2024 International Conference on Applied Mathematics & Computer Science (ICAMCS) | 979-8-3503-6380-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAMCS62774.2024.00027

A procedure with MySQL Python for developing data warehouse and analytics using data from a greek soft drinks company

Sotirios Tsakiridis
Dept. Computer, Informatics and Telecommunications Engineering
International Hellenic University
Serres, Greece
sotitsak@ihu.gr

Alkiviadis Tsimpiris
Dept. Computer, Informatics and Telecommunications Engineering
International Hellenic University
Serres, Greece
atsimpiris@ihu.gr

Athanasios Angeioplastis
Dept. Computer, Informatics and Telecommunications Engineering
International Hellenic University
Serres, Greece
aagiop@gmail.com

Nikolaos Papaioannou
Dept. Computer, Informatics and Telecommunications Engineering
International Hellenic University
Serres, Greece
npapaioannou@ihu.gr

Paris Mastorocostas
Dept. Informatics and Computer Engineering
University of West Attica
Egaleo, Greece
mast@uniwa.gr

Dimitrios Varsamis
Dept. Computer, Informatics and Telecommunications Engineering
International Hellenic University
Serres, Greece
dvarsam@ihu.gr

***Abstract*—The research focuses on the process of creating a data warehouse to meet the decision-making needs of a Greek beverage company. The data cover the period from 2018 to 2022. The developed data warehouse schema follows the star schema, with one fact table and five-dimension tables. Based on this structure, a web application was developed in PHP, which displays reports, charts, and analyses on the company's performance through interactive queries. Additionally, prediction functions tailored to the company's data were developed in Python. The process followed, the queries for creating the data warehouse, as well as the commands for creating analyses and predictions, serve as a useful guide for other companies that wish to upgrade the capabilities of their ERP systems.**

***Index Terms*—Data analysis, decision-making, MySQL, Data Warehouse, Business Intelligence**

I. Introduction

The progress of technology in recent years has resulted in tremendous advancements and improvements across various sectors, [1], [2], [3], [4], [5]. Industrial activities are using specialized and constantly changing technical equipment on a regular basis, [6]. The development and marketing departments benefit from these technologies, which also make it easier to execute industrial production and efficiently arrange necessary duties inside the operational framework. Furthermore, they facilitate the fastest and most efficient digital connections amongst departments inside an industrial unit as well as between units around the world, enhancing communication between branches worldwide.

In today's competitive business environment, making decisions based on accurate and timely information is critical to business success and sustainability. Data are essential to this process because it allows businesses to analyze market trends in-depth, understand consumer behavior, maximize operational effectiveness, and find new business prospects, [7]. Through data analysis, businesses can make informed decisions that mitigate risk, improve efficiency and increase profitability, [8]. Accurate and timely data enable businesses to quickly respond to market changes, maintain a competitive edge and make strategic decisions that support long-term growth and sustainability, [9].

However, handling large amounts of data presents enormous challenges. The enormous amounts of data that are collected daily require specific tools and techniques for effective analysis and interpretation, [10]. Data warehouses play a pivotal role in managing big data, [11], providing a comprehensive and centralized platform for storing, organizing, and managing vast amounts of data, [12]. These repositories make it possible to effectively combine, classify, and retrieve data from various sources, which makes it easier to conduct in-depth analysis and extract insightful information, [13]. Additionally, data warehouses enhance the performance of analytical processes and ensure the accuracy and reliability of data, empowering businesses to make more informed and strategic decisions, [14].

In summary, they allow businesses to aggregate data from multiple systems and organize them in a way that makes it easy to extract useful information. By using business intelligence techniques such as reporting, charting, and forecasting, companies can improve their efficiency and competitiveness. Beyond data collection and storage, data analysis is a key element of the process. Developing forecasting functions, enables companies to predict future trends and adjust their strategies accordingly, [15], [16]. Additionally, developing a

979-8-3503-6380-7/24/$31.00 ©2024 IEEE
DOI 10.1109/ICAMCS62774.2024.00027

web application, allows users to access real-time reports and graphs, thereby enhancing decision-making capabilities.

This study aims to address the decision-making needs of a Greek beverage company by developing a comprehensive data warehouse. The data cover the period from 2018 to 2022. The developed data warehouse schema follows the star schema, with one fact table and five-dimension tables. The study also includes the development of a web application in PHP, that provides interactive reports, charts, and performance analyses. Furthermore, prediction functions tailored to the company's specific data are implemented using Python. The aim of this study is not only to illustrate the process of creating and utilizing a data warehouse, but also to provide a useful guide for other companies that wish to upgrade the capabilities of their ERP systems.

The rest of the paper is structured as follows: Section II introduces the methodology employed. Section III presents the results and the discussion and finally, Section IV offers the conclusions.

II. Methodology

In this section, a proposed solution to the above problem is described.

The first step in developing the data warehouse was to collect and prepare the data from the company's existing ERP system. The data included sales transactions, inventory levels, customer information, and financial records spanning from 2018 to 2022. Numerous preparation procedures were carried out before the data were loaded into the data warehouse. Initially, in order to ensure data accuracy, data cleaning was conducted to identify and rectify inconsistencies and missing values, remove duplicate records, and correct erroneous entries. Subsequently, the data format was standardized through data transformation, which included converting data types, normalizing data values, and creating derived columns to record crucial metrics. Finally, data integration created a single coherent dataset by aligning and combining information from many sources.

To create the data warehouse and perform analyses, specific SQL queries and Python commands were employed. The data warehouse was designed using the star schema, which consists of a central fact table and five dimension tables. This schema optimizes query performance and simplifies reporting and analysis. Specifically, the cleaned and transformed data were loaded into the data warehouse using the ETL (Extract, Transform, Load) procedure. MySQL and Python scripts were used to automate this process. During the extract phase, SQL queries and Python connectors were used to retrieve data from source systems, while in the transform phase, the data were processed and restructured according to the schema design. Finally, the data were loaded into the corresponding tables in the data warehouse after being converted.

To offer users with an interface to interact with the data warehouse, a PHP web application was developed. This application includes features such as interactive reports, allowing users to generate and view detailed sales and financial performance reports. It is also includes dynamic charts, offering visual representations of data through various chart types like line graphs, bar charts, and pie charts, as well as performance analysis tools, which allow the evaluation of company's performance metrics and trend identification. Additionally, prediction functions tailored to the company's data were developed in Python.

III. Results ans Discussion

In this section the results are presented. As mentioned above, the primary focus of this research is to address the decision-making needs of a Greek beverage company by developing a comprehensive data warehouse, but also to provide a useful guide for other companies that wish to upgrade the capabilities of their ERP systems. The study also includes the development of a web application in PHP, that provides interactive reports, charts, and performance analyses. A thorough analysis derived from aggregated data is presented in the Results section, which follows our methodical data processing approach. The following graphic representations provide important performance indicators for the Greek soft drink company, providing insights on costs, profits, and production amounts from 2018 to 2022.

The subsequent visual depiction, showcased in Figure 1, illustrates the central menu interface of the application designed for the Company.

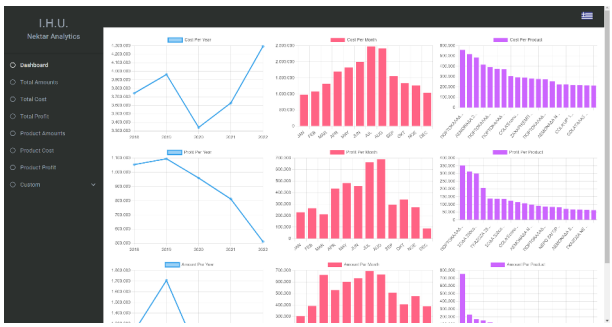

Fig. 1. The main menu of the application

The developed user interface offers an extensive and detailed navigation of various business data categories. Users can access a general overview, providing a broad summary of all available data, including product information, total business costs, and profit.

In addition to these general categories, the interface includes specialized categories tailored to individual products, such as specific data on Coke, soda, and other products. Users can view detailed information pertaining to each product category.

The interface additionally includes an advanced search feature that enables a thorough examination of product details over various time periods. Data on an annual, monthly, weekly, and even daily basis can be viewed by users. This feature makes it easier to perform a thorough temporal analysis,

enabling the examination of trends and performance metrics over various time frames.

Taking advantage of these capabilities, users can efficiently gather and analyse relevant information, leading to well-informed conclusions and strategic decisions.

Figure 2 presents a comprehensive monthly profit analysis. It is apparent from Figure 2 that total monthly sales reach their highest levels during the summer months, especially in July and August, with profits of approximately 665.000 and 689.000, respectively. In contrast, December records the lowest profits, amounting to 88.000. This pattern is likely due to the high temperatures in Greece during the summer and the substantial increase in tourism. This detailed analysis highlights the significance of seasonal trends in the company's profitability, offering a deeper understanding of monthly performance.

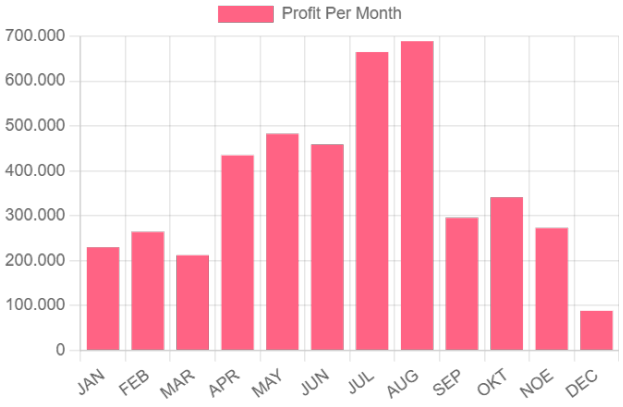

Fig. 2. Profit per month

The annual cost incurred by the company is shown in Figure 3. The data shows notable variations during the period of the analysis. The expenses increased from 3.700.000 in 2018 to a peak of almost 4.000.000 in 2019. This increase could be attributed to higher production volumes and increased operational expenses. But in 2020, expenses show a significant decline, coming down to almost 3.350.000. The following years exhibit a slow but steady recovery in expenses, rising to approximately 3.600.000 in 2021 and then sharply to 4.300.000 in 2022, presumably as a result of rising raw material costs and inflationary pressures.

Figure 4 illustrates the annual profits. After 2019, the profits have consistently decreased. Starting at 1.050.000 in 2018, the profits rose slightly to 1.100.000 in 2019. However, the profits decreased significantly in the following years, arriving below 1.000.000 in 2020, 800.000 in 2021, and 500.000 in 2022. The steady decline in profits indicates challenges in maintaining profitability, possibly due to increasing costs, competitive pressures, and market disruptions.

Figure 5, which depicts a drop in sales quantities in 2020 followed by a gradual increase, further supports these findings. This pattern runs counter to the reported decline in earnings

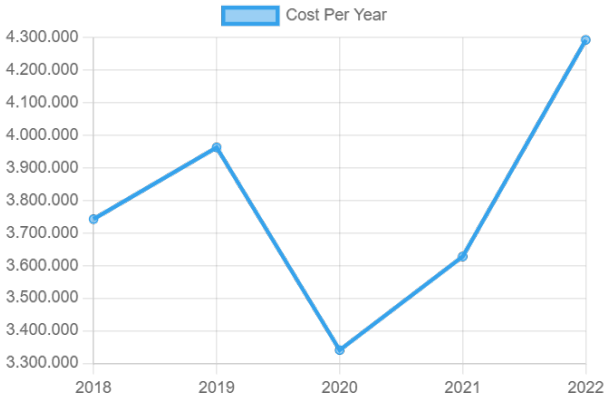

Fig. 3. Cost per year

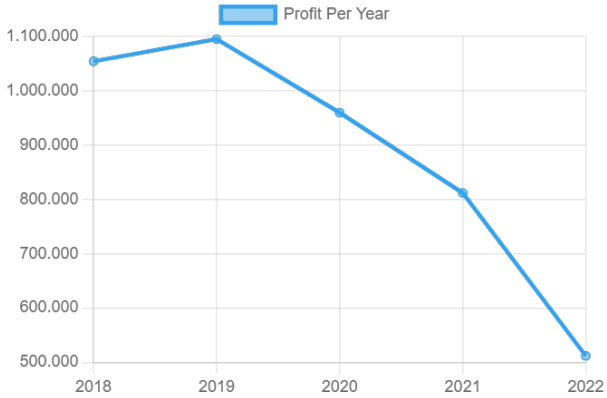

Fig. 4. Profit per year

during the same time frame. Specifically, the highest sales were recorded in 2019, with approximately 1.750.000 units. In 2020, however, sales clearly show a major decline, falling to approximately 1.000.000 units, followed by a gradual increase and ending at about 1.200.000 units in 2022.

It appears that valuable conclusions can be drawn regarding the company's overall performance, future trajectory, and potential changes. However, as previously mentioned, the application also allows for a detailed analysis of the company's individual products.

Figure 6 and Figure 7 illustrate the Costs and the profits per product respectively.

Numerous important insights are revealed by the cost and profit per product study. The "PORTOKALADA" variants and "LEMONADA" are among the most expensive to produce, with costs approximately 500.000. However, "PORTOKALADA" also yields some of the highest profits, reaching up to 350.000. This indicates that despite its high production costs, this product is an important source of income for the company.

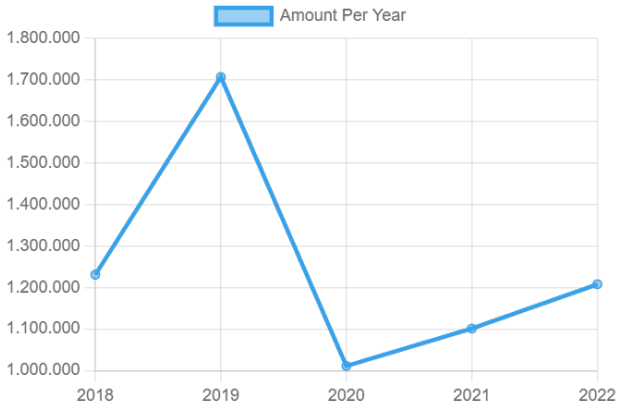

Fig. 5. Amount per year

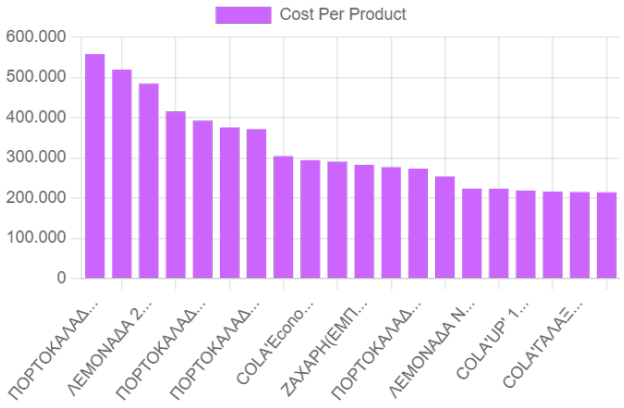

Fig. 6. Cost per product

On the other hand, certain goods have high manufacturing costs but low profit margins. For instance, while ”LEMONADA” incurs a high cost, it is not among the most profitable products. On the other hand, products such as ”SODA 250cc” and ”SODA 330cc” have relatively lower production costs, but generate high profits. This discrepancy suggests opportunities for cost management and profit optimization.

Products like ”PORTOKALADA” and ”COLA Econo” demonstrate a balance of high profitability and manageable production costs. These products serve as strategic assets for the business, pointing out areas in which resources should be concentrated to increase total profitability. Conversely, products that have high production costs but yield low profits, such as ”LEMONADA”, present opportunities for cost reduction strategies. By improving production efficiency or enhancing the market appeal of these products, the company could improve its profitability.

Additionally, products that yield high profits despite lower production costs, such as ”SODA 250cc”, should be considered for scaling opportunities. Increasing the production and market reach of these products could significantly enhance the company’s overall profit margins.

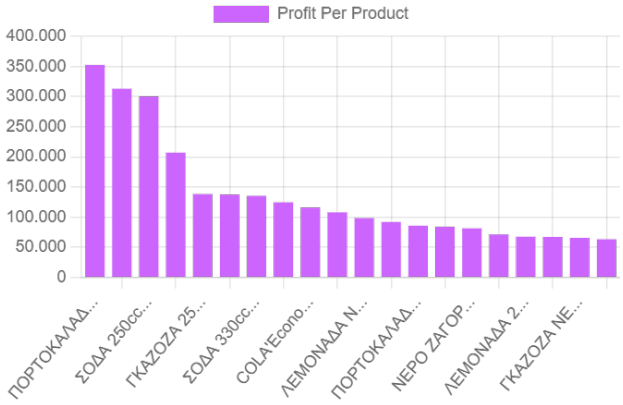

Fig. 7. Profit per product

In conclusion, the company’s product lineup shows varied performance in terms of cost and profitability. By focusing on high-profit products with manageable costs and addressing discrepancies where high costs do not align with high profits, the company can optimize its strategy for improved financial performance.

However, users can refine their search to a more specialized level, allowing them to monitor fluctuations in profits, costs, or quantities of specific products over any desired period, such as monthly, weekly, or even daily. Below two examples are presented. Figure 8 illustrates the monthly earnings of ”PORTOKALADA” in 2021 and Figure 9 shows the daily earnings of ”PORTOKALADA” in August 2022.

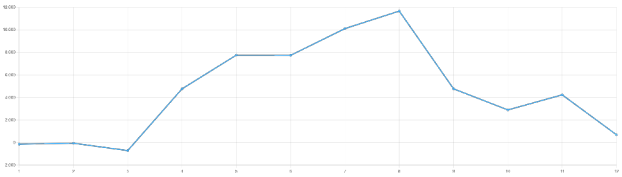

Fig. 8. Monthly earnings of ”PORTOKALADA” in 2021

Figure 8 depicts the ”PORTOKALADA”’s monthly earnings for the year 2021. The data indicates that the company experienced zero earnings or losses during the first three months. Subsequently, earnings gradually increased over the following months, peaking in the eight month, when earnings reached 12.000. However, this peak is followed by a sharp decline, with earnings falling even below 2.000 in the subsequent months.

Figure 9 depicts the daily profits of ”PORTOKALADA” for August 2022. The profits exhibit significant variability throughout the month. A notable peak is observed on the 24th day, where the profit is approximately 2.500. Conversely, several days, particularly towards the end of the month, showed a marked decline in earnings, with some days recording minimal or zero earnings.

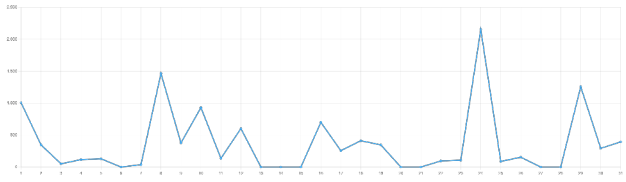

Fig. 9. Daily earnings of ”PORTOKALADA” in August 2022

These visual representations provide valuable insights into the temporal variations in ”PORTOKALADA” earnings, highlighting periods of high and low profitability. Such detailed analysis enables more informed decision-making regarding production and marketing strategies to optimize earnings.

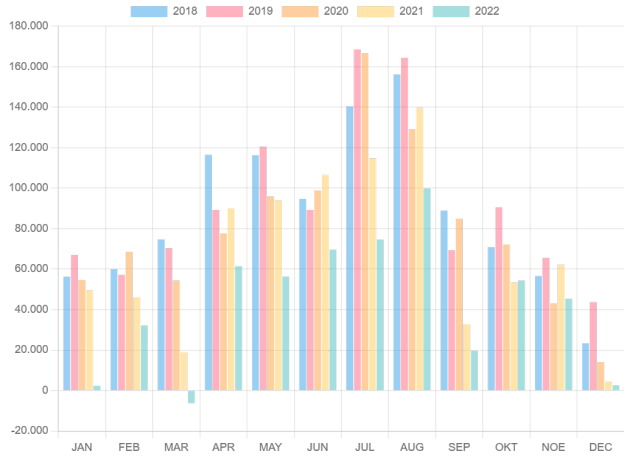

Fig. 10. Monthly earnings from 2018 to 2022

Additionally, users can compare costs, product volumes, and profits within combined charts. Figure 10 presents a comparative analysis of monthly earnings from 2018 to 2022. The chart demonstrates significant variability in earnings across different months and years. A clear seasonal trend is observed, with earnings generally peaking in the summer months, particularly in July and August. Throughout the observed years, earnings tend to be lower in the winter months, particularly in December and January, which aligns with expected seasonal consumption patterns. The comparison across different years reveals both growth and volatility, suggesting that while the company benefits from seasonal peaks, it also faces challenges that impact earnings consistency. The present analysis underscores the significance of strategic planning in mitigating the impact of seasonal fluctuations and maximizing profitability during periods of peak.

In summary, the analysis provides critical insights into the financial and operational performance of the Greek soft drinks company. The findings underscore the volatile character of expenses, a concerning reduction in earnings, and the seasonal fluctuations in sales volumes. These results are critical for strategic planning and decision-making that aims to maximize cost control and increase profitability.

IV. Conclusions

This study successfully developed a comprehensive data warehouse to address the decision-making needs of a Greek soft drinks company, covering data from 2018 to 2022. The data warehouse, structured using the star schema with one fact table and five-dimension tables, facilitated the creation of a web application in PHP. This application offers interactive reports, charts, and performance analyses, significantly enhancing the company’s data analysis capabilities. Additionally, prediction functions were implemented using Python, allowing for tailored forecasting based on the company’s specific data.

Multiple important insights regarding the company’s success were uncovered by the analysis. Seasonal trends were evident, with peak earnings occurring during the summer months, particularly in July and August, likely driven by high temperatures and increased tourism. Conversely, the winter months showed lower earnings, indicating a need for strategic planning to mitigate these seasonal impacts.

The cost and profit analysis per product highlighted key areas for improvement. Products like ”PORTOKALADA” demonstrated a balance of high profitability and manageable production costs, marking them as strategic assets. Conversely, some products with high production costs, such as ”LEMONADA,” exhibited lower profitability, suggesting opportunities for cost reduction and efficiency improvements.

Yearly trends showed a concerning decline in profits from 2019 onwards, despite fluctuations in costs. This decline underscores challenges such as rising production costs and market competition, necessitating strategic adjustments to maintain profitability.

In conclusion, the developed data warehouse and accompanying web application could provide a robust framework for enhancing the Greek beverage company’s decision-making processes. By leveraging detailed data analysis and prediction functions, the company can optimize its operations, manage costs more effectively, and improve profitability. These tools and insights could serve as a valuable guide for other companies seeking to upgrade their ERP systems and enhance their decision-making capabilities.

Acknowledgment

This research work was carried out as part of the project ”Optimization of placement and counting products in large industrial areas using UAV” (Project code: KMP6-0083129) under the framework of the Action ”Investment Plans of Innovation” of the Operational Program ”Central Macedonia 2014 2020”, that is co-funded by the European Regional Development Fund and Greece.

References

[1] M. Figliozzi, “Analyzing the impact of technological improvements on the performance of delivery drones,” *Transportation Research Procedia*, vol. 79, pp. 68–75, 2024, city Logistics 2023.

[2] K. Chen, F. Abtahi, J.-J. Carrero, c. Fernandez-Llatas, and F. Seoane, "Process mining and data mining applications in the domain of chronic diseases: A systematic review," *Artificial Intelligence in Medicine*, vol. 144, p. 102645, 08 2023.
[3] K. Strikas, N. Papaioannou, I. Stamatopoulos, A. Angeioplastis, A. Tsimpiris, D. Varsamis, and P. Giagazoglou, "State-of-the-art cnn architectures for assessing fine motor skills: a comparative study," *WSEAS TRANSACTIONS ON ADVANCES in ENGINEERING EDUCATION*, vol. 20, pp. 44–51, 04 2023.
[4] H. Xueliang, T. X. Lin, and W. Xiao, "Mitigate cross-market competition caused by the risk of uncertainty and improve firm performance through business intelligence," *Heliyon*, p. e34542, 2024.
[5] N. Papaioannou, A. Tsimpiris, A. Kounani, I. Stamatopoulos, A. Angeioplastis, and D. Varsamis, "Comparative analysis of convolutional neural networks for early diagnosis of plant diseases and pest in a multiclass dataset," *International Journal of Computing and Optimization*, vol. 10, pp. 41–53, 01 2023.
[6] S. Tsakiridis, A. Papakonstantinou, A. Kapandelis, P. Mastorocostas, A. Tsimpiris, and D. Varsamis, "Optimizing uav-based inventory detection and quantification in industrial warehouses: A lidar-driven approach," *WSEAS Transactions on Systems*, vol. 23, p. 121–127, 2024.
[7] Z.-X. Huang, K. Savita, and J. Zhong-jie, "The business intelligence impact on the financial performance of start-ups," *Information Processing and Management*, vol. 59, no. 1, p. 102761, 2022.
[8] M. Afsharian, "Data science essentials in business administration: A multidisciplinary perspective," *Decision Analytics Journal*, vol. 11, p. 100442, 2024.
[9] J. Wang, A. H. Omar, F. M. Alotaibi, Y. I. Daradkeh, and S. A. Althubiti, "Business intelligence ability to enhance organizational performance and performance evaluation capabilities by improving data mining systems for competitive advantage," *Information Processing and Management*, vol. 59, no. 6, p. 103075, 2022.
[10] A. C. Yoshikuni, R. Dwivedi, D. Zhou, and S. F. Wamba, "Big data and business analytics enabled innovation and dynamic capabilities in organizations: Developing and validating scale," *International Journal of Information Management Data Insights*, vol. 3, no. 2, p. 100206, 2023.
[11] B. Martin and K. C. Davis, "Multi-temperate logical data warehouse design for large-scale healthcare data," *Big Data Research*, vol. 25, p. 100255, 2021.
[12] L. W. Santoso and Yulia, "Data warehouse with big data technology for higher education," *Procedia Computer Science*, vol. 124, pp. 93–99, 2017, 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia.
[13] J. Luo, J. Xu, O. Aldosari, S. A. Althubiti, and W. Deebani, "Design and implementation of an efficient electronic bank management information system based data warehouse and data mining processing," *Information Processing and Management*, vol. 59, no. 6, p. 103086, 2022.
[14] G. Garani, D. Tolis, and I. K. Savvas, "A trajectory data warehouse solution for workforce management decision-making," *Data Science and Management*, vol. 6, no. 2, pp. 88–97, 2023.
[15] J. A. Macas and C. R. Borges, "Monitoring and forecasting usability indicators: A business intelligence approach for leveraging user-centered evaluation data," *Science of Computer Programming*, vol. 234, p. 103077, 2024.
[16] S. F. Ahamed, A. Vijayasankar, M. Thenmozhi, S. Rajendar, P. Bindu, and T. Subha Mastan Rao, "Machine learning models for forecasting and estimation of business operations," *The Journal of High Technology Management Research*, vol. 34, no. 1, p. 100455, 2023.
[17] Ahmed S. Abutaleb, Michael G. Papaioannou, "Optimal Exchange Rates: A Stochastic Control Approach", *WSEAS Transactions on Business and Economics,* vol. 19, pp. 2027-2039, 2022
[18] A. D. Zisopoulos, G. K. Broni, N. D. Kartalis, K. G. Panitsidis, "Research Articles - Invention Patents Equilibrium; Research Integration, Spatiotemporal Development Strategy, and Circular Economy", *WSEAS Transactions on Business and Economics, v*ol. 19, pp. 1956-1966, 2022