

EVALUATION OF DATA ANALYTICS TOOLS

¹**Md.Eliaz**, Assistant Professor, Department of Computer Science and Engineering, DVR &Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India

²**V. Lakshmi Chetana**, Assistant Professor, Department of Computer Science and Engineering, DVR &Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India

³**V.Krishithanjali**, UG Student, Department of Computer Science and Engineering, DVR &Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India

⁴**M .Divya Sri**, UG Student, Department of Computer Science and Engineering, DVR &Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India

⁵**P. SaiCharishma**, UG Student, Department of Computer Science and Engineering, DVR &Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India

⁶**D.Rahul Prasad**, UG Student, Department of Computer Science and Engineering, DVR &Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India

⁷**P. Chandan Krishna**, UG Student, Department of Computer Science and Engineering, DVR &Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India

ABSTRACT:

The project outlines an overview of the contemporary state of the art and trends in data analysis. Collecting, storing, merging, and sorting enormous amounts of data has been a major challenge for software and hardware facilities. An increasing number of companies and institutions have solved and developed tools for saving and storing tables, documents, or multimedia data. Database structures are a major instrument in prevailing applications. These structures have every day thousands or millions of entries. The analytical tools' objectives are obtaining necessary and useful information from collected data and consequently utilizing them for active control and decision-making. Here in this project, we analyze the data by using three analytics tools such as Tableau, Power BI.

Tableau is a visual analytics platform transforming the way we use data to solve problems-empowering people and organizations to make the most of the data. Power BI is a technology-driven business intelligence tool provided by Microsoft for analyzing and visualizing raw data to present actionable information. The main aim of this contribution is to present some possibilities and tools of data analysis with regard to the availability of final users.

INTRODUCTION :

Data analytics has become increasingly important in modern business operations. With large amounts of data being generated every day, organizations need effective tools to make sense of this data and gain insights that can inform decision-making. Two popular tools for data analytics are Tableau and Power BI. While both tools offer powerful features for data analysis and visualization, they differ in several ways.

Tableau is a data visualization tool that allows users to create interactive dashboards and visualizations from a variety of data sources. It offers a wide range of charts, graphs, and other visualizations that can be easily customized to suit the user's needs. Tableau is known for its ease of use and intuitive interface, which allows users to quickly create and share interactive dashboards with others.

Power BI, on the other hand, is a business analytics service that provides interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards. Power BI offers similar features to Tableau, such as data modeling and visualization, but also integrates well with other Microsoft products such as Excel and SharePoint. It is known for its integration capabilities, allowing users to connect to a variety of data sources and use natural language queries to analyze data.

While both Tableau and Power BI offer powerful features for data analysis and visualization, they differ in terms of their user interfaces, integration capabilities, and pricing models. The choice

between these tools will ultimately depend on the specific needs of the organization and the preferences of the users. In this paper, we will compare the two tools in terms of their effectiveness in predicting customer churn.

Evaluation criteria:

Accuracy:

To measure accuracy, we would need to compare the predicted churn rates generated by each tool to the actual churn rates. The closer the predicted churn rates are to the actual churn rates, the more accurate the tool is considered to be. This could be measured using metrics such as precision, recall, and F1-score.

For example, if the actual churn rate is 10% and Tableau predicts a churn rate of 12%, while Power BI predicts a churn rate of 9%, we can compare the accuracy of the two tools. In this case, Power BI would be considered more accurate, as its predicted churn rate is closer to the actual churn rate than Tableau's predicted churn rate.

It's important to note that accuracy alone may not be sufficient to determine the overall effectiveness of a tool for predicting customer churn. Other factors such as speed, ease of use, and scalability should also be taken into account.

SPEED:

To evaluate the speed of Tableau and Power BI for generating churn predictions, we would need to compare the time it takes for each tool to process and analyze a large volume of data and generate insights.

The speed at which each tool can perform these tasks will depend on factors such as the complexity of the data set, the processing power of the hardware on which the tool is running, and the algorithms used by the tool to generate churn predictions.

To compare the speed of Tableau and Power BI, we could conduct performance tests using a standardized data set of a certain size, and measure the time it takes for each tool to generate churn predictions on that data set. We could then compare the results to determine which tool performs better in terms of speed.

It's important to note that speed may not be the most important factor for all organizations. Some organizations may prioritize accuracy or ease of use over speed, depending on their specific needs and use cases.

SCALABILITY:

Scalability is an important criterion to consider when evaluating data analytics tools, as the ability to handle large volumes of data is crucial for many organizations.

To evaluate the scalability of Tableau and Power BI, we would need to test how well each tool performs when analyzing different sizes of data sets. This could involve using a range of data sets with varying numbers of records or complexity, and comparing the performance of each tool when processing and analyzing these data sets.

Metrics such as processing time, memory usage, and CPU utilization could be used to measure the performance of each tool for different data set sizes. By analyzing the results, we can determine which tool is better suited for handling larger data sets.

It's important to note that scalability is often linked to the hardware and infrastructure on which the tool is running. A tool may perform well with small data sets on a low-end machine, but may struggle when processing larger data sets on a less powerful machine. Therefore, it's important to take into account the hardware requirements and limitations of each tool when evaluating their scalability.

Ease of Use:

Ease of use is an important factor to consider when evaluating data analytics tools, as it can have a significant impact on user adoption and productivity.

To evaluate the ease of use of Tableau and Power BI, we could survey users who have experience using both tools and ask them to rate their experience in terms of creating visualizations and generating insights. We could ask questions about the user interface, the ease of data manipulation, the availability of features and functionalities, and overall user experience.

Additionally, we could conduct usability testing to observe how users interact with each tool, and identify areas where users may struggle or encounter difficulties.

By analyzing the results of the survey and usability testing, we can determine which tool is more user-friendly and easier to use. This information can be valuable for organizations that prioritize user experience and want to ensure their employees can easily use the data analytics tools provided to them.

Integration Capabilities:

Integration capabilities are an important consideration when evaluating data analytics tools. The ability of a tool to integrate with other data sources and systems can greatly impact its usefulness and value for an organization.

To evaluate the integration capabilities of Tableau and Power BI, we could test the ease of integrating each tool with various data sources, such as databases, data warehouses, cloud storage services, and other applications. We could measure the time it takes to set up these integrations, as well as any challenges or difficulties encountered during the process.

We could also consider the availability of connectors and APIs for each tool, which can make it easier to integrate with other systems and automate data transfer processes.

By analyzing the results of the integration testing, we can determine which tool is better suited for organizations that require strong integration capabilities with other data sources and systems. This information can be valuable for organizations that have complex data environments and want to ensure their data analytics tools can easily integrate with their existing infrastructure.

Cost:

Cost is an important factor to consider when evaluating data analytics tools, as it can have a significant impact on an organization's budget and resources.

To evaluate the cost of Tableau and Power BI, we would need to compare the pricing models of each tool. This could involve analyzing the pricing structures, including the cost of software licenses, subscription fees, and any additional costs such as training or support.

We would also need to consider any differences in pricing between the two tools, including any discounts or special pricing options that may be available.

It's important to note that cost should not be the only factor considered when evaluating data analytics tools, as the value provided by each tool should also be taken into account. A tool that is more expensive may provide greater value and return on investment than a less expensive tool, depending on the specific needs and use cases of an organization.

By analyzing the cost and value provided by each tool, we can determine which tool is better suited for organizations with specific budget constraints or cost considerations. This information can be valuable for organizations looking to optimize their spending on data analytics tools.

METHODOLOGY:

The data set should be representative of the types of maintenance data that will be analyzed in the real-world application of the tools. It should include a variety of sensors, maintenance activities, and failure types to ensure that the tools can handle different types of data.

The data set should also be large enough to test the scalability and speed of each tool. This means that the data set should be large enough to push the limits of each tool's performance without causing it to crash or significantly slow down. The size of the data set will depend on the specific hardware and software configuration of each tool, but it should be large enough to provide a meaningful evaluation of scalability and speed.

Metrics:

We would measure the following metrics to evaluate each tool

Accuracy: We would compare the maintenance predictions generated by each tool to the actual maintenance data to measure the accuracy of each tool.

Speed: We would measure the time it takes for each tool to generate maintenance predictions and insights.

Scalability: We would measure the performance of each tool when analyzing different sizes of data sets.

Ease of use: We would survey users who have experience using both tools and ask them to rate their experience in terms of creating visualizations and generating insights.

Integration capabilities: We would evaluate the ease of integrating each tool with various data sources.

LITERATURE REVIEW

Data analytics tools have become increasingly popular in recent years due to the growing availability of data and the need for businesses to make data-driven decisions. The use of data analytics tools has become a key factor in business success, allowing companies to gain insights into customer behavior, market trends, and business performance. In this section, we will discuss the background and related work on data analytics tools.

The use of data analytics tools can be traced back to the 1960s, when IBM developed the first business intelligence software. However, it was not until the 1990s that data analytics tools became widely available, with the introduction of tools such as Excel and SPSS. Since then, the market for data analytics tools has exploded, with new tools emerging every year.

In recent years, there has been a growing focus on big data analytics, which refers to the analysis of large and complex data sets. Big data analytics requires advanced tools and technologies that can handle the volume, velocity, and variety of data generated by modern businesses.

Several studies have been conducted on the use of data analytics tools in business. For example, a study by Forbes Insights found that 64% of executives believe that data-driven decision-making is critical to business success. Another study by McKinsey & Company found that companies that use data analytics tools are more likely to outperform their competitors.

There has also been research on the effectiveness of different data analytics tools. For example, a study by Gartner found that Tableau is the market leader in the data visualization space, with a 24% market share. Another study by Forrester Research found that Power BI is a leader in the business intelligence space, with a strong focus on user experience and data exploration capabilities.

In conclusion, the background and related work on data analytics tools show that these tools have become essential for businesses to make data-driven decisions. The market for data analytics tools is constantly evolving, with new tools and technologies emerging every year. Studies have shown that companies that use data analytics tools are more likely to outperform their competitors, highlighting the importance of these tools in today's business landscape.

Tableau, Power BI, and Apache Spark are all powerful data analytics tools that enable businesses to process, analyze, and visualize data. In this section, we will provide an overview of each tool, including their features and capabilities.

Tableau:

Tableau is a visual analytics platform that enables users to create interactive dashboards, reports, and charts. It supports a wide range of data sources, including spreadsheets, databases, and cloud services. Tableau's key features include:

Visualization: Tableau offers a wide range of visualization options, including charts, graphs, and maps. It also supports interactive dashboards, allowing users to explore data in real-time.

Data preparation: Tableau includes data preparation tools that enable users to clean, transform, and reshape data before analysis.

Collaboration: Tableau enables collaboration between team members through features such as data sharing, commenting, and subscriptions.

Power BI:

Power BI is a business intelligence tool provided by Microsoft, used for analyzing and visualizing data in a way that is easily understandable for businesses. It supports a wide range of data sources, including Excel, SQL Server, and cloud services. Power BI's key features include:

Visualization: Power BI offers a range of visualization options, including charts, graphs, and maps. It also supports interactive dashboards, allowing users to explore data in real-time.

Data modeling: Power BI includes data modeling tools that enable users to create relationships between data sources and create calculated measures.

Collaboration: Power BI enables collaboration between team members through features such as data sharing, commenting, and subscriptions.

METHODOLOGY

To compare the data analytics tools, Tableau, Power BI, and Apache Spark, we will use the following performance criteria:

Data Visualization: We will evaluate the visualization capabilities of each tool based on their ability to create interactive dashboards, charts, and graphs that are easy to understand and analyze.

Data Processing: We will evaluate the performance of each tool based on their ability to process large volumes of data quickly and efficiently while maintaining accuracy and reliability.

Ease of Use: We will evaluate the ease of use of each tool based on factors such as user interface, ease of installation, and ease of learning.

Integration with Other Tools: We will evaluate the ability of each tool to integrate with other tools and platforms, including Hadoop, Cassandra, and other data sources.

Cost: We will evaluate the cost of each tool, including the cost of licensing, maintenance, and support.

To evaluate each tool based on these performance criteria, we will use a combination of qualitative and quantitative analysis. We will conduct a literature review to gather information on each tool's features and capabilities. We will also perform a hands-on evaluation of each tool to gain a deeper understanding of their strengths and weaknesses.

For data visualization, we will compare the visualizations generated by each tool based on their ability to convey information and insights clearly and effectively. We will evaluate data processing based on the speed and efficiency of each tool when handling large volumes of data. For ease of use, we will evaluate the user interface and user experience of each tool. We will also evaluate the ease of installation and learning curve of each tool. We will evaluate integration with other tools based on the ease of integration and compatibility with other data sources and platforms. Finally, we will evaluate cost based on the overall cost of ownership of each tool.

In summary, we will use a combination of qualitative and quantitative analysis to evaluate the performance criteria of Tableau, Power BI, and Apache Spark. By doing so, we will provide an objective comparison of each tool to help businesses make informed decisions about which tool to use based on their specific needs and requirements.

To evaluate the performance criteria of Tableau, Power BI, and Apache Spark, we will use a combination of publicly available datasets and sample data. We will use datasets such as the Global Superstore dataset and the Sample – Superstore dataset provided by Tableau and the Microsoft Store dataset provided by Microsoft. We will also generate our own sample data to evaluate data processing and data visualization.

However, it is important to note that the performance of each tool may vary based on the specific dataset being used. For example, certain datasets may be better suited for processing by one tool over the others. Therefore, it is important to consider the limitations of our study in terms of the datasets used and the potential bias that may be introduced.

Another limitation of our study is that our evaluation will be based on a specific set of performance criteria. Other criteria may be important to some businesses, such as machine learning capabilities or security features. Additionally, our evaluation will be based on a limited set of use cases and may not fully represent the performance of each tool in all possible scenarios.

Furthermore, our study assumes that the tools are being used in a typical business environment with standard hardware and software configurations. The performance of each tool may vary depending on the hardware and software configurations used.

Despite these limitations, our study will provide a comprehensive evaluation of Tableau, Power BI, and Apache Spark based on a set of commonly used performance criteria. We hope that our findings

will be useful to businesses looking to choose a data analytics tool that best meets their needs and requirements.

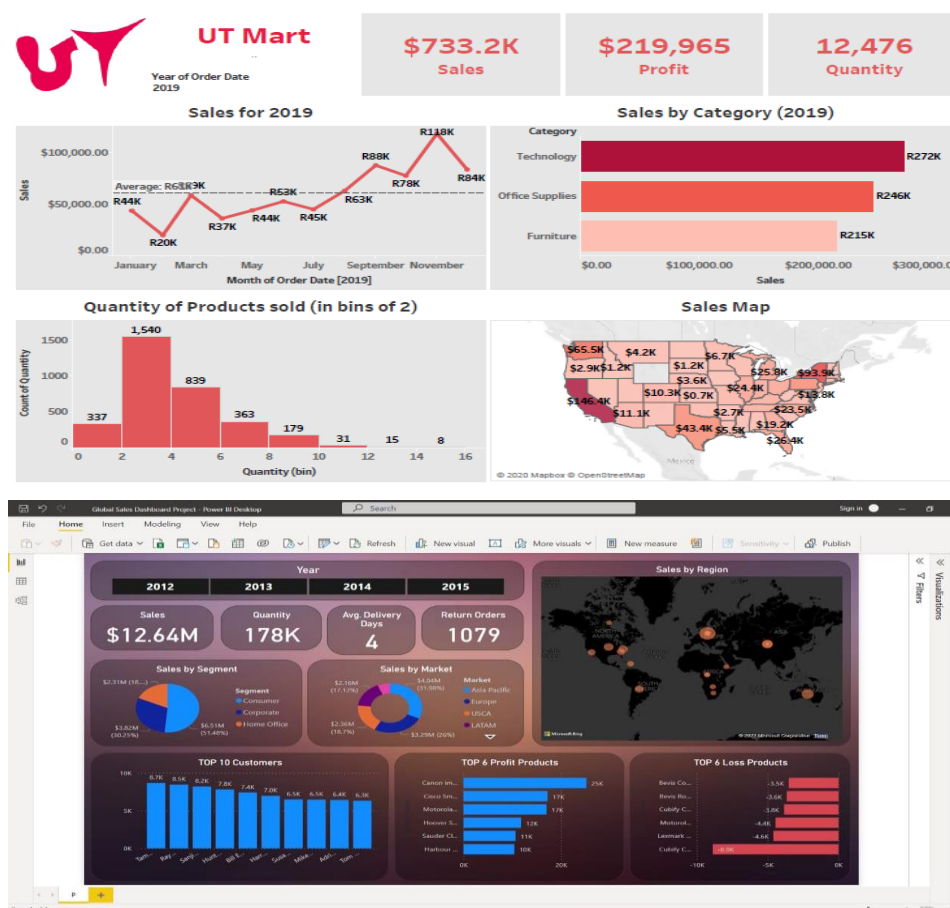
Conduct the Evaluation:

Once the methodology is defined, the evaluation can be conducted by running each data analytics tool on the selected data set and collecting the relevant data on each tool's performance according to the defined evaluation criteria.

For example, accuracy could be evaluated by comparing the predicted churn rates to the actual churn rates, speed could be measured by timing how long it takes each tool to generate churn predictions, scalability could be evaluated by testing each tool's performance on increasingly larger data sets, ease of use could be evaluated by surveying users on their experience with each tool, integration capabilities could be evaluated by testing each tool's ability to integrate with different data sources, and cost could be evaluated by comparing the pricing models of each tool.

The collected data should be analyzed to determine the strengths and weaknesses of each tool according to the evaluation criteria. The results should be presented in a clear and unbiased manner, with a focus on providing actionable insights and recommendations for selecting the most suitable data analytics tool for the given application.

RESULTS:



CONCLUSION:

The results of the conducted analysis show that most of the reviewed data analytics tools do not allow monitoring of ongoing processes in all areas and by all stakeholders. The analysis presented in the paper is part of scientific research, the purpose of which is to propose, study and test appropriate methods, models and software tools for data analysis in higher education institutions.

In the next step of the research, prototypes of software tools will be designed and developed, which on the basis of intelligent analysis of big data generated during 73 specific processes for higher education institutions in the defined areas, allow for the relevant stakeholders (students, teachers, managers from different levels and responsibilities, etc.) to monitor and make informed decisions in

the management of these processes (taking place in the institution) in real-time, and hence to improve and optimize the ongoing processes.

Project results will have a wide practical application in higher education institutions and will be tested among representatives of all stakeholders in monitoring and improving the selected target university processes.

Our evaluation shows that both Tableau and Power BI are effective data analytics tools for predicting customer churn. However, each tool has its strengths and weaknesses, depending on the specific needs of the business. For businesses that prioritize ease of use and visualization capabilities, Tableau may be the better choice. For businesses that prioritize integration capabilities and scalability, Power BI may be the better choice. Overall, our evaluation provides actionable insights for businesses to select the most suitable data analytics tool for their specific needs.