

Széchenyi István University
Faculty of Mechanical Engineering,
Informatics and Electrical Engineering
Department of Informatics

MASTER THESIS

Syed Rafi Ahmed
MSc in Business Informatics

2024-2025



MSc Business Informatics

Faculty of Mechanical, Electrical Engineering and Informatics

Comparative Analysis of Business Analytics Tools

Master Thesis

Consultant

Dr. Erdős Ferenc (Associate Professor)

Prepared by

Ahmed Syed Rafi (FNAT2H)

Declaration

I, **Syed Rafi Ahmed (FNAT2H)**, MSc Business Informatics' student, declare that the thesis entitled **Comparative Analysis of Business Analytics Tools** is my own work, and that I have used only the indicated sources and to the extent indicated, in accordance with the rules of citation, with the exact indication of the references.

My results are based on my own work, calculations, research, real measurements, and are credible to the best of my knowledge.



Győr,

Syed Rafi Ahmed

DATA SHEET FOR AUTHORIZATION OF THE DIPLOMA THESIS

Student information:

Name: Ahmed Syed Rafi
Address: K1/812, Egyetem Ter 1
Phone: +36205612646
Programme: Business Informatics
Degree: MSc
Training form: full-time

Neptun ID: FNAT2H

e-mail: syedrafi.cse31@gmail.com

Diploma thesis data:

Beginning academic year and semester: 2023/2024 autumn
Expected submission: 2023/2024 spring
Topic: Comparative Analysis of Financial Predictive Analytics Tools
Language: English
Type: public

Short description:

Data analytics is the most important activity of decision support systems. The main target of data analytics is to extract useful values, provide suggestions and support decision-making. The thesis aims to provide a comparative analysis of today's available predictive analytics software packages in the field of finance.

Subtasks

- Literature overview of decision support systems.
- History of financial analytics software.
- Market overview of financial analytics software vendors.
- Comparison of different tools using the same business scenario and dataset.
 - dataset selection
 - data cleaning and preprocessing
 - predictive modeling (clustering, decision tree, ARIMA/ SARIMA)
 - Result comparison of the different analysis models.
- Conclusions.

Internal supervisor:

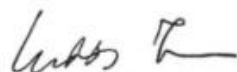
Name: Dr. Ferenc Erdős
Department: Department of Informatics
Position: associate professor

External supervisor:

Name: Dr. Buics Laszlo
Workplace: Szechenyi Istvan University
Job title: assistant professor
Address:
Phone: +36701234567 format

e-mail: buics.laszlo@ga.sze.hu

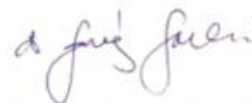
Győr, 2023. június 28.



Internal supervisor




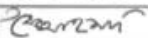

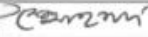




External supervisor



Head of department

Annex 7


THESIS CONSULTATION FORM

Date	Topics Covered	Signature
04/08/2023	Clarifying the thesis goal	
06/09/2023	Thesis structure and further tasks	
01/11/2023	Structure of the thesis and literature review	
07/05/2024	Literature review, referencing and further tasks	
26/03/2025	Assesment of the Methodology	
03/05/2025	Data Analysis and interpretation	
10/05/2025	Framework validation, Conclusion and Recommendations	
13/05/2025	Final corrections	

I have revised the thesis. It can be submitted / cannot be submitted.

Date: 2025.05.15. 
external supervisor

The thesis can be / cannot be accepted for evaluation:

Date: 2025.05.15

internal supervisor

External examiners	name:	
	position:	
	Recommended mark:	Signature

Abstract

We are living in a data-driven world where selecting the most appropriate business analytics software is crucial for a user. With a wide range of commercial and open-source BI tools available on the market, a user finds it difficult to position himself to align with technical aspects, business needs, and decision-making capabilities. To address this issue, my thesis investigates those challenges by developing a structured methodology that consists of testing and decision-making pipeline to evaluate three widely used analytics tools: Kibana, JMP, and Power BI. The methodology includes a sales dataset, which is used to evaluate tools, focusing on Ease of use, interactivity, customizability, cost and Third-party plugin. The final result shows that Kibana offers scores higher in evaluating criteria in comparison to other tools. While Kibana is the best choice for users in terms of flexibility, real time analytics, and cost effectiveness, JMP is still the best choice only for statistical purposes with advanced built-in statistical features. In a nutshell, my thesis offers a pipeline based on a practical decision-making framework for new users to choose the correct tools from the market available.

Keywords: Business Intelligence, Data Analytics, Visualization Tools, Kibana, JMP, Power BI, Competitor Profile Matrix, Data Integration,

Table of Contents

1.	Introduction.....	4
1.1	Objectives (End goal)	4
1.2	Structure of thesis	5
1.3	Problem Statement and Research Questions	5
1.4	Investigated Software by interview	6
2.	Literature Review.....	8
2.1	Difference between tools	8
2.2	Significance of Data analysis for business	9
2.3	Data visualization overview	10
2.4	Related Work	10
3.	Business Analytics Software Packages.....	14
3.1	JMP statistical software Overview	14
3.2	Kibana.....	16
3.2.1	Working process of Kibana	17
3.3	Power BI	18
4.	Methodology	20
	Table 2: Formula of Competitor Profile Matrix for scoring.	28
5.	Results and Analysis.....	29
5.1	Clustering analysis for segmentation.....	29
5.2	Sales trend analysis.....	34
5.3	Growth Rate analysis.....	37
5.4	Result evaluation	43
6.	Conclusion.....	47
6.1	Limitations and Recommendations	47
6.2	Future works	48
	Bibliography.....	49
	Appendix.....	53

List of Tables

Table 1: Different types of Analytics tools (Somani & Deka, 2017).....	09
Table 2: Formula of Competitor Profile Matrix for scoring.....	28
Table 3: Competitor profile matrix with experimental score.....	45

List of Figures

Figure 1. Extend of (DeLone & McLean, 2016) model used by (Al-Okaily et al., 2023)..	11
Figure 2. Graphs created in Graph Builder (JMP 18.1 Help, n.d.)	15
Figure 3. A workflow diagram of Kibana.	17
Figure 4. A basic workflow of how Power BI works (Machiraju & Gaurav, 2018).....	18
Figure 5. A workflow process for testing and evaluating Business analytics software	20
Figure 6. An overview of survey responses for criteria weighing.	22
Figure 7. Pie chart of demographic overview of participants.	23
Figure 8. Overview of Post processed dataset.	29
Figure 9. Clustering by deal size for segmentation.	30
Figure 10. Dataset Preprocessing Interface. CSV file loaded with 2823 documents ingested in 5 steps.	31
Figure 11. Field statistics of loaded dataset in Data view interface.	32
Figure 12. Clustering by deal size for segmentation.	33
Figure 13. Clustering by deal size for segmentation.	34
Figure 14. Sales Trend over MONTH_ID	35
Figure 15. Sales Trend over MONTH_ID.	36
Figure 16. Sales Trend over MONTH_ID.	37
Figure 17. Growth Rate% over YEAR_ID.	38
Figure 18. Summary of preprocessed data in JMP.	39
Figure 19. Formula Box in JMP.	40
Figure 20. Growth Rate% over YEAR_ID.	41
Figure 21. Growth Rate% over YEAR_ID.	42
Figure 22. CPM table visualization as a Radar Chart.	45

1. Introduction

Today's world is fueled by data. Every day a large portion of business data, including both quantitative and qualitative resources, is generated by different business sectors. Financial data is crucial for management and governance in the business sector. (Sun et al., 2019) For decision-makers, data analytics is crucial. Such analysis facilitates better decision-making, raises responsibility, advances financial stability, and tracks the effectiveness of the organization (Intellify Solutions, n.d.). According to a survey conducted by Gartner, 87% of corporations have low BI (enterprise intelligence) and analytics maturity, primarily at basic or opportunistic levels, which creates obstacles to their ability to fully understand data assets and modern analytics systems. In order to enhance BI maturity, Gartner suggests four steps: developing holistic strategies, creating flexible organizational structures, implementing data governance, and adopting integrated analytics platforms (Gartner, 2018). Due to characteristics of business data of different financial domains, there are several challenges to analyzing business data. It is perplexing and challenging for a new user to find suitable software to work on the specific challenges and dataset type they have. This is because each software which is available on the market is diverse in features, usability, functionality and user interface. This often leads a new user to try out multiple software while learning the basics of each. This often leads to spending so much time and money, which effect the timeline.

1.1 Objectives (End goal)

The main objective of this thesis is to study and analyze different data analytics tools. My work will allow the user to find suitable software which is compatible to their dataset type and allow to complete their task effectively. This approach will reduce deployment time of a project significantly when a user already has a dataset and knows what type of analytical task is required. In order to compare software packages in the field of data analytics I need to focus on specific features and capabilities. Here are some key aspects to focus on during comparison: 1. Dataset preprocessing: (Tabular Data) Numerical data, Categorical data, Time series data.

2. Data Visualization Capabilities: Compare intuitiveness and ease of use the data visualization for given software.

3. Data Integration and Connectivity: Evaluate how each software import data from different sources. Comparing their features and types of data this software can import.

4. Collaboration and Sharing: Evaluate the collaboration features of each software. Consider how easily you can share reports and collaborate with team members. The goal is to facilitate effective collaboration within your team.

5. Scalability: Determine whether the software can manage large size datasets and preprocess time. Find the best one in comparison process.

1.2 Structure of thesis

This thesis is written into six chapters where each chapter focuses on systematic exploration towards comparative analysis of Business Informatics Tools. The chapters are structured as given below:

Chapter 1 contains problem statements, formulation of research questions, objectives and a short interview section to explain the rationale for selected software.

Chapter 2 reviews existing literature on business analytics tools, research gaps, and the difference between business analytics tools.

Chapter 3 contains discussion on Business Analytics software packages including JMP, Kibana and Power BI.

Chapter 4 provides research methodology for selecting the best business analytics tools, dataset selection, experimental setup and evaluation criteria.

Chapter 5 presents how analysis has been done and the result with visual evidence and detailed scoring outcomes.

Chapter 6 contains key findings, the limitations of study and future work.

1.3 Problem Statement and Research Questions

There are several challenges from both technical and administrative point of view during analyzing business data. One most prominent issue is defining business goals in terms of

organizational needs. Selection of right tool according to end user will hinder accurate analysis which may no align with the final business goals. (Abdul Rahman et al., 2016) From a technical standpoint, problems such as third-party plugin integration capabilities, difficulty of handling preprocessing dataset and scalability effect tool's usability.

Inadequate utilization of business analytics information can lead to suboptimal decision-making. Appropriate datasets are an important factor in the realm of analytics. Without the appropriate datasets, analytics outputs may provide irrelevant information to decision-makers which leads to failure to adapt to modernized inventory management systems. (Liu et al., n.d.)

To address these problems, this thesis aims to conduct a controlled analytical environment using a standardized dataset to investigate selected software based on interviews. The following research questions have been formulated to make the comparison:

RQ1: Which business analytics tool offers the best usability?

RQ2: Which business analytics tool offers the best visualization features?

1.4 Investigated Software by interview

I conducted a short discussion with three different people who are engaged in different industries. These short discussions helped me to identify figured out most widely used software in the industry up to date. To conclude Kibana, Power BI and JMP are the most common and widely used software in the industry for task.

Interview 1: In a detailed conversation with M. Asif Hasan, who is a research assistant, University of Freiburg. When asked about business intelligence tools, he effectively expressed the importance of open-source software. He recommended that I explore Kibana from the Elastic Stack, supported by accessible online data and practical applications. Kibana is a robust, open-source analytics and visualization platform that, from his perspective, possesses significant potential in the domain of business data analysis. In considering cost-effectiveness, he prioritized the open-source aspect of Kibana.

Apart from that, he underscores other commercial business analytics tools due to subscription model and scalability issues. He asserts that Kibana's open-source nature fosters a

substantial and engaged community, complete documentation, a plethora of tutorials, community forums, and other online resources that facilitate ongoing learning and professional advancement. He suggested running Kibana on Linux based platforms as he found difficulties running Kibana on Windows.

Interview 2: In an online meeting with Mrs. Ruman Khanchi, Procurement Analyst at Nokia, who has experienced supply chain, a well informative interview was conducted. Her preferred data analytics tool is Power BI. She emphasized its intuitiveness and user-friendly interface. With exceptional ease of use, she put Power BI her favorite data analytics tool for her daily use. This tool helps her to track KPI, and effective data-driven decision making. She mentioned that Power BI offers various licensing options which is helpful for personalized projects as well.

Interview 3: In an interview with Mr. Jewel Kumar Roy, a knowledgeable person from Szechenyi Istvan University. He is a dedicated researcher and academic specializing in Financial Technology (Fintech), machine learning, and artificial intelligence in finance. When I asked about data analytics software that is preferable to him, he suggested me to try JMP. He mentioned that JMP has advanced statistical analysis capabilities and predictive modeling capabilities. User friendly interface is a crucial factor to him. According to him, JMP can understand the nature of different algorithms and preprocess data accordingly while ensuring accuracy and reliability in analytical outcomes.

2. Literature Review

The literature review of my thesis contains several subchapters. In this part, I showcase the difference between tools in the first part, importance of data analysis for business, improving client experience and decision making in the second part. Thereafter, I focus on data visualization in general. The last subchapter contains related work regarding my thesis and a brief overview of each work.

2.1 Difference between tools

Data analysis is the methodical use of logical and statistical techniques to describe the scope of the data, modularize the data structure, condense the data representation, illustrate using tables, graphs, and images, assess statistical inclinations, probability data, and draw meaningful conclusions. These analytical techniques require various types of tools to meet different criteria, including cloud based or local deployments, AI integration, automation capabilities. Based on deployment there are two possibilities:

1. Cloud-based Tools: Google BigQuery, Amazon Redshift, Microsoft Power BI (online version), Tableau Online.

2. Local/ On-premises Tools: Excel, Apache Hadoop, Tableau Desktop, SAS.

Based on AI integration there are several tools for example: IBM Watson Analytics, Salesforce Einstein Analytics.

Based on Automation Capabilities there are some tools including Alteryx, Apache Nifi, Microsoft Power Automate.

The below mentioned table provides a brief discussion on several tools including IBM Infosphere, IBM SPSS, Apache Mahout, Azure Machine Learning Studio, Halo, Tableau, SAP InfiniteInsight, @Risk, Oracle Advanced Analytics, TIBCO SpotFire, R, Mathematica from different analytical understandings. (Somani & Deka, 2017)

Analytics Tool	Descriptive	Diagnostic	Predictive	Perspective
IBM Infosphere	✓			
IBM SPSS	✓	✓	✓	
Apache Mahout			✓	
Azure Machine Learning Studio	✓	✓	✓	
Halo	✓	✓	✓	✓
Tableau	✓			
SAP InfiniteInsight			✓	
@Risk				✓
Oracle Advanced Analytics	✓	✓	✓	
TIBCO SpotFire			✓	✓
Mathematica	✓		✓	

Table 1: Different types of Analytics tools (Somani & Deka, 2017)

2.2 Significance of Data analysis for business

By using data analysis, your business can scale up in several ways. Here is some example of:

Developing more effective promotional plans:

Based on gathered data, it is easy for a business to develop creative and unique marketing strategies. By analyzing trend and finding out customer expectations, it is easier for business to develop policies allined with customer expectations. (Fahl, 2017)

Sorting out pain points:

Data can help to detect disruptions in business that is driven by a set of processes and patterns. The abrupt rise in customers complaints, the decline in sales, or the drop in productivity might all be attributed to these little variances. With the help of data analysis, it will be possible for anyone to identify these disruptions and take necessary actions. (Fahl, 2017)

Improving client experience:

Feedback data provided by clients are important for businesses owner. As mentioned earlier, based on pain points, owners can modify products or service in a improved style. Some companies, for example, offer their clients personalized private emails. This communicates the firm's sincere concern for its customers and desire to meet their needs. Effective data management is the only thing that makes this possible. (Fahl, 2017)

Decision making:

Making critical business decisions requires data. For example, before launching a novel product in the market, gathering information on the size of the consumer base, the rival's prices, and the current market trends are necessary. When a firm makes decisions without considering facts, it will cost the firm a lot. In terms of decisions about the function of departments, managing personnel, etc. can also benefit from the use of data. For instance, data can help to determine the number of employees needed to run a division in accordance with business requirements. (Fahl, 2017)

2.3 Data visualization overview

Data visualization is the way of presenting data using familiar visual elements, such as charts, graphs, infographics, and even animations. This considers critical step in the data science process and decision makers (IBM, n.d.). One of the easiest way to visualize data is creating a dashboard from multiple sources. By definition a dashboard is a way of showcasing different types of visual data in a particular window (Tableau, n.d.). Decision makers from different business domain can set policy, detect KPI (Key performance indicator) by visualizing data from dashboards. Dashboards have some traditional visualization techniques including Tables, Pie charts and stacked bar charts, line charts and area charts, histograms, scatter plots, heat maps, tree maps (IBM, n.d.).

Two problems appear while trying to visualize data:

1. How much information the human eye can handle.
2. The types of tools used to show data. (Paczkowski, 2021)

2.4 Related Work

This paper (Al-Okaily et al., 2023) addressed effectiveness of BI implementation in Jordanian companies by assessing the organizational effectiveness of data analytics-oriented BI technologies. This research has been conducted in the DeLone and McLean IS success Model (DMM) (DeLone & McLean, 2016), which includes analyzing system quality, information quality, data quality, training quality, user satisfaction to measure effectiveness of BI technology. To address this limitation, in my thesis I filled the gap by focusing on tool-specific features and creating test environment for data analysis simulation.

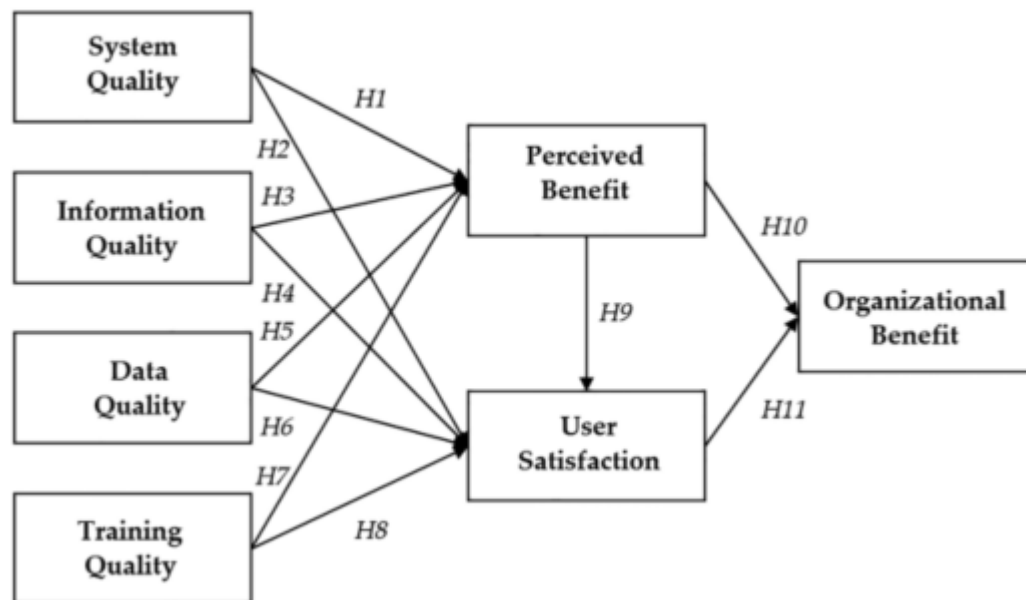


Figure 1. Extend of (DeLone & McLean, 2016) model used by (Al-Okaily et al., 2023)

The survey paper by (Khalajzadeh et al., 2022) evaluates qualitative, tool comparison-based methodology focusing on three categories: Functionality (AI-SDLC coverage), Usability, Cognitive Effectiveness. Furthermore, use-case based evaluation has been done by investigating property price prediction (Azure ML) and stroke prediction (AWS Sagemaker). The limitation of this study includes limited domain diversity and dataset, absence of computational process data and focusing on one-time evaluation of investigating software. The main research gap between (Khalajzadeh et al., 2022) and my work is more tool-evaluation oriented while I provide a technical and applied dimension, focusing on hand-on-data transformation and scoring system to finalize best software.

The conference paper by (Sharma et al., 2021) highlights BI, data analytics and data science topics in general while comparing paid and open-source data analytics tools including Power BI,

Google Analytics, Quilkview, ThoughtSpot, R and Python. Lack of quantitative evaluation such as data-driven performance benchmark, proper methodology to investigate comparison of Business Intelligence tools are the main research gap.

(Vijayaraj et al., 2016) covered Hadoop, HPCC, Strom, HBase, GridGain which are categorized as Big Data analytics tools to perform feature-based comparison. This study does not evaluate empirical testing or specific dataset based evaluation to compare between tools while my work narrows down specific BI tools with real world dataset and follows a structured methodology. However, (Vijayaraj et al., 2016) uses holistic approach like Three V's of big data including Volume, Variety, Velocity and Five C's framework including Connection, Conversion, Cyber, Cognition, Configuration for comparison.

(Biju & Mathew, 2017) conducted a study which showcased a quantitative case study using different real datasets to make a comparison among five selected analytical tools including R, Minitab, IBM Watson, SAS and SPSS. During the comparison ten parameters including Size of Data, Type of Data, Data Loading, Prediction Capabilities, User Interface, Additional Tools, Result Presentation, Independent Installation, Exporting Output and Review data have been investigated. For each cases, Prediction capabilities for five software have been labeled as NA which suggests a disconnect between analysis segment and comparison chart.

The author (Vidhya et al., 2014) reviewed big data analytics tools like Pentaho, Jaspersoft, Splunk, Tableau, and Karmasphere. These tools were investigated on different dimensions, including data type support (structured/unstructured/semi-structured), integration with data sources, OS compatibility, ETL and dashboard capabilities, user interface and developer extensibility. This study lacks empirical validation and only focused on demonstrating big data analytics tools' usage and creating alertness among users.

This paper (Rajeswari et al., 2017) compares two popular tools – R and Tableau – under three different datasets. By examining their capabilities, this study aims to highlight the strengths and weaknesses of each tool in data visualization and analysis. Furthermore, insights will be drawn regarding their respective performances, providing valuable guidance for practitioners selecting the most suitable tool for their specific data needs. The author followed a methodology which includes three steps: data collection, analyzing the data using tools and comparing the performance. The histogram of the Blood Transfusion Service Centre dataset, the scatterplot of the Forest Fire dataset and the pie chart of the Crime dataset have been checked for both R and

Tableau. While experimenting with these datasets under a formulated mechanism, the author stated his experience through considering the data visualization capability of both of these software based on ease of use, interactivity, speed, licensing, programming knowledge requirement, graphical library and statistical analytical capabilities.

3. Business Analytics Software Packages

3.1 JMP statistical software Overview

As product of SAS Institute JMP is categorized as a statistical software that activates data exploration and discovery through statistically designed experiments. It provides analytical tools and a graphical user interface (GUI) that prioritizes smooth and natural workflow for data analysis. JMP's formula editor allows users to create new variables as functions of other variables, facilitating data transformation and further investigations. (Jones & Sall, 2011)

JMP Capabilities:

There are several capabilities of JMP software which is given below

1. Advanced statistical Modeling: Statistical modeling refers to the data science process of applying statistical analysis to datasets. A statistical model is a mathematical relationship between one or more random variables and other non-random variables (HEAVY.AI, n.d.).

2. Automation and Scripting: An automation script consists of a launch point, variables with corresponding binding values, and the source code (IBM, n.d.). JMP can automatically save scripts to reproduce any data table or analysis in its current state (JMP, n.d.).

3. JMP internet open command: According to the JMP User Community (2018), the Internet Open Command is a feature in JMP statistical discovery software that allows users to search a website for tables and reformat them as JMP tables. This feature is particularly useful for accessing external data sources and incorporating them into the JMP environment for further analysis.

4. Data Exploration and Visualization: JMP offers wide range of graphing, data summary, tabular summaries , geographic maps, filtering, interactive dashboards and text data processing. JMP uses Graph Builder to create interactive data visualization and share results with other users.

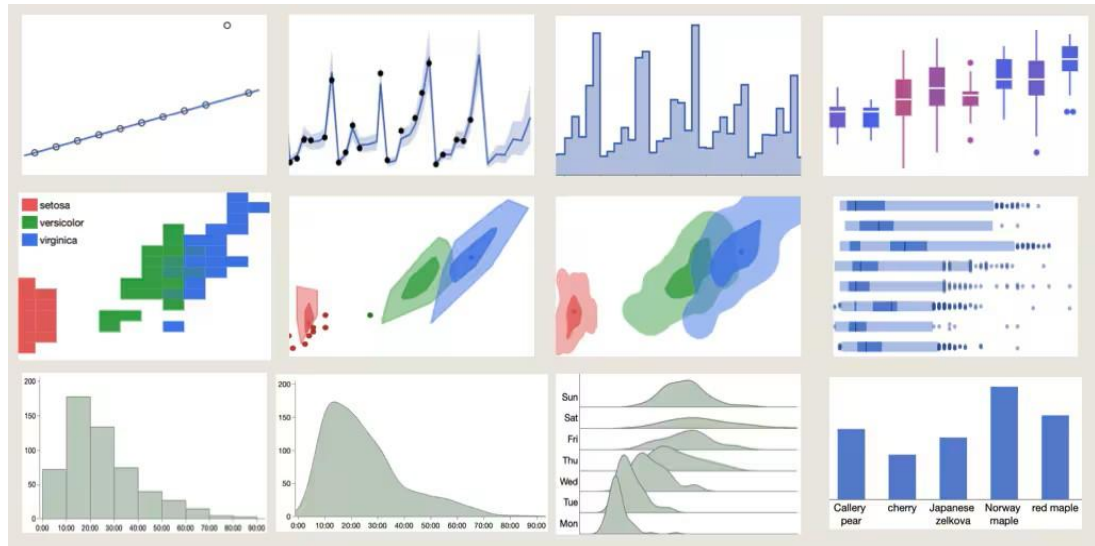


Figure 2. Graphs created in Graph Builder (JMP 18.1 Help, n.d.)

5. Predictive Modeling and Machine Learning: Predictive Modeling is a statistical method that utilizes historical data to forecast the result of future events. The process involves making a mathematical model that takes historical data as variables and predicts output variable based on pertinent input variables. In order to train and enhance these models Machine learning algorithms are used. Thus, by using these models a user can make well informed decisions (Qlik, n.d.).

Types of Predictive model	Techniques for predictive modelling
Regression	Linear regression, polynomial regression, and logistic regression, Elastic Net (pro), Lasso (pro)
Decision Trees	Recursive partitioning, Bootstrap forest(pro), Random forest(pro), Partition
Clustering	K-NN (pro), Naïve Bayes (pro), SVM (pro),
Time Series	ARIMA, Seasonality, Autocorrelation, Moving Average

3.2 Kibana

Kibana is an open source user interface that is used to visualize and analyze Elasticsearch data. The interface is accessible through a browser, and kibana ships with a built-in web server. The kibana server communicates with an Elasticsearch cluster to retrieve its data.

Kibana consists of a number of applications. Within its menu we can see a number of apps but also this contains use case of specific solutions. Here I am providing a table that contains descriptions of apps.

Apps	Description
Discover	It is used to run ad hoc queries against data. It is shortcut to apply filters, inspect, matching documents.
Visualize	Lens, Area, Controls, Data Table, Gauge, Heatmap, Metric, Vertical Bar etc.
Dashboard	It is a collection of all visualization elements for user finding scope.
Canvas	It is highly customizable dashboard including options of CSS, images, etc.
Maps	Visualizing location based data.
Machine Learning	For detecting anomaly, forecast future values etc.
Graph	An interface for visualizing connection in data
Logs	Provides data of application's behavior.
Metrics	To monitor the health and performance of services and applications.
APM	This refers to the process of monitoring how a given application performs.
Uptime	Check availability of hosts and services.
Dev Tools	This contains several development tools.
Ingest Manager	Currently experimental. This offers an UI for ingesting data into Elasticsearch cluster.
Stack Monitoring	This monitors elastic stack.

3.2.1 Working process of Kibana

In this segment, I represent a workflow diagram which illustrates interaction between client, database, server, elasticsearch and kibana.

1. Client: This is UI, like a web browser, where user can request for data. This exchange data with the server.
2. Database: A storage system where data has been stored. It interacts with the server to retrieve or store data according to the requests received from the client.
3. Server: The server acts like a bridge between the database and elasticsearch. The server handles client requests, interacts with the database to create, read, update, delete operation.
4. Elasticsearch: A search and analytics engine that enables quick and flexible searching. The server sends data to Elasticsearch which it index to do fast and efficient search operation.
5. Kibana: A visualization tool that connects to Elasticsearch. Kibana offers a user friendly interface for creating dashboards and visualizations.

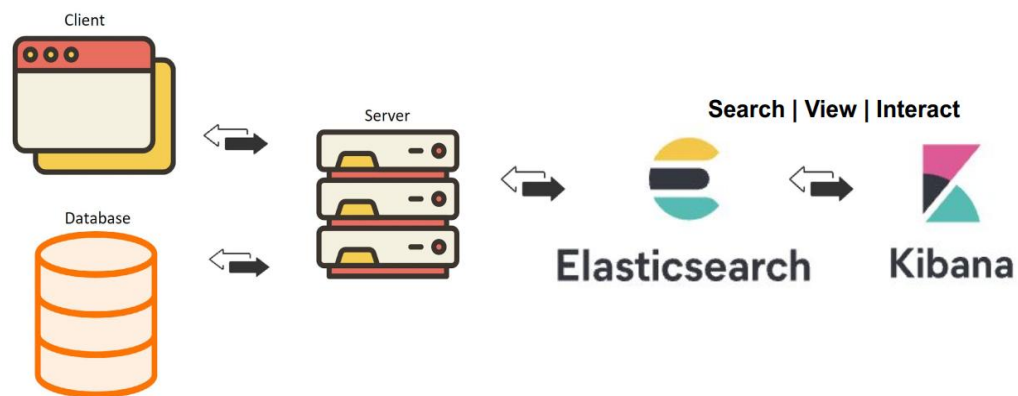


Figure 3.A workflow diagram of Kibana (Jung, n.d.).

This workflow allows users to take optimal decisions by utilizing the powerful searching and analytical tools of Elasticsearch and the interactive visualization features of Kibana (Jung, n.d.).

3.3 Power BI

Microsoft Power BI is a standalone Microsoft business intelligence product that includes both desktop and web-based applications for loading modeling and visualizing data. Here I provide a workflow on how Power BI works.

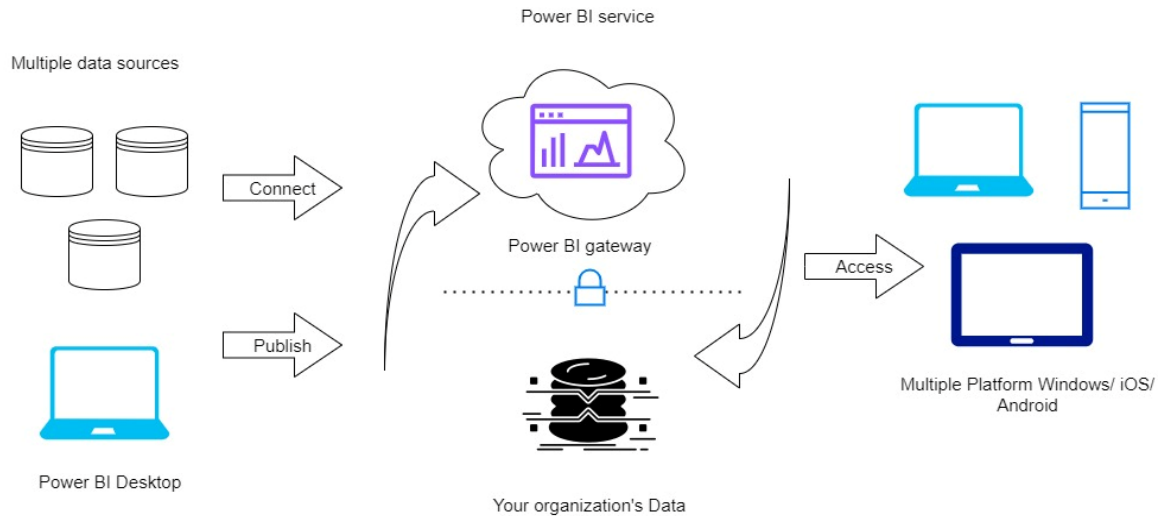


Figure 4. A basic workflow of how Power BI works (Machiraju & Gaurav, 2018)

Power BI has lots of potential integration capabilities. It can be integrated with hundreds of data sources (Machiraju & Gaurav, 2018). A user can be embedding power bi reports into a web app such as Microsoft Azure.

The basic operations of Microsoft Power BI begin with collecting data from the single or multiple data sources. Then analyze the data through connectors and gateways within the organization. After that, based on analysis, a user constructs reports by using different visuals and filters. Then publish the report to the web using power BI desktop. (Bhargava, 2018)

There are two variants of Power BI, which are:

- Power BI Desktop.
- Power BI Service.

Power BI Desktop: The name illustrates its functionality. It is an on-premises version of Power BI that enables users to perform reports, queries, slicing etc. This software also can have integration capacity to Power BI service, which makes data insights easier to build and share.

Power BI service: Power BI Service/Power BI Online is a business intelligence service that hosts reports on the cloud, specifically on Microsoft Azure. Power BI Desktop and Power BI

Service have distinct focuses. Power BI Desktop is primarily used for data creation, while Power BI Service is designed for data sharing. There is a distinction between the two when it comes to their interface (Machiraju & Gaurav, 2018).

4. Methodology

In this chapter, I developed a structural methodology for evaluating business analytics tools. This approach is followed to guarantee academic validity and scientific rigor. The methodology consists of three phases combining Data Collection and Criteria Development, Experimental Design and Implementation, and Analysis and Conclusion Below, I provide a diagram which illustrates the full procedure.

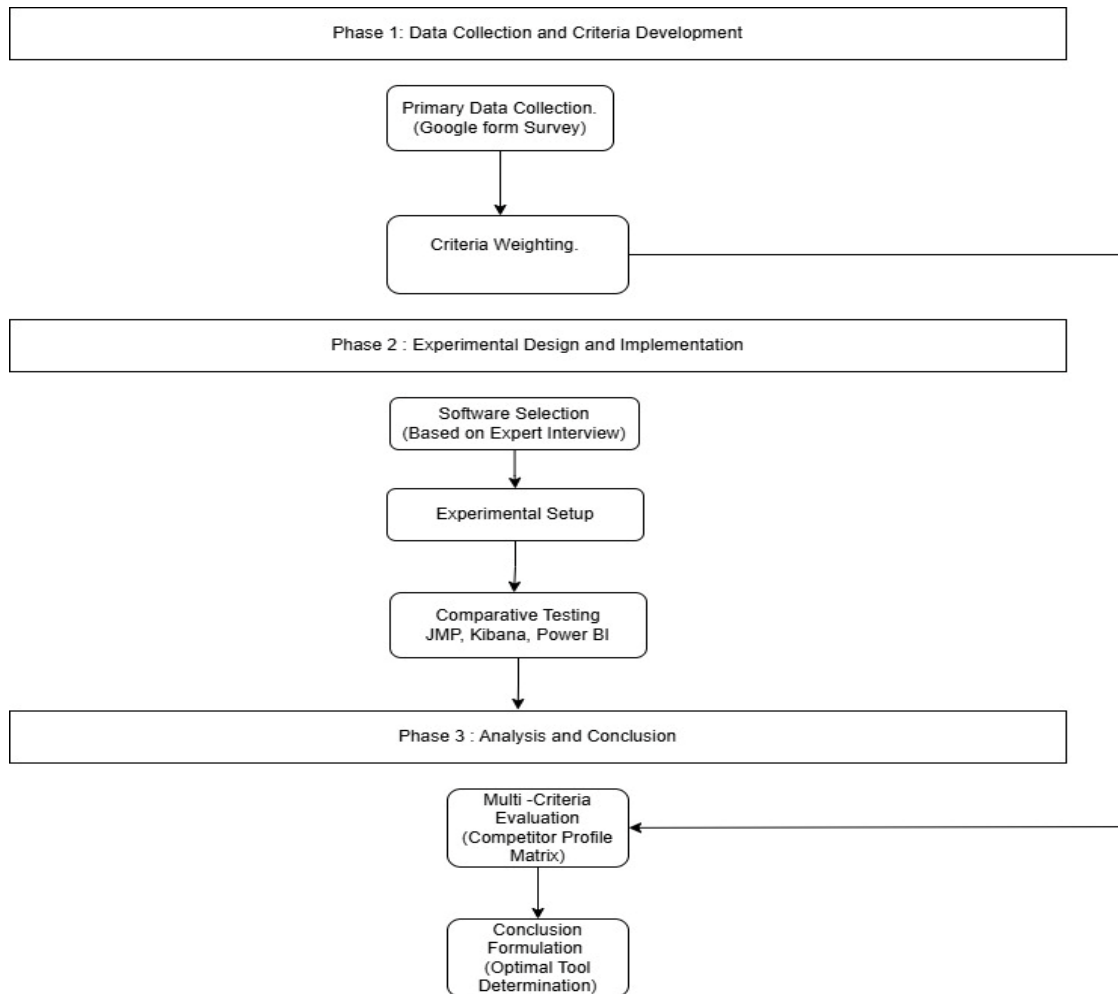


Figure 5.A workflow process for testing and evaluating Business analytics software

Phase 1: This phase consists of Data Collection and Criteria Development

Primary Data collection: The initial steps involve gathering primary data through a structure survey via google forms to both professionals and students in different fields related to business analytics. The survey includes rank matrix to capture criteria score. I am using the competitor profile matrix to select the winner software. Weight will be developed based on data collected from the survey.

Criteria Weighting: Here I am providing evaluation criteria of business intelligence software:

The software packages will be evaluated based on certain criteria. Each criterion will be scored a specific weighted to represent its importance in the overall competitive analysis. Here is a detailed explanation of ranked elements:

Criterion	Explanation & Examples
Ease of Use	This illustrates how easily an analyst visualizes data effectively.
Interactivity	This portrays easy drilling-down, filtering, highlighting and hovering data.
Customizability	This indicates no code or low code approach.
Cost	Price model of software.
Third Party Plugin	The plugin was developed by other company which increase software's functionality.

This is the form; I got 21 responses. Here I am providing random 5 samples.

Response Number	Profession (Expert on)	Most important Criteria	2 nd Important Criteria	Least Important Criteria
2	Researcher	Ease of Use	Interactivity	Customizability
5	Researcher	Ease of Use	Interactivity	Cost
9	Professional Person	Ease of Use	Customizability	Cost

10	Professional Person	Ease of Use	Customizability	Cost
21	Professional Person	Interactivity	Customizability	Third party plugin

General Information:-

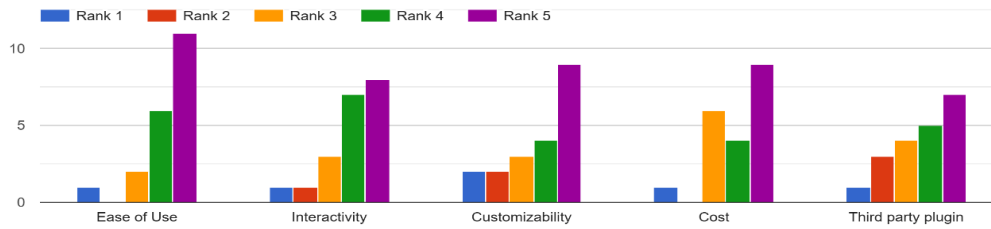


Figure 6. An overview of survey responses for criteria weighing.

This bar chart depicts responses which are collected from the survey, consisting of 21 responses. This chart shows distribution of five key dimensions of business analytics tools: ease of use, interactivity, customizability, cost, and availability of third-party plugins. Each dimension is rated on a scale from Rank 1 (lowest) to Rank 5 (highest). The result indicates the most important criteria is high usability. Similarly, “Interactivity” received higher ranking (Rank 4 and Rank 5), suggesting that users prefer interactive visualizations in business analytics tools.

Overall, the survey result highlights that users prioritize usability, interactivity, and cost effectiveness chronologically as the highest factors when choosing software. The respondent seemed less concerned about “customizability” and “third party plugin” factors, which might indicate discomfort with coding approaches.

What is your profession
21 responses

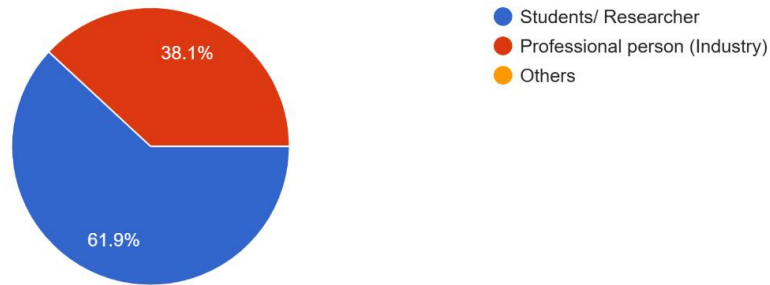


Figure 7. Pie chart of demographic overview of participants.

The pie chart illustrates the demography of respondents to the survey evaluating business analytics tools. Among the participants, a significant majority of respondents, 61.9% , identified themselves as students/ researchers. This indicates academic focus and relevance of analytic tools within an educational institution setting. Whereas 38.1% of participants were professional individuals from industry settings, which suggests a significant industry attachment and relevance of the study findings in real-world business environments. The lack of respondents in “Others” category respondent indicates that the results has been collected from both theoretical and practical perspectives, guaranteeing that the findings are accurate, robust and only include both academic and industry- oriented participants.

Phase 2: Experimental Design and Implementation

Software Selection: The selection of analytics tools for comparison is developed by semi-structured interviews with experts. The interview protocol includes inquiries regarding tool prevalence, perceived strengths and limitations, and contextual performance in financial analytics tasks. Content analysis of interview transcripts revealed three predominant tools warranting comparison: JMP Statistical Software, Kibana (Elastic Stack), and Microsoft Power BI. The selection criteria prioritized tools with established market presence, distinct architectural approaches, and reported application in financial analytics contexts.

Experimental Setup:

A controlled testing environment is established to minimize extraneous variables that might influence comparative performance. Key control measures include:

- Standardized hardware specifications (processor, memory, storage).
- Identical dataset implementation across platforms.
- Investigating business models and finding potential analytical approaches from given dataset.
- Determining analytical tasks to perform in selected platform.
- Documented tool versions and configurations.

Standardized hardware specifications: Windows 11 operating system runs below the mentioned hardware specifications for Power BI and JMP.

Hardware	Description
Processor	Intel(R) Core (TM) i5-8250U CPU
Memory	16.0 GB DDR 3
Hard Disk space	Disk 0 : 931.51 GB, Disk 1: 238.35
Other	Keyboard, mouse

Ubuntu operating system is running below mentioned hardware specifications for Kibana.

Hardware	Description
Processor	Intel(R) Core (TM) i5-13420H CPU
Memory	32.0 GB DDR 5
Hard Disk space	512GB SSD
Other	NVIDIA GeForce RTX 4060 8GB

Identical dataset implementation across platform:

Sample sales data, including order information, sales figures, customer details, and shipping data, utilized for segmentation, customer analytics, clustering, and further purposes has been used in this thesis. Motivated by retail analytics. This was first utilized for Pentaho DI Kettle;

nevertheless, Gus Segura discovered that the set could be advantageous for Sales Simulation training. This dataset is licensed under CC0: Public domain (Kyanyoga, n.d.).

Number of Rows: The dataset contains multiple entries, each representing a sales order.

Number of Columns: There are 25 columns in the dataset, which includes:

- ORDERNUMBER: The unique identifier for each order.
- QUANTITYORDERED: The number of items ordered.
- PRICEEACH: The price of each item.
- ORDERLINENUMBER: The line number of the order.
- SALES: The total sales amount for the order.
- ORDERDATE: The date on which the order was placed.
- STATUS: The status of the order (e.g., Shipped).
- QTR_ID: The quarter in which the order was placed.
- MONTH_ID: The month in which the order was placed.
- YEAR_ID: The year in which the order was placed.
- PRODUCTLINE: The product line of the ordered items.
- MSRP: This field contains Manufacturer's Suggested Retail Price.
- Customer and shipping information columns (ADDRESSLINE1, CITY, STATE, etc.).
- COUNTRY: The country to where the order was shipped.
- TERRITORY: The sales territory.
- CONTACTLASTNAME and CONTACTFIRSTNAME: The name of the customer contact.
- DEALSIZE: The size of the deal (e.g., Small, Medium, Large).
- Missing Values: Some columns like ADDRESSLINE2 and STATE may have missing values, indicated by NaN.
- Data Types: The dataset includes a mix of numeric (e.g., QUANTITYORDERED, PRICEEACH, SALES), categorical (e.g., STATUS, COUNTRY, DEALSIZE), and date (ORDERDATE) data types.

Determining analytical tasks to perform in selected platform:

Selected dataset refers the business model of Wholesale distribution of collectible products. This product includes categories like Classic Cars, Motorcycles, Planes, Ships, Trains, Trucks and

Buses, and Vintage Cars. Given this business model, several analytical strategies can be formulated.

1. Customer Segmentation: A cluster can be done based on purchasing behavior, deal size, and geographic location to conduct market research.
2. Sales Performance Analysis: Evaluation of sales trends over time, across different territories, and among various product lines to measure sales performance.
3. Inventory Optimization: Analyze order patterns to forecast demand and manage inventory effectively.

To demonstrate software capabilities through comparison, I select to perform Clustering for segmentation, Sales trend over time and Growth rate analysis for given software.

Documented tool versions and configurations:

Based on interviews, I downloaded these software to my hardware to do research. Details are provided in tabular format which is given below.

Software name	Version	Configuration/System Requirements	Source
JMP	18 Student Edition	Windows 10/11, macOS (minimum 4GB RAM, 1GB disk space recommended)	https://store.jmp.com/
Kibana	8.x	Elasticsearch 8.x.x, Java (JDK) 17+, minimum 8GB RAM, 2GB disk	www.elastic.co/kibana
Power BI	Desktop April 2025	Windows 10/11, min. 4GB RAM (8GB recommended), .NET Framework 4.8	powerbi.microsoft.com

Phase 3: Analysis and Conclusion

Competitor Profile Matrix: The competitor profile matrix (CPM) is a strategic tool used in market research to evaluate and compare competitors based on critical success factors that are important for success in a specific industry (Everett, 2008). This approach usually helps to understand competitive positioning of selected software.

How the competitor profile matrix works:

Step 1: Select the most important criteria for data visualization that you need to consider. These criteria will be key factor to differentiate between selected software and competitor software.

Step 2: Assign weight for each criterion.

Step 3: Assign a value for each criterion based on result analysis. Use five-point scale for the critical success factor evaluation.

5- Best choice.

4- Good choice.

3- Moderate choice.

2- Negative choice.

1- Very negative choice.

The scoring formula is calculated as:

$$\text{Score} = \sum(W_i \times P_i)$$

Where,

W_i = Weight of criterion i, and P_i = Performance score for criterion

Critical Success factors	Weight	JMP	Power BI	Kibana
Ease of use	W_1	$W_1 * P_1$	$W_1 * P_1$	$W_1 * P_1$
Interactivity	W_2	$W_2 * P_2$	$W_2 * P_2$	$W_2 * P_2$
Customizability	W_3	$W_3 * P_3$	$W_3 * P_3$	$W_3 * P_3$
Cost	W_4	$W_4 * P_4$	$W_4 * P_4$	$W_4 * P_4$
Third party plugin	W_5	$W_5 * P_5$	$W_5 * P_5$	$W_5 * P_5$
Total Score	1.0	Result 1	Result 2	Result 3

Table 2: Formula of Competitor Profile Matrix for scoring.

5. Results and Analysis

5.1 Clustering analysis for segmentation

a. **JMP:** The clustering analysis started in JMP by importing the dataset sales_dataset.csv. The variable Deal_Size is a categorical variable with three distinct record- 'Small', Medium' and 'Large'. This Deal_Size field was selected for clustering. To make an indicator column for the clustering process I had to navigate to Cols> Utilities> Make Indicator Columns, selecting Deal_Size. This process generated three new binary columns named 'DEALSIZE_Large', 'DEALSIZE_Medium', 'DEALSIZE_Small' encoding the presence of each distinct record as (1) or absence (0).

Thereafter, the K means clustering algorithm was achieved by navigating to Analyze > Clustering > K-Means Cluster. With this process, JMP automatically detected the optimal number of cluster sizes as 3.

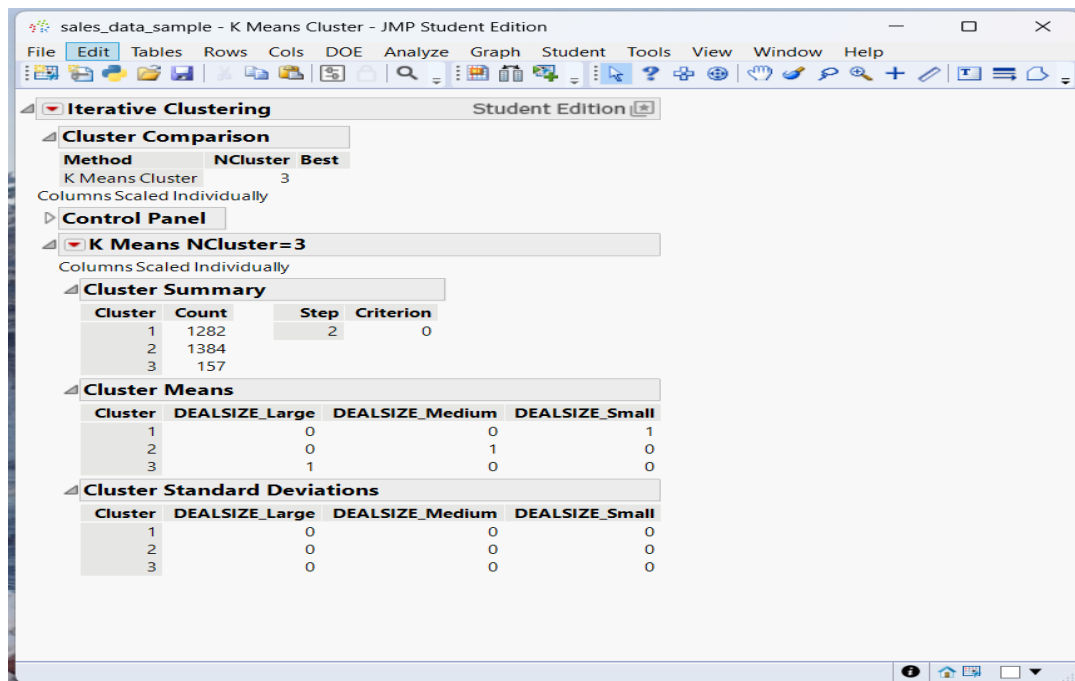


Figure 8. Overview of Post processed dataset.

Upon completion of this step, JMP appended a new column to the dataset, indicating the cluster assignment for each record. This divided each record into three group, which helped visualize the data for the next stage.

To visualize, the Graph Builder tool was selected by navigating to Graph > Graph Builder. This is an interface where targeted visualization can be achieved by a simple drag and drop method.

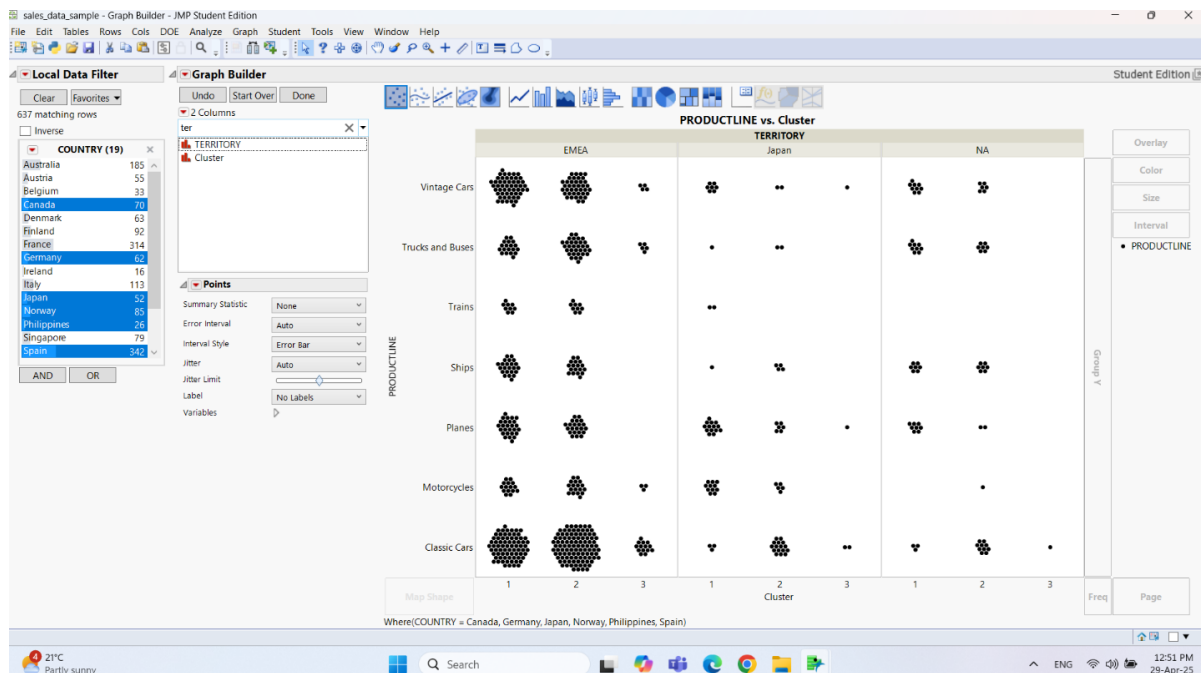


Figure 9. Clustering by deal size for segmentation.

Here, the X axis was assigned to the cluster variable, which represents each cluster ranging from 1 to 3. The Y axis was assigned to the ProductLine variable. This process segmented product lines based on deal size. To enhance the visualization, the territory variable was dragged and dropped into Group X, which segmented the data by geographical regions.

This process provides a detailed visualization of how product lines are distributed within each cluster across various territories. Furthermore, JMP's Local Data Filter was tested to do dynamic data filtering based on different countries. This, in turn, amplifies the potential for exploratory analysis.

b. Kibana: For Kibana, preprocessing datasets is not straightforward. To facilitate exploratory data analysis The sales_data_sample.csv dataset was uploaded through the integrations→Upload file section. Kibana's find_file_structure API automatically analyzed the file to infer its structure. After the structural analysis, Kibana created an Elasticsearch index. At the same time, an ingest pipeline can perform various transformations, including field renaming, data type conversion, and enrichment operations, to ensure the data conforms to the desired schema.

After that the dataset is ingested into Elasticsearch. Additionally, an index pattern is defined within Kibana in order to query and visualize data smoothly.

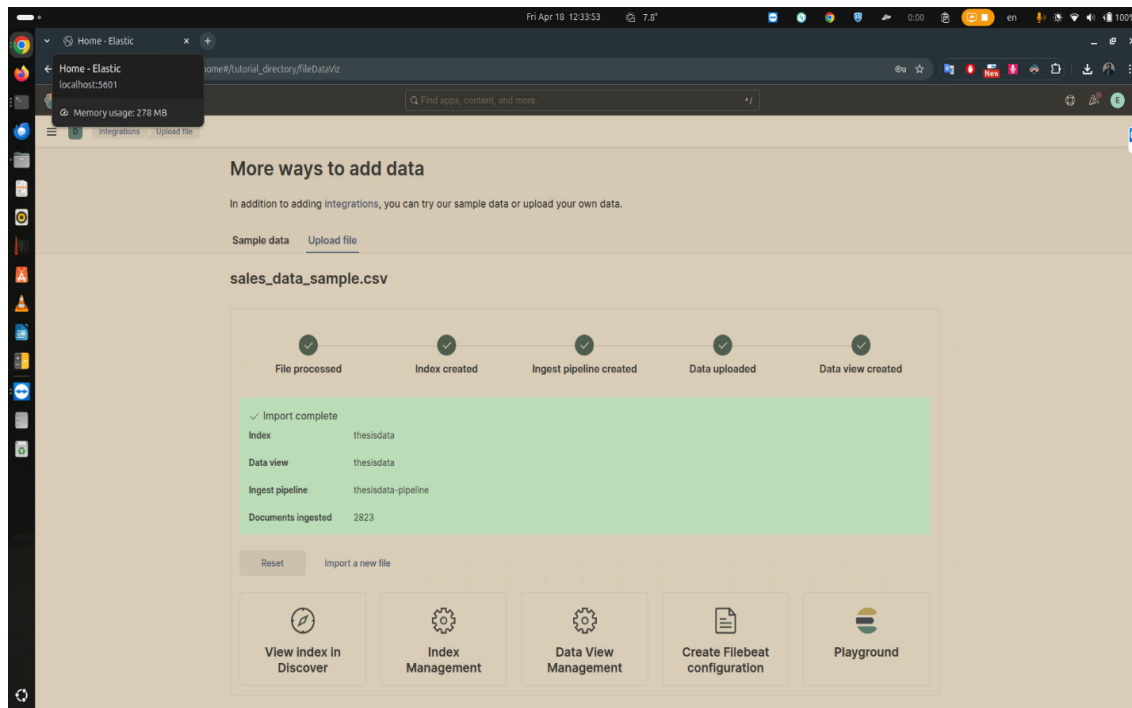


Figure 10. Dataset Preprocessing Interface. CSV file loaded with 2823 documents ingested in 5 steps.

At the post-ingestion phase, the dataset is accessible under Kibana's Discover tab, allowing seamless exploration. A user can perform KQL to filter data alongside graphical filter options without coding skill, quick check distribution of each available field.

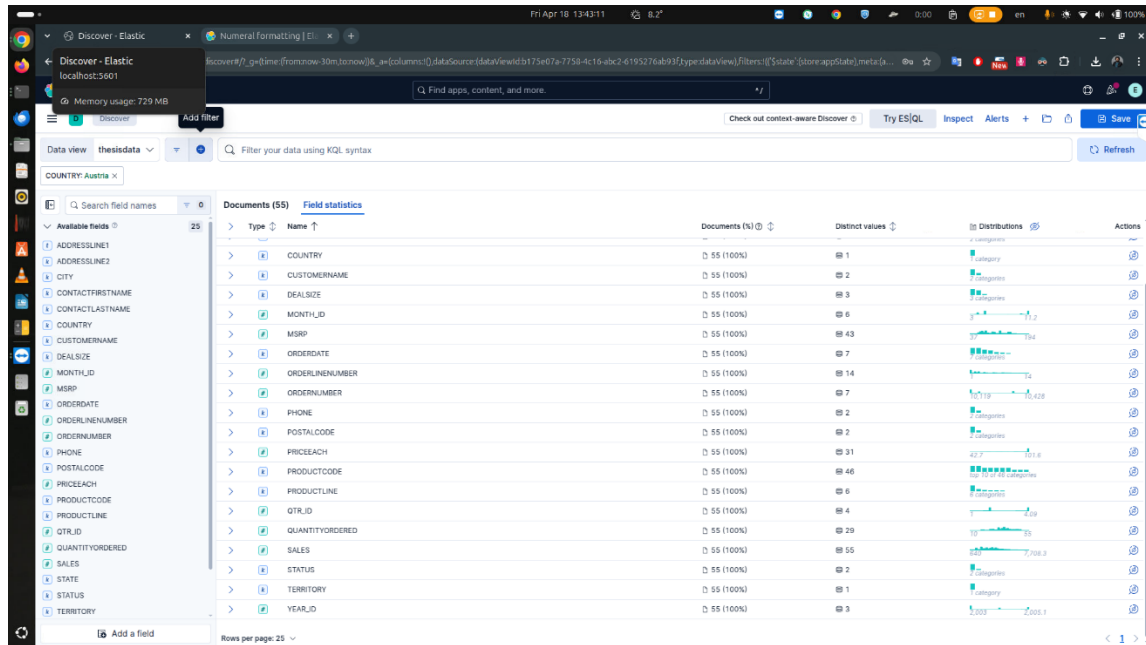


Figure 11. Field statistics of loaded dataset in Data view interface.

After the dataset preprocessing data analysis Clustering analysis has been performed for Kibana. The cluster will be created based on DEAL_SIZE of the given dataset. To conduct the same, Create Visualization option has been selected under Dashboard section. This invoked Lens editor, which is a feature of kibana, provides a workspace where visualization can be created interactively. In comparison to JMP, Kibana has some chart limitations. Lens editor does not contain scatter chart. Due to this limitation, Bar char (stacked) has been selected. The field deal_size was assigned to the horizontal axis by dragging from the field panel. By default, the lens editor assigned Count of records as vertical axis. Both case, I did not need to configure axis data from the field. User can drag-drop different field from the dataset to the editor and filter data interactively from the workspace. The visualization was later saved and added to the dashboard for ongoing analysis and generating report through further configurations including add panel, add

from library, controls options.

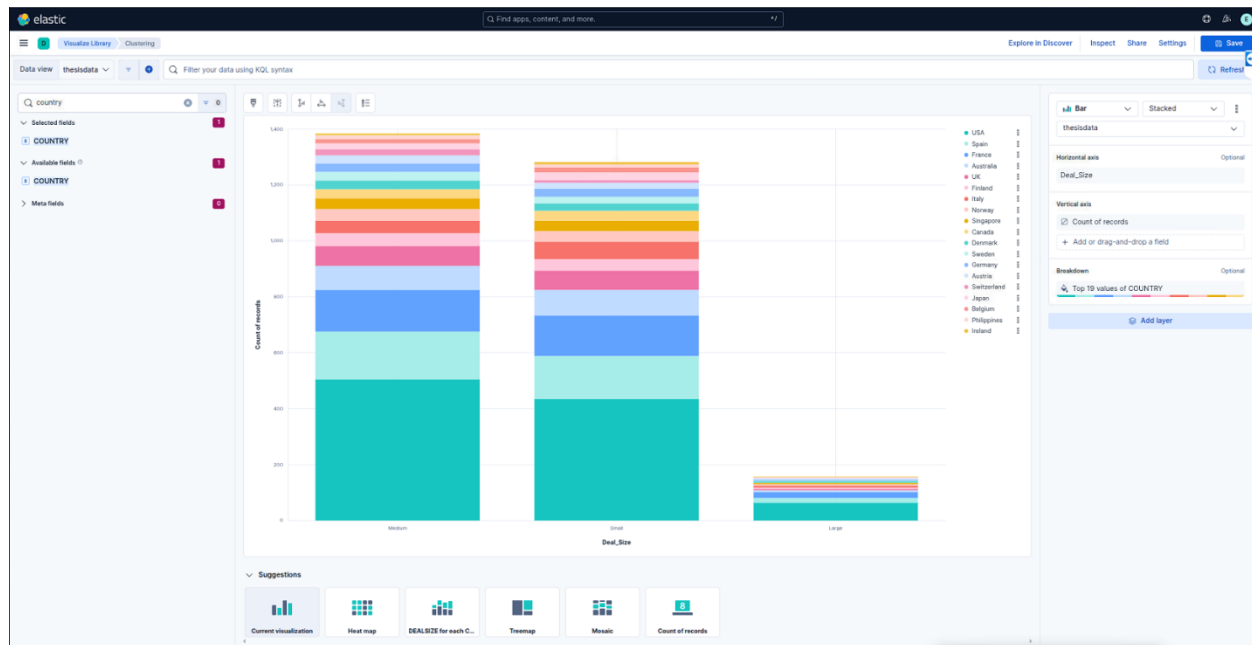


Figure 12. Clustering by deal size for segmentation

c. Power BI: In order to replicate cluster analysis similar to the section, in Power Bi, sales dataset has been preprocessed. The initial step involves importing sales dataset into Power BI to transform it. The power query editor is used to transform dataset by creating conditional columns. A new column has been created titled Cluster based on categorical field deal_size. This was achieved by assigning values as follows: 'Small' -> 1, 'Medium' -> 2, 'Large' -> 3. This step is done to transform categorical data into numerical for quantitative analysis and visualization of deal sizes.

Next step, an index column has been created as unique identifier to detect each datapoint in the visualization process. Upon completion of these transformations, now dataset is ready for performing clustering.

To visualize the clustering, scatter chart was picked by assigning cluster to the X axis. This represents encoded deal sizes. For the Y axis, the index column was assigned to effectively distribute data points vertically to distinguish each record.

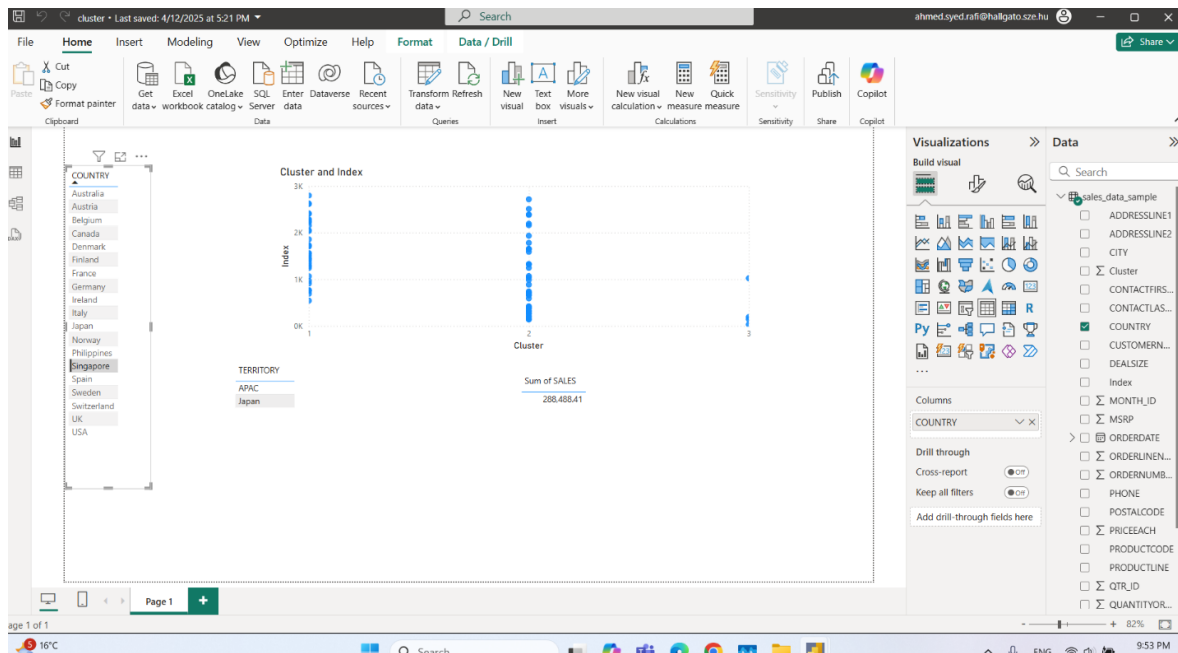


Figure 13. Clustering by deal size for segmentation.

To check maximum visualization capabilities, from different countries and territory fields were integrated to the dashboard. The country field was added to the legend area and the territory field was assigned to the tooltips section to enhance visualization capabilities.

5.2 Sales trend analysis

a. JMP: To do Sales trend analysis over time, Sales dataset was imported into JMP. The Month_ID column represents a temporal component, while the Sales column represents as sales figure against Month_ID. In this particular analysis, no dataset preprocessing is required.

Next from the main menu, Graph-> Graph Builder was selected to open the Graph Builder Interface. The Month_ID variable was dragged and dropped to the X-axis of the Graph builder. This created a sequence of months on X axis. The Sales variable was dragged and dropped to the Y-axis. Line chart has been selected to visualize relationships between Month_ID and Sales. To aggregate Sales data by month, summery statistics within Line property has been selected as Sum.

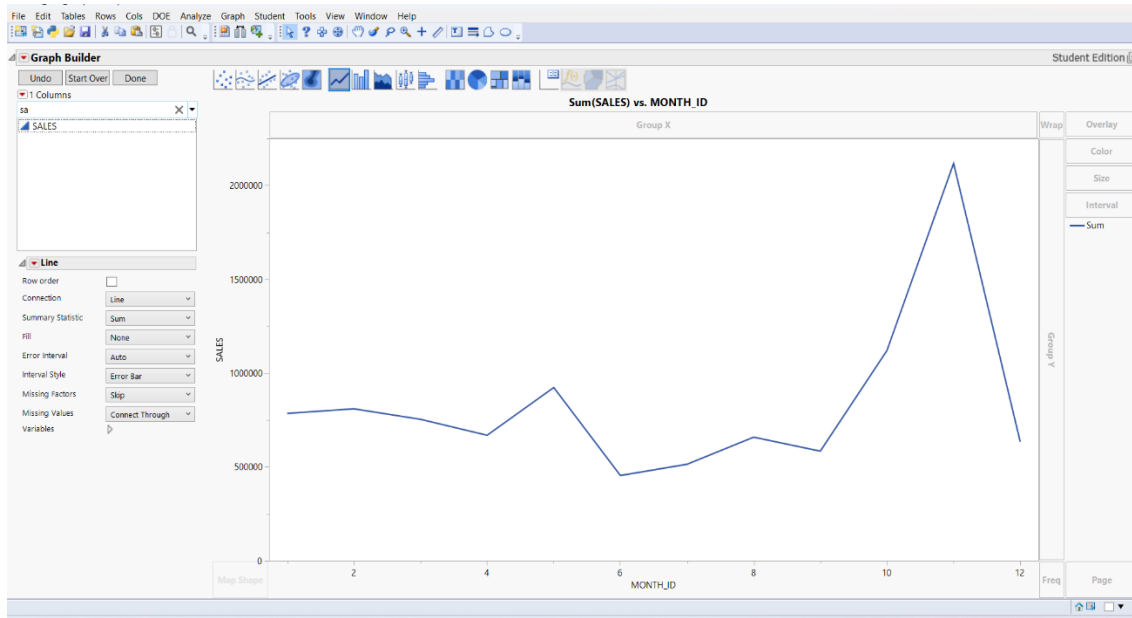


Figure 14.Sales Trend over MONTH_ID

The line chart provides information on patterns and trends of sales over time. This will help user to observe fluctuations and seasonal decomposition, and detect anomalies in the overall sales performance.

b. Kibana: To replicate the same analysis performed in JMP in Kibana, no dataset preprocessing is required in this step, as dataset was pre-loaded from the above-mentioned section 5.1.b. Within Kibana, the Create Visualization option was selected. On Edit Visualization plane, Line chart has been selected where MONTH_ID field is dragged into horizontal axis and Sales into vertical axis. Vertical axis configuration was required to convert SALES column into ‘Sum

of Sales' by using the quick function of Kibana 'Sum'.

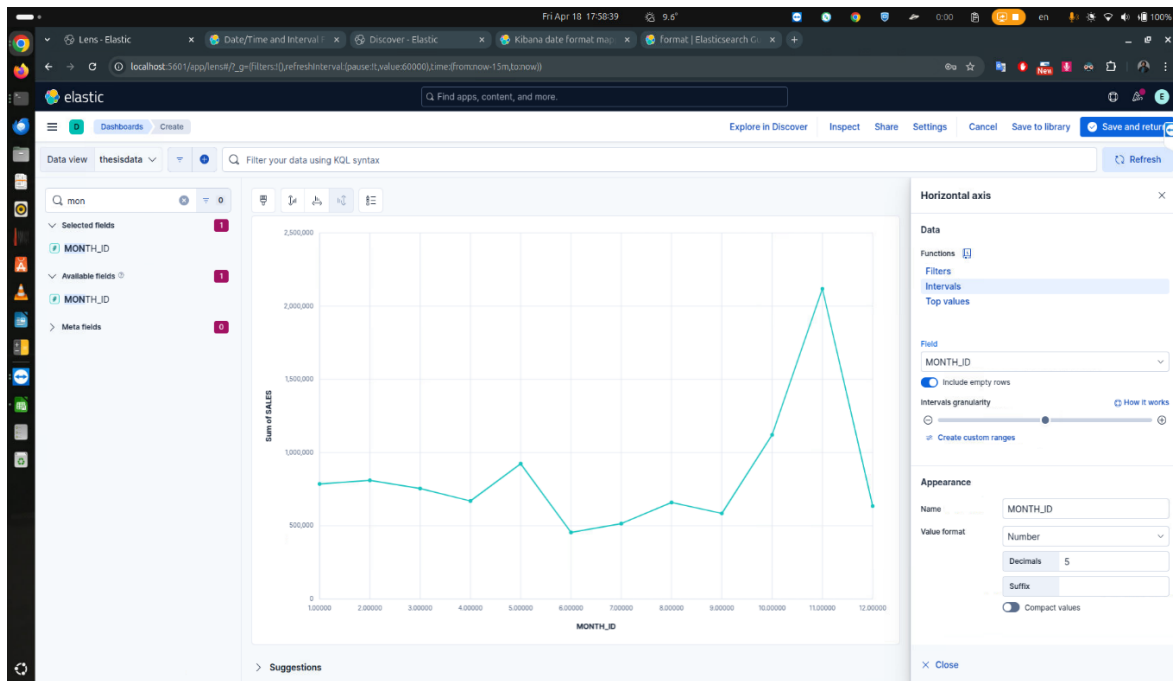


Figure 15.Sales Trend over MONTH_ID.

The line chart depicts sales trend over time by aggregating sales data per month, which enables users to visualize fluctuations, seasonal effects, and potential anomalies.

c. Power BI: The sales dataset was imported into Power BI. At this stage, no additional transformation was required, as the dataset was already structured for this analysis. Within the visualization interface, a line chart was selected. After that, month_id field was dragged to the X axis and the sales field was dragged to Y axis. A sum aggregation was done for sales.

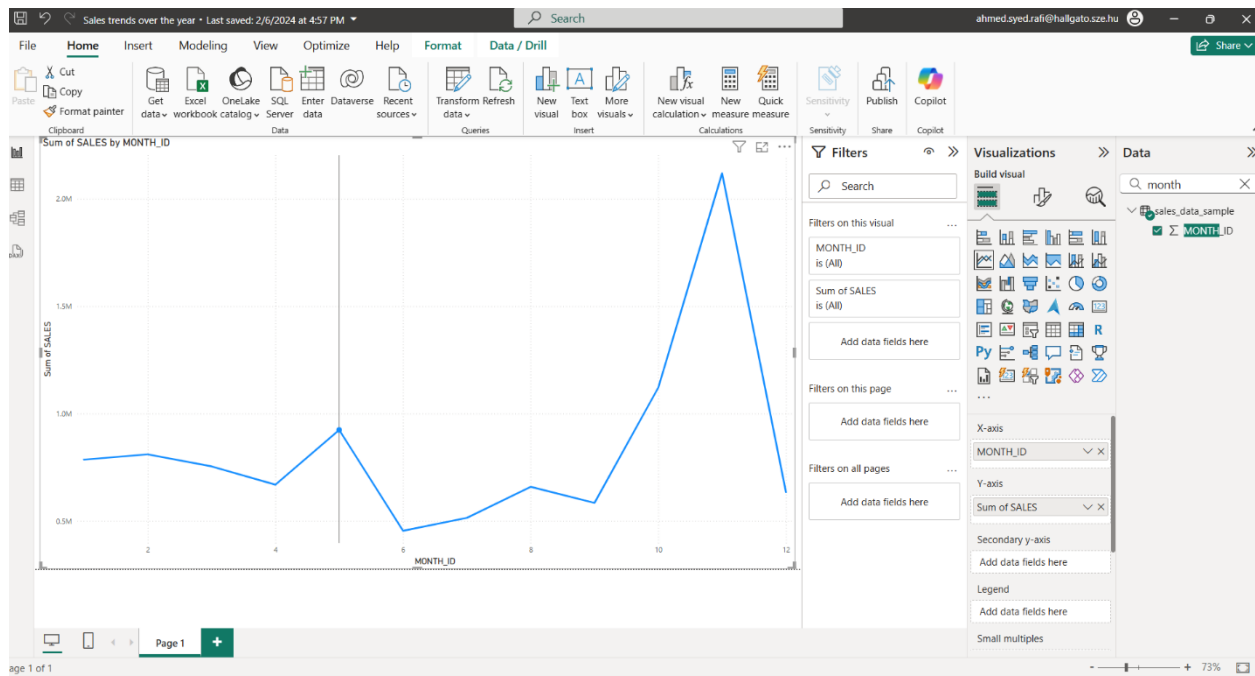


Figure 16.Sales Trend over MONTH_ID.

The line chart depicts the identification of patterns and trends over monthly sales.

5.3 Growth Rate analysis

a. Kibana: Initially, the approach involved using Kibana's TSVB visualization to aggregate sales data by year_id and subsequently apply mathematical formulas for calculating growth rates. The reason to choose this approach is TSVB's math function, where Growth rate formulation can be done easily. However, in Kibana 9.0.0 version, this approach proved unfeasible due to the absence of custom mathematical computation.

In order to address this limitation, Custom Visualization of Kibana has been selected. This visualization is featured by Vegalite visualization, which is a grammar, a declarative language for interactive design creation. (Vega, n.d.).

Inside the coding segment of the visualization platform, data transformation, calculation and type of visualization as 'Bar' chart has been done.

```
BEGIN DataRetrieval
CONNECT to Elasticsearch (or relevant database)
QUERY dataset: 'salesforgrowth'
AGGREGATE by YEAR_ID
  COMPUTE sum of SALES as 'total_sales'
  SORT years ascending
RETRIEVE aggregated results:
FOR EACH year:
  STORE 'YEAR_ID' as year
```

```

STORE 'total_sales' for each year
END DataRetrieval

```

Source: Author's own coding on Kibana (Appendix 1)

Above-mentioned pseudo-code represents data retrieval from the source, where dataset has been preprocessed for transformation.

```

BEGIN DataTransformation
FOR EACH year IN retrieved data:
  SET sales = total_sales.value
  SET year_label = key (YEAR_ID)
  SORT all data by year ascending

FOR EACH year, STARTING from second entry:
  SET prev_sales = sales value of previous year
  IF prev_sales is NOT NULL AND greater than 0 THEN:
    COMPUTE growth_rate = ((sales - prev_sales) / prev_sales) * 100
  ELSE:
    SET growth_rate = 0
  END loop
END DataTransformation

```

Source: Author's own coding on Kibana (Appendix 1)

Above mentioned pseudo-code performs calculations to compute growth rate over the year. The calculated growth rate was visualized by using Vega's visualization grammars, including mark, encoding, condition, value, and tooltip function. This approach requires Json coding knowledge and provides user flexibility during visualization. The final visualization is given below.

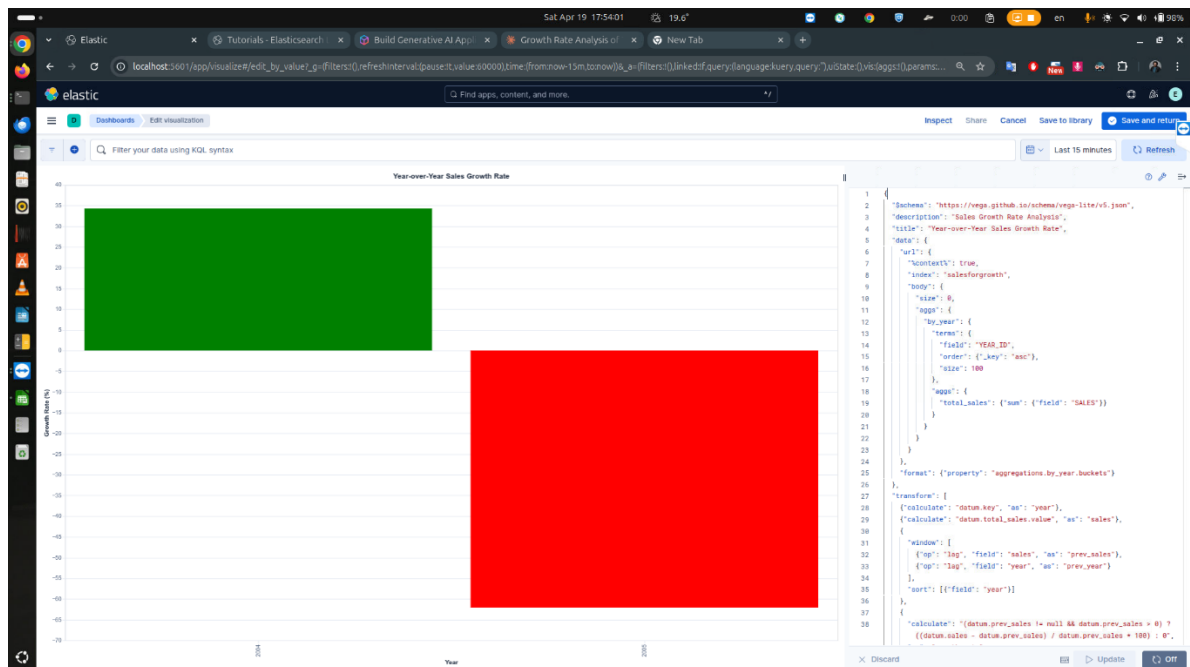
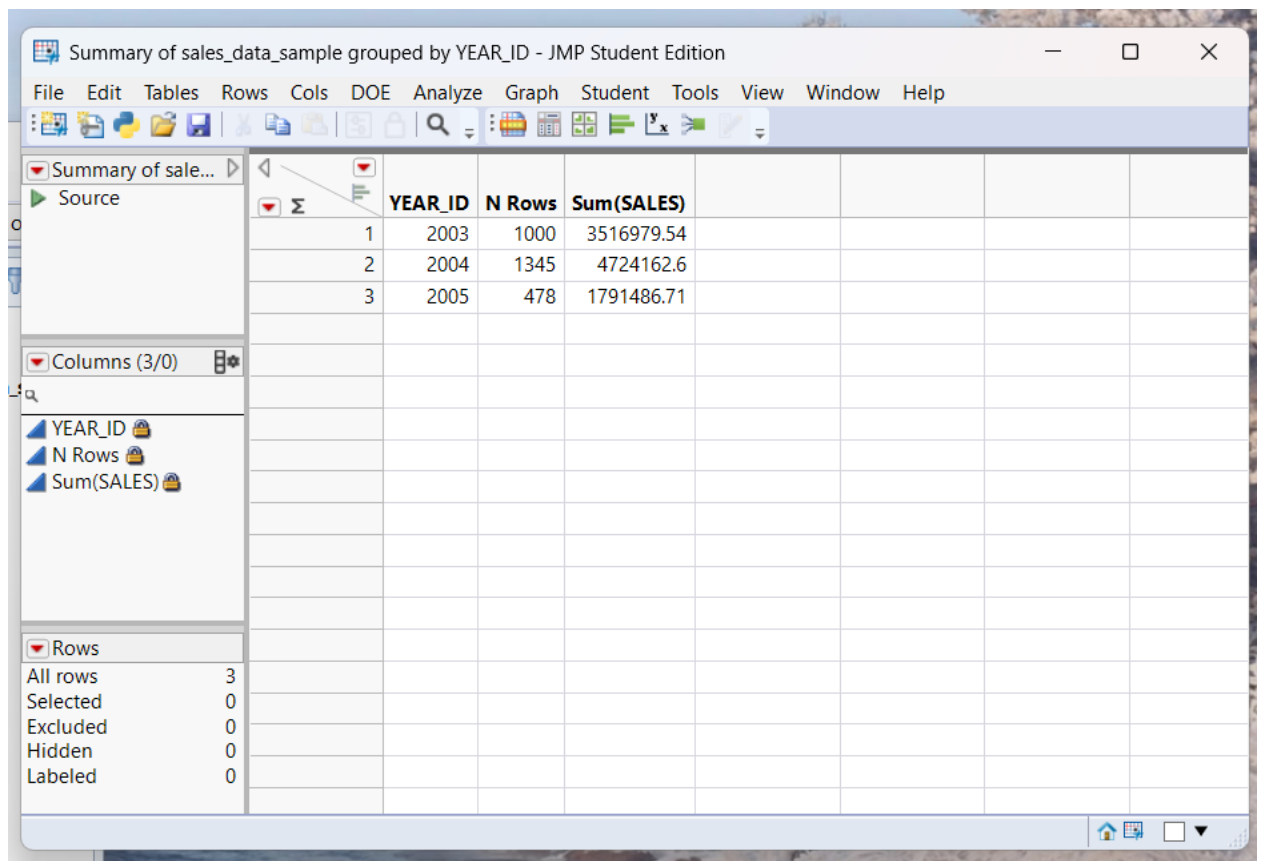


Figure 17. Growth Rate% over YEAR_ID.

b. JMP: To analyze Growth rate of Sales over the year, JMP's data summarization and formula functionalities were used. First, sales_data_sample.csv has been imported to JMP. No preprocessing was required at this step. Next, Tables-> Summary has been selected to open Summary dialog where YEAR_ID was selected as grouping variable. Then, SALES field has been selected to do statistical operations to compute total sales per year. A new summary table has been generated.



Summary of sales_data_sample grouped by YEAR_ID - JMP Student Edition

	YEAR_ID	N Rows	Sum(SALES)
1	2003	1000	3516979.54
2	2004	1345	4724162.6
3	2005	478	1791486.71

Columns (3/0): YEAR_ID, N Rows, Sum(SALES)

Rows: All rows (3), Selected (0), Excluded (0), Hidden (0), Labeled (0)

Figure 18. Summary of preprocessed data in JMP.

After that, a new column has been created as 'YOY_Growth%' and set the data type as Numeric. Under Column Properties, the below-mentioned formula was written.

```
If(
Row() == 1, 0,
((:Total_Sales - Lag(:Total_Sales, 1)) / Lag(:Total_Sales, 1)) * 100
)
```

Source: Author's own coding (appendix 2)

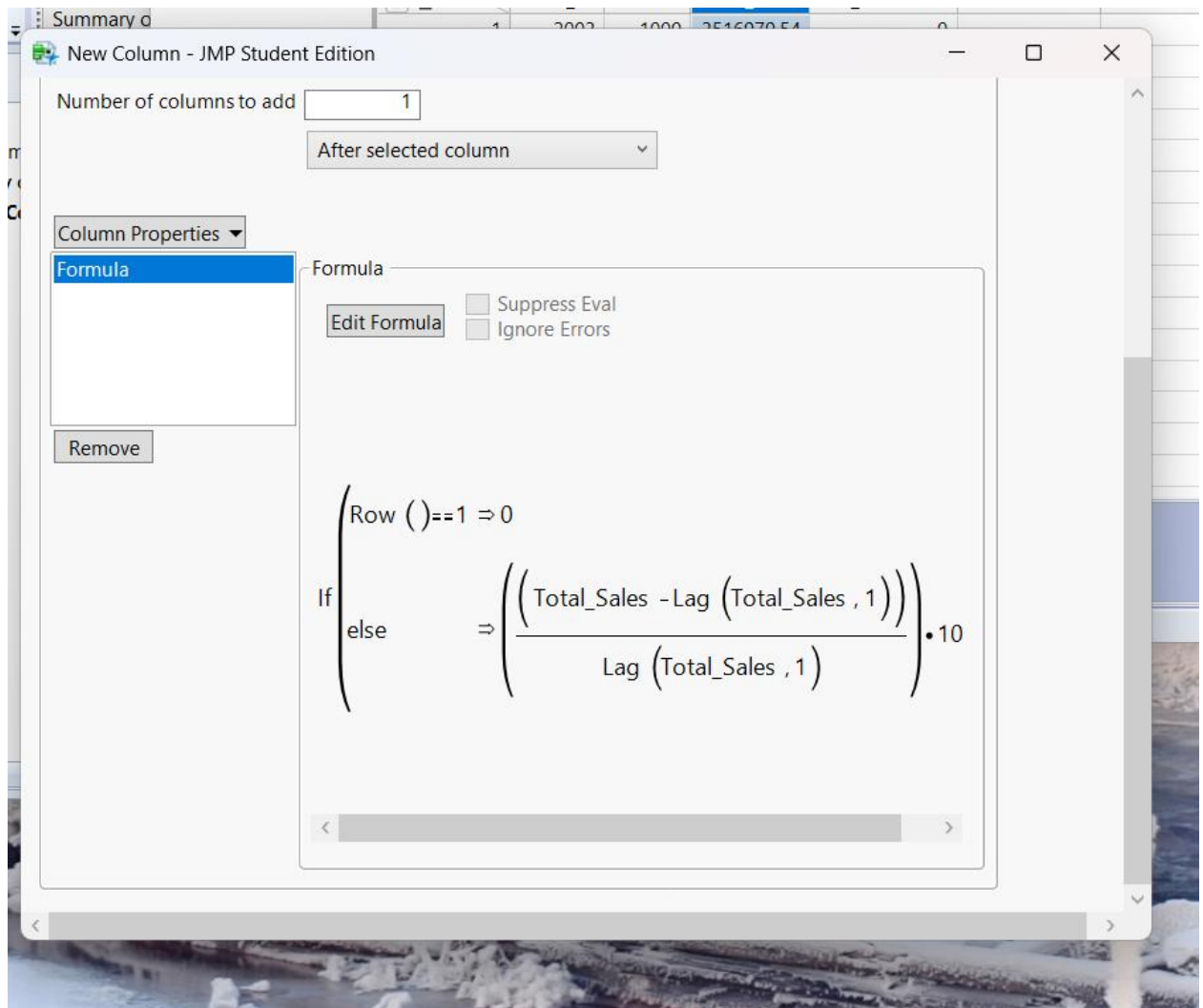


Figure 19. Formula Box in JMP.

The formula dialogue box automatically generates codes into understandable statistical formula for users with no prior coding knowledge. This formula calculates the percentage change in Total_Sales. As for the first row, where no record exists, the growth rate is set to 0.

At the final step, for visualizing the YOY Growth rates, Graph -> Graph Builder has been selected which opens a new window where visualization has been performed by dragging Year_ID to the X-axis and YOY_Growth% to the Y axis. Bar chart icon was selected for this visualization.

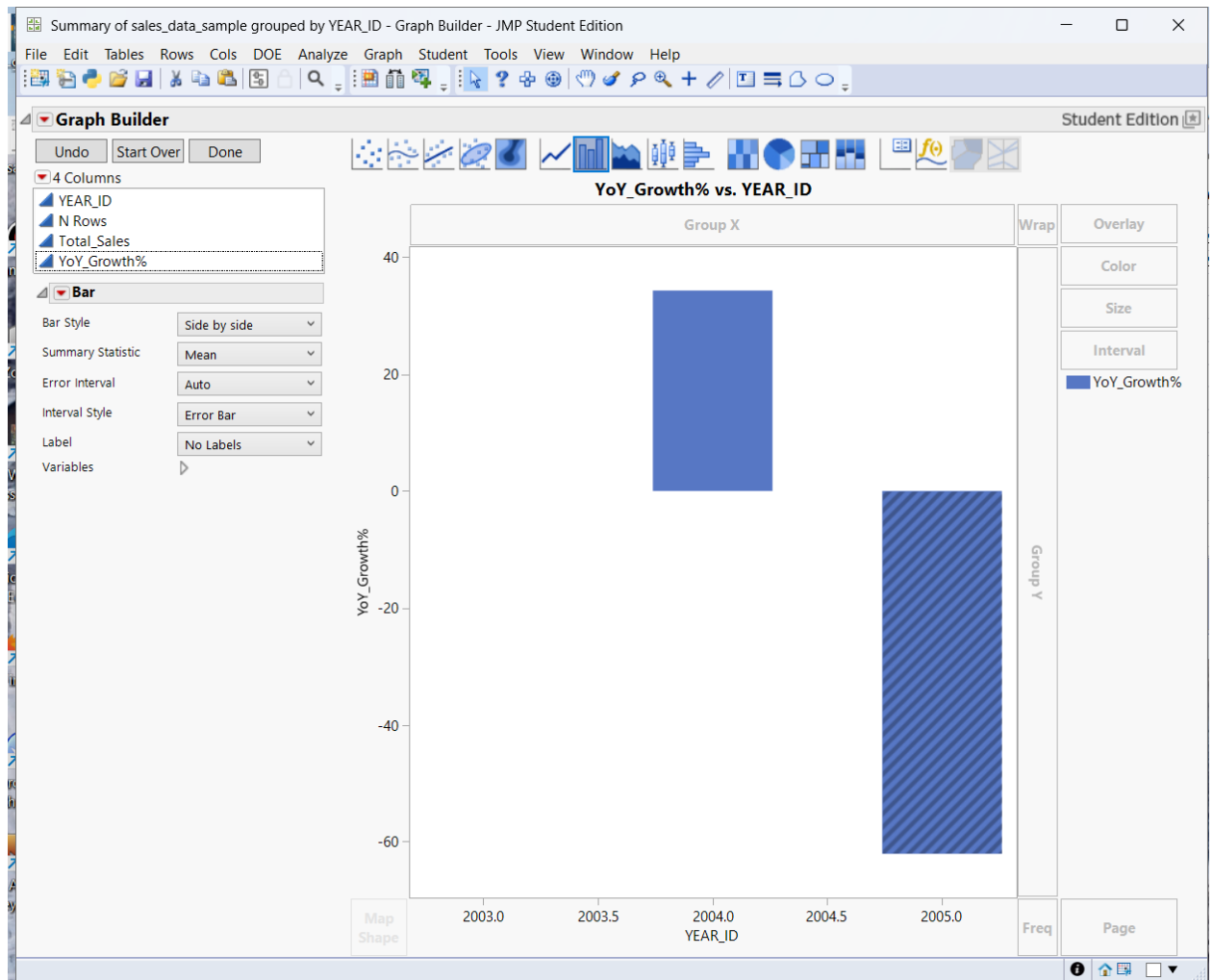


Figure 20. Growth Rate% over YEAR_ID.

Settings of Bar chart are mentioned in the figure 20. The resulting bar chart provides necessary visualization of annual sales growth rates.

c. Power BI: To analyze growth rate of sales over the year, Power BI's Power Query Editor was used to preprocess the sales_data_sample.csv. First, the sales dataset was imported into Power BI, and then this preprocessing went through several steps:

Grouping Data: The group by function was applied to the YEAR_ID column and then the aggregation operation was selected Sum on the Sales Column. This operation created a new column named Total_Sales that contains the total sales for each year.

Adding Index Column: An index column from 0 was added to the table. This index number has been added to calculate previous year's sales.

Previous year's sales calculation: A custom column was generated by using following power Query formula:

```
= Table.AddColumn("#Added Index", "Prev_Year_Sales", each try #"Added Index"{{[Index]-1}[Total_Sales] otherwise null)
```

This formula selected the Total sales value from the previous row based on the index. If the previous row does not exist, it returns null.

Growth rate calculation: Additionally, a custom column was added to the table with the following Power Query formula:

```
= Table.AddColumn("#Added Custom", "YoY_Growth%", each if [Prev_Year_Sales] = null or [Prev_Year_Sales] = 0
then 0
else ([Total_Sales] - [Prev_Year_Sales]) / [Prev_Year_Sales] * 100)
```

This code calculates the percentage change in Total_Sales against the previous year.

Finally, index column was removed as this step is only required to calculate Previous year's sales. The dataset preprocessing steps were done after selecting Close & apply to load the process data into Power BI.

To visualize YoY growth rates a bar chart has been created in the dashboard. YEAR_ID column was assigned to the X-axis and YoY_Growth_% column was assigned to the Y axis.

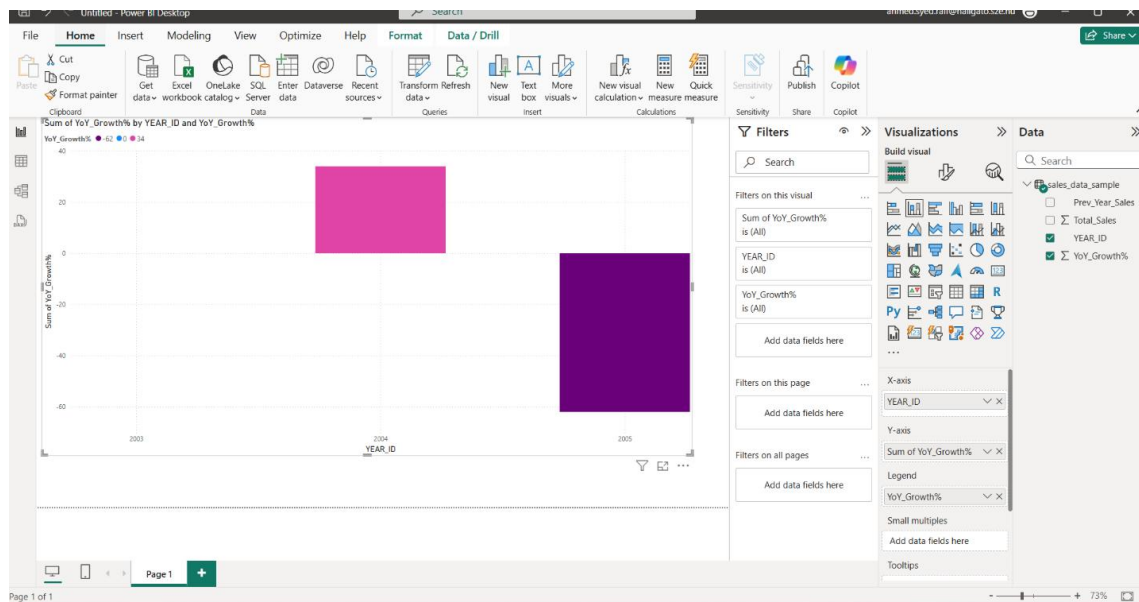


Figure 21. Growth Rate% over YEAR_ID.

This approach is highly dependent on dataset preprocessing. Power query is advantageous for performance optimization when a user has to deal with a large dataset. However, in this case,

this method limits the visualization capabilities, as other necessary fields were absent for further filtration and investigation for more insights.

5.4 Result evaluation

To summarize experiences while performing clustering, sales trend analysis and growth rate analysis for each software, a comparative chart is presented below, mentioning strong points and weak points that were obtained during the operation.

Tool	Experiment	Strong Points	Weak Points
JMP	Clustering	Less steps in Preprocessing, Easy to apply clustering with Statistical focus.	No presence of dashboard feature (Student Edition), commercial license restrictions.
	Sales Trend analysis	Easy way of dataset import, statistical ability.	Lack of dynamic dashboard features, real time multi-view filtering options.
	Growth rate analysis	Preprocessing with data summarization and formulation, straightforward coding in formula editor.	Less options for visualization in Graph Builder.
Kibana	Clustering	One time task in dataset preprocessing, Real time filtering, Strong visualization with color coding alongside hovering effects.	Prior Programming knowledge required specially Json, limited built-in chart, requires Linux operating system for smooth run time operation.

	Sales Trend Analysis	Real time visualization, interactivity during operation.	No visible weak points.
	Growth rate analysis	Vegalight custom visualization with more visualization options and flexibility.	Coding knowledge is required and less accessible for non-technical users.
Power BI	Clustering	Highly interactive UI, Scope for customization, Wide range of transformation options.	Complexity in dataset preprocessing, limited statistical feature.
	Sales Trend Analysis	Customizable design during visualization	No visible weak point.
	Growth rate analysis	Flexible in Power Query editor for statistical formulation.	Required preprocessed data and no native support for direct real time visualization.

This comparison chart provides overall information on understanding each tool's alignment with different analytical task. This experiment reveals level of complexity and diversity across the task. In terms of JMP, statistical analytical tasks were executed without prior coding knowledge and minimal effort due to its integrated formula editor. While, Kibana offers a wide range of customizability with more technically demanding approaches, including proficiency in scripting language JSON and understanding of Vega visualization grammar. Concurrently, Power BI required considerable amount of effort during preprocessing stage, utilizing DAX formulas and Power Query transformations. However, once preprocessing was finalized, the following analysis and visualization process were conducted without prior problems and with high interactivity.

Based on my evaluation criteria, I scored JMP, Power BI and Kibana below mentioned table in five different categories.

Critical Success Factor	Weight	JMP	Power BI	Kibana
Ease of Use	0.22	(0.22*5)	(0.22*3)	(0.22*2)
Interactivity	0.20	(0.20*4)	(0.20*5)	(0.20*3)
Customizability	0.19	(0.19*3)	(0.19*5)	(0.19*5)
Cost	0.20	(0.20*1)	(0.20*1)	(0.20*5)
Third-party Plugin	0.19	(0.19*2)	(0.19*4)	(0.19*5)
Total	1	3.05	3.57	3.94

Table 3: Competitor profile matrix with experimental score.

To understand the CPM table, I developed a radar chart to visually inspect the scoring system. The scores were derived from my experimental insights and weights were collected from industry experts and researchers in different domains.

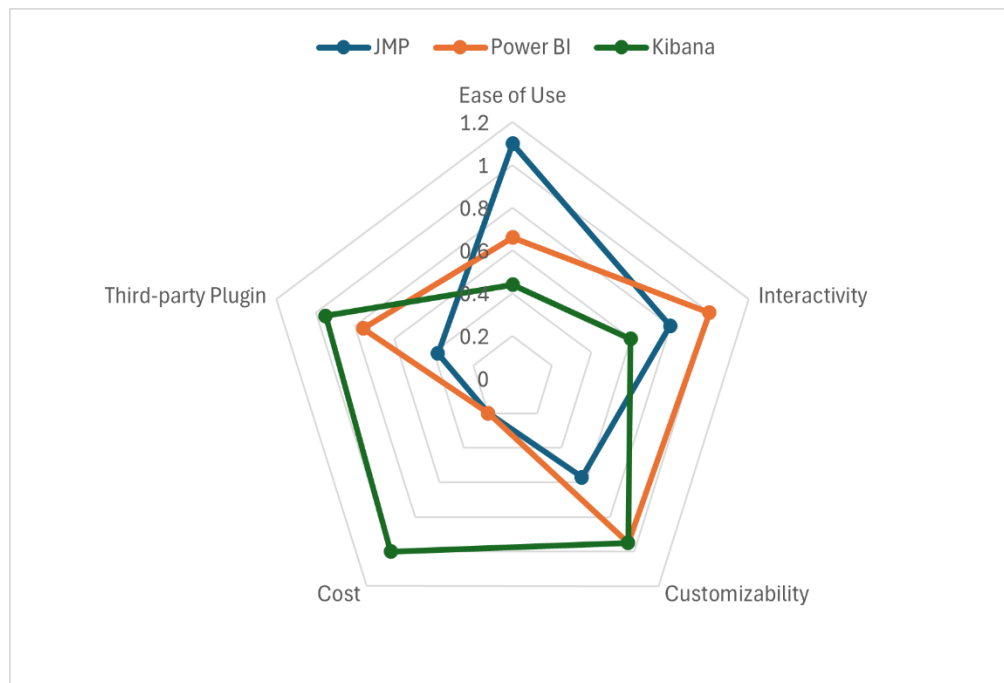


Figure 22.CPM table visualization as a Radar Chart.

This chart illustrates that JMP scored highest in Ease of Use (1.1), indicating its strong ground toward statistically driven analysis and simplified workflows. Additionally, this tool scored

less in cost (0.2) due to high cost of pricing models. Similarly, Kibana demonstrated its stronghold in Cost (1.0) and third-party plugin support (0.95) due to its open-source platform and integration of Elastic Stack. However, it underperforms in Ease of Use (0.44) due to its requirements of programming language and Linux OS experience from user. Meanwhile, Power BI demonstrated its capabilities in interactivity (1.0) and Customizability (0.95) which support dynamic dashboard and business centric report development.

Overall, Kibana is the winner with a score of 3.94 out of 5 and Power Bi is the runner up in this comparison. Here, costing is the main determining factor between the Kibana and Power BI which made the difference during the comparison.

6. Conclusion

The thesis aims to evaluate three Business Intelligence Software -Kibana, JMP, and Power BI. Depending on the requirements of the user, this presented pipeline will help to identify the most suitable tools for their specific data analysis. This study has been conducted structurally and systematically through critical evaluation criteria including usability, visualization capabilities, cost, and third-party plugin integration. This work also demonstrates that the decision to choose software vastly depends on the dataset. The capabilities of each software may drastically change based on the types of datasets. However, my thesis explores investigated software properly from end user perspective under a simulated experiment. Apart from this the visualization capabilities depend mainly on dataset which can be boosted by different types of plug-ins. Here, Kibana excels all requirements due to its accessibility of third-party plug-ins and visualization libraries.

Additionally, the empirical findings of this thesis indicate that open-source software with highly customized features, plug-in-based tools like Kibana, distinguished itself from other tools. This tool values itself by showcasing real-time analytics capabilities, cost-effectiveness, and strong integration features powered by Elasticsearch.

6.1 Limitations and Recommendations

While I covered some aspects of the comparison pipeline, there are some scopes for future research. To start with, while creating a CPM chart, using a small sample size as a determining factor may lead to some bias in result. Secondly, solely focusing on interviewing specialized within the EU region may misrepresent the choice of Third World countries expert demands for Business intelligence software features.

Apart from this, the usage of a limited dataset does not represent the complex real world business setup where a user may need to do a large range of analysis where intuitiveness and effectiveness is important in decision making landscape. Finally, all the software I used is consistently being updated with features, plug-ins and changes to the user interface. Because of this, in the future the usability and functionality may vastly change.

The limitation of my work includes the participant pool size for evaluation, which will improve the overall result while expanding to a more diverse and larger dataset.

Another limitation is the timeframe for the study that has been developed in a short period of time. More interviews could have been conducted and based on that open-source and paid Business Intelligence tools could have been categorized and compared. These changes will further narrow down the research questions.

6.2 Future works

In the case of AI-driven solutions, automated recommendations can be developed where a user can interact with an AI agent using natural language to provide their needs and dataset structure to the AI agent, where the AI agent investigates and simulates user experience and provides decisions with detailed explanations to the user.

To reduce user bias in scoring, my future work may include standardized evaluation tools such as the Technology Acceptance Model (TAM) (Aljarrah et al., 2016) and ISO 9241-11 metrics to improve the pipeline (ISO, 2018).

My thesis only focused on a single dataset. In case of future work, I will test across multiple datasets for example, financial data, healthcare data, geospatial data, etc. These changes will definitely improve the scope and usability of the software comparison.

In the future, to develop robust workflow management, which will track every navigation step and user hesitation time and based on that, a new pipeline can be created. Also, I can analyze user behavior quantitatively by collecting these data and creating clickstreams and heatmaps.

Bibliography

B., & Sall, J. (2011). JMP statistical discovery software. *WIREs Computational Statistics*, 3(3), 188–194. <https://doi.org/10.1002/wics.162>

JMP User Community. (2018, August 1). *How to use and automate Internet Open in JMP*. <https://community.jmp.com/t5/JMPPer-Cable/How-to-use-and-automate-Internet-Open-in-JMP/ba-p/63906>

Intellify Solutions. (n.d.). *Top 5 data analytics challenges*. <https://intellifysolutions.com/blog/top-5-data-analytics-challenges/>

Gartner. (2018, December 6). *Gartner data shows 87 percent of organizations have low BI and analytics maturity*. <https://www.gartner.com/en/newsroom/press-releases/2018-12-06-gartner-data-shows-87-percent-of-organizations-have-low-bi-and-analytics-maturity>

Delen, D., & Ram, S. (2018). Research challenges and opportunities in business analytics. *Journal of Business Analytics*, 1(1), 2–12. <https://doi.org/10.1080/2573234x.2018.1507324>

Fahl, J. (2017). *Data analytics: A practical guide to data analytics for business, beginner to expert (Data analytics, prescriptive analytics, statistics, big data, intelligence, master data, data science, data mining)*. CreateSpace Independent Publishing Platform.

HEAVY.AI. (n.d.). *What is statistical modeling? Definition and FAQs*. <https://www.heavy.ai/technical-glossary/statistical-modeling>

IBM. (n.d.). *Automation scripts*. IBM Documentation. <https://www.ibm.com/docs/en/control-desk/7.6.1.2?topic=tasks-automation-scripts>

JMP. (n.d.). *What JSL can do for you*. JMP Help. <https://www.jmp.com/support/help/en/17.0/#page/jmp/what-jsl-can-do-for-you.shtml#ww204748>

IBM. (n.d.). *What is data visualization?* <https://www.ibm.com/topics/data-visualization#:~:text=the%20next%20step-,What%20is%20data%20visualization%3F,that%20is%20easy%20to%20understand>

Tableau. (n.d.). *What is a dashboard? A complete overview*. Retrieved April 16, 2024, from <https://www.tableau.com/learn/articles/dashboards/what-is>

JMP. (n.d.). *Graph Builder examples*. JMP Help. Retrieved May 3, 2025, from <https://www.jmp.com/support/help/en/18.1/index.shtml#page/jmp/graph-builder-examples.shtml>

Qlik. (n.d.). *Predictive modeling: What it is & why it matters*. <https://www.qlik.com/us/predictive-analytics/predictive-modeling>

Jung, L. H. (n.d.). *Part 1: Intro to Elasticsearch and Kibana* [GitHub repository]. GitHub. <https://github.com/LisaHJung/Part-1-Intro-to-Elasticsearch-and-Kibana>

Machiraju, S., & Gaurav, S. (2018). *Power BI data analysis and visualization*. De Gruyter.

Kyanyoga. (n.d.). *Sample sales data* [Data set]. Kaggle. <https://www.kaggle.com/datasets/kyanyoga/sample-sales-data/data>

Everett, R. F. (2008). *The Entrepreneur's Guide to Marketing*. United Kingdom: ABC-CLIO.

Abdul Rahman, A., Abu Hasan, N., & A. Lahad, N. (2016). ISSUES AND CHALLENGES IN BUSINESS INTELLIGENCE CASE STUDIES. *Jurnal Teknologi*, 78(8–2). <https://doi.org/10.11113/jt.v78.9554>

Al-Okaily, A., Teoh, A. P., & Al-Okaily, M. (2023). Evaluation of data analytics-oriented business intelligence technology effectiveness: An enterprise-level analysis. *Business Process Management Journal*, 29(3), 777–800. <https://doi.org/10.1108/BPMJ-10-2022-0546>

Bhargava, M. G. (2018). *Analysis and Design of Visualization of Educational Institution Database using Power BI Tool*.

- Biju, S. M., & Mathew, A. (2017). Comparative Analysis of Big Data Analytics Software in Assessing Sample Data. *Journal of International Technology and Information Management*, 26(2), 2–22. <https://doi.org/10.58729/1941-6679.1301>
- DeLone, W. H., & McLean, E. R. (2016). Information Systems Success Measurement. *Foundations and Trends® in Information Systems*, 2(1), 1–116. <https://doi.org/10.1561/29000000005>
- Jones, B., & Sall, J. (2011). JMP statistical discovery software. *WIREs Computational Statistics*, 3(3), 188–194. <https://doi.org/10.1002/wics.162>
- Khalajzadeh, H., Abdelrazek, M., Grundy, J., Hosking, J., & He, Q. (2022). Survey and Analysis of Current End-User Data Analytics Tool Support. *IEEE Transactions on Big Data*, 8(1), 152–165. <https://doi.org/10.1109/TBDDATA.2019.2921774>
- Liu, Y., Han, H., & DeBello, J. E. (n.d.). *The Challenges of Business Analytics: Successes and Failures*.
- Paczkowski, W. R. (2021). *Business Analytics: Data Science for Business Problems*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-87023-2>
- Rajeswari, C., Basu, D., & Maurya, N. (2017). Comparative Study of Big data Analytics Tools: R and Tableau. *IOP Conference Series: Materials Science and Engineering*, 263, 042052. <https://doi.org/10.1088/1757-899X/263/4/042052>
- Sharma, K., Shetty, A., Jain, A., & Dhanare, R. K. (2021). A Comparative Analysis on Various Business Intelligence (BI), Data Science and Data Analytics Tools. *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 1–11. <https://doi.org/10.1109/ICCCI50826.2021.9402226>

- Somani, A. K., & Deka, G. C. (Eds.). (2017). *Big Data Analytics: Tools and Technology for Effective Planning* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b21822>
- Sun, Y., Shi, Y., & Zhang, Z. (2019). Finance Big Data: Management, Analysis, and Applications. *International Journal of Electronic Commerce*, 23(1), 9–11.
<https://doi.org/10.1080/10864415.2018.1512270>
- Vidhya, S., Sarumathi, S., & Shanthi, N. (2014). *Comparative Analysis of Diverse Collection of Big Data Analytics Tools*. 8(9).
- Vijayaraj, J., Saravanan, R., Victor Paul, P., & Raju, R. (2016). A comprehensive survey on big data analytics tools. *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, 1–6. <https://doi.org/10.1109/GET.2016.7916733>
- Vega. (n.d.). *Vega: A visualization grammar*. <https://vega.github.io/vega/>
- International Organization for Standardization. (2018). *ISO 9241-11:2018 – Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>
- Aljarrah, E., Elrehail, H., & Aababneh, B. (2016). E-voting in Jordan: Assessing readiness and developing a system. *Computers in Human Behavior*, Advance online publication. <https://doi.org/10.1016/j.chb.2016.05.078>

Appendix

(Semi-structured interview)

1. What kind of software are you using for data analysis?
2. In your experience, which feature do you like most about this software?
3. Can you describe any specific difficulties you've faced while using this software?
4. How intuitive do you find the software when creating data visualizations? Please explain your experience.
5. Does this software easily support collaboration within your team? If so, how?
6. How effectively does the software handle large datasets and complex analyses?

Questionnaire for determining factors of software: (Structured Interview or survey)?

What is your profession?

- a) Students / Researcher.
- b) Professional person (Industry)
- c) Others

Below mentioned factors consisting of 5-point Likert scale.

Ease of use	1	2	3	4	5
Interactivity	1	2	3	4	5
Customizability	1	2	3	4	5
Cost	1	2	3	4	5
Third party plugin	1	2	3	4	5

1. Kibana Code:

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "description": "Sales Growth Rate Analysis",
  "title": "Year-over-Year Sales Growth Rate",
  "data": {
    "url": {
      "%context%": true,
      "index": "salesforgrowth",
      "body": {
        "size": 0,
        "aggs": {
          "by_year": {
            "terms": {
              "field": "YEAR_ID",
              "order": { "_key": "asc" },
              "size": 100
            },
            "aggs": {
              "total_sales": { "sum": { "field": "SALES" } }
            }
          }
        }
      }
    },
    "format": { "property": "aggregations.by_year.buckets" }
  },
  "transform": [
    { "calculate": "datum.key", "as": "year" },
    { "calculate": "datum.total_sales.value", "as": "sales" },
    {
      "window": [
        { "op": "lag", "field": "sales", "as": "prev_sales" },
```

```

    {"op": "lag", "field": "year", "as": "prev_year"}
  ],
  "sort": [{"field": "year"}]
},
{
  "calculate": "(datum.prev_sales != null && datum.prev_sales > 0) ? ((datum.sales - datum.prev_sales) / datum.prev_sales * 100) : 0",
  "as": "growth_rate"
},
{"filter": "datum.prev_sales != null"}
],
"mark": {"type": "bar"},
"encoding": {
  "x": {"field": "year", "type": "nominal", "title": "Year"},
  "y": {"field": "growth_rate", "type": "quantitative", "title": "Growth Rate (%)", "axis": {"grid": true}},
  "color": {
    "condition": {"test": "datum.growth_rate < 0", "value": "red"},
    "value": "green"
  },
  "tooltip": [
    {"field": "year", "title": "Year"},
    {"field": "prev_year", "title": "Previous Year"},
    {"field": "sales", "title": "Sales", "format": ",.2f"},
    {"field": "prev_sales", "title": "Previous Year Sales", "format": ",.2f"},
    {"field": "growth_rate", "title": "Growth Rate (%)", "format": ".2f"}
  ]
}
}

```

2. JMP Code:

```

If(
  Row() == 1, 0,
  ((:Total_Sales - Lag(:Total_Sales, 1)) / Lag(:Total_Sales, 1)) * 100
)

```