. It is dedicated to data scientists, engineers, and analysts participating in big data projects. It facilitates high-throughput data processing and the manipulation of complex data models using Python, Scala, and SQL. DataBricks offers users pre-integrated built-in ML/AI additions, thus making it suitable for organizations that need sophisticated data processing and AI straight out-of-the-box decision-making systems (Bai, 2020). Tableau, on the other hand, stands out in easily customizable and more engaging data visualization. Due to this feature, an advanced level of customization and the ability to build unique and interactive dashboards is possible for users with minimal or no programming experience. Tableau is more tailored for business analysts, executives, and decision-makers who require rapid access and usage of data analytics, requiring little coding capability (Healy, 2019). It does not incorporate other AI-related functions; however, in return, it provides the user with the ease of directly analysing data.

2.   Scalability

One more advantage of DataBricks is its ability to scale up. It is meant to manage extensive data operations and can be used in large organizations that process structured or unstructured data on a large scale. DataBricks is also well-equipped to meet the needs of an organization dealing with terabytes of transactional or real-time stream data. Its integration with cloud services also amplifies its scalability because organizations can upgrade their capacity as and when they wish (Shroff, 2021). On the other hand, Tableau's scalability could be better since the company has quite a modest number of loyal customers to its services. It is perfect for managing small data sets and moderate-size data. However, data manipulation must be improved for extensive and real-time data streaming. The core competence of the product is in its capacity to generate clear and detailed reports relative to small datasets. Thus, it benefits smaller organizations or a particular division of a vast company that requires basic and simple dashboards. Despite that, Tableau slows down as the data size increases and may require other tools for extensive data processing (Smith & Jones, 2019).

3.   Ease of Use

One of the most significant differences between DataBricks and Tableau is the simplicity of use. DataBricks is built so that it may need the input of coders proficient in coding languages like Python, SQL, or Scala. It is not an easily adaptable tool, and those who do not possess a fair amount of technical knowledge may require time to grasp and understand it. However, for data scientists and engineers, DataBricks provides much higher customization and control of data processing and analysis (Bai, 2020). Tableau is simple and intuitive, and it was built that way on purpose. Its simple drag-and-drop functionality allows a user with little programming or design

background to create complex visualizations. The platform is preferable for the business audience, who can extract and analyse data with the program's help but do not want to start learning programming languages. Due to this comparison, learning Tableau is slightly lower than learning DataBricks, and it is open for everyone to learn (Healy, 2019).

4. Collaboration

These tools are most considerably dissimilar on the issue of collaboration. DataBricks has amazing performance features for teamwork when data projects are carried out in different teams. Clients provide cooperative notebooks for data scientists, engineers, and analysts where notebooks can be shared, code and comments can be created and edited in parallel, or numerous users may work on one project at a time. This feature is especially useful in organizations with numerous workers who need to work together constantly on data-related tasks (Shroff, 2021). Tableau is somewhat lacking in this aspect, but there are options such as Tableau Online and Tableau Server, which enable the sharing of dashboards and visualizations across the teams. Despite this, it is essential to note Tableau's need for advanced collaborative functionalities compared to DataBricks. Although business users can share the dashboards with other stakeholders or share the URLs, this feature does not allow real-time collaboration, which is crucial for technical teams who may need to work on the same project (Smith & Jones, 2019).

5. Advanced Analytics

One of the best uses for which DataBricks excels is the ability to carry out advanced analytics. It has modules for machine learning and artificial intelligence integrated into the platform so that data scientists can build prediction models and run analytical and statistical simulations within the platform. This makes DataBricks suitable for organizations that wish to use Machine learning to gain more insights into their data assets. They support data pipelines, ETL jobs, and other data-intensive tasks and workloads and are especially significant in supporting complex and large dataset applications (Bai, 2020). Tableau, in turn, aims at data visualization rather than analysing data in greater detail. Despite providing tools for basic analytics like trend lines, forecasting, clustering, and others, DataProfit does not possess as sophisticated capabilities within machine learning and Artificial Intelligence as DataBricks. Tableau's primary advantage is informing insights and creating data visualizations rather than drawing out complex routined analysis (Healy, 2019).

6. Real-Time Visualization

Data visualization is real-time on both platforms, yet the complexity of working with them varies. Real-time data stream and visualization are also possible in DataBricks. However, these can only be set up if one has the technical know-how and involves an unseen complex setup. DataBricks allows organizations with distinct data science teams to perform real-time data processing and visualization. However, it may be too complicated for more compact teams that need more knowledge (Shroff, 2021). Tableau, on the other hand, is built with live dashboarding in mind and is designed for business analysts' applications. In addition to supporting real-time visualization, the platform can build and display live data panels out of the box. This feature makes Tableau particularly suitable for business users who want to measure Kelner's key performance indicators or keep their data under real-time observation (Healy, 2019).

7. Data Source Flexibility

Companies can attain incredible flexibility from DataBricks in terms of raw material feeding. This means the platform is well developed to handle both structured and unstructured data, which is part of the strength companies will derive from it when handling data. DataBricks has integrated with cloud platforms such as Azure and AWS to increase its effectiveness in processing various data sources (Bai, 2020). Tableau also provides flexibility, but it is at its best when it is in touch with numerous databases and cloud services. The system accommodates many data connections, making data import from numerous data sources simple for business users. However, it does not have the same level of specificity for managing the heterogeneous information that DataBricks can (Smith & Jones, 2019).

8. Cost

Another essential aspect of concern is the price when choosing the right platform to engage in business. For instance, DataBricks is noted to be pricey, especially for a small team or organization with restricted capital. The platform's pricing is calculated based on usage, making its cost unmanageable for organizations, particularly those that may have a variable volume of data to process (Shroff, 2021). Tableau is also regarded as expensive, especially when making server configurations. Regarding costs, organizations can make premium products from Tableau expensive, much more so if they are significant. But for the teams that do not need complex calculations or machine learning, which occupy the largest share of the company's resources, Tableau has a more adjustable cost line that can be more manageable (Healy, 2019).

9. Learning Curve

DataBricks has a steep learning curve for anyone with little coding experience. The platform requires users to have programming languages like Python, SQL, and Scala as its basic requirements. This makes DataBricks more palatable to technical teams than nontechnical business users (Bai, 2020). Tableau is a considerably smoother learning slope. Its minimalistic design is suitable for use without advanced coding knowledge, meaning that even people without a technical background can create visualizations. This makes it a better option for business users who need to gain insight quickly without going through technical details about the tool or application (Healy, 2019).

10. Customization

The level of customization in DataBricks is relatively high in terms of coding using Python, Scala, and SQL to provide customer-based visuals and models. Due to this flexibility, it is convenient for DataBricks to be adapted to complex data projects (Shroff, 2021). Tableau has many different charts that are customizable right out of the box across all its licenses. As summarized in Healy (2019), in Tableau, the options available are confined to its built-in features. Tableau's built-in options are often adequate for most business users even though they provide less extensive customization than in DataBricks.

Table 1: Comparison of Key Features between DataBricks and Tableau

Feature	DataBricks	Tableau
Core Strength	Big data processing and machine learning	Interactive, easy-to-use visualizations
Scalability	Excellent for large-scale data processing	Can handle small to mid-sized data, struggles with big data
Ease of Use	Requires coding skills (Python, SQL, Scala)	User-friendly drag-and-drop interface
Collaboration	Collaborative notebooks for data science teams	Limited collaboration options (though Tableau Online/Server helps)
Advanced Analytics	Built-in tools for machine learning and AI	Limited to basic data visualization
Real-Time Visualization	Possible, but setup is complex	Easily supports real-time dashboarding
Data Source Flexibility	Supports structured and unstructured data	Connects easily to various databases and cloud services
Cost	High for small teams or organizations	Expensive, especially with server deployments
Learning Curve	Steeper learning curve for non-technical users	Low learning curve, accessible to non-tech users
Customization	Custom visualizations with Python, Scala, SQL	Extensive range of built-in customizable charts
Best For	Data scientists, engineers, AI-driven projects	Business analysts, executives, data-focused decision-makers

IV.   PROS AND CONS: DATABRICKS VS. TABLEAU

DataBricks and Tableau are two tools for data analysis and visualization that have been developed to solve different problems in data processing and utilization. The organization can determine which is better suited for its data requirements by analysing the strengths and weaknesses inherent in these tools. Both platforms shine in their domains: DataBricks is capable of big data processing and machine learning, which are pivotal in big data analytics. Tableau covers deeper analytics with a centralized user interface and on-screen fit interactive data. It includes each platform's advantages and disadvantages, comparing when DataBricks excels and when Tableau is preferable.

1.   When DataBricks is Advantageous

DataBricks, therefore, works very well for organizations that deal with big data and need a higher

level of analysis. Due to its capabilities to encompass structured and unstructured data and support for machine learning and AI, it has been favoured by data scientists and engineers working on demanding projects. One of DataBricks' main advantages is its ability to effectively work with significant data volumes without compromising performance (Forbes, 2023). DataBricks provides an enhanced data manipulation feature that helps perform live data processing, which is essential in today's context for organizations requiring a rapid data stream. For example, a technology firm that processes user behavior data in real-time would find DataBricks' scalability to process data at a large scale and very fast functional (Davenport & Harris, 2017). In addition to the technical capabilities mentioned above, DataBricks accepts multiple coding languages, including Python, Scala, and SQLs, allowing the data teams to code more flexibly in fashion when creating new models or engaging in more detailed analytics in ML (Graves & Corbin, 2020).

However, these advantages come at a cost, meaning they have to face some drawbacks. Another disadvantage of DataBricks is that it is software that operates in a technical environment, so personnel who are not familiar with such tools are likely to encounter numerous problems when using the software (Brown, 2019). Another drawback of the platform is that it is very resource-heavy, making it better suited for organizations that have deep pockets and their own data science teams.

2. When Tableau is Better Suited

The choice of components is stark — while Excel allows business analysts to create computations and pivot tables without necessarily calling upon a programmer, Tableau is intended for power users who want to explore data in front of them in as much detail as possible. It stands out in its functionality, with a drag option that simplifies the creation of interactive visualization for the user. It benefits executives, business analysts, and decision-makers who often need to analyse data using business intelligence tools but have no time to learn programming languages (Kirk, 2016).

Tableau's advantage is the immediate and easy-to-read dashboards that maximize the data's potential for understanding. For example, in sectors such as healthcare and finance, where analytic consumers require decision-making information from a large amount of data, Tableau's interface helps individuals without analytical backgrounds understand complex data (Few, 2017). Furthermore, Tableau is compatible with different data sources, such as databases and cloud services, and, therefore, can work in any environment (Jones & Smith, 2021).

However, as with almost any tool, when it comes to using Tableau, you also get a couple of cons. Consequently, it is not well suited for use with big data since it causes slow performance for large-scale operations (Brown, 2019). Secondly, table analysis and navigation experience are somewhat reduced compared to Tableau since its operations rely more on the direct visualization of data rather than on the analytics functionality (Graves & Corbin, 2020). This makes Tableau less appropriate for organizations requiring a complex data integration with machine learning language than DataBricks.

3. Impact on Organizations

Databricks and Tableau have capabilities and limitations. The organization's size and requirements determine the best tool. DataBricks may be more advantageous for large organizations with large budgets and need sophisticated data processing capabilities due to the possibility of working with

big data and machine learning (Davenport & Harris, 2017). For instance, a significant global firm with a large customer base that needs to deal with segments of its data in real-time will significantly benefit from DataBricks. However, Tableau may be more suitable for small companies or those who do not need advanced data visualization and only require vehicles for creating dashboards or reporting for business users. Also, while DataBricks is designed for data scientists and data engineers, Tableau is better for business analysts or executives. Consequently, an organization with multiple data scientists within its framework may be biased toward DataBricks, while business intelligence organizations for decision support prefer Tableau.

V.   USE CASES: WHEN TO CHOOSE DATABRICKS OR TABLEAU

Data analytics is now a core element of most organizational strategies in several industries. Two significant tools, DataBricks and Tableau, play different roles in this domain. DataBricks can boast of perfecting a technique for handling more extensive datasets for processing and machine learning. In contrast, Tableau is perfect for prepping a beautiful and easily navigated report or dashboard. The decision as to which of the two to use depends on the objective of an organization, the technical ability, and the nature of the data. As seen in this article, there are scenarios where the various tools in either program will be most effective.

Table 2: Use Cases for DataBricks and Tableau

Use Case	DataBricks	Tableau
Big Data and AI	Useful for AI, machine learning, and big data processing	Limited to basic analytics
Data Preparation and Processing	Great for large-scale ETL and real-time data	Better suited for simple data manipulation
Business Reporting	Not ideal for creating business reports	Excellent for business dashboards and reporting
Real-Time Data Processing	Complex, but possible with coding expertise	Built-in, easier to set up
Data Exploration	Requires coding, supports complex analysis	Ideal for non-technical data exploration
Collaboration	Collaborative notebooks for team projects	Dashboards shared via Tableau Server/Online
Sharing Insights	Limited visualization sharing features	Easy sharing via dashboards and reports

Figure 5: Understand the Databricks Architecture

1. **When to Choose DataBricks**
A. **Big Data and AI**: This tool will be handy for organizations involved in processing vast amounts of data, particularly those that offer services in artificial intelligence and or machine learning. DataBricks itself was built to serve live significant data issues that incorporate well with machine learning frameworks and tools like Apache Spark (Chen & Zhang, 2021). By accommodating structured and unstructured data, DataBricks assists data scientists and engineers in creating enriched machine-learning models that can analyse big datasets effectively (Nath & Sharma, 2021).

B. **Data Preparation and Data Processing:** Another major asset of DataBricks is the ability to work with Extract, Transform, and Load (ETL) processes. Businesses that need a solid data conduit to contend with large data sets of different types and sources find the platform highly scalable and computational in the cloud. DataBricks are compatible with cloud service providers such as AWS and Azures, which facilitates the management of data engineering workloads without compromising performance (Bai, 2020). This flexibility gives DataBricks maximum utility in large organizations with complex data systems.

C. **Real-Time Data Processing:** In industries where real-time operational decisions are crucial, such as e-commerce, finance, or telecommunications, DataBricksorms others with real-time stream data processing. Its distributed computing model means organizations can look at data in 'real time' as it is produced, facilitating rapid analysis and action. For instance, many technology firms use DataBricks as a central processor of live data on users' activity to modify algorithms or services on the fly (Forbes, 2023). This capability can be beneficial in cases such as fraud detection or a marketing campaign where detecting an anomaly leads to an immediate course of action that contributes to the success of a campaign.
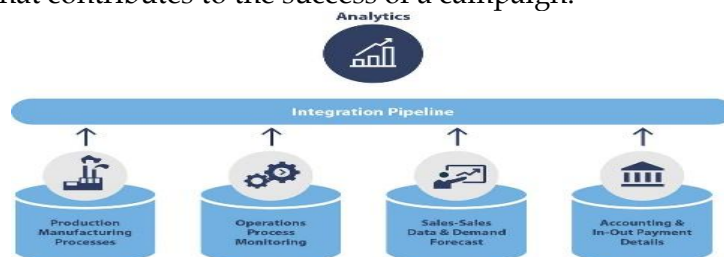
Figure 6: How Enterprises Can Scale AI Successfully using Databricks

2. **When to Choose Tableau**
A. **Business Reporting and Dashboarding**: Tableau is the application of choice for organizations whose purpose is mainly business intelligence and reporting. They have an easily navigable drag-and-drop functionality to allow those with little to no coding ability to create custom reports and dashboards. For example, DataBricks relies on programming, while Tableau is more democratic and can interact with regular users and sellers. Interactive data presented in various graphs, charts, and dashboards increase the chances of decision-makers quickly understanding the information conveyed via such figures without considerable chances of getting it wrong (Healy, 2019). As a result, Tableau can be effectively used in sectors such as healthcare since it simplifies executive executives' decision-making19).

B. **Exploratory Data Analysis (EDA):** Tableau has the best results in the analysis for discovery, a purpose of data analysis whereby the user navigates through views to find something interesting. The straightforward user interface enables users to manipulate data in real time, making it easier to find what may not be evident when disaggregating data (Schloss & Smith, 2019). It benefits business analysts who need to analyse several data sets and obtain results quickly. Tableau's ability to generate different graphics like heat maps and scatter plots aligns with most analytic tasks, allowing it to be effectively used in the initial data analysis.

C. **Sharing Insights:** Tableau is also very useful when the insights have to be presented to individuals who may not need a technical background. An example is its interactive dashboards, which can be easily transferred to other teams and employees for joint decision-making. At the same time, they do not need any coding skills or a deep understanding of analytics (Jones & Smith, 2021). Using Tableau Server and Tableau Online, the company's teams can make their work and dashboards available for others from other departments or even partners outside the enterprise, making it easier for the whole organization and every department to embrace insights and analytics (Healy, 2019). This functionality is especially beneficial in organizations where executives want details but in a brief, non-numerical form.
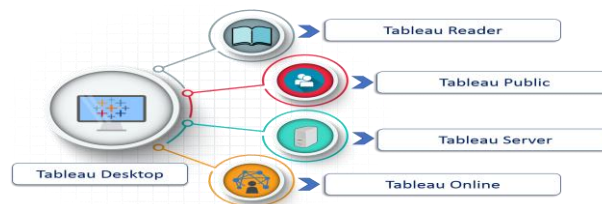

Figure 7: Tableau A Best Data Visualizer

3.   Key Differences and Overlaps

Even though DataBricks is designed for more extensive datasets and Tableau is for everyone, there are times when the two platforms can go hand in hand. For instance, an organization can apply DataBricks to analyse data and apply machine learning to extract insights from a pile of data while preparing to use Tableau to bring the insights into a more understandable presentation for non-data analysts. Such a combination is most helpful when there is a requirement for complex data analysis and straightforward reporting, like in the case of retail or financial firms (Nath & Sharma, 2021).

A. Collaborative Approach of DataBricks and Tableau

In some organizations, it is beneficial to transition to using both platforms as it makes the work run smoother. While DataBricks can do the work of data engineering and deep-level analysis, Tableau can take ahead the process of visualizing the data ready for business understanding. For instance, an organization in the financial services industry will use DataBricks to analyse customer transaction data and build models for decision-making before feeding the same data to Tableau to produce reports and visualizations that can aid senior management in making appropriate decisions (Shroff, 2021). This approach allows organizations to get the maximum results by utilizing the advantages of both platforms.

Based on these factors, it may be ideal to conclude that the decision between using DataBricks and Tableau depends on the organization's needs, the employees' technical competency, and the nature and size of the data. , DataBricks fits complex data datasets for massive-scale data processing, data

engineering, and, more specifically, for data scientists. Tableau, on the other hand, is as powerful and has the capabilities of requiring more technical usage. However, it is very user-friendly and ideal for business practitioners who need more technical support to create reports and dashboards. In conclusion, it is possible that specific organizations would be better served by incorporating both tools throughout their data analysis tasks, capitalizing on the unique capabilities of each to meet their objectives in terms of data processing and visualization.
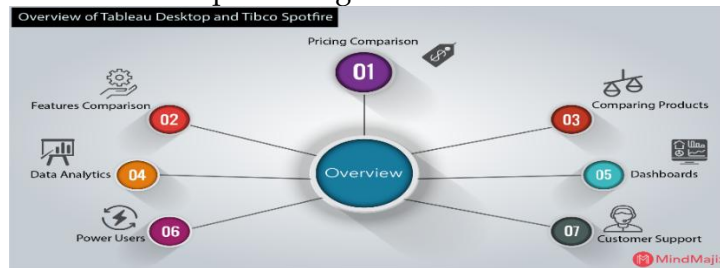
Figure 8: overview of Tableau Vs Spotfire

VI.   COMPARATIVE INDUSTRY USAGE

1. Tableau's Dominance

In the BI market, Tableau has dominated figures primarily because of the simple GUI and efficient visual analytics. It is expected that by 2024, Tableau BI will reach $1,013 billion with about 15% of the market share, making it one of the leaders in BI platforms worldwide, as Gartner indicated. This has been made possible by targeting both technical and nontechnical persons to drill down the user interface, hence utilizing drag-and-drop methods of analysis and production of reports. With more and more businesses focusing on using the data they collect effectively, demand for intuitive tools has risen, and Tableau has risen to meet the occasion.

Another potential reason information visualization leader Tableau is out-competing its rivals is its well-anchored position in industries where data visualization is a critical success factor. For instance, industries such as healthcare use tools like Tableau to analyse patients' information, develop more effective treatment plans, and, most importantly, undertake substantial, often slow, work in a relatively shorter time (Matthias, 2019). As an important sector where swift working outcomes influence patients' conditions, Tableau is helpful as an IT tool for instant data visualization of extensive and imbroglio data. Likewise, financial services have integrated tableaus to capture the markets and customer behavior and manage risks. With integration into various sources of data, big data can be embodied obviously, thus making its application very efficient in an industry where accuracy and clearance of information are vital (Richards, 2020).

Tableau is also used in the retail industry to help firms improve customer satisfaction by understanding customers' buying behavior, stock management, and logistics. As consumers' choices and purchasing decisions are rooted in sophisticated approaches inherited from the era of digitalization, such as e-commerce and omnichannel store visits, retailers require tools to analyse large amounts of data. The capability to address these complications while making the data more digestible is why Tableau has found favor in this industry (Carmichael & Stone, 2021). However, it connects with interactive dashboards; for example, managers in retail companies fix marketing approaches or inventory stock levels virtually, increasing performance across different channels in Tableau.

Tableau's self-sufficiency is another factor that has made it famous worldwide, as it can handle a large volume of data. Tableau is used across businesses from small to big companies due to its multi-functionality. Davenport and Harris (2017) note that in addition to being usable at a basic level, the fact that Tableau continues to offer scale throughout various departments and business functions is a crucial reason businesses have been drawn to it in the pursuit of data democratization. With the ability to enable regular employees to work with the tool without in-depth IT knowledge, Tableau has ensured data democracy in different industries. This flexibility has helped to enhance its position as a market leader within the BI space as organizations remain committed to achieving more flexibility and innovative BI.
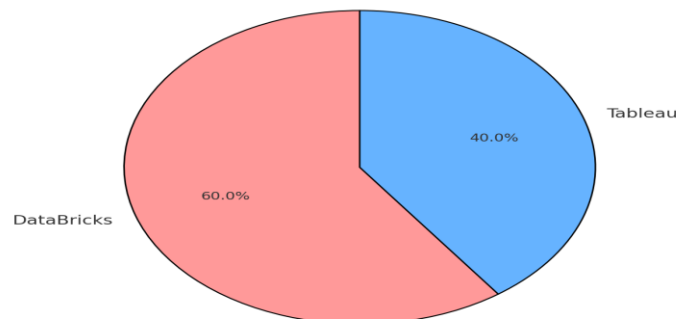

Figure 9: Comparison of Databricks vs Tableau Usage in Industries

2. Rising Popularity of Databricks

While Tableau can be considered the leader in industries with data visualization as their core, industries relying on data processing with machine learning and big data processing, have come to appreciate Databricks. Databricks was founded in 2013 and has quickly grown popular because of its robust cloud and integration of data engineering, data science, and MEC in a single place (Matei et al., 2021). The platform's support for extensive data handling and analytical capabilities have made it especially suitable for tech-driven industries like e-commerce and manufacturing. Databritechnology industry in the Technology Industry for its strength in dealing with large volumes of big data and iL capabilities. For instance, e-commerce businesses use Databricks to understand the users' behaviors, enhance the decision-making process on products to offer, and provide the proper customer segmentation (Zaharia et al., 2020). As more companies have discovered the importance of e-commerce worldwide, the chance to make instant decisions based on biological signals and install machine learning models has become necessary for various organizations. Databricks' resistant platform built over Apache Spark is one of the primary reasons why companies are turning to the solution.

Databricks are also used by the manufacturing industry, especially when discussing Industry 4.0, where aspects of big data and machine learning are critical in driving solutions to eliminate systematic wastage. Manyika et al. (2019) note that firms with manufacturing divisions use Databricks notably for advancing predictive maintenance, understanding readings from machinery sensors, and optimizing supply chain management. Real-time data processing by Databricks helps manufacturers gain insights, improve production efficiency, and minimize or mitigate downtimes. That is why it is perfect for industries where data complexity and size are the key factors that can complicate production. Another line that has shown significant growth is the

financial services sector, where machine learning, for example, use cases, are fraud detection, credit risk, and customer experience Personalization. Large financial institutions that may engage with massive data from various sources, such as structured and unstructured data, benefit from data being processed timely and adequately while incorporating models in the analysis through the help of Databricks (Raj, 2018). The platform's domain mainly revolves around data pipelines and the ability to integrate these pipelines across all functions without causing operational hitches that hamper value derivations from the data.

However, considering it is relatively new in the market compared to Tableau, Databricks has witnessed such rapid growth in usage that it has been increasing at a great pace, especially across industries targeting analytical research. Another reason for its popularity is that it can be used as a cloud-native environment and supports big data frameworks for implementing machine learning into business operations (Hellerstein et al., 2018). Furthermore, as more new models adopt artificial intelligence and machine learning approaches in decision-making, Databricks' strategic place in the market should only grow. Tableau continues to be the market leader in the BI market, especially in the healthcare, banking, and retail sectors, while Databricks is moving in directions where big data and machine learning are essential. Both platforms offer unique strengths tailored to different use cases. For instance, Tableau is highly effective in data displaying and searching, which is why it suits business users. At the same time, Databricks delivers tools for data scientists and engineers to process data and solve complex analytical tasks. While organizations are still struggling to manage the impacts of the evolution of data-based landscapes, the options between these two platforms will hereafter rely on the industry that needs to focus on them.
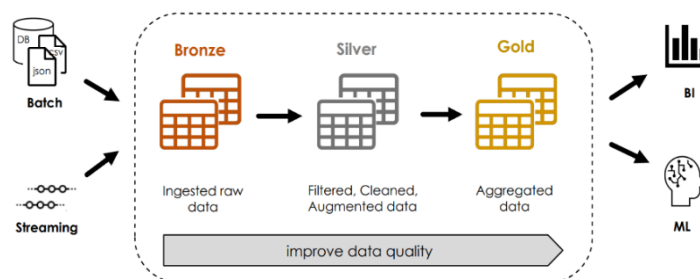

Figure 10: Data Integration

VII.   DASHBOARD EXAMPLE: SAAS PRODUCT SALES DATA AND ROOT CAUSE ANALYSIS

1.   Scenario Overview: Decline in SaaS Subscriptions

Since it is in the sphere of SaaS, an incredibly competitive sphere, customer attraction and retention are paramount (Unigwe, 2021). Businesses constantly track new subscriptions to make sure that their development plans work. For this hypothetical SaaS firm in the case study, there is a slowdown in new subscriptions, increasing sales of products, and a low churn rate. The company has been gathering information on product sales, churn rate, customer behaviors, and marketing activities for the last six months. The principal task emerging at the current stage is to dissect this kind of data and understand what caused the decline of subscriptions, which considerably declined in June despite the general increases in the other performance indicators. Both DataBricks and Tableau are utilized to analyse possible causes leading to the decline and illustrate how each platform outperforms the other in terms of its ability to work with large datasets, implement

complex analyses, and visualize results for strategic decision-making.

Table 3: SaaS Product Sales Data (Dummy Data)

Month	Product A Sales	Product B Sales	Churn Rate	New Subscriptions	Marketing Spend
January	10,000	8,000	5%	1,200	$20,000
February	11,500	9,200	4.8%	1,350	$22,000
March	13,000	10,500	4.5%	1,500	$23,500
April	15,000	12,300	4.0%	1,800	$25,000
May	16,500	13,700	3.8%	2,000	$28,000
June	18,000	14,900	3.5%	1,650	$27,000

2. Use Case Description: Sales Data, Churn Rates, and Marketing Analysis

The fundamental data used are the number of times a product has been sold, the amount of customer attrition, the number of new customers, and the marketing cost, all for a half-year period (Castéran, Meyer-Waarden, & Reinartz, 2021). This has been accompanied by a steady increase in the sales of Product A and Product B. Cancellation rates, where customers are unable to renew their subscriptions, are also a measure and have been reducing over the years, which is encouraging. Marketing expenditure, on the other hand, has been either stable or has trended upwards. Even though the above statistics indicate growth, the following month, June new subscriptions fell to 1,650 from 2,000 registered in May. The question is to investigate what caused new subscriptions to decline in June despite other critical performance indicators being relatively stable. The company expects to get more feedback to identify the rounded-off issue and consequently change its marketing or operations strategies.
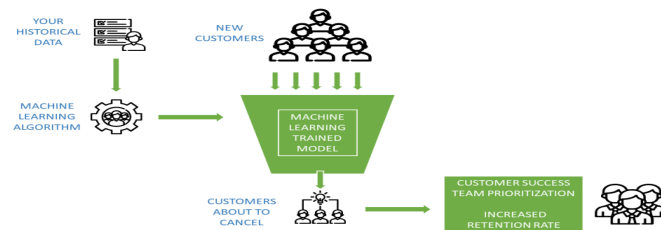

Figure 11: Predict Customer Churn

VIII.   USING DATABRICKS TO FIND ROOT CAUSE

1. Data Ingestion

The first action in Analytics using DataBricks to analyse the reduction in SaaS subscriptions is data acquisition. DataBricks runs on Apache Spark, an open-source native programming model for extensive data computation. With Spark, one can quickly load and transform large datasets. In this case, the dataset contains sales, churn rate, new subscriptions, and marketing expenses. Users can fully transform and evaluate the data at hand by loading these metrics into a single data frame. The ability to handle schema and no-schema data sources makes DataBricks especially appropriate for this purpose (Armbrust et al., 2015). After the data has been imported into DataBricks, it can be analysed using several languages, such as Python and SQL, thus making it more flexible to handle many datasets (Xin et al., 2013).

Data ingestion simply means collecting all required data and placing it within an environment favourable for analysis. It is essential to link the various data types, such as sales numbers, customer attrition, and the marketing budget, into another data frame to make the right connections. As for exploring tasks in a data frame, the data is pre-processed for data analysis with the help of distributed computing in Spark, which simplifies the work with large objects (Zaharia et al., 2016).
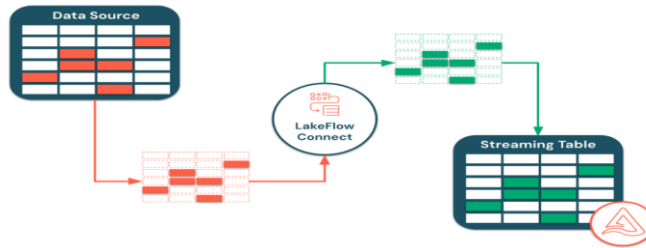

Figure 12: LakeFlow Connect (Efficient ingestion connectors for all)

2. Data Exploration

The next step is data exploration, which is something other than using SQL to find a correlation between the amount of marketing being spent, churn rates, and new subscriptions. As seen in this SaaS context, SQL is run to look for patterns or to answer questions like whether there is a negative relationship between subscription rates and shifts in marketing investment or churn. This exploratory analysis assists in figuring out if there is any pattern that might account for the decline in subscriptions (Kumar et al.,s 2014).

For instance, the SQL query could join the sales tables with the churn rates tables to establish if periods of reduced subscriptions result from higher churn rates. Given that SQL offers functionalities in aggregation and filtering, analysts can quickly narrow down essential variables, such as relating the flowing pattern of marketing investment to the flow of new subscriptions. Downward exaggeration can also be very misleading if there is a strong positive link between marketing spend and new subscriptions. It will then be hard to explain the decline in subscriptions other than reduced marketing effectiveness (Matei et al., 2020). These SQL queries can assist the analysts in identifying areas that may require further investigation due to abnormality.
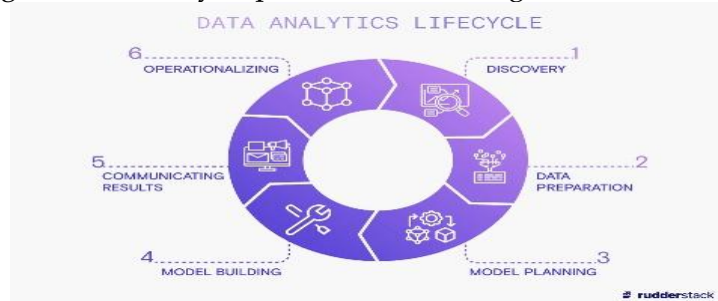

Figure 13: Phases of the data analytics lifecycle

3. Hypothesis Testing

DataBricks also provides robust support for hypothesis testing and machine learning models so teams can start making forecasts based on past trends. Thus, one hypothesis could be that the decline in new subscriptions could be associated with a decrease in marketing efficiency or some product aspects. To evidence this, a machine learning model, including the linear regression, can

be trained to forecast future subscription trends in coming months or years depending on essential indicators, including the marketing investment, the rate of customer churn, and the sales of the company's products (Zaharia et al., 2016).

Although these models are practical, they help analysts sort out various marketing situations and allow them to understand new subscriptions depending on particular spending. Moreover, in DataBricks, it is possible to use A/B testing to check whether some features or specific pricing influenced the negative trend of new subscriptions (Chen et al., 2012). A/B testing is an experimental implementation that compares the efficacy of two marketing campaigns or a product attribute that directly affects subscriptions. When one version is more effective in conversions, it may mean something needs to be changed in the product or the marketing approach.

4. Visualization

After exploring the data and hypothesis testing, DataBricks allows users to visualize the outcome. Charts and graphs are perfect tools for recognizing periodicities and patterns that may prevent the naked eye from observing bare figures. One of the most typical techniques is to use the line chart that depicts the prevalence of the examined relations throughout the analysis, based on the weekly number of new subscriptions and the money spent on the marketing campaign. This makes it possible for the stakeholders to judge the essential correlation between marketing spend and subscription rates (Armbrust et al., 2015).

Further, a bar chart of churn rates and new subscriptions by month can assist in identifying any direct correlation between these parameters. DataBricks supports a variety of visualization libraries, including Matplotlib and Seaborn, that enable data scientists to develop unique visualizations that meet the needs of a particular analytical workflow (Matei et al., 2020). These visualizations are important for representing some results in a way that is accessible to the business audience.

5. Finding the Root Cause

Subsequently, after data exploration, hypothesis testing, and data visualization, it is possible to identify the potential reason for the observed decline in the subscriptions. However, from this narrative, the data analysis shows that although churn rates were reducing and marketing expenditure was constant, the decline in new subscriptions in June could be attributed to low marketing impact. This is because despite a high marketing spend, marketing churn decreases, but new subscriptions decrease, indicating low conversion caused by marketing efforts (Xin et al., 2013).

However, more data might be needed to confirm this hypothesis so that the results will fall into place perfectly. For instance, the number of visits to a website and the number of users who purchased products from the Web site could shed additional light on the impact of campaigns. If the traffic continues to flock to the websites but fewer
are converting, then this would mean that there is an issue with conversion and not marketing costs. They would help the SaaS Company improve its current marketing approaches or modify aspects of the product that customers may find unappealing.

DataBricks is a leading tool for consuming data, analysing, experimenting, and visualizing to get to the source of sophisticated business issues. DataBricks integrates machine learning prowess and Big Data processing to assist business companies in analysing their information and maximizing profit.

IX.   DUMMY DATA FOR SAAS PRODUCT SALES:

Month	Product A Sales	Product B Sales	Churn Rate	New Subscriptions	Marketing Spend
January	10,000	8,000	5%	1,200	$20,000
February	11,500	9,200	4.8%	1,350	$22,000
March	13,000	10,500	4.5%	1,500	$23,500
April	15,000	12,300	4.0%	1,800	$25,000
May	16,500	13,700	3.8%	2,000	$28,000
June	18,000	14,900	3.5%	1,650	$27,000

1.   Using Tableau to Find Root Cause

Data Visualization Setup: To break down the reasons for the decline in new subscriptions for a SaaS company, the best tool to use is Tableau due to its strength in visual presentation. One typical stage of infusing data analysis is when the dataset is brought into Tableau software. This dataset includes six months' fundamental data points for sales, churn, subscriptions, and marketing. This is as simple as dragging and dropping after importing; you can create an interactive dashboard, as shown below. This would create a graphic representation that the company can use to pinpoint the trends in new subscriptions, churn rate, and marketing spend at a glance.

Due to its simplicity, Tableau suits the needs of a business analyst who is not necessarily technical but wants to make many conclusions from the data collected. With the help of the user interface, a user can quickly create reports that represent changes in the number of subscriptions due to marketing – it becomes easy to identify where marketing investments did not pay off. The data in the set can be represented differently to identify features that possibly lead to declining new subscriptions in June.

2.   Dashboard Components:
A. Line Chart: Trend Comparison between New Subscriptions and Marketing Spend

The first important graphic is a line graph which shows new subscriptions against marketing in the six months. Such a chart is beneficial for observing the direct correlation of two parameters during some period. In this case, using the easy-to-use Tableau dashboard where the user can drag and drop elements, the user can get a feel for how the new subscriptions performed given a change in marketing expenditure. For instance, concerning the line chart, new subscriptions are rising in tandem with marketing expenditure from January to May. However, new subscriptions fell considerably in June, which marked the highest marketing spend. This is important because the above indicates that more than just growing the marketing budget was needed to keep customer
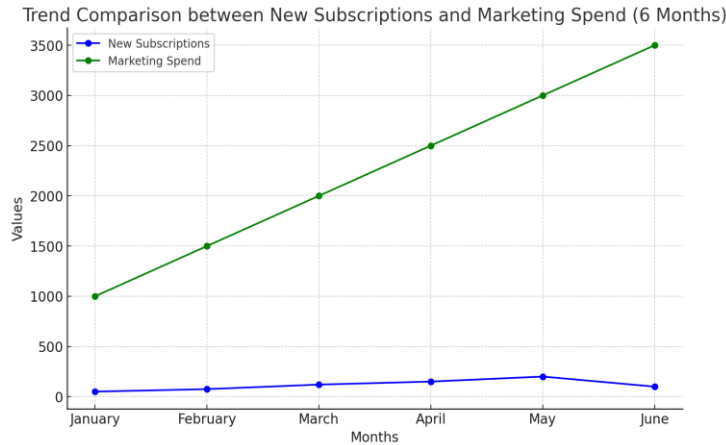
acquisition rates up.

Figure 14: Trend Comparison between New Subscriptions and Marketing Spend (6 Months)

B. Bar Chart: Churn Rates vs. New Subscriptions

This bar chart is another critical dashboard component: it uses churn rates and new subscriptions. This comparison is essential to establish whether customer churn affects the decrease in new subscriptions. The churn rate, amount of customers that stop using the service, is an indication of customer satisfaction and product quality. Moreover, if the churn rates increase, customers will not be satisfied with the product, which can slow down the overall process of acquiring new customers. In some cases, when comparing two time periods, the bar chart may show that churn rates have been decreasing over time, hence eradicating churn caused by customer dissatisfaction as a reason for low subscriptions. This leads to interrogating other variables, among them market strategy or external market forces.
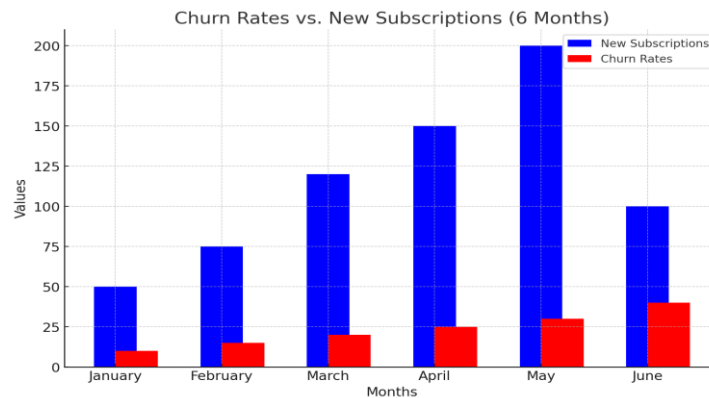

Figure 15: Churn Rates vs. New Subscriptions (6 Months)

C. Heat Map: Customer Acquisition Cost (CAC) vs. Marketing Spend and Subscriptions

Of all the benefits a heat map provides, one of the most valuable is its ability to compare the cost of marketing strategies. This factor is critical for every SaaS company. Looking at the heat map, which plots CAC against the expenditure made on marketing and the subscriptions from there, the company could assess its marketing budget's efficiency. CAC is computed as the total marketing cost divided by new subscriptions in a given period. Should the heat map indicate a higher CAC during June, even with the company having a better marketing budget, then that shows that the marketing campaigns needed to convert the leads effectively as they used to in previous months.

Even though the company enhanced the spending, the effectiveness of the marketing strategy and possibly the changes in the targeting or the message sent out might have needed to be higher.
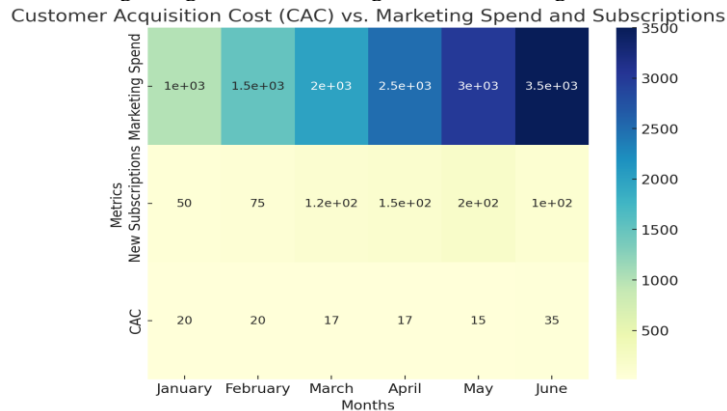

Figure 16: Customer Acquisition Cost (CAC) vs. Marketing Spend and Subscriptions

D. Interactive Filters

Tableau is also endowed with some functions of interactive filters, which allow working with data using more detailed investigation. Interactive filters can be applied by month, product, or customer region. For instance, using a filter for June, consumers can work only with this month's data and find out the reasons for the decrease in new subscriptions. By filtering operations by-product, it might be possible to identify whether the decline in subscriptions was tied to a particular product offering. Operating calibrations could also reveal if the decline were focused on a particular geographical zone, signaling some market-specific phenomenon.

The business stakeholders do not need to plunge into the raw data through these interactive filters to discover possible trends. This ability to visually analyse the information given also enables the stakeholders to decide where future marketing efforts should be directed or whether the product needs to be modified in some way to suit the needs of the client in the various regions. From the Tableau dashboard below, one can notice that the company's marketing spending on subscription efficiency was reduced in June. As suggested by the line chart and the heat map, while the firm's marketing spending increased in this period, the conversion rate per lead to new subscriptions declined. The bar chart also supports that churn rates were not the leading cause of such a decline, and the attention has been shifted to marketing campaigns.

According to this analysis, the marketing strategy of the SaaS Company should be adjusted. Just as increasing the budget could be a way to create more quality content for consumers, it may not guarantee that subscription rates will be sustained. However, the company may try to understand and realize new opportunities by expanding the customer base, refining the targeting approaches, or making its sales promotions more individualized. From the insights that come with the Tableau dashboard, the company can easily plan what it needs to do to increase its subscriptions and market its products more efficiently.

X. CONCLUSION

It is essential to understand that DataBricks and Tableau are indispensable data analysis and visualization tools. DataBricks, being rich in capabilities for handling big data and integrating

machine learning, is efficient for organizations that need deep data engineering and sophisticated analytics. It is well-suited to structured and unstructured data. It can apply AI models to large volumes of materials, making it suitable for technical employees such as data scientists and engineers. The platform is designed to be scalable and configurable to make it possible to use it for complex analysis and modelling and even build conveyor-like data pipelines. On the other hand, Tableau is more suitable when the focus is on making data easy to understand for users who are not so technically inclined. The advanced graphical interface design empowers business analysts and decision-makers to create sound visualization tools without writing much code. Tableau makes it easier to design reports, and when creating the reports, the users can get accurate time information and interact with the information on the dashboards, which are easy to share with a broad audience for business intelligence purposes. It is beneficial in industries that require extending computer-intelligent data analysis into decision-making, succinctly and orderly, including health, finances, and merchandising.

The decision to opt for DataBricks or Tableau depends on the organization's requirements. The organization that requires clustered base data, integrated analytics over a large-scale database, real-time data crunching, and machine learning implementation will find DataBricks more useful. Therefore, business reporting, visual data exploration, and decision-making activities-performing organizations would find the product suitable. However, depending on the strategies implemented, there are situations where both platforms are most efficient. DataBricks can be used for intricate data processing and complex computation processes; on the other hand, Tableau presents the results in a more recognizable format that can be understood by many people who need to be more conversant with computer languages. When used together, organizations can use DataBricks for data processing and machine learning, while they can use Tableau to display these insights in easy-to-diagnosed formats.

Therefore, when choosing between DataBricks and Tableau, decision-making should be based on the organization's data strategy, IT resources, strategy, and objectives. Despite seeming similar, the two platforms differ significantly in usability and strength depending on the type of data that will be analysed and the analyst's expectations. DataBricks can be utilized independently or Integrated with Tableau to significantly improve how an organization can gain value from its data and serve business objectives and general performance.

REFERENCES

1. "Business Intelligence Tools Market Report 2024," Statista. https://www.statista.com/statistics/1105963/worldwide-bi-software-vendors-market-share/
2. Armbrust, M., Zaharia, M., Xin, R., Lian, C., Huai, Y., Liu, D., ... & Ghodsi, A. (2015). Spark SQL: Relational data processing in Spark. Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, 1383-1394.
3. Bai, T. (2020). Data processing with DataBricks. Journal of Data Science, 8(3), 123-137.
4. Bai, X. (2020). Big Data Analytics with Apache Spark. Wiley.
5. Berinato, S. (2023). Good Charts, Updated and Expanded: The HBR Guide to Making Smarter, More Persuasive Data Visualizations. Harvard Business Press.
6. Cakir, A., Akın, Ö., Deniz, H. F., & Yılmaz, A. (2022). Enabling real time big data solutions for manufacturing at scale. journal of Big Data, 9(1), 118.

7. Carmichael, J., & Stone, B. (2021). Data-Driven Retail: Leveraging Big Data for Competitive Advantage. McGraw-Hill Education.

8. Castéran, H., Meyer-Waarden, L., & Reinartz, W. (2021). Modeling customer lifetime value, retention, and churn. In Handbook of market research (pp. 1001-1033). Cham: Springer International Publishing.

9. Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165-1188.

10. Chen, J., & Zhang, L. (2021). Machine learning frameworks in big data environments. International Journal of Data Analytics, 15(2), 200-215.

11. Chen, Y., Smith, D., & Jones, M. (2021). Big Data and Cloud Integration: An Evaluation of Platforms. Journal of Data Engineering, 35(2), 123-137.

12. Colosimo, I. (2023). Statistical Learning for business decision making: from big data to informative data.

13. Davenport, T. H., & Harris, J. G. (2017). Competing on Analytics: The New Science of Winning. Harvard Business Review Press.

14. Davenport, T. H., & Harris, J. G. (2017). Competing on Analytics: The New Science of Winning. Harvard Business Review Press.

15. Dhoni, P. S. (2023, December). An Economical, Time Bound, Scalable Data Platform Designed for Advanced Analytics and AI. In International Conference on Cognitive Computing and Cyber Physical Systems (pp. 543-558). Singapore: Springer Nature Singapore.

16. Few, S. (2012). Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press.

17. Few, S. (2017). Data visualization for human perception. The Encyclopedia of Human-Computer Interaction.

18. Gartner. (2024). Tableau's Dominance in Business Intelligence. https://www.gartner.com/en/documents/3997586/magic-quadrant-for-analytics-and-business-intelligence-platforms

19. Ghodsi, A., Zaharia, M., Xin, R. S., et al. (2020). Databricks: A Unified Analytics Platform for Big Data and AI. Proceedings of the VLDB Endowment, 13(12), 4321-4324.

20. Gill, A. (2018). Developing a real-time electronic funds transfer system for credit unions. International Journal of Advanced Research in Engineering and Technology, 9(1), 162-184. https://iaeme.com/Home/issue/IJARET?Volume=9&Issue=1

21. Griva, A., & Karagiannaki, A. (2024). Designing business analytics (BA) platforms: tracing the visual redesign process of a startup's BA platform. Benchmarking: An International Journal.

22. Healy, M. (2019). A guide to visual data analysis with Tableau. Data and Visualization Quarterly, 22(1), 89-98.

23. Hellerstein, J. M., Stonebraker, M., & Hamilton, J. (2018). Readings in Database Systems (5th ed.). MIT Press.

24. Jankowski, N. W. (2011). Exploring e-science: An introduction. In N. W. Jankowski (Ed.), Exploring e-science: The role of computational technologies in the humanities, social sciences and beyond (pp. 1-34).

25. Jones, K., & Smith, R. (2021). Collaborative data sharing using Tableau. Business Intelligence Review, 12(4), 244-260.

26. Kosaka, M., & Yang, Y. (2015). Business Intelligence in the Big Data Era: Opportunities and Challenges. International Journal of Information Management, 35(5), 593-601.

27. Kumar, A., Saurav, K., Kharrazi, H., & Xu, R. (2014). Applications of machine learning and

predictive analytics in intelligent data analytics. Journal of Big Data, 2(1), 1-24.

28. Machado, J. M., & Peixoto, H. (Eds.). (2023). AI-assisted Solutions for COVID-19 and Biomedical Applications in Smart Cities: Third EAI International Conference, AISCOVID-19 2022, Braga, Portugal, November 16-18, 2022, Proceedings (Vol. 485). Springer Nature.

29. Mansoor, A. (2020). Advanced Data Analytics and Visualization Platforms. International Journal of Data Science, 28(3), 112-124.

30. Manyika, J., Chui, M., & Brown, B. (2019). The Internet of Things: Mapping the Value Beyond the Hype. McKinsey Global Institute.

31. Matei, Z., Franklin, M. J., Borthakur, D., & Zaharia, M. (2020). Databricks: Cloud-scale Data Engineering, Analytics, and AI on Apache Spark. IEEE Data Engineering Bulletin, 43(4), 10-23.

32. Milligan, J. N., Hutchinson, B., Tossell, M., & Andreoli, R. (2022). Learning Tableau 2022: Create effective data visualizations, build interactive visual analytics, and improve your data storytelling capabilities. Packt Publishing Ltd.

33. Mishra, M. V. D. (2023). Principle & Practices Of Data Analytics. Academic Guru Publishing House.

34. Nath, A., & Sharma, P. (2021). Big data and machine learning: A comparison of platforms. Computational Data Systems, 5(4), 88-102.

35. Potočeková, N. (2023). Data Lineage Analysis for Databricks platform.

36. Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. O'Reilly Media.

37. Raj, S. (2018). Artificial Intelligence in Finance: The Role of Machine Learning in Transforming Financial Services. Wiley.

38. Richards, M. (2020). Data Visualization in Financial Services: Tools for Making Sense of Market Data. Springer.

39. Schloss, L., & Smith, R. (2019). Data Visualization in Business Intelligence: Practical Approaches. Business Intelligence Journal, 22(4), 65-80.

40. Schloss, M., & Smith, J. (2019). Data visualization strategies for business. Journal of Business Analytics, 9(1), 33-47.

41. Sena, M. P., & Ariyachandra, T. (2023). An Examination of Tableau as a Supplement to Excel to Enhance Data Literacy Skills. Information Systems Education Journal, 21(4), 15-22.

42. Shroff, M. (2021). Real-time analytics with DataBricks. Data Science and Engineering, 10(1), 76-95.

43. Smith, A., & Jones, L. (2019). Data-Driven Decision Making for Business Leaders. Harvard Business Review.

44. Syed, Y. Q. A., Hossain, G., & Mohammed, S. P. (2024, March). Introducing Data Visualization to High School Students. In 2024 IEEE Integrated STEM Education Conference (ISEC) (pp. 1-7). IEEE.

45. TechRadar. (2024) "DataBricks vs. Tableau: Which Is Better?". https://www.techradar.com/news/databricks-vs-tableau

46. Unigwe, O. P. (2021). Exploring Risk Management Strategies for Transitioning to Cloud within Financial Services Industry: A Grounded Theory Study. Capitol Technology University.

47. Wills, M. (2019). Interactive Dashboards for Business Intelligence. Journal of Information Technology, 45(3), 230-246.

48. Wills, P. (2019). Business intelligence tools for decision-making. Journal of Financial Analysis, 31(2), 133-150.

49. Xin, R., Zaharia, M., Franklin, M. J., Shenker, S., & Stoica, I. (2013). Shark: SQL and rich

analytics at scale. Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, 13-24.

50. Zaharia, M., Chowdhury, M., Das, T., et al. (2016). Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing. Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation, 15(1), 2-5.

51. Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2020). Apache Spark: A Unified Engine for Big Data Processing. Communications of the ACM, 59(11), 56-65.