STUDENT SESSION

UTILIZING BUSINESS ANALYTICS, BIG DATA, AND VISUALIZATION FOR SALES PERFORMANCE OPTIMIZATION

Aleksandar Radivojević*,
[0009-0008-0573-2362]
Miloš Mravik,
[0000-0001-5442-3998]
Marko Šarac
[0000-0001-8241-2778]

Singidunum University,
Belgrade, Serbia

Abstract:

This research paper examines the role of business analytics and reporting in the sales sector, utilizing big data. The main goal is to analyse sales data, identify key insights, and present the results through visualization using Tableau.

The introductory section discusses the research subject, objectives, and challenges that may arise when applying business analytics in sales. Types of content explores the technologies used for handling big data, which are crucial for the efficient storage, processing, and analysis of large datasets. These technologies include distributed systems like PIG, data warehouses such as Red Lake, and tools for data analysis and integration, including SQL Server, SSIS, and SSAS. The section also highlights the role of data visualization tools like Tableau and Power BI in presenting key insights and supporting business decision-making. The methodology describes the steps for collecting and analysing sales data, along with an overview of data visualization tools such as Tableau. A dedicated section provides a detailed description of the database, including its structure, and data types.

The data analysis covers various types of sales performance, customer, and product analyses. Data visualization and interpretation are presented through Tableau, focusing on result interpretation and their application in business decision-making. The thesis concludes with a summary of key findings.

Keywords:

Business Analytics, Business Reporting, Tableau, Aata Visualization, Distributed Systems, Big Data.

INTRODUCTION

The subject of this research is the application of business analytics and reporting, working with big data, all within the context of the sales sector. The focus is on the analysis and visualization of sales data and improving the presentation of the final solution in the form of data products. Today, business analytics is achieved through two main groups of tools: business intelligence tools and big data analytics tools. Business Intelligence (BI) utilizes techniques such as data mining, data visualization, and specialized tools to enable organizations to make informed decisions. Big Data analytics tools allow organizations to collect, store, process, and analyse large amounts of data to gain valuable insights. These tools help extract valuable information from complex and unprocessed data.

Correspondence:
Aleksandar Radivojević

e-mail:
aleksandar.radivojevic.20@singimail.rs

DOI: 10.15308/Sinteza-2025-503-510

The aim is to explore and demonstrate how business analytics and reporting can significantly impact the analysis of sales data and improve sales strategies. Specifically, it will focus on the use of data visualization tools such as Tableau. It includes identifying key patterns and insights from the data that can contribute to a better understanding of sales trends, support informed decision-making, and enhance sales strategies.

2. TYPES OF CONTENT

2.1. BUSINESS ANALYTICS

Business analytics is the practice of analysing and interpreting data to gain insights into business performance and identify areas for improvement. This process can be enhanced with the use of various techniques and tools, such as predictive analytics, machine learning, and natural language processing, which empower analysts and business users to glean insights from structured and unstructured data sources. [1]

The importance of business analytics lies in the fact that it is crucial for every organization to have informed decisions to improve performance. When businesses have insights into their data, they can automatically optimize their processes to reduce costs, enhance customer experiences, identify new growth opportunities, and even predict future trends and anticipate customer needs.

2.2. BUSINESS REPORTING

Business reporting refers to the practice of gathering, analysing, and presenting data related to organizational activity and performance. In today's business world, where competition is rampant, the role of business reporting cannot be overemphasized. It allows for informed decision-making by providing accurate information that is critical to the success of any corporation.

Through reporting, companies can easily track their performance, identify trends, and make data-driven decisions that can impact organization-wide operations. It is essential, therefore, that organizations allocate sufficient resources to this process to ensure that they have access to real-time insights that enable them to stay competitive in the dynamic business environment. [2]

2.3. BIG DATA AND DISTRIBUTED SYSTEMS

With the rapid growth of Internet users, data generation has increased exponentially. This data comes from millions of messages sent and received via platforms like WhatsApp, Facebook, and Twitter, as well as trillions of photos taken and countless hours of videos uploaded to YouTube every minute. According to a recent survey, 2.5 quintillion (2,500,000,000,000,000,000 or 2.5×10^{18}) bytes of data are generated daily. This massive volume of data is referred to as "big data." However, big data is not just about size—it is a broad term for data that is vast, complex, and can be structured or unstructured, arriving at high velocity. Notably, 80% of today's data has been generated in just the past few years. The continuous expansion of big data is driven by the ever-increasing volume of produced information and the growing need to capture and analyse it. [3]

The defining characteristics of Big Data are commonly represented by the 5V model:

1. Volume – the vast amount of data.
2. Variety – the diverse types of data.
3. Velocity – the speed at which data is generated and processed.
4. Veracity – the trustworthiness and quality of data.
5. Value – the usefulness of the data for decision-making and insights.

Data Warehouses: Red Lake is a data warehouse (Data Warehouse) used for processing and analysing large volumes of data (Big Data). Although it is not as widely known as traditional solutions like Snowflake, Google Big Query, or Amazon Redshift, Red Lake is increasingly utilized in DataOps and analytical environments where fast data storage and real-time data processing are required. It is designed to handle the challenges of Big Data by providing a platform that can scale quickly and support high-speed processing, essential for modern data-driven applications.

Distributed Systems: Apache Pig is a distributed system for processing large datasets, used for data analysis and transformation within the Hadoop ecosystem. Developed by Yahoo!, Pig enables users to write complex data processing tasks using a simple scripting language called Pig Latin. This language is designed to abstract the complexity of traditional MapReduce programming, making it easier for developers to process large datasets. How Apache Pig Works:

Data Analysis and Integration Tools: SQL Server, SSIS (SQL Server Integration Services), and SSAS (SQL Server Analysis Services) are tools that enable data integration, analysis, and transformation. SQL Server – A relational database developed by Microsoft. It enables data storage, management, and querying using the T-SQL language. SSIS (SQL Server Integration Services) – A tool for ETL processes (Extract, Transform, Load). It is used for data integration, cleaning, and migration from various sources. SSAS (SQL Server Analysis Services) – A service for analytical data processing. It enables the construction of OLAP cubes and data models, which facilitates advanced analytics and BI reporting.

Visualization Tools: Tableau and Power BI are the most popular data visualization tools that enable users to create visual representations of data easily. These tools simplify the analysis of large datasets and allow for informed decision-making through interactive visualizations. Tableau – A data visualization tool that enables interactive and intuitive charts, dashboards, and analytics. It is suitable for data exploration and storytelling. Power BI – A Microsoft BI tool for reporting and analytics, allowing connection to various data sources, creation of interactive dashboards, application of DAX expressions, and integration with SQL Server and other Microsoft tools. [4]

2.4. STRUCTURE OF DATABASE

The database used for analysis and reporting is orders_frostonline.xlsx. It uses the .xlsx format, which stands for Excel Spreadsheet in XML format and was first introduced in Excel 2007 as a replacement for the older .xls format. .xlsx files are compressed ZIP archives containing multiple XML files and other types of data organized in a structure that represents worksheets, tables, charts, formulas, and other Excel functionalities.

It contains eight tables: Categories, Customers, Employees, Orders, Orders details (details of each individual order), Products, Shippers, and Suppliers. These tables are interconnected through primary keys. When we want to link two or more tables in databases, they need to have a key, meaning a column or set of columns that will allow the tables to be connected. The primary key is a unique identifier within a table. Each row in that table has a unique value in the column representing the primary key.

Table 1 shows information about Table Categories:

Table 1. Table Categories

CategoryID	CategoryName	DescriptionText
1	Beverages	Soft drinks, coffees, teas, beers, and ales
2	Condiments	Sweet and savoury sauces, relishes, spreads, and seasoning
3	Confections	Desserts, candies, and sweet breads
4	Dairy Products	Cheeses

The table below, Table 2, shows information regarding Table Customers:

Table 2. Table Categories

Customer ID	Customer-Name	ContactName	Adress	City	PostalCode	Country
1	Alfreds Futterkisste	Maria Anders	Obere Str. 57	Berlin	12209	Germany
2	Around the Horn	Thomas Hardy	120 Hanover Sq.	London	WA1 1DP	UK
3	Berglunds snabbkop	Christina Berglund	Bergusvagen 8	Lulea	S-958 22	Sweden

Table 3 is showing information regarding Table Employees:

Table 3. Table Employees

EmployeeID	LastName	FirstName	BirthDate	Photo	Notes
1	Davolio	Nancy	1968-12-08	EmplD1.pic	Education includes…
2	Fuller	Andrew	1952-02-19	EmplD2.pic	Andrew received…
3	Leverling	Janet	1963-08-30	EmplD3.pic	Janet has…

In Table 4, we can see information about Table Orders:

Table 4. Table Orders

OrderID	CustomerID	EmployeeID	OrderDate	ShipperID
10248	90	5	1996-07-04	3
10249	81	6	1996-08-04	1
10250	34	4	1996-09-08	2
10251	84	3	1996-10-08	1

Table 5 is showing information regarding Table OrdersDetails:

Table 5. Table OrdersDetails

OrderDetail ID	OrderID	ProductID	Quantity
1	10248	11	12
2	10248	42	10
3	10248	72	5
4	10249	14	9

In Table 6, we can see information about Table Products:

Table 6. Table Products

ProductID	ProductName	SupplierID	CategoryID	Unit	Price
1	Chais	1	1	10 boxes x 20 bags	18
2	Chang	1	1	24 – 12 oz bottles	19
3	Aniseed Syrup	1	2	12 – 550 ml bottles	10
4	Chef Antons Cajun Seasoning	2	2	48 – 6 oz jars	22

Table 7 and Table 8 are showing information regarding Table Shippers and Table Suppliers:

Table 7. Table Shippers

ShipperID	ShipperName	Phone
1	Speedy Excpress	(503) 555-9831
2	United Package	(503) 555-3199
3	Federal Shipping	(503) 555-9931

Table 8. Table Suppliers

SupplierID	SupplierName	ContactName	Adress	City	Postal-Code	Country	Phone
1	Exotic Liquid	Charlotte Cooper	49 Gilber St.	London	EC1 4SD	UK	(171) 555-2222
2	New Orleans Cajun Delights	Shelley Burke	P.O. Box 78934	New Orleans	70117	USA	(100) 555-4822
3	Grandma Kellys Homestead	Regina Murphy	707 Exford Rd.	Ann Arbor	48104	USA	(313) 555-5735

2.5. WORK IN TABLEAU

In this research, the connection and integration of sales-related datasets enable comprehensive analysis and visualization. By utilizing Tableau, the data from multiple tables is combined, processed, and presented through interactive dashboards. This tool allows for seamless data exploration, facilitating the identification of key sales trends, performance metrics, and correlations. The ability to create dynamic visual representations enhances data-driven decision-making, making Tableau an essential component of modern business analytics and reporting. [5]

2.5.1. The physical layer of connection

For everything to function properly, each of these tables must be physically connected through common keys via inner joins. The Customers table is connected to the Orders table through the CustomerID field, which is present in both tables; the Orders table is connected to the Employees table through EmployeeID; to the Order Details table via OrderID; and to the Shippers table through ShipperID. The Order Details table is connected to the Products table via the ProductID field, while the Products table is connected to the Categories and Suppliers tables through the CategoryID and SupplierID keys, respectively.

2.5.2. Dashboards

By combining dimensions and measures in Tableau, you can create meaningful visualizations:

- Rows often contain “Measures”, displaying quantities (e.g., total sales).
- Columns often contain “Dimensions”, categorizing data (e.g., months or regions).

For example, if you want to see sales by month, the Month (Dimension) would be placed in Columns, while Sales (Measure) would be in Rows. This creates a chart where the X-axis represents months, and the Y-axis represents sales figures. Some measures in Tableau are automatically generated, while others need to be created using SQL queries, depending on what you want to display. Figure 1 shows a chart named "Customers in Country", representing the total number of customers in each country. Visually, it is displayed as a world map, where each country with customers is shaded in dark brown or lighter brown, depending on the number of customers. Countries without customers are shown in grey colour. When hovering over a country shaded in brown, a small tooltip appears, displaying the country's name and the total number of customers from that country.

Figure 2 shows the table “Customer Spent” that displays key information about customers related to their spending. It includes the following details: Customer ID, Full name, Residential address, and total amount spent by the customer. Different colours in the Sum (Price) column visually highlight customer spending – a darker colour indicates a higher spending amount. This visualization allows for the quick identification of high-spending customers, making it easier to target marketing and sales efforts more effectively.

In Figures 3 and 4, different data analysis techniques are illustrated. Figure 3 presents forecasting, which is useful for predicting trends in sales, revenue, inventory, weather conditions, and other areas where planning ahead is essential. In this study, forecasting was used to predict the number of customers from 1999 to 2002 based on data from 1997 to 1999. This means that a longer historical period is used to make predictions for an upcoming period. The curve illustrates the customer trend, showing an initial decline, followed by a stable phase, and then another decline in the following years.

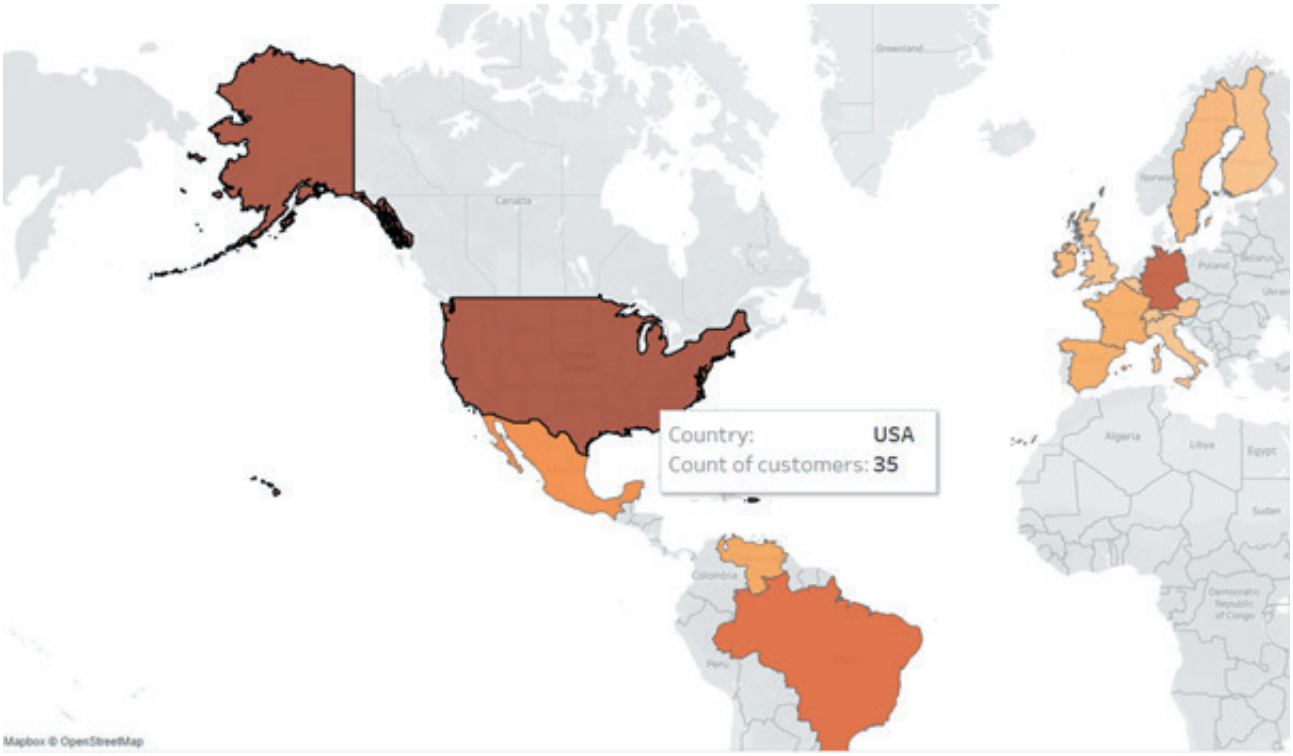

Figure 1. Graph – Customers in Country

Cust..	Custome..	Address
2	Ana Truji..	Avda. de ..
5	Berglund..	Berguvs..
7	Blondel p..	24 place ..
11	Bs Bever..	Fauntler..
13	Centro c..	Sierras d..
14	Chop-sue..	Hauptstr..
15	Comérci..	Av. dos L..
18	Du mond..	67 rue de..
20	Ernst Ha..	Kirchgas..
24	Folk och ..	Åkergata..
25	Frankenv..	Berliner ..
30	Godos Co..	C/ Romer..
33	GROSELL..	5ª Ave. L..
34	Hanari C..	Rua do P..
35	HILARIÓ..	Carrera 2..
37	Hungry ..	8 Johnst..
38	Island Tr..	Garden H..
39	Königlich..	Maubels..
44	Lehmann..	Magazin..
46	LILA-Sup..	Carrera 5..
48	Lonesom..	89 Chiar..
49	Magazzi..	Via Ludo..
52	Morgens..	Heerstr. ..
55	Old Worl..	2743 Ber..
58	Pericles ..	Calle Dr. ..
61	Que Delí..	Rua da P..
63	QUICK-St..	Tauchers..
65	Rattlesn..	2817 Mil..
66	Reggiani ..	Strada P..
67	Ricardo ..	Av. Copa..
68	Richter S..	Grenzac..
69	Romero ..	Gran Vía 1
71	Save-a-lo..	187 Suff..
75	Split Rail..	P.O. Box ..

SUM(Price)
2,5 515,9

Figure 2. Table – Customer spent

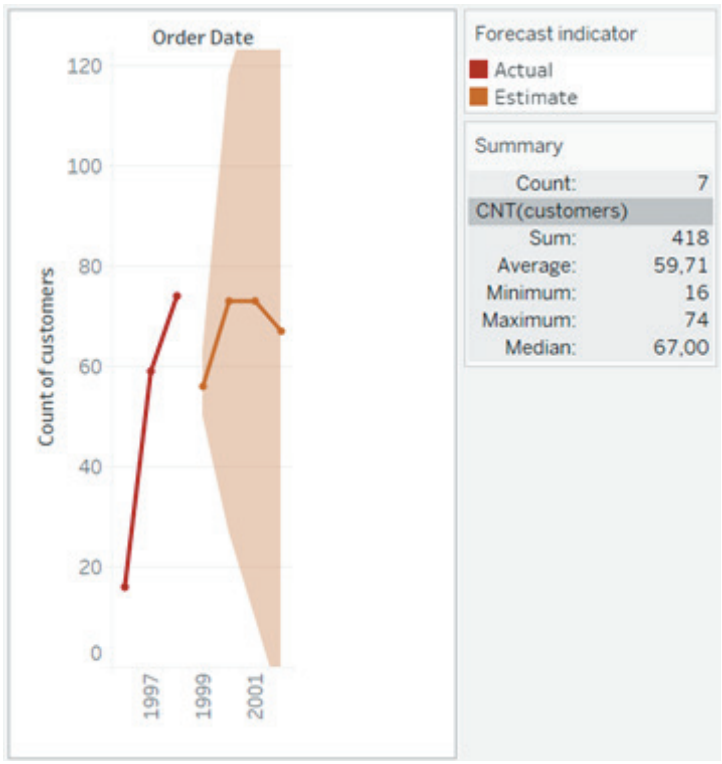

Figure 3. Forecasting

Figure 4. Clustering

On the other hand, Figure 4 demonstrates clustering, which is valuable for customer segmentation, identifying patterns in large datasets, and creating targeted marketing strategies. In this case, clustering is divided into 15 small clusters, where each cluster represents a single country. Essentially, the data is grouped by country, with each country forming a separate unit. When hovering over a circle, a tooltip appears providing clear insights into total profit, number of products sold, and country name. This approach significantly helps users analyse key metrics by specific groups, making it easier to interpret. [6]

3. CONCLUSION

As seen in the provided visualizations, data presented in a visual format is much easier to interpret compared to raw Excel tables. This approach enables faster and more intuitive insights, making it easier to spot trends, patterns, and anomalies. By transforming raw data into interactive dashboards, businesses can make more informed decisions.

Working with SSAS (SQL Server Analysis Services) and SSIS (SQL Server Integration Services) allows for more advanced data management. SSAS helps create multidimensional models, optimizing data analysis and query performance, while SSIS automates data integration (ETL), ensuring data is processed and ready for analysis at scheduled times, without manual intervention. These tools—SQL Server, SSIS, SSAS, and visualization platforms like Tableau and Power BI—work together as an integrated system, improving data processing efficiency and enabling actionable data insights.

For example, data collected from a data warehouse can be processed using distributed systems like Apache Pig or Spark, cleaned, and stored in an accessible format like Excel or an SQL database. SQL is then used for queries, and SSIS and SSAS integrate and model the data for optimized analysis. Finally, Tableau or Power BI visualizes the processed data, making it easier for decision-makers to interpret and act upon. This process illustrates how these tools simplify big data management and enhance decision-making. Effective data visualization and reporting help businesses make data-driven decisions, improve customer relationships, and gain a competitive edge by identifying trends and optimizing resources.

Future research could explore integrating AI and machine learning for predictive analytics, which could lead to more accurate forecasting, personalized customer experiences, and automated decision-making processes, enhancing business intelligence in an evolving digital landscape. [7] [8]

REFERENCES

[1] "What is business analytics? The complete guide for analysts," [Online]. Available: https://www.thoughtspot.com/data-trends/business-analytics. [Accessed 04 03 2025].

[2] "What is Business Reporting?," [Online]. Available: https://www.jaspersoft.com/articles/what-is-business-reporting. [Accessed 04 03 2025].

[3] B. Balusamy, N. R. Abirami, S. Kadry and A. H. Gandomi, Big Data: Concepts, Technology and Architecture, Hoboken, NJ: John Wiley and Sons, Inc., 2021.

[4] D. Liu and V. Y. Yoon, "Developing a goal-driven data integration framework for effective data analytics," *Decision Support Systems,* vol. 180, p. 114197, 2024.

[5] A. Ambasht, "Real-Time Data Integration and Analytics: Empowering Data-Driven Decision Making," *International Journal of Computer Trends and Technology*, vol. 71, no. 7, pp. 8-14, 2023.

[6] A. Upadhye, "Enhancing business strategy with sales data visualization," *International Journal of Data Science Research and Development (IJDSRD)*, vol. 2, no. 1, pp. 27-36, 2023.

[7] M. Mravik, T. Vetriselvi, K. Venkatachalam, M. Sarac and N. Bacanin, "Diabetes Prediction Algorithm Using Recursive Ridge Regression L2," *Computers, Materials & Continua,* vol. 71, no. 1, pp. 457-471, 2022.

[8] M. Šarac, M. Mravik, J. Dijana, Š. Ivana, Ž. Miodrag and B. D. Nebojša, "Intelligent diagnosis of coronavirus with computed tomography images using a deep learning model," *Journal of Electronic Imaging*, vol. 32, no. 2, pp. 1-10, 2022.