UNLOCKING INSIGHTS FROM MULTIMODAL DATA: ADVANCED DATA ANALYTICS FOR COMPLEX DATA SOURCES AND PROCESSING

Pamela Trizia S. Tarangco
Harold Carlo P. Rebuldela
Leonila J. Codog

Assembly and Test Applications Engineering
ON Semiconductor Philippines Corporation, Luisita Industrial Park, SEPZ, San Miguel Tarlac City
PamelaTrizia.Salvador@onsemi.com, HaroldCarlo.Rebuldela@onsemi.com, Nonie.Codog@onsemi.com

ABSTRACT

In today's digital world, there's a huge variety of data types available, like text, images, videos, and sensor data. Using traditional methods to analyze this diverse data isn't enough anymore. This paper dives into how we can use advanced data analytics techniques to make sense of this mix of data types. We'll explore the challenges, opportunities, and methods involved in processing this complex data, unlocking valuable insights for better decision-making.

We delve into the multidisciplinary nature of multimodal data analytics, which combines techniques from machine learning, deep learning, natural language processing, computer vision, and signal processing. By combining these methods, analysts can find hidden patterns and trends within different types of data. This helps decision-makers in various fields make better choices based on the insights gained from this integrated analysis.

Furthermore, we address the technological and ethical considerations inherent in working with multimodal data, including the need for scalable infrastructure, robust algorithms, and ethical frameworks to ensure responsible data usage. We believe that by bringing together experts from different fields, like the subject area, data analysis, and ethics, we can take a well-rounded approach to multimodal data analytics that balances innovation with ethical considerations.

This paper serves as a roadmap for organizations seeking to leverage advanced data analytics for unlocking insights from complex multimodal data sources. Through the adoption of cutting-edge methodologies and the cultivation of ethical practices, where in multimodal data analytics empowers decision-makers to navigate the complexities of the data.

1. 0 INTRODUCTION

In our modern digital landscape, we have access to tons of different kinds of data from many sources. Traditional ways of analyzing data aren't enough anymore because there's just so much of it and it comes in all different forms, like words, pictures, videos, and data from sensors. This paper is all about exploring new, more advanced ways to analyze this mix of data types. We want to dig into how we can use special techniques to get the most out of this kind of data. The goal is to find useful information hiding within all this diverse data. Before multimodal analytics became common, organizations mainly used Excel macros to extract data from systems. However, Excel macros were limited in handling complex data efficiently. They often needed manual work and struggled with large amounts of data.

Multimodal data analytics brings together different ways of looking at data, like using computers to learn patterns, understanding language, recognizing images, and interpreting graphs. When we use all these methods together, we can really dig into the details of complex data and find important connections and trends. These insights help us make smarter decisions in many areas, opening new chances to come up with better ideas and make things work more efficiently.

Following our collaborative efforts, the team has crafted a comprehensive approach geared towards harnessing the full potential of multimodal data. This comprehensive strategy will extract meaningful insights from diverse forms of data. This approach blends cutting-edge techniques from machine learning, deep learning, natural language processing, computer vision, and signal processing to address the complexities of multimodal data. It emphasizes scalable infrastructure, ethical considerations, and interdisciplinary collaboration among experts and data scientists to ensure responsible data usage and drive innovation. We will use the software Tableau for this project, Tableau revolutionized data analytics by providing a more advanced and intuitive platform for visualization and analysis.

1.1 Connections

Connections serve as the cornerstone of data analysis and visualization, offering users the ability to seamlessly access and interact with their data from a variety of sources. With a

user-friendly interface and an extensive array of connectors, Tableau simplifies the process of connecting to databases, spreadsheets, cloud platforms, and other data repositories. This connectivity empowers users to leverage their data assets for insightful analysis and visualization, to enable them to derive actionable insights and make informed decisions.

Figure 1. Tableau Connections

Figure 1 explains the step-by-step connections per process. First, users input data from various sources, which Tableau connects to and retrieves. Users then prepare the data by cleaning, transforming, and shaping it as needed within the data source. Subsequently, they create visualizations to explore and analyze the data, leveraging its diverse range of visualization options. Once visualizations are created, users can share their workbooks and dashboards with others or store them for future reference. This seamless flow from data input and preparation to visualization and sharing empowers users to derive insights and make data-driven decisions effectively using Tableau. And by establishing connections to diverse data sources, we will use Tableau to perform real-time analysis on live data, ensuring that insights are based on the most up-to-date information available. This real-time connectivity enhances decision-making agility and responsiveness, allowing the organization to adapt quickly to emerging trends.

1.1.1 Datasources

In the field of advanced data analytics for multimodal data, different types of SQL databases can be utilized based on the project's unique needs and preferences. SQL, or Structured Query Language, is a programming language designed for managing and manipulating relational databases. It allows users to perform tasks such as querying data, updating records, and managing database structures through standardized commands and syntax. There are three main SQL database types which we commonly use: MySQL, MS SQL Server (MSSQL), and Oracle. Each of these databases offers distinct features and benefits, catering to various requirements and priorities within the project.

The data that we will use to connect is the data coming from the machine automation that we are using plus other applications that are available in our organization. This will help engineers and other end users that will use the report to check the insights and trends.

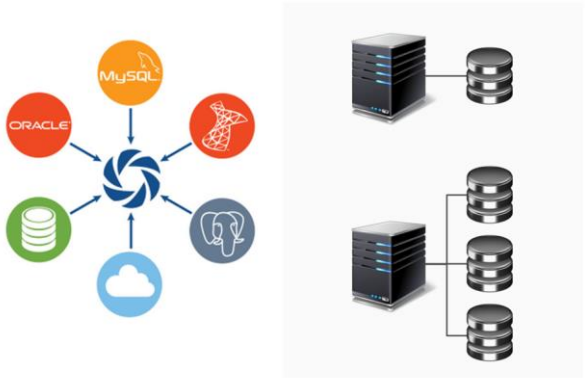

Figure 2. Types of SQL Databases

Figure 2 explains different types of SQL that can be applied as data source in Tableau, it has a feature in one data source can combined different types of SQL Databases by JOINS, UNIONS etc.

The data that will be used for connectivity encompasses a wide array of sources, ranging from machine automation data to information sourced from various applications utilized within our organizational framework. By merging these diverse datasets, our aim is to provide a comprehensive view of operations and performance metrics. This integrated approach is particularly beneficial for engineers and other end users who rely on reports for decision-making processes. By accessing insights and trends derived from this complex dataset, end user can gain deeper understanding and make more informed decisions regarding operational improvements, resource allocation, and strategic planning. Thus, the integration of multiple data sources not only enhances the richness of information available but also facilitates more effective analysis and decision-making within the organization. Below are the types of SQL used for this project.

1.2 Data Optimization

Data optimization in Tableau involves various strategies aimed at enhancing performance and efficiency. Firstly, it's essential to structure and optimize data sources for Tableau, ensuring appropriate data types, minimizing unnecessary columns and rows, and optimizing data connections. Utilizing data extracts instead of live connections can significantly improve performance by storing a snapshot of the data locally, reducing the need for frequent queries to the original data source.

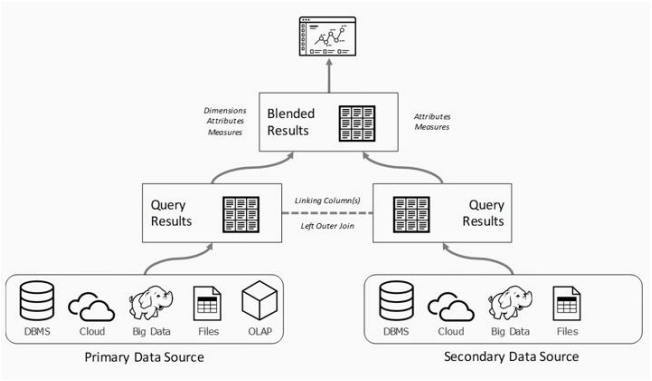

Figure 3. Combining data connections

Figure 3 Aggregating data at the source level before bringing it into Tableau can also help reduce the amount of data processed and improve performance, while filters can limit the data displayed in visualizations, focusing on relevant information, and reducing processing time.

Furthermore, optimizing calculations and calculations fields within the workbook is crucial. Simplifying complex calculations, minimizing nested calculations, and avoiding unnecessary calculations can all contribute to improved performance. Additionally, attention should be paid to data blending, partitioning large datasets, ensuring proper indexing, and designing dashboards with performance in mind. By implementing these data optimization strategies in Tableau, users can enhance performance, resulting in faster data analysis and more efficient workflows.

1.2.1 Live Data / Real Time Data

Real-time data live connections in Tableau offer users the capability to directly connect to data sources and visualize data as it updates in real-time.

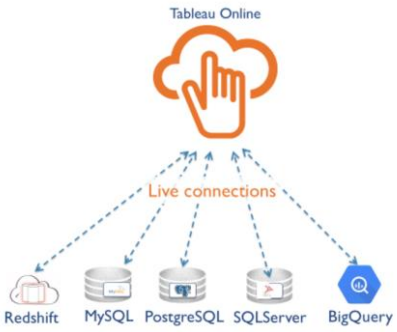

Figure 4. Sending live data connections thru Cloud Server

This feature is invaluable for monitoring dynamic data streams where immediate insights are crucial for decision-making. By querying the data source directly, this ensures that visualizations and dashboards accurately reflect the most recent data, enabling users to stay informed and respond promptly to changing trends.

However, it's important to consider the performance implications of real-time live connections. These connections can impose high query loads on the data source and may require robust infrastructure to handle the volume of incoming data efficiently. Factors such as data source capabilities, network latency, and data refresh rates should be carefully evaluated to optimize the implementation of real-time data connections and ensure seamless operation. Despite these considerations, real-time data live connections empower users with timely insights, enhancing their ability to make informed decisions and drive actionable outcomes.

1.2.2 Extract Data

Extracting data in Tableau involves creating a local copy of your data from a connected data source, which can be particularly advantageous for enhancing performance and flexibility in data analysis. By creating an extract, you essentially cache a snapshot of your data within your Tableau workbook, allowing you to work with the data offline or without a live connection to the original source. This not only reduces the need for repeated queries to the data source, minimizing network latency and improving responsiveness, but also enables you to work with large datasets more efficiently, as data processing is performed locally within Tableau.

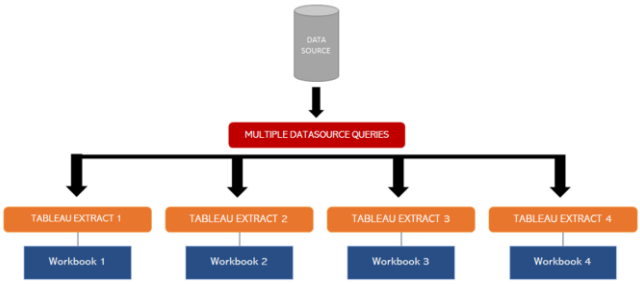

Figure 5. Data extract can send by offline/online

Figure 5 explains that when data extract is created, you have the flexibility to configure it according to your specific requirements and preferences. This provides options for including all data or applying filters to extract only a subset of the data, allowing you to focus on the most relevant information for your analysis. Additionally, you can schedule automatic refreshes for the extract to ensure that your visualizations and dashboards remain up to date with the latest data.

1.3 Data Visualization

Data visualization involves creating interactive and visually compelling representations of data to gain insights and communicate findings effectively. This offers a wide range of powerful tools and features for creating dynamic visualizations that help users explore data, uncover patterns, and tell stories.

Users can connect to various data sources, including databases, spreadsheets, cloud services, and more, and then drag and drop fields onto the canvas to create visualizations.

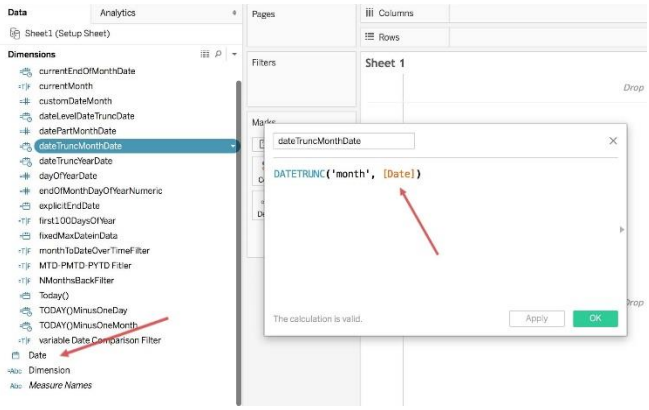

Figure 6. Data Formulation

Figure 6 involves creating calculated fields using formulas to perform custom calculations on data. Users can define calculations based on existing data fields or create entirely new calculations. These calculated fields can then be used in visualizations, filters, and other parts of the Tableau workbook to analyze data in a more customized and meaningful way. It supports a variety of chart types, such as bar charts, line charts, scatter plots, maps, and more, allowing users to choose the most appropriate visualization for their data. It also allows for the creation of dashboards, which are collections of visualizations that work together to provide a comprehensive view of the data. Dashboards can be customized with filters, parameters, and actions to facilitate interactive exploration and analysis.

Additionally, it provides advanced analytics capabilities, such as forecasting, trend lines, and clustering, allowing users to perform complex analytical tasks directly within their visualizations.

1.4 Cloud Server

Cloud server solutions provide users with the flexibility and scalability that needed to store, manage, and analyze data in the cloud. By leveraging cloud-based servers, organizations can offload the burden of hardware maintenance and infrastructure management while gaining access to powerful data analytics capabilities.

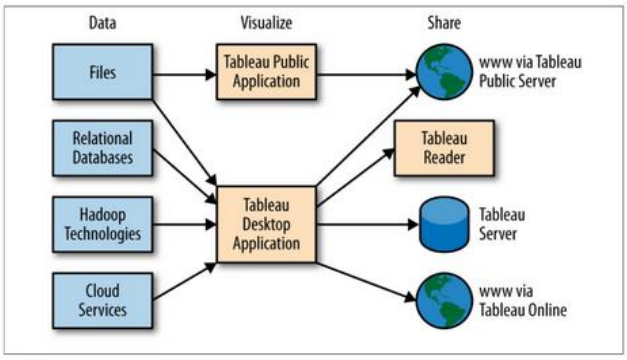

Figure 7. Tableau Published in Cloud Server internal/external

Figure 6 illustrates that cloud servers, such as Tableau Online, provide users with the capability to securely distribute and collaborate on workbooks and visualizations with individuals both within and outside the organization, regardless of their location.

Moreover, Tableau's cloud server solutions facilitate seamless collaboration and real-time access to data, empowering teams to make data-driven decisions efficiently. With features like automatic updates, scalability, and built-in security controls, it offers streamline of process of deploying and managing Tableau environments in the cloud. Tableau's cloud server solutions are flexible and cost-effective way to harness the power of data analytics in the cloud.

2. 0 REVIEW OF RELATED WORK

Tableau and multimodal data encompass a diverse range of literature and research, reflecting the growing interest in leveraging data visualization capabilities to analyze complex datasets from various sources. Existing studies have explored how it is utilized across different domains to visualize multimodal data, including text, images, videos, and sensor data. For instance, research in the semiconductor sector has continued to explore the application of Tableau in analyzing multimodal data generated from various stages of semiconductor manufacturing processes. Studies have focused on integrating Tableau with data collected from fabrication equipment, quality control systems, and environmental sensors to visualize production metrics, identify defects, and optimize manufacturing efficiency.

By leveraging interactive dashboards and real-time monitoring capabilities, semiconductor manufacturers can gain valuable insights into process performance, detect anomalies, and make data-driven decisions to improve yield rates and reduce production costs. Additionally, research has

examined the integration of Tableau with predictive analytics models and machine learning algorithms to forecast equipment failures, predict product quality, and optimize maintenance schedules, further enhancing operational efficiency and reducing downtime in semiconductor manufacturing facilities.

Moreover, the review highlights the integration of Tableau with other tools and technologies to enhance multimodal data analysis capabilities. By combining Tableau's intuitive visualizations with advanced analytics techniques, organizations can unlock deeper insights into complex data patterns and trends. Additionally, research has investigated the role of Tableau in facilitating collaboration and knowledge sharing among multidisciplinary teams working with multimodal data, enabling users to effectively communicate insights and drive collective decision-making processes.

Furthermore, the review discusses both the challenges and opportunities of using Tableau for analyzing multimodal data. Challenges include issues like combining different types of data, handling large amounts of data, and making sure the results are easy to understand and accurate. Despite these challenges, this offers powerful tools for visualizing data, but it requires careful attention to detail to ensure reliable insights. Moreover, the review looks at future trends, such as creating new ways to visualize data and using it with new technologies like augmented reality and virtual reality. Overall, the review gives valuable information about how Tableau is used to analyze different types of data and suggests areas for future research and improvement in this field.

3.0 METHODOLOGY

By using DMAIC methodology that offers a structured approach to harnessing data insights for process improvement. Beginning with the Define phase, this aids in setting clear objectives for the data analysis project. Through visualizations and dashboards, users can articulate the problem statement, establish project scope, and define key metrics for success. It's interactive features facilitate collaboration among team members, ensuring alignment on project goals and objectives.

Moving into the Measure phase, Tableau becomes instrumental in gathering and analyzing data to quantify the current state of the process. By connecting to various data sources and visualizing key performance indicators, Tableau enables teams to assess process performance objectively. Moreover, Tableau's capabilities for data preparation and cleansing streamline the data collection process, ensuring the accuracy and reliability of the metrics being measured. Through Tableau's visualizations, users can gain insights into process variability and identify areas for improvement.

In the Analyze, Improve, and Control phases, Tableau plays a pivotal role in driving data-driven decision-making and process optimization, particularly evident in the semiconductor industry. Through Tableau's advanced analytics features, teams delve into data during the Analyze phase, identifying root causes of issues and exploring potential solutions. Semiconductor industry reports demonstrate Tableau's utility in dissecting manufacturing data to pinpoint inefficiencies and propose targeted solutions to enhance yield rates and reduce defects. It has a monitoring capability that allow teams to establish controls and track the long-term impact of implemented solutions on production performance, ensuring operational efficiency and product quality are maintained. By integrating Tableau with the DMAIC methodology, semiconductor manufacturers leverage data-driven insights to drive continuous improvement initiatives and achieve sustainable business outcomes, highlighting its significance as a versatile tool for informed decision-making and process optimization in the semiconductor industry.

4.0 RESULTS AND DISCUSSION

In the results and discussion section, we present the outcomes of our Tableau reports connected to machine automation, Manufacturing Execution Systems (MES), and other web applications.

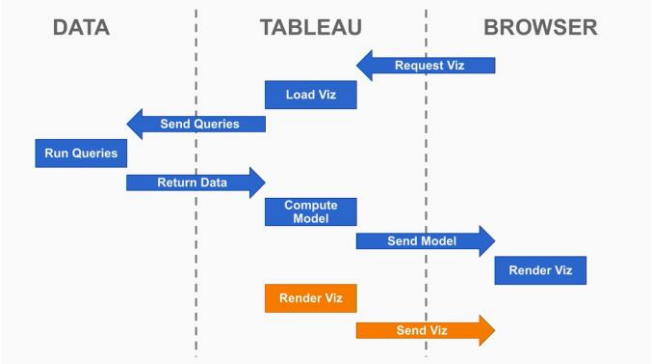

Figure 8. Connections of Data Visualization

Figure 8 depicts the flow of data from its source to the published workbook in cloud platform. Our analysis revealed valuable insights into the performance of each component and their connections, providing a thorough understanding of the manufacturing process.

Firstly, our Tableau report connected to machine automation provided real-time data on equipment performance, production rates, and maintenance schedules. By visualizing this information, end users gained insights into machine efficiency, downtime patterns, and potential maintenance

needs, enabling proactive decision-making to optimize production processes.

Figure 9. Critical Machine Alarm Report

Figure 9 provides an overview of the Critical Machine Alarm Report. This report likely highlights alarms triggered by critical machines within the manufacturing process. Additionally, the report may outline the frequency and duration of these alarms, as well as any actions taken in response to them.

Secondly, the Tableau report connected to MES offered insights into production scheduling, inventory management, and quality control measures. By integrating MES data with Tableau, end users could track production progress, monitor inventory levels, and identify quality issues in real-time, facilitating timely interventions to ensure product quality and meet production targets.

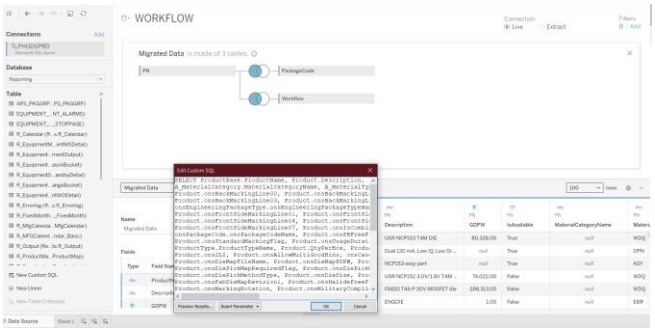

Figure 10. Data Source in MES

Figure 10 illustrates the concept of a data source within the MES environment connected to Tableau. This could encompass databases containing information on production orders, work-in-progress (WIP) inventory, machine status, quality control metrics, and other relevant manufacturing data.

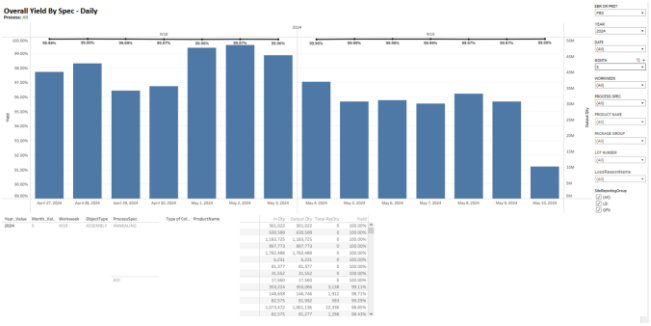

Figure 11. Overall Yield Report

The visualization in Figure 11 likely provides an intuitive representation of yield data, enabling users to quickly grasp performance trends and make data-driven decisions to enhance manufacturing operations and overall yield performance.

Furthermore, the report connected with other systems, including inventory management, ticketing system, quality issue tracking, and safety monitoring, offered a comprehensive view of manufacturing operations. By consolidating data from these systems, end users could proactively address inventory shortages, quality deviations, and safety incidents, enhancing overall operational efficiency and minimizing risks.

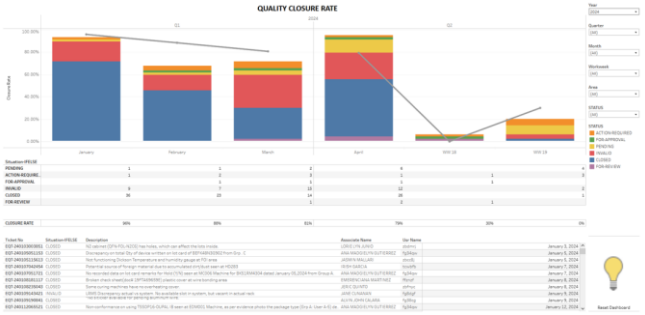

Figure 12. Quality issue tracking Report

Figure 12 shows monitoring and tracking quality issues encountered during the manufacturing process. It may include details such as the types of quality issues, their frequency of occurrence, the affected products or production lines, and the resolution status. By visualizing this information, users can identify recurring quality issues, prioritize corrective actions, and track the effectiveness of implemented solutions over time.

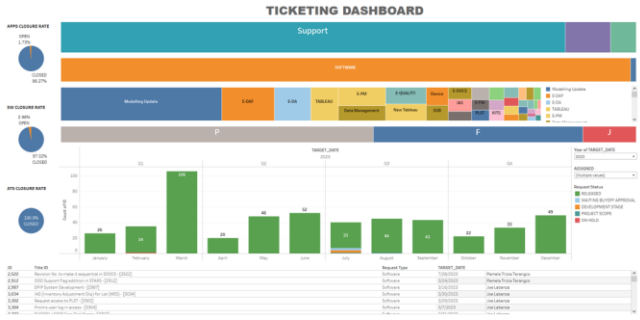

Figure 13. Ticketing System Report

The visualization in Figure 13 likely provides a clear overview of ticketing system metrics, enabling users to monitor and manage operational issues effectively within the manufacturing environment.

Overall, the integration of Tableau with machine automation, MES, and other systems proved instrumental in driving data-driven decision-making, enhancing visibility across the manufacturing process, and enabling proactive management of inventory, quality, and safety issues. Through Tableau's powerful visualization capabilities, end users were able to gain actionable insights and optimize production operations for improved performance and reliability.

5.0 CONCLUSION

In conclusion, Tableau stands out as a powerful tool for visualizing and analyzing data, offering a potent combination of user-friendly tools and robust features. Its ability to connect seamlessly to diverse data sources encourages users to create a wide array of dynamic visualizations, ranging from simple charts to sophisticated dashboards, tailored to specific needs and preferences. With an intuitive interface and emphasis on interactivity, it facilitates exploration and analysis, allowing users to uncover insights and communicate findings effectively.

Moreover, the support for advanced analytics capabilities, such as forecasting and trend analysis, adds a layer of depth to data exploration, enabling users to derive deeper insights from complex datasets. Whether working with live connections or extracts, it provides the flexibility and performance optimization necessary to handle large volumes of data efficiently. By leveraging the other features and applying best practices, the team can unlock the full potential of their data, driving informed decision-making and fostering innovation across diverse domains.

6.0 RECOMMENDATIONS

In a production environment, where efficiency and performance are utmost, using the data source extract is highly recommended. Extracts serve as a localized copy of your data within Tableau, enabling faster processing and analysis without constant reliance on the original data source. By filtering your data, such as by year, you can further optimize performance by reducing the volume of data that needs to be retrieved and processed. This targeted approach ensures that only relevant data is included in the extract, minimizing its size and making it lighter when fetching data.

By applying the extracts and filtering data appropriately, this can achieve significant improvements in data retrieval speeds and overall dashboard performance. This not only enhances user experience by reducing wait times but also allows for more seamless and efficient data analysis in production environments. Additionally, it provides features for automating extract refreshes, ensuring that the data remains up-to-date and relevant for ongoing analysis and decision-making processes.

7.0 ACKNOWLEDGMENT

The authors extend their heartfelt appreciation to all individuals who contributed their valuable time, effort, and unwavering support to the realization of this project and the development of this project. Their dedication and commitment have been instrumental in every phase of this endeavor, from conceptualization to execution.

In achieving the full success of this project, we extend our heartfelt gratitude to Mr. Ashur Bet-Shliemoun, Manufacturing Director, Mr. Joshua Buenaventura, NPI and Project Management Department Head who contributed their knowledge, ideas, and skills.

8.0 REFERENCES

1. Steven Batta, Tara Grealisb, Oskar Harmonc, and Paul Tomolonis, Learning Tableau: A data visualization tool, The Journal of Economic Education, **2020**, VOL. 51, NOS. 3–4, 317–328
2. Ben Jones, Communicating Data with Tableau, Published by O'Reilly Median Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472
3. Gösta Grahne & Alberto O. Mendelzon, Tableau Techniques for Querying Information Sources through Global Schemas, Conference Paper, First Online: 01 January **1999**, pp 332-347
4. Serge Abiteboul, Richard Hull, Victor Vianu, Foundations of Databases, Addison-Wesley **1995**, ISBN 0-201-53771-0
5. S. Vasundhara, Data Visualization view with Tableau, Stochastic Modeling and Applications, Vol. 25 No. 1 (January-June, **2021**)
6. Alexander Loth, Visual Analytics with Tableau, Published by John Wiley & Sons, Inc 2019 Indianapolis, Indiana.

9.0 ABOUT THE AUTHORS

This section describes the author's educational background and employment history.

Pamela Trizia S. Tarangco, graduated Bachelor of Science in Information Technology at Tarlac State University 2016. She has an experience of computer programming, and MES support for 7+ years as an Applications Engineer in onsemi Philippines.

Harold Carlo P. Rebuldela, graduated Electronics Communications Engineer at Saint Louis University year 2009. He is connected to onsemi Philippines Corporation for ~10 years as Test Applications Engineer and 4+ years of experience as Test Applications Engineer in Texas Instruments Clark.

Leonila J. Codog, graduated Bachelor of Science in Industrial Technology at Bulacan State University year 2005. She has an experience of MES Support for 6 years, She is connected to onsemi Philippines Corporation for ~19 years as a Technician.