iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44400

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Comparative Study of Data Visualization Tools in BigData Analysis for Business Intelligence

Aryamaan Pandey[1], Ishita Sharma[2], Akshat Sachan[3], Dr. P. Madhavan[4]
[1, 2, 3]*BTech CTech SRMIST, Kattankulathur*
[4]*Assoc. Prof., Dept of CTechSRMIST, Kattankulathur*

Abstract: ***In this era of information, the development of figures has radically expanded, where a colossal amount of data is being delivered from various sources. Because of this huge collection, the worth of information turns into a significant component in each perspective. Data communication is very important to any business - be it small, midsize or Brobdingnagian. Businesses need Data Visualizations to identify data trends at a rapid pace, which would otherwise be tedious. Data Visualization is a robust technology capable of presenting a large dataset in a graphical format. Its centrality goes far beyond merely introducing information to clients. It also assists customers in comprehending information and making decisions. It involves the use of charts, graphs, diagrams, and other info-graphics to communicate data. Surprisingly, a plethora of Data Visualization tools and methodologies have evolved in recent years. This paper conducts an assessment of the mostwidely used and distributed visualization tools for massive data sets, presenting a synoptic of the main functional and non-functional characteristics of the surveyed tools.***

Keywords: Big Data, Data Visualization, Data VisualizationTools, Business Analytics, Business Intelligence

I. INTRODUCTION

Data visualization and translation of information are gettingvital expertize in today's trade world [1]. The distinction between what is relevant and what isn't becomes increasingly blurred, as the amount of data available grows. Undoubtedly, the consideration of information visualization tools and dashboards available permits officials to form choices in a briefsum of time.

Data visualization plays a crucial role within the voyage ofchanging crude information into a more refined form. It helps to present data constructively regardless of what business or vocation you select. One of the essential steps in any business is taking raw data, modeling it, and presenting it in order to draw conclusions [2]. Data visualization tools have the capacity to prepare and show the information in an absorbable manner. It is effective in making solid information reports and dashboards by accumulating enormous amounts of data from distinct sources thereby helping in the discovery of patterns and the formulation of an opinion.

This study distinguishes the data visualization tools on the basis of the format of output, need for coding knowledge and commercial part of the tool [10]. Massive volume of data is displayed in a way that is both accessible and intelligible.

In this paper, we examine several prominent Data Visualization tools, including Tableau, Power BI, ZohoAnalytic and others.

II. METHODOLOGY

Exploration of shipment dataset (From Kaggle) is done using six different data visualization tools which are - Tableau, Power BI, Zoho Analytic, Dataiku DSS, Celonis and MS Excel. This dataset contains 10999 observations of 12 variables *viz.* Warehouse Block, Mode of shipment, Customer rating etc. The dataset is examined to classify the different modes of transportation through which the item was delivered. It is created in order to obtain an exact count of shipments using three different modes of transportation: flight, road, and ship. To learn about client opinions on shipment services, a pie chart is utilized to examine customer ratings by category. Finally, to provide a broad picture of consumer satisfaction, we calculated the average of customer ratings. A line graph is utilized to show the importance of the product in order to count the prioritized items. Furthermore, a bar chart is created to track the number of things dispatched from various warehouses that arrived on time. To aid the users, a histogram is created to provide information on the number of products that fall under a given discount range. Last but not least, a donut chart is created to correlate the number of customer service calls with the number of things that did not arrive at their destination. This is done in order to maintain track or to have a proper check on the customer service.

In this study, we are evaluating and comparing the above-mentioned Data Visualization tools on the basis of various factors such as Free Version availability, Sharing Dashboards, On-boarding/UI feasibility, Customizable Visualization features and charts, Coding, Pricing (Cloud), Pricing (On-Premises), Integration with Other Tools/Scripts, Storage Limit, Data Cleaning, File Upload Size, CustomerSupport and Data Sources.

III. TOOLS ANALYSIS

A. Tableau

Tableau is an effective tool for Data Visualization that is rapidly growing in popularity. With Tableau, data analysis is very fast, and the dashboards and worksheets that are created are very visually appealing. Tableau's data is straightforward to comprehend by experts at all tiers of a business. A variety of visualizations can be created to display data and showcase insights in a magnificent manner.

1) Pros

- Complex table calculations can be performed more rapidly with scripting languages such as R and Python.
- Can handle large amounts of data efficiently.

2) Cons

- Difficulty in importing custom visualizations.
- Reports cannot be embedded easily into other applications.

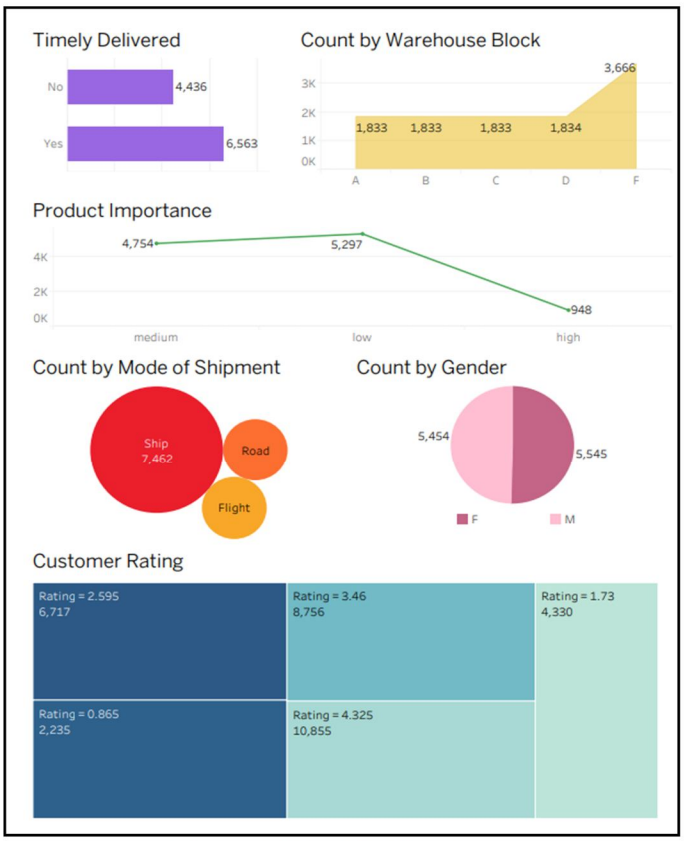


Fig. 1. Tableau Dashboard

B. Power BI

This analytic and visualization tool by Microsoft lets you analyze and visualize data, extract insights, and share insights within your organization. This tool can handle data from a wide variety of sources. Furthermore, it offers the ability to generate adhoc reports which help in the analysis of data.

1) Pros

- Reports and dashboards pre-built for SaaS solutions.
- Data can be scheduled for refresh, without manual intervention after publishing to Power BI web service.

2) Cons

- Users with the same email domain can only use the dashboard and report sharing feature.
- The file size of a Power BI report cannot exceed 1GB.

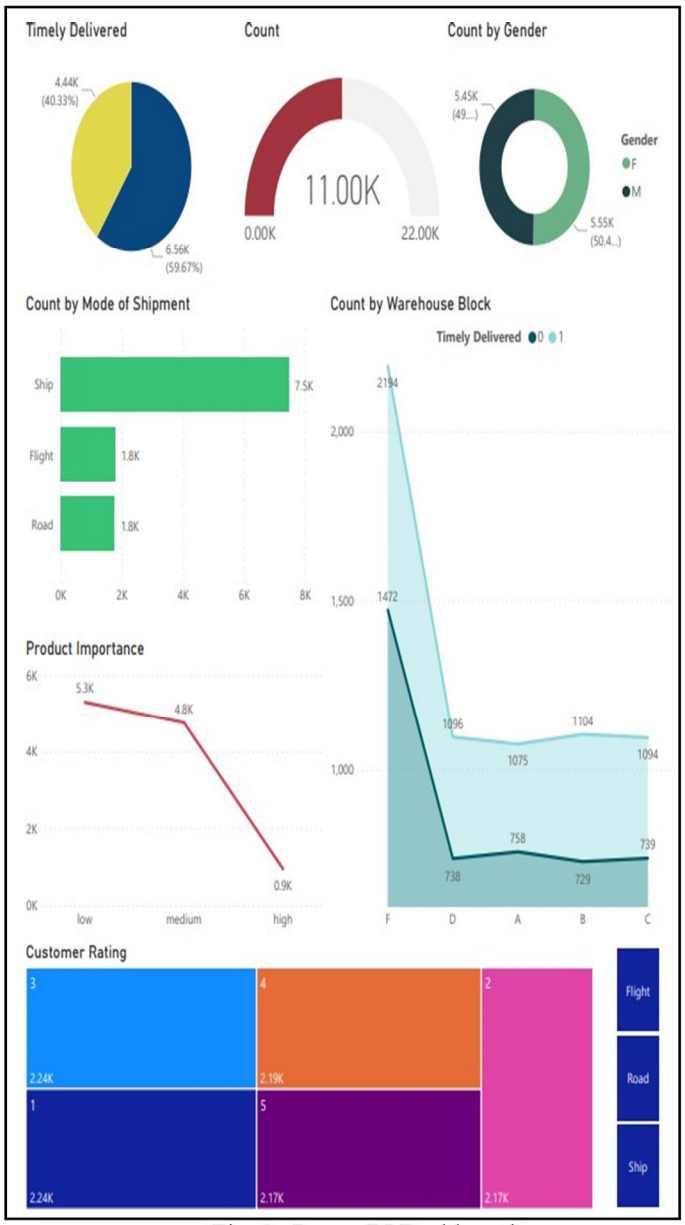


Fig. 2. Power BI Dashboard

C. Zoho Analytic

Zoho Analytic is a business intelligence (BI) platform thatwas launched in 2009. It provides state-of-the-art analytical solutions that converts uncooked facts into actionable insights.The platform allows users to retrieve data from any data source and visually analyze it to make informed data-driven decisions. Users can also share insights and collaborate.

1) Pros

- Easy to collaborate and share insights. (Report sharing and collaboration).
- ZIA (AI Assistant) provides ready-to-use embeddedBI solutions.

2) Cons

- API connections to non-Zoho sources are not reliable.
- Outdated dashboard and charts design with atimestamp.

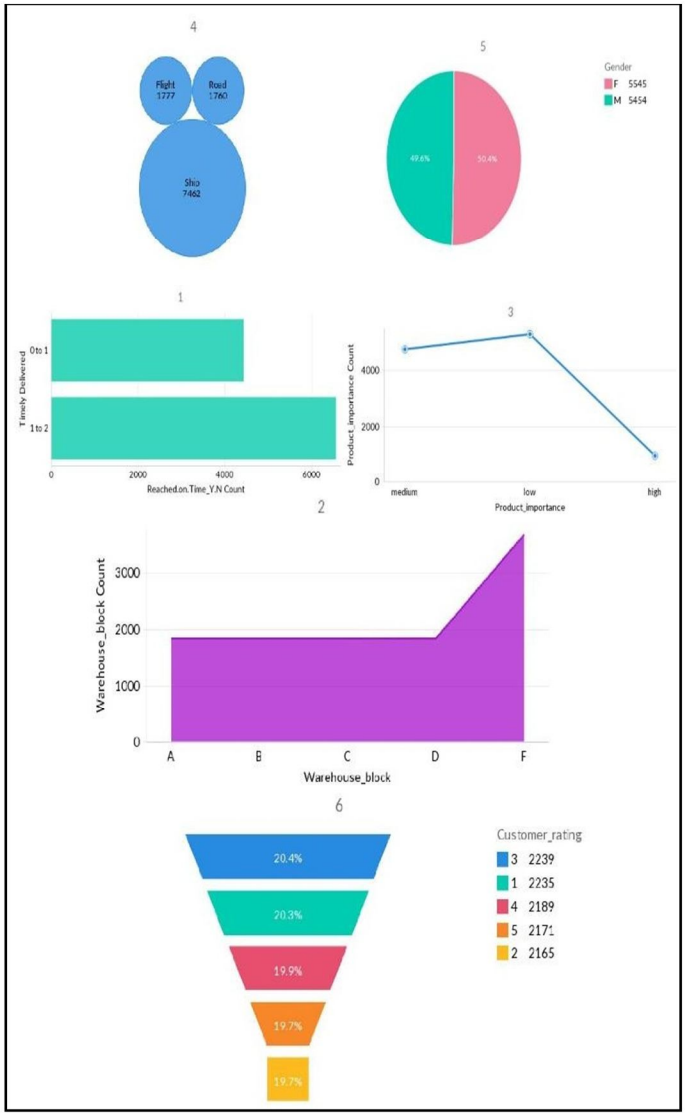


Fig. 3. Zoho Analytic Dashboard

D. *Dataiku DSS*

Dataiku DSS is data science software specifically designedfor engineers, analysts and scientists to effectively generate, explore and deploy data products. It is a systematic data science and analysis tool that includes a visual user interface and inbuilt coding. It promotes the use of popular notebooks like Spark, Hive, Python, and R. From wrangling - to analysis to modeling, drag-and-drop can be utilized anywhere within the data-flow prototype process.

1) *Pros*

- Ability to build models in Python or R.
- Very intuitive and easy to use UI.

2) *Cons*

- Lags while working with a large amount of data.
- Lacks proper documentation.

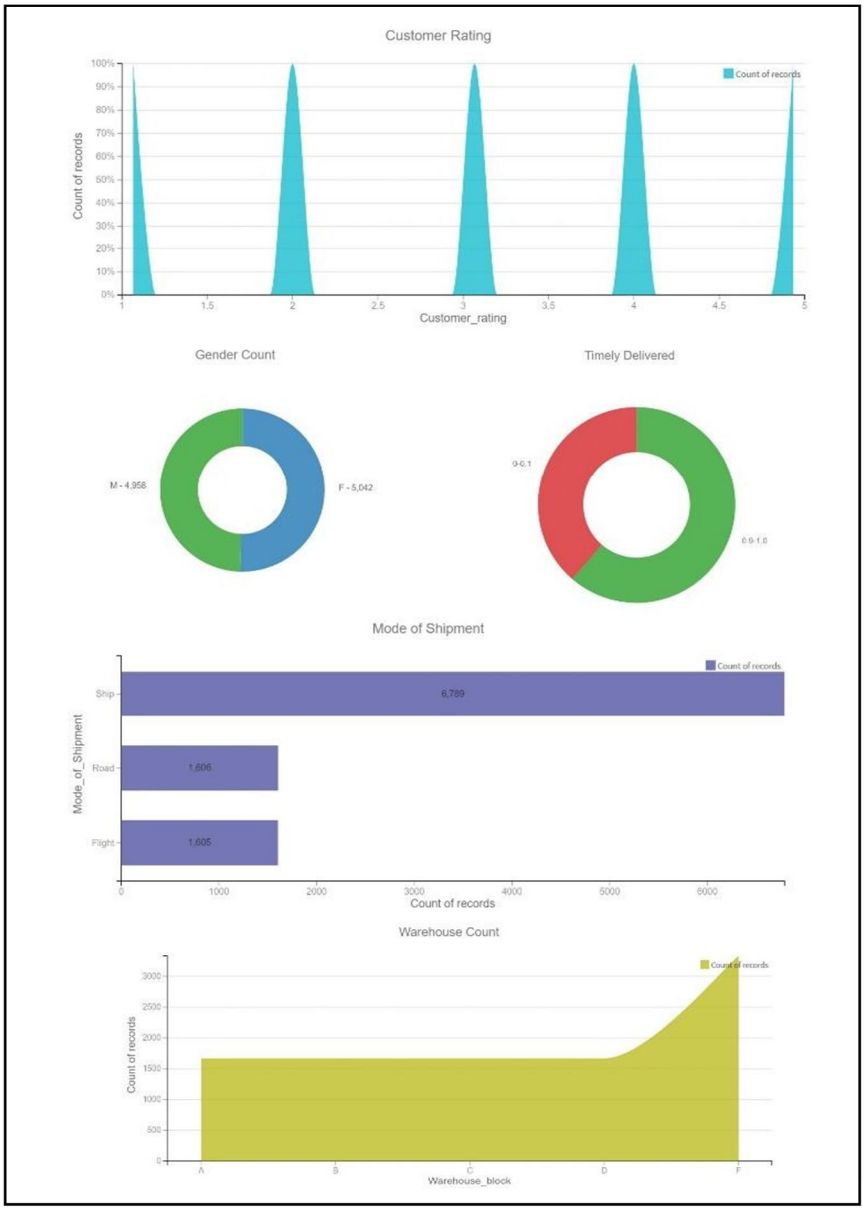


Fig. 4. Dataiku DSS Dashboard

E. Celonis

Celonis Execution Management Platform enable businesses to execute everything from data ingestion to process and task mining to planning and simulation to visual and daily management. It has launched an Execution Management System to eliminate system complexityaltogether.

1) Pros

- Automatization of known issues (Action Flows / Action Engine).
- Level of granularity of the information.

2) Cons

- Complexity of PQL programming and inadequate support for it.
- Using date picker in Celonis is a big hurdle in itself(It wipes out one date on alternating the other)

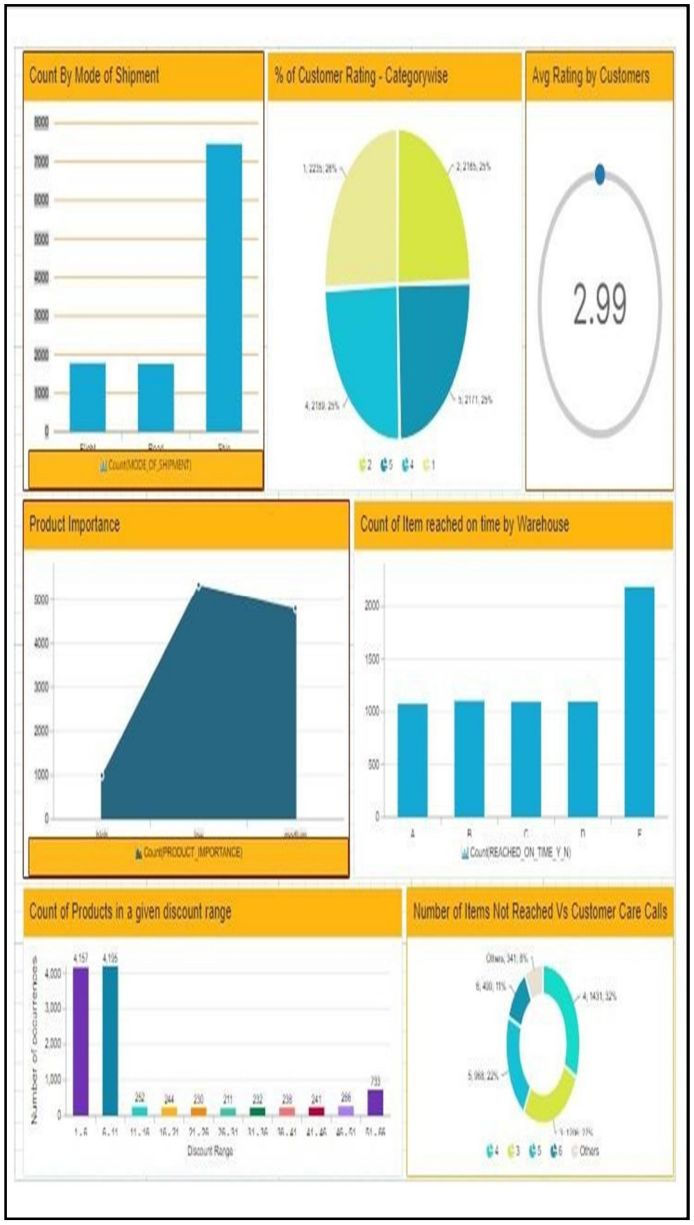


Fig. 5. Celonis Dashboard

F. MS Excel

For organizing, filtering, and displaying humongous amount of data, MS Excel is widely utilized. It is most frequently used by accountants or professionals to manage large, heavy datasets, but it could be used by almost everyone. Balance sheets, budgeting, editorial calendars, and data calculators are examples of such applications.

1) Pros

- It can work with SQL exports.
- The data is easier to segment.

2) Cons

- It can be a time-consuming venture.
- There are no automatic updates. the user's linguistic skills.

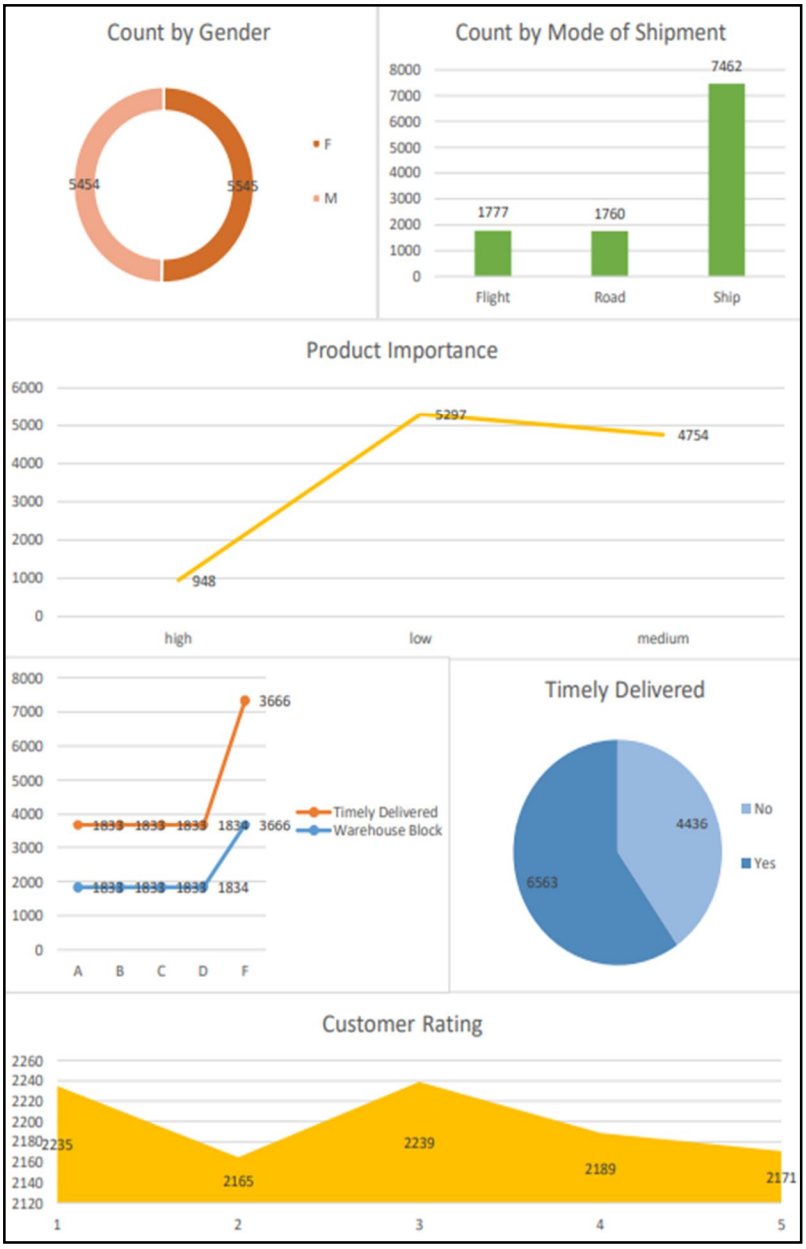


Fig. 6. MS Excel Dashboard

IV. MERITS OF DIGITAL DATA VISUALIZATION

1) Information presented graphically is simple to comprehend.
2) It enables information to be communicated regardless of
3) Large amounts of data can be seen at one go thereby reducing the time taken in reading the information in text or table formats otherwise.
4) This type of data aids in the development of a company decisions.
5) It provides a rapid overview of complicated data that would have taken long time span to comprehend.
6) The user does not rely on others for assistance as the photos are self-explanatory.
7) When it comes to translating, the user may not need any special expertise for interpreting the visualized data.
8) It conveys a more accurate picture of the data than words alone.
9) The user's interaction with the image becomes more user-friendly.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Table I. Tableau vs Power BI vs ZOHO Analytic

Features	Data Visualization Tools		
	Tableau	*Power BI*	*Zoho Analytic*
Free Version	Yes	Yes	No
Dashboard Sharing	Yes	Yes	Yes
On-boarding/UI	Easy	Easy	Easy
Customizable Visualization Features	Yes	Yes	Limited
Coding	Both	Yes	No
Pricing(Cloud)	Starts at $70 for 1 creator per month	Starts at $20 for 2 users/month	Starts at $30 for 2 users per month
Pricing(On-premises)	Starts at $150 for 5 users per month	Part of Power BI Premium plan costs $4,995/month	Starts at $150 for 5 users per month
Integration with other tools/scripts	Integrates greatly with the R language, C, C++, Java, and Python - can be used to implement the Tableau Software Development Kit.	Models data using Data Analysis Expression and the M language. It also supports R, but it is only available to enterprise-level users	It allows integration with other Zoho Tools and APIs only.
Storage Limit	100 GB For Base Version	10 GB For Base Version	10GB For Free Version
Data Cleaning	It can be achieved by checking "Use Data Interpreter" on the left-hand pane of sheets section	Can be done using the Transform Data tab in queries window.	On the statistics tab we can complete the data cleaning process.
File Upload Size	Supports file uploading of size upto 10 GB	Supports upload upto 1GB file of various type	100 MB only in the free plan or having less than 100000 columns
Customer Support	Yes	Limited	Yes. It is available 24*7 for paid subscriptions only
Data Sources	Excel, Text File, PDF, JSON, statistical file, Amazon Redshift, Cloudera Hadoop, Google Analytics, dropbox, google sheets, google drive etc.	Microsoft Excel, Text/CSV, Folders, MS SQL Server, Access DB, Oracle Database, IBM DB2, MySQL database, PostgreSQL database, etc.	Excel, AWS, SQL, Salesforce, Zoho CRM, Xero, Stripe, Google Ads, Mail Chimp, Survey Monkey, YouTube, Jira, Bigin, Alchemy, LinkedIn Pages, Zen desk, Microsoft azure, Panoply, MongoDB

Table II. Dataiku DSS vs Celonis vs MS EXCEL

Features	Data Visualization Tools		
	Dataiku DSS	*Celonis*	*MS Excel*
Free Version	Yes	Yes	Yes
Dashboards Sharing	Yes	Yes	Yes
On-boarding/ UI	Complex	Easy	Complex
Customizable Visualization Features	Yes	Yes	Yes
Coding	Both	Yes	Both
Pricing(Cloud)	Starts at $499 for 10 users per month	Starts from $0.01/ Per Feature	Free
Pricing(On-premises)	Starts at $ $ 2499 for per 30 users per month	Starts from $0.01/ Per Feature	Free
Integration with other tools/scripts	It integrates with the popular notebooks like Spark, Hive, Python, and R	It has great support for the python language. We can use the machine learning workbench to automate the processes and it also helps in getting more insights	It supports python SQL exports
Storage Limit	No Storage limits	No Storage limits	2 GB for office suite
Data Cleaning	High lights the uncleaned data but have to be cleaned manually	Data cleaning can be done with transformation step or with the help of pycelonis python package	Data Cleaning can be performed manually. We can't hardcode it
File Upload Size	Supports a Single file of size upto 1GB regardless of its type	1 GB but bigger files can be splitted and pushed with the help of Data Push API	An excel sheet of size upto 2GB can be used to work on
Customer Support	Limited. Chat Bot and email service is available	Yes, it is available 24*7 at an additional fee	Only Documentation and 'Microsoft support' website is available
Data Sources	Excel, CSV, Snowflake, Amazon Redshift, Azure Synapse, Google Big Query , PostgreSQL, MySQL, Amazon S3	Microsoft Excel, Text/CSV, Folders, MS SQL Server, Access DB, Oracle Database, IBM DB2, MySQL database, PostgreSQL database, etc.	Workbook, Text/CSV, JSON, XML, Sharepoint, SQL Server, SAP Hana Database, Azure SQL Database, Online Services, etc

V. RESULTS

Each dashboard is generated and compared to how each software package is perceived by the user in terms of dashboard development. The ultimate outcomes are listed below.

A. Tableau

Tableau lacks features that a full-fledged business intelligence tool would have, such as large-scale reporting, datatable creation, and static layouts. The solution's potential to share outcomes is also limited.
But at the same time, it can handle complex calculations with the help of the scripting languages, like R and Python.

B. Power BI

Power BI desktop application is not available for MAC users. They can work only on the cloud environment which requires a strong internet connection otherwise it becomes quite slow to visualize the data.On the contrary, it has 'summarize' function through which we can get summary of any visual in Power BI.

C. Zoho Analytic

While visualizing the shipment dataset we discovered that there are insufficient charts and graph templates in the zoho analytics tool. As we wanted to create a heatmap for customer ratings, but owing to a lack of resources(graphs/charts), we were forced to settle for a funnel graph.
On the other hand, it provides a well versed AI Assistant (ZIA), which provides ready-to-use embedded BI solutions andreports.

D. Dataiku DSS

It doesn't have the feature to label the graphs and charts while visualizing the dataset, because the free version restricts it to do so.
However, it has the feature of monitoring and controlling the version. As a result, if something goes wrong, the user can easily roll back to a previous version.

E. Celonis

We utilized PQL (Product Query Language) to create the KPI for the number of things that arrived on time that were dispatched from multiple warehouses. Since there are few resources for PQL on the web, it was challenging to build code for it under particular limits and conditions.
Whereas, with the aid of powerful process mining capabilities, user can analyze the whole process flow with the help of process explorer, case explorer, user analysis tracker, conformance tracker, etc.

F. Microsoft Excel

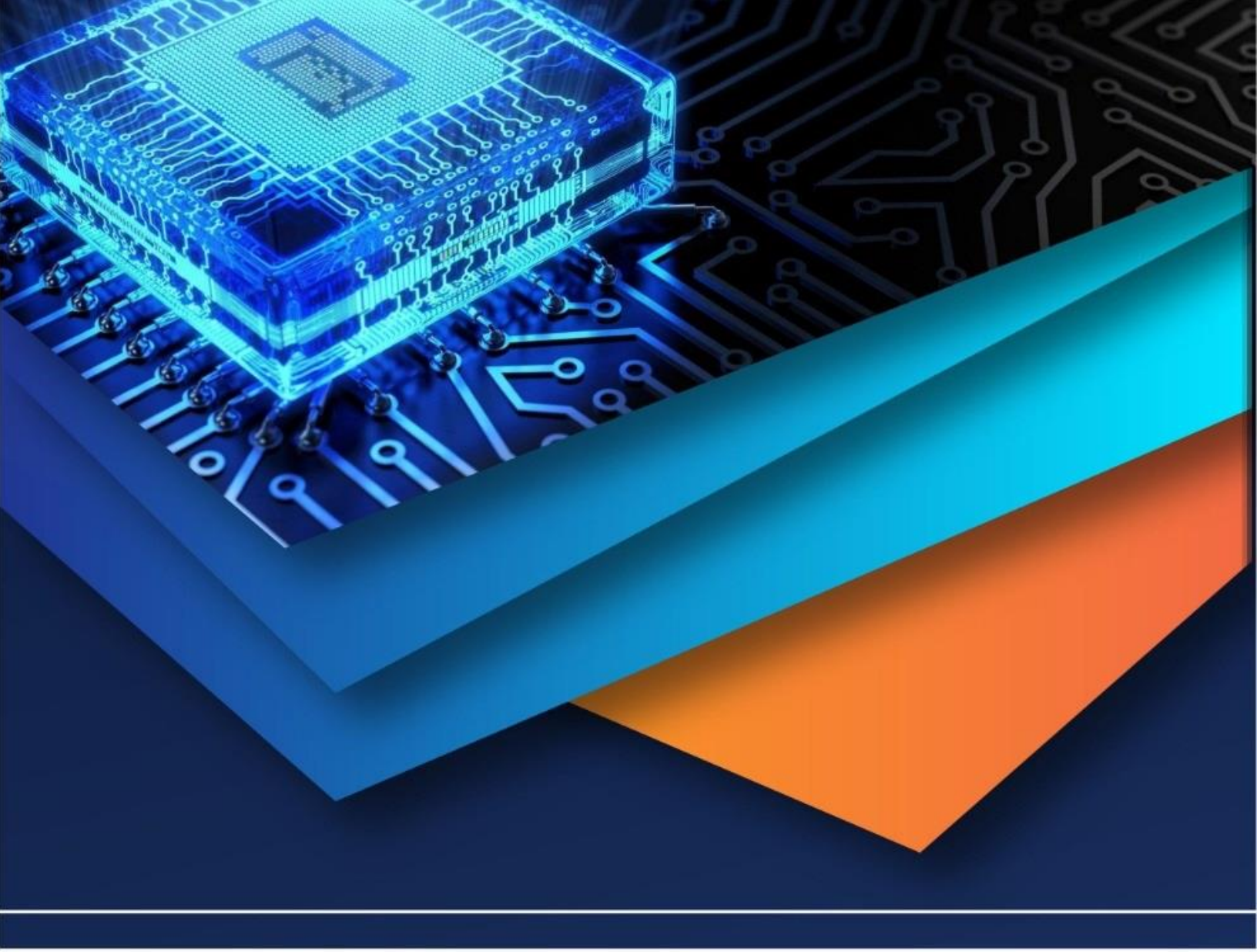
Along with 0 and 1, an extra row labelled 'Blank' was introduced while visualizing a timely delivered pie chart. This 'Blank' was redundant for the chart. By filtering the blank label from the field button, the 'Blank' had to be manually removed.
Although, it has Pivot Charts, that are used to visually summarize and investigate complex data. It also provides interactive filtering controls directly on the chart, allowing theuser to easily study a portion of the data.

VI. CONCLUSION

In the Big Data field, visualization is a crucial aspect. Thispaper is focused on various Data Visualization applications, which are utilized in several fields like - Big Data Analytics, commercial, enterprise, websites, mobile platforms, etc., and deliberately examines the tools. As we are able to see, it's pretty tough to choose from Power BI, Tableau, Zoho Analytic,Celonis, MS Excel, and Dataiku DSS, but we believe we have done so. However, one has to look at their requirements to determine which one to go along with. For instance, a small and medium commercial enterprise (SMB) employer would possibly be afflicted by licensing price from Tableau regardless of its exceptional features. Power BI and Dataiku DSS are extremely good for hybrid surroundings with computers restrained to customers. Celonis provides larger options for integrations from different tools and scripts. For automated anddetailed analysis Zoho Analytic is quite good, considering it has limited sets of chart representations. MS Excel is good for small datasets which are easy to handle manually. Last but notthe least, Power BI has great user-friendly UI in the Data Visualization business. The software's drag-and-drop functionality, combined with access to over 85 data visualization programs, provides an unrivalled user experiencethat results in an aesthetically beautiful and useful report.

REFERENCES

[1] Michael Diamond and Angela Mattia, Jacksonville University, "Data Visualization: an exploratory study into the software tools used by business", Journal of Instructional Pedagogies, vol. 17 pp. 1-3,November 2015.
[2] Mohammed Kemal, University of Liverpool, "Data Visualization Tools in Action Choosing a Visualization Software", Research-gate., pp. 2, April 2019.
[3] Syed M., Noopur G., and Rakesh K. IIIT Bhubaneshwar. "Big Data Vizualization: Tools and Challenges". pp.709-711, December 2016.
[4] Rajeswari C, Dyuti Basu and Namita Maurya, VIT University "Comparative Study of Big data Analytics Tools: R and Tableau", IOP Conference Series: Materials Science and Engineering, November 2017.
[5] Aman Velpula, Anirudh Suhas Venkatayogi and Nitesh Kumar Yadav, "Compare and Contrast Cloud based CRM software: Salesforce.com Vs Zoho crm", February 2015.
[6] Matthew N. O. Sadiku1, Adebowale E. Shadare, Sarhan M. Musa and Cajetan M. Akujuobi, Prairie View A&M University, "Data Visualization", Volume. 02, December 2016.
[7] Soly Mathew Biju and Alex Mathew, "Comparative Analysis Of Selected Big Data Analytics Tools", Journal of International Technology and Information Management, Volume 26, December 2017.
[8] Arockia Panimalar.S, Komal M.Khule, Karthika.S, Nirmala Kumari.T, "Data Visualization Tools and Techniques For Datasets In Big Data",Sri Krishna Arts and Science College, International Research Journal of Engineering and Technology (IRJET), Volume: 04, August 2017.
[9] U. S. Junghare, Shri Shivaji Science college, "Analytical Evaluation towards Big Data Visualization Applications, Tools and It's Issues", The International journal of analytical and experimental modal analysis, September 2021.
[10] Leena Gokhale and Kirti Nilesh Mahajan,Bharati Vidyapeeth Deemed University, "Comparative Study of Data Visualization Tools" Research gate., pp. 48-49, May 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY