University of Economics - Varna

Department of Informatics

Daniela Pencheva Pencheva

BUSINESS INTELLIGENCE IN RETAIL TRADE WITH FAST MOVING CONSUMER GOODS

A B S T R A C T

of a dissertation

for obtaining an educational and scientific degree “Doctor” in a professional field 3.8. Economics, scientific specialty “Application of computer technology in economics”

Advisor: Assoc. Prof. Silvia Parusheva, PhD

Varna

2022

The dissertation consists of 233 pages, of which: Main text 175 pages, 5 Applications, 76 Figures, 13 Tables, List of literature sources 229 titles and internet sources.

The defence of the dissertation will take place on 31.05.2022 from 10 am in the hall 320 of the University of Economics - Varna at a meeting of the Scientific Jury, appointed by Order № RD-06-65/04.04.2022 of the Rector of the University of Economics - Varna.

The materials on the defence are available to those interested in the University of Economics - Varna, www.ue-varna.bg.

I. GENERAL CHARACTERISTICS OF THE DISSERTATION

1. Relevance of the topic

In recent years, companies are increasingly focusing on the development and modernization of their information systems with the help of innovative information technologies. These include Internet of Things technologies, text mining, computational intelligence techniques, Artificial Intelligence, social robots, etc. The growing importance of information technology for business process management is seen in the context of processes for digital innovation and digital transformation. They are related to the capabilities of digital technologies to support business processes to create a reliable and advanced digital business model. The most frequently asked key questions before taking action to build a digital business model are related to the benefits of implementing information technology and digitalization of business processes and the ability of company departments and related business processes to be better organized to bring more added value. The leading benefits of the implementation of information technology can be summarized in the following groups:

- Parallel processing of work tasks from different departments in the company, which leads to higher productivity and added value in performance.
- It is easier for companies to communicate with each other to exchange information of different nature.

The trade sphere, on the other hand, does not lag behind these trends and it also has processes related to striving to increase the efficiency of trade business processes through automation and quality control, attracting new customers, optimizing costs and more. In the context of these opportunities is the desire in the digitalization of trade activity to apply the latest information technologies, incl. related to intelligent methods, big data analysis and business intelligence, the concept of the “Internet of Things” and many others. It is in the direction that binds the opportunities for improving the digitalization of business processes in the trade sphere with the help of business intelligent methods that the research interest in the current scientific work is directed.

The relevance of the research topic is determined by the tendency of compa-

nies in the trade sector to invest in improving their current information systems in order to achieve higher productivity, minimize manual labor and reduce costs.

Technological advances have had a positive impact and provided an opportunity for digital transactions to be digitized through the application of the latest information and communication technologies. Efforts are aimed at improving the individual elements in the supply chain in wholesale and retail trade and their interaction, incl. with the possibilities of the business intelligence approach, machine learning methods, etc. The aim is to provide greater added value in retail chains and specifically in the management of orders to suppliers.

2. Thesis

The main thesis of the study is that the inclusion of a business intelligent module for order management to suppliers in the retail chain of fast moving consumer goods helps to optimize their planning and improve related business processes in supply chains, improve quality of the decisions taken and achieving greater efficiency of trade activity.

3. Purpose and tasks of the dissertation

The aim of the dissertation is to apply a business intelligent approach in the retail trade of fast moving consumer goods, proposing a model of a business intelligent module for order management to suppliers in the retail chain and develop a prototype.

To achieve this goal, the following **tasks** are set:

1) To explore the possibilities for improving trade information systems using methods and technologies for business intelligence (BI).

2) To make a proposal for a conceptual model of *a business intelligent module for managing orders to suppliers* in the retail chain using business intelligent and analytical methods and tools.

3) To develop a *component for improving the quality of* data in the Trade Information System (TIS), necessary in the context of the proposed business intelligent module.

4) To propose *basic functionalities* provided by the business intelligent module

to the trade information system, as well as to develop its *functional prototype*.

4. Object and subject of research

The object of research in the dissertation is the management of orders to suppliers in retail chains for fast moving consumer goods (FMCG). **The subject of the research** is the development of a model of an intelligent module and its prototype for order management to suppliers with the application of the business intelligent approach, incl. the methods of machine learning and business intelligent and analytical methods and tools, as well as the object-oriented approach and visual methods of modeling.

5. Methodology of the research

The research in the dissertation uses several research methods and approaches, including systematic analysis, comparative analysis, systematic approach, economic analysis, object-oriented approach, and modeling, as well as survey and semi-structured interview methods. In order to test the results of the research in the dissertation, the methods and techniques of visualization and prototyping are used.

6. Approbation

One study, two articles and one conference paper have been published on the topic of the dissertation.

In the dissertation a conceptual model of a business intelligent module for managing orders to suppliers in retail chains has been developed. The proposed model was used as a starting point in the development of the prototype, which is based on real data from the retail chain.

II. STRUCTURE OF THE DISSERTATION

The dissertation has a total volume of 233 pages and includes an introduction, three chapters, a conclusion, a list of references and 5appendices Applications, 76 Figures, 13 Tables, References list with 229 titles and internet sources.

The content of the dissertation is as follows:

III. SUMMARY OF THE DISSERTATION

Chapter 1. Improving trade information systems through BI technologies

As a result of the research in the **first chapter** it was found that the improvement of trade information systems in their part of the management of orders to suppliers is possible with the application of methods and tools for business intelligence. Key technologies that have the potential to support the process of optimizing order planning include the concept of data warehousing, OLAP technology, knowledge extraction and more. The main directions in which BI methods and tools can be applied in the context of achieving flexibility, accuracy, and dynamism in the process of managing orders to suppliers are indicated. A study was made of software systems for retail trade management with FMCG, offered on the Bulgarian market, and three leading software systems were presented in detail, incl. in the context of supplier procurement management, using a system of indicators. Based on the derived BI technologies and tools and the possibilities for modeling, a conceptual model of a business intelligent module for order management to suppliers in the FMCG retail chain is offered.

The **first paragraph** presents some theoretical statements in the context of improving trade information systems through the application of business intelligence technologies. The possibilities for digitalization of the tasks and processes in retail trade are considered, and a number of author's opinions are summarized.

With the increasing globalization of retail, more effective control of data, information and market knowledge is key to maintaining the competitive advantage of trade companies. The main trend in the retail sale of consumer goods is the consolidation of the positions of major market players and the replacement of independent retailers by retail chains.

Summarizing the various activities, processes, operations, and related data in retail is a challenge that has attracted the research interest of a number of authors. Some of them propose to **cover the tasks and sources of basic data** in retail information systems. They are presented in fig. 1.

The tasks performed by the companies operating in the retail trade, incl. and with FMCG, can be classified into 4 main groups: technically engineered, value-adding

core, administrative and decision-making tasks. The tasks belonging to three of the mentioned groups - technically engineered, administrative and decision-making, are of a general nature and do not differ regarding the different companies in the retail trade. *Value-added tasks* are of major interest. These include the following subtasks: goods management; ordering goods; Customer Service; distribution of goods; transportation of goods; financial and accounting tasks (invoicing of goods, liabilities / receivables, and audit). The brief description of these subtasks includes the following:

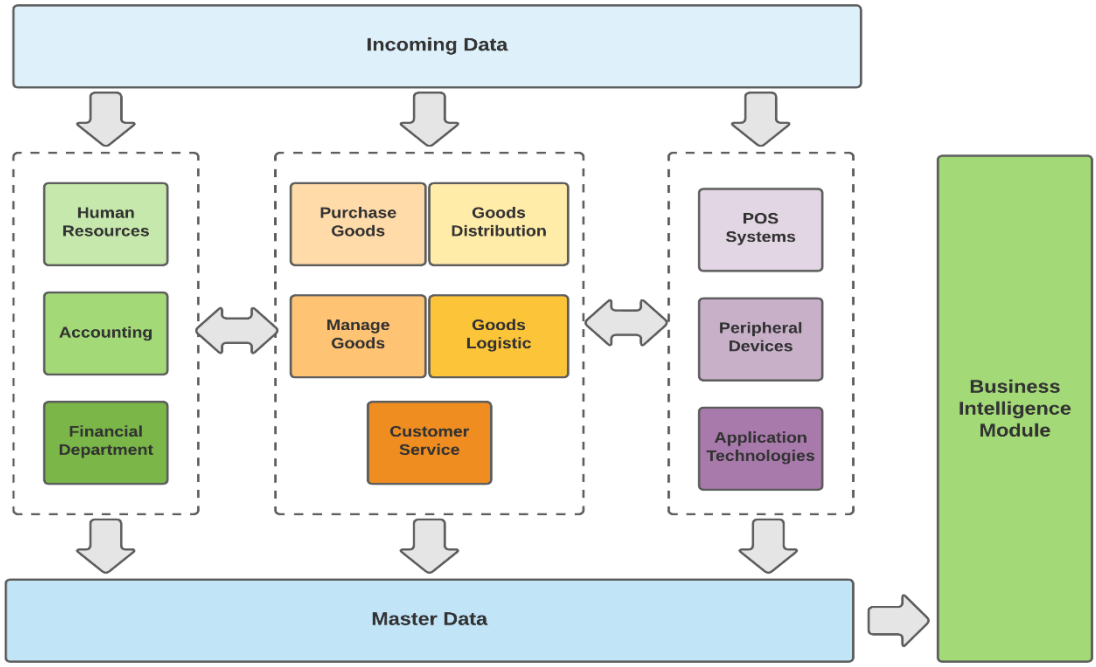

Fig. 1. Tasks and related activities and data in retail

Goods management refers to commercial marketing, covering analytical processes, formulation of objectives, choice of strategy, composition, and control of the marketing mix in the company. The four areas of the marketing mix introduced by Prof. McCartney and known as 4P: product, price, distribution, and promotion are essential to commercial marketing decisions.

From the point of view of the trade process, placing an order for goods includes all activities related to planning the product range, replenishment of inventories in stores and the availability of sufficient goods in the sales hall in response to customer demand. The processes between the central warehouses and the shops, between the suppliers and the warehouses and between the suppliers and the shops also participate (in the presence of direct deliveries). These activities are the basis of our research

interest in the dissertation. After choosing the source of supply and form of delivery, an order is made by the store based on a contract. It includes and describes all the conditions of the order and its implementation.

The composition and execution of an order for the supply of goods depends on factors such as: type of store, complexity of the product range, frequency of demand for the ordered goods, value of goods and more. The listed characteristics can be considered as a starting point for creating an incoming data flow, which is the basis of trade information systems for retail sales. At the heart of the system are the core tasks and related activities that bring added value.

The second paragraph examines the information systems that support the activities of commercial companies, as well as the management of orders to suppliers. The two main groups of systems used in trade companies for inventory management, supply chains and distribution channels, incl. and supplier procurement management, human resources and payroll management, multi-channel and omnic marketing, product portfolio management, etc. - *trade information systems* and *corporate systems*.

1. The trade information systems are specific and closely specialized for the companies in the retail trade and specifically with FMCG. They support the work tasks of the management of the trading company in making decisions related to FMCG, sales, defining a trading strategy and more. The following types of TIS are included in this category: electronic Point of Sales (EPOS) systems; Sales Force Automation systems; Category management systems; Product information management systems, etc.

2. Enterprise systems – they are used in companies, regardless of the sector in which they work. In the context of commercial companies, they are considered as complex systems, covering all ongoing business processes. There are several definitions in the literature for corporate systems, characterized as centralized systems that manage large volumes of data, with easy access to them at any functional or managerial level. These include Enterprise Resource Planning systems (ERP systems), Customer Relationship Management systems (CRM systems), Human Resource Management systems (HRM systems), Management Information Systems (MIS), Financial Management Systems, Supply Chain Management systems (SCM systems),

Supplier Relationship Management systems (SRM systems), etc.

The **third paragraph** examines the possibilities of using business intelligence in the context of trade information systems and order management to suppliers. Business intelligence is a general term covering company assets such as: company infrastructure, applications, data, gained practical experience to achieve optimal management decisions in a competitive business environment. Some of the main factors influencing the use of BI include large volumes of data available in companies and their minimal application in subsequent processes, as well as the need to improve and optimize existing company systems. Business intelligence can be used to describe the process that organizations implement, to collect data and analyze it with the intention of gaining a competitive advantage. Another interpretation of the concept focuses on focusing the efforts of the company's management on effective management through the application of business intelligence.

BI technologies use decision support tools such as generating queries for specific data or information, reports, and multidimensional analysis. The received information is presented to the user with detailed or summarized reports. Varieties of reports range from spreadsheets to a variety of graphs and charts. In the dissertation, it is argued that these technologies can be successfully applied as part of **the vision for further development and improvement of trade information systems and specifically the process of managing orders to suppliers**. The idea of proposing an intelligent module, which improves the order management process by applying intelligent methods and techniques, is being developed.

The important stages, part of **the application of the BI approach in trade activity**, include the following: collection and storage of data from various sources, their processing and transformation into information and knowledge, analysis, interpretation of generated knowledge to achieving improved decision-making. All steps in the process are realized with relevant technologies, including data warehouses, real-time analysis (Online Analytical Processing /OLAP/), Knowledge Management Systems /KMS/, Decision Support Systems /DSS/). The main aspects and stages of BI are presented in fig. 2.

The fourth paragraph presents a study of software systems offered on the

Bulgarian market - I-Cash, IS Grocery and Excepta, which also have functionality for managing orders to suppliers.

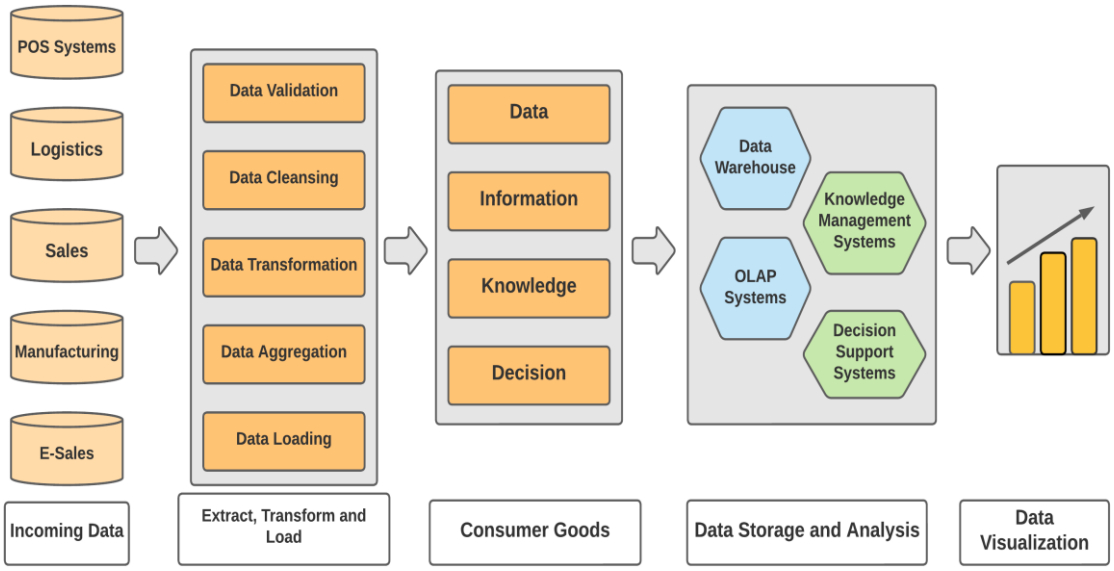

Fig. 2. Main aspects and stages determining the BI approach in the context of trade activity

A summary of the key functionalities of the considered systems is systematized in Table 1.

Table 1.

Comparative characteristics of software systems, including management orders to suppliers

Indicators	I-Cash	IS Grocery	Expecta
Software system type	• Web-based system • Own development	• Web-based system • Uses MS Dynamics as a basis	Desktop application with client-server architecture • Own development
Organization of user groups	• Creation of user groups according to the position held • Set different access levels for each group		
Create a new order	• Manual filling of the desired quantities • Sending orders by email	• Manual filling of the quantities • Sending orders by email • Suggestions for future sales • Data visualization	• Manual filling of the desired quantities • Sending orders by email and other channels
Preparation of reports	• Preparation of documents accompanying the trade activity	• Functionality for preparing reports	• Module for preparation of reports

Based on the three software systems for management of trade with FMCG, it **can be summarized** that they have a complex nature and manage complex business processes related to the diverse activities in trade companies.

The author received trial access to all three software systems. As a result of their detailed research, the opinion is that the most complete and valuable coverage of the typical retail processes with FMCG business processes is available in the Expecta system. Regarding the management of suppliers' orders, it can be pointed out that in two of the systems - Expecta and IS Grocery, some elements of application of business intelligent approaches are available. The Expecta has a module for preparing reports and opportunities for creating different types of reports, which creates added value for them. All three software systems do not offer a full-featured business intelligence module that optimally manages and synchronizes complex vendor order management processes.

Chapter 2. Conceptual model of a business intelligent module for managing orders to suppliers

The first paragraph examines the application of modeling in the context of TIS improvement. The leading visual standards for the development of visual models - UML and BPMN - have been studied.

The UML standard offers specific graphic notation, with a high level of abstraction, with established rules, reflecting different points of view. The technology is defined as very suitable for modeling object-oriented systems with a high level of complexity. The leading advantages of UML make it a suitable choice for visual presentation of an advanced model of intelligent module from TIS. *UML is independent of other programming languages* or specific techniques and methods. The diagrams describe functionalities well and can present aspects of TIS regardless of the software used.

The application of a well-established standard in modeling, created by a set of widely accepted object-oriented software design methods, *reduces model development time* by focusing on the elements involved rather than how to compile it. By using a single standard in modeling, communication between the teams involved in project development is more effective. UML is suitable for modeling both large and complex

systems, as well as small ones. It is also used to design modifications to existing systems, because it has the *ability to upgrade and expand the created models*.

The main goal of the BPMN standard is to create a kind of bridge between designers, developers, and end users of the company's system. It is focused on recreating business processes in a simple way with a well-defined notation. Between the two standards there are similarities in terms of visual interpretation of processes and their decomposition into successive tasks, but UML has wider possibilities for graphical representation of the states and structure of objects from other categories participating in the system. Table 2. presents a comparative characteristic of the two standards.

Table 2.

Comparative characteristics of the two standards – UML and BPMN

Indicators	UML	BPMN
Essence	• General purpose visual modeling language. • Used to specify, visualize, construct and document information system artifacts. • **Complete modeling of information systems**.	• Graphic illustration of business processes. • Notation is easily understood by business users. • **Business process modeling**.
Visualization	• Data, events, conditions are modeled**.** • Low capacity for event modeling.	• Events are modeled only**.** • High capacity for event modeling.
Graphic representation	• UML notation is focused mainly on **modeling information systems.** • Used when modeling various aspects.	• BPMN notation is focused on **business process modeling.** • Applied in in-depth business modeling processes.
Approach used	• Applies an object-oriented approach to application modeling.	• Applies a process-oriented approach to system modeling.
Main advantages	• Established notation with a high level of abstraction. • Scalability and structuring. • Presentation of different points of view.	• Simplified presentation of business processes. • Easy to understand standard. • It is characterized by great flexibility.
Existing shortcomings	• Development of complex models. • Sophisticated maintenance.	• Representation of only an aspect of business structures. • Limited notation.

Based on the above characteristics, it can be argued that the UML standard has the potential to illustrate the structure of the business intelligence module for order

management to suppliers in retail chains of FMCG more fully, considered in the dissertation, and its adjacent objects, their states, and existing relationships, and BPMN has a complementary function in the study, illustrating the more complex processes occurring in the module.

The second paragraph presents the developed conceptual model of the business intelligent module for management orders to suppliers. Its description is preceded by the creation of diagrams using UML, graphically reproducing various aspects of the functionalities of the TIS. These include the following*: conceptual business model, business scenario diagram, class diagram, sequence diagram, and state diagram*.

The *conceptual business model* is an integral part of the business modeling process. It shows the specific purpose of the system in question as a whole. The developed conceptual model of the process is the basis for the object-oriented presentation of elementary processes, which cannot be decomposed into simpler ones. They define the roles and activities performed, as well as the artifacts obtained in the development of information systems.

The objects involved in the proposed module for management orders to suppliers are summarized in a conceptual model presented in fig. 3. The main emphasis is placed on the application of business intelligent methods, which are expressed in the preparation of automated forecasts for the required quantities of goods, organization of key data in OLAP cube, application of Data Mining methods for deriving Key Performance Indicators (KPI) required in trade activity.

The diagram illustrates objects without their associated properties, emphasizes the relationships between them and their organization, and does not define fields, attributes, or methods.

1) **Data cleansing** – the essence is characterized by methods for enrichment, cleansing and validation of the entire incoming data stream. After successful application of the methods the data passes to the entity “OLAP”.

2) **OLAP** – the entity is responsible for systematizing the incoming data flow in a multidimensional cube and its storage. It is the starting point of the two main classes in the order management module – “Forecasting” and “Data Mining”.

3) **Forecasting** – this entity has an adjacent “External Sources” attribute. As this attribute carries additional content related to the preparation of a forecast model, it is depicted as an independent entity. The “Forecasting” entity is determined by a forecasting method, the result of which is reflected in the “Order” entity.

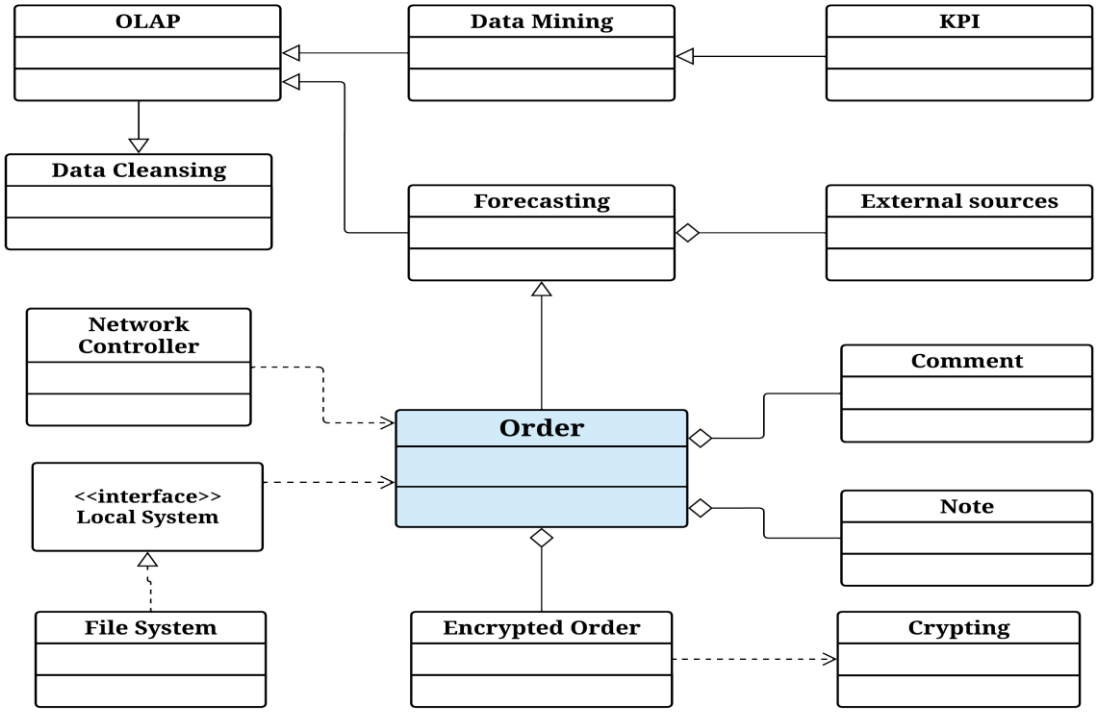

Fig. 3. Conceptual model of a management orders to suppliers module (class diagram)

4) **Order** – with adjacent attributes “Content”, “Comment” and “Note”. The last two attributes, “Note” and “Comment”, can have additional content and are shown as separate entities. The “Order” is characterized by methods for setting and adding additional content.

5) **Data Mining** – the entity is characterized by a method for extracting knowledge from data sets for time period, sales data, product and customer data. The results obtained are reflected in the “KPI”.

6) **KPI** – this entity is characterized by methods for calculating key indicators for trade activity, which show the current state of the trading company and in particular the movement of orders and sales. The following indicators are monitored: duration of the sales cycle, sales, customer loyalty, outflow of customers, assessment of customer experience, reporting of monthly profit.

7) **Secure order** – with an associated attribute for storing an encrypted password. The entity is directly related to the “Encryption” entity, which includes a method for setting and encrypting a password.

8) **Local provision** – an interface-type entity that recreates the system kernel. The operations it contains are related to the basic commands when working with orders - create, save, modify, and delete. The “Local Security” interface of the system is invariably linked to the “File System”, which is shown as a separate entity.

9) **Network controller** – entity with methods for managing the creation, modification, deletion, and access of an order. The “Network Controller” manages system requests from its users.

The third paragraph presents the components involved in the model of business intelligence module for management orders to suppliers: data quality improvement component, key data management component, component for forecasting the required quantities of goods for sales, component for visualization and editing of orders to suppliers, a component for preparing reports according to user-defined criteria. All components are summarized in a complete logical model of the business intelligence module, presented graphically in fig. 4. The architecture of the data warehouse is also considered.

The process of improving the quality data begins with the entry of data in the TIS. Profiling of the input data is then performed to provide a comprehensive description of their condition. The analyzed data are directed to the component for improving the quality of the data, and for their improvement individually developed rules are applied according to approved techniques.

Cleansing techniques are among the traditional ways to improve data quality. They are applied after the profiling of data on values categorized as ineligible for the respective set. It is considered that data cleaning solves the existing problems, such as duplicated and missing values, discrepancies in functional dependencies, discrepancies in the sequence of values in the order, etc. Inaccuracies in the data are corrected by correcting the violated values - the value is made equal to or different from a preset value or a value is removed or generated.

According to the requirements of the system and the specifics of business data,

the last two of the above actions, namely the **technique of data enhancement** and **data transformation** can be distinguished as two separate groups of operations for improving data quality. At the next stage, using the **data validation technique**, a check is made for compliance of the values stored in the system with the eligible domains. The domains are determined in accordance with the needs of the company and are implemented in the system as a catalog containing only current and correct values. Once the rules for improving the quality of the data have been applied, the next step is to check their eligibility.

The key data management component acts as an intermediate link between the purified data loaded into the data warehouse and its transfer to the multidimensional OLAP cube. aimed at the key data grouping module. It involves the following operations: grouping data, finding similarities in the data and merging the data.

At next stage, the aggregated data is redirected to the OLAP loading cube, which completes the process in the component and checks their eligibility. If the result is positive, the data is recorded in the TIS database (operational data), otherwise the flow continues to another branch, where it is stored for further processing. The processes related to the subsequent processing of data of "low" quality are not the subject of the dissertation research.

After considering the **influence of parameters** such as: input data, considered specifics of the trade sphere, reporting of consumer behavior, and expected results from the application of business intelligence, the forecasting component is planned to include intelligent methods of analysis, which are studied by many authors. The main place in the present dissertation is occupied by the methods of data analysis related to the *processing of time series data*. This type of data is defined as a sequence of observations reflecting the change of a specific variable at regular intervals, in sequential order. When analyzing time series, it is important to consider the nature of the studied data, their specifics and relevance to a specific period. Characteristic of the data participating in time series is the existence of a high probability of a relationship between current and previous values, i.e., for the presence of autocorrelation. The study of the autocorrelation function is a starting point in the analysis of time series. It can be established that the current value is based only on its previous historical value,

i.e., the time series is the realization of an autoregressive process of first order AR (1) or depends on its previous two values and the time order is the realization of an autoregressive process of second order AR (2), etc.

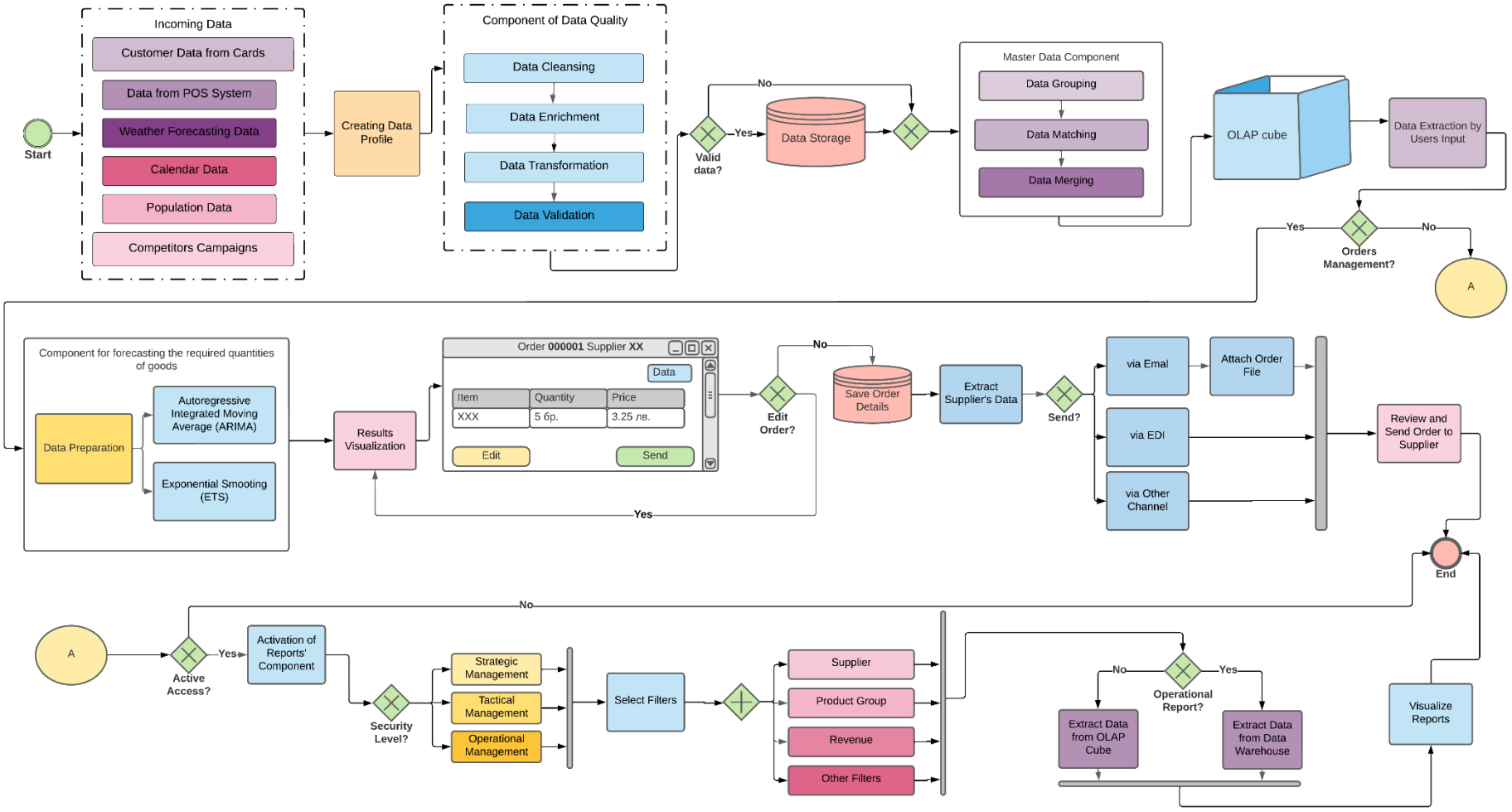

Fig. 4. Complete logical model of a business intelligent order management module

Extracting data from the multidimensional OLAP cube is the first step performed in the component for forecasting the required quantities of goods included in the preparation of orders to suppliers. The transformation of cleansed data into information and solutions based on it is a multi-component process combining machine learning techniques based on statistical and machine learning algorithms, as well as individually developed business rules according to the specifics of FMCG trade.

The order visualization and editing component performs a control function and allows the user to check the results for the required quantities of products calculated by the forecasting component. The sequence of the performed functions and the participating objects are presented graphically in fig. 4. Thanks to business intelligence, the results obtained are expected to meet the needs of the trading company, but nevertheless provides for the possibility of manually changing the quantities offered.

The component for reporting according to set criteria does not directly concern the process of management orders to suppliers. The main functionalities are related to the visual presentation of data and information in the TIS, offering users the ability to create individual reports, dashboards, export information from the system and more. To activate this component, the user of the system must have the appropriate authorization to work with visual reports. If there is access, the user's affiliation to the respective user group is explicitly checked.

For the operation of the business intelligent order management module, its information support is important. The underlying information base uses a data warehouse with appropriate architecture. There are several leading data warehouse architectures, including Kimball Bus Architecture, Inman Hub-and-Spoke Architecture, and Hybrid Architecture[12]. Hybrid architecture combines the benefits of Kimball and Inman architectures. In it the bus of the general dimensions is preserved, as well as the corporate data warehouse with accompanying obligatory normalization. The users' access of the system is limited to the applications for data visualization.

[1] Kimball, R. and Ross, M. (2013). The Data Warehouse Toolkit: The Defi nitive Guide to Dimensional Modeling. 3rd Ed., Indianapolis: John Wiley and sons.

[2] Yessad, L. and Labiod, A. (2016). Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. 2016 International Conference on System Reliability and Science (ICSRS), IEEE, pp. 95-99.

Among the leading advantages of the hybrid architecture are *speed and accuracy in loading input data, atomicity of data, application of analytical structures, speed in data visualization and execution of requests, high level of security due to limitations access of system users.*

We believe that according to the subject area and purpose of the business intelligence module, the hybrid architecture is suitable for use in the intelligence module. The motives for choosing a specific data warehouse architecture are also supported by the results of a focus group survey of retailers conducted using the unstructured interview method. Respondents view trade data as one of the largest assets of companies and state that it is organized into a relational data warehouse using a hybrid architecture of common facts and dimensions. The logical model of the data warehouse, developed according to the adopted architecture, as well as the structure of the tables, are included in the third chapter.

Chapter 3. Functional prototype of business intelligent module for managing orders to suppliers

In the **first paragraph** of the third chapter the current state of the management of goods' orders to suppliers of representatives of the chains for retail with FMCG operating on the Bulgarian market is studied. The study is focused on retail chains, representatives of supermarkets and discounters, as statistics show that their sales revenue forms 55% of the retail trade with FMCG on the Bulgarian market.

Representatives of three of the retail chains – Lidl, BulMag and Dar, responded to the invitation to participate in the study. It is accomplished using **several survey methods**. The first one is based on face-to-face **conversations with representatives of the management teams** of retail companies, using **online video conferencing** (with video conferencing software), as well as **telephone conversations**. The method of the **semi-structured interview** with a specially designed questionnaire is also applied. A **personalized questionnaire** was sent to each of the representatives electronically, containing questions common to all participants and questions directly related to the specifics of the retailer and the results of its preliminary detailed study.

The profile of the three retail chains, presented in a summarized form, includes the following: Dar retail chain has 10 supermarkets and BulMag retail chain has 18

retail stores. The Lidl retail chain, represented nationally, has 107 stores in 50 cities across the country. According to the ranking of Capital „Regal 100“ for 2020, all three retailers are positioned in the Top 30 of the chains for FMCG in Bulgaria. The Lidl retail chain is in second place with sales revenues of BGN 1.015 billion, BulMag is in eleventh place with revenues of BGN 94.063 million, and Dar is in twentieth place with revenues of BGN 34.403 million.

Local retail chain Dar

The retail information system of the Dar retail chain is semi-centralized, as each of the supermarkets can place an independent order to a supplier. The retail system has developed a module for automated preparation of orders and the required quantities of goods are calculated according to an individually developed algorithm. It includes checking sales for the same period of the previous year, checking the available quantity of goods, determining the minimum stock of goods, as well as the period until the next delivery. After preparing the order values, it is reviewed by an employee and sent to the supplier. There are two types of data transfer channels in the Dar retail chain - electronic data interchange (EDI) and e-mail. A module has been implemented in the automated order management module, which monitors the daily orders schedule. It sends periodic reminders to the employees to start new order processes.

Regional retail chain BulMag

The company has a centralized retail information system. All supermarkets and the e-shop are connected to it. The automation of the processes for managing the orders to suppliers has been accomplished by using technologies for calculating the required quantities of goods.

The process is presented graphically in fig. 5. by means of the BPMN standard. Performed operations are as follows: selection of the supplier to which the order is sent and activation of the module; visualization of an electronic document with recorded quantities for order from the respective supplier; manual adjustment of the proposed values; adding the order as an attachment to e-mail with filled in data of the supplier; last review for correctness of the data and sending the order.

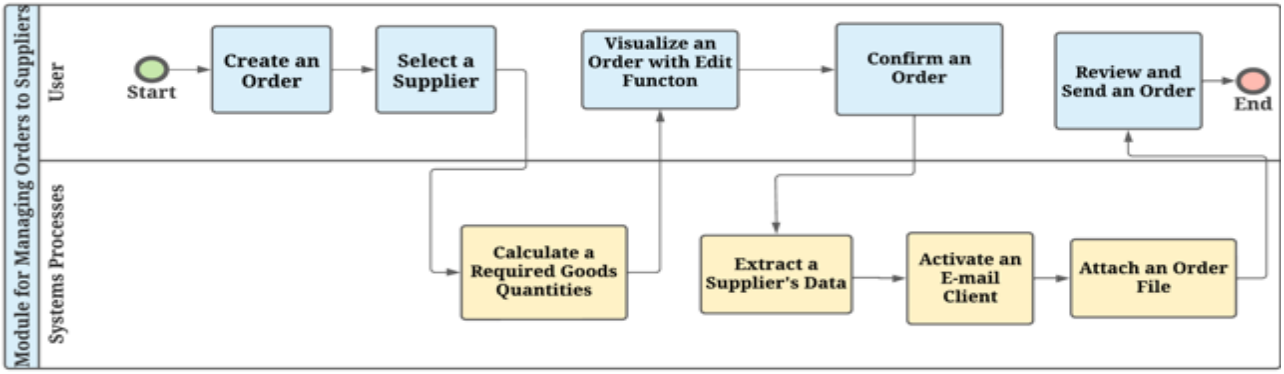

Fig. 5. Diagram of the processes in the automated order management module in the BulMag retail chain

The algorithm in the module includes the following variables: time range between orders, quantity of product sales for the past 7 days, warranty stock, currently available quantity of goods, participation of goods in promotion during the period. The centralized order applies to all sites in the chain together with the e-shop.

Nationally represented retail chain Lidl

The international retail chain Lidl currently has 107 self-service supermarkets in 50 cities in Bulgaria. Lidl's order management system to suppliers is characterized by a two-factor scheme. Initially, each store prepares an order for the required quantities of goods, then it is sent to the Central Warehouse. Orders are summarized into one grouped order which is sent to a supplier. The described process uses many subsystems that store different types of data. For this reason, system processes have been developed for timely synchronization of data at all levels, monitoring the accuracy of data; data enrichment; validation of input data; check for the availability of the required minimum data for each consumer product.

The graphical interpretation of the described activities for placing orders to suppliers is presented in fig. 6.

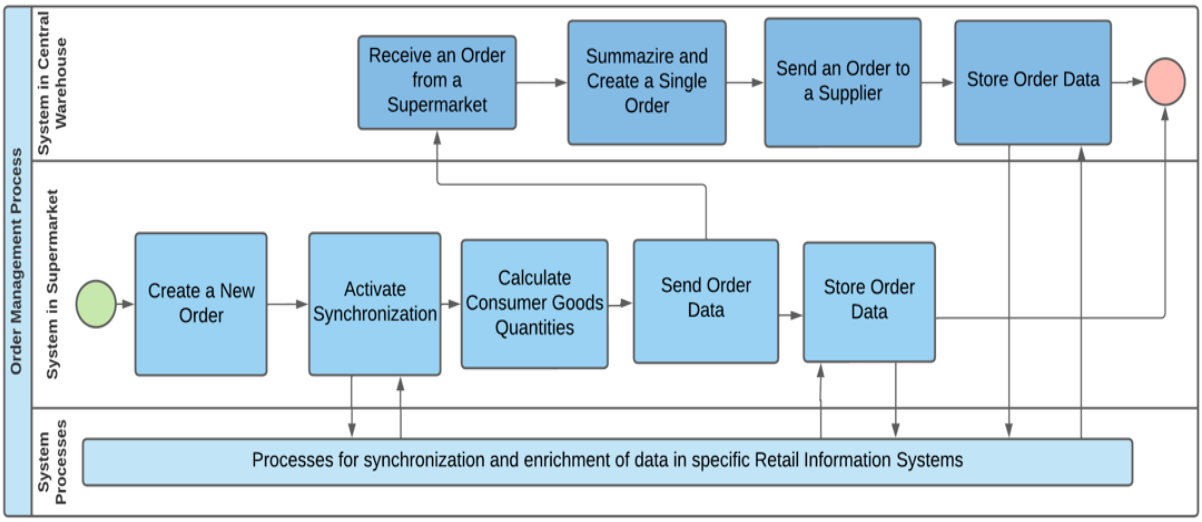

Fig. 6. Two-factor order management process to suppliers in the Lidl retail chain

Automated process for sending orders to suppliers has some elements of business intelligence, such as considering the necessary quantities of goods according to the meteorological situation. It is implemented in the systems of the individual branches when preparing orders to the Central Warehouse. To calculate the required quantity of goods, many indicators and data are processed, such as compliance with the limits for minimum and maximum shelf availability, data on previous sales for different time periods, planned promotion campaigns, logistics data for orders, etc. Forecast information is also available, which is compiled based on statistical data for past periods, a set of mathematical models and proposals for future sales. It serves as a reference in the final preparation of the order before it is sent for aggregation in the Central Warehouse. After receiving all orders from the supermarkets, an authorized employee performs actions for creating a consolidated order to make an informed decision. In the second stage of the process of order management to suppliers, automation is minimized for security reasons.

The Lidl, BulMag and Dar retail chains are a positive example for FMCG retailers who follow global technology standards to automate order management processes to achieve higher business process productivity and minimize manual labor. Representatives of all three retail chains say that the systems are subject to constant improvement to minimize manual labor, reflect the legislative changes related to retail in the country, and update the software. Problems include the storage and analysis of

big data generated daily, the need to respond quickly to consumer demand, the availability of manual activities that slow down the overall process.

Retail systems store a variety of information about purchases, orders, customer suppliers and other related information, but its aggregation for knowledge acquisition and automation can be considered a challenge. The enrichment of the algorithms related to the calculation of the estimated quantities of goods through intelligent technologies is also among the tasks that companies set themselves.

In this context, it is appropriate to develop a prototype of a business intelligent module for order management to suppliers in retail chains. It offers optimization and automation of the process of preparing orders to suppliers, helping to reduce manual labor and increases the speed of retail information systems in the management of retail business processes.

The **second paragraph** presents the prototype developed by the author of a business intelligent module for managing orders to suppliers. An in-depth study of the leading software technologies for automation and business intelligence has been conducted to form a set of technologies involved in development.

For the purposes of the research, a software tool was chosen for the implementation of the intelligent approach in data processing and order preparation, focusing on **Business Intelligence and Analytical Platforms** (BIAP). After considering the capabilities and advantages of Microsoft Power BI and its leading positions in research by researchers and consulting companies such as Gartner, its use in the preparation of the prototype is envisaged. Another platform involved in the development is **Alteryx Designer**. Its use is necessary due to the need for the Power BI application to be supplemented with functionalities for the design and implementation of data transfer processes.

Such processes are: “Extract, Transform, Load – ETL”, maintenance of functionalities for standardization, cleansing and validation of data, and built-in analytical methods. In addition to using Power BI and Alteryx Designer, the **Ataccama Data Quality Analyzer** software product is also used. It is used in connection with the activities of initial research and profiling of the incoming data stream to detect possible data inaccuracies. The developed relational data warehouse

was designed with the tools of **ERwin Data Modeler**, and its implementation is carried out with **MS SQL Server**, and both software products have proven over time leaders in their fields.

According to the structure of the data and the object of their research, machine learning methods are applied, suitable for time series analysis. The selected methods refer to time series forecasting. They are strongly represented in the selected Alteryx Designer platform. Their main purpose is to study a series of past events (past sales) and to look for hidden patterns in them (seasonality) in order to make the most reliable forecast for future events (sales). A necessary condition is that the studied time series be consistent, e.g., reflect sales for day, week, month, quarter, or year only. When creating the forecast, future values are added to the values that reflect the past and present, and thus a single horizon is formed. It is possible that there are some gaps in the analyzed data sets and for this reason various algorithms have been developed in Alteryx Designer, which enrich the studied data to obtain a reliable forecast. The purpose of the methods used for analysis of time series in the development environment is to predict the future value of the studied quantity (quantity of goods for sale). By examining the relationship of a variable in a series of values, models can be found to extrapolate future values.

The **third paragraph** of the third chapter presents the approbation of the developed prototype of a business intelligent module in the Dar retail chain.

The chain's retail information system is semi-centralized, with each of the supermarkets being able to place an independent order to a supplier. The current practice of the supply management chain involves the use of a developed module for their automated preparation as part of the retail system. In it the required quantities of goods are calculated according to an individually developed algorithm, which includes the following variables: **verification of sales for the same period of the previous year, verification of the available quantity of goods, determination of minimum inventory, period until next delivery**. Communication with suppliers is carried out mainly through the EDI communication channel, but it is possible to send an order by e-mail or fax.

The need to refine the result values and fully automate the process of managing

orders to suppliers determines the need for the application of intelligent methods for forecasting the required quantities of consumer goods. Appropriate visual presentation of data in the form of reports and dashboards greatly supports the work of specialists in the Sales Department.

These reasons necessitate the development of a prototype of a business intelligent module for managing orders to suppliers and its testing in the Dar retail chain. With its help it is possible to determine as accurately as possible the required quantities of consumer goods involved in managing orders to suppliers, as well as visual interpretation of data and information as interactive dashboards and reports.

Specific tasks for the implementation of the prototype include the following:

1) Development of a data warehouse that stores the data involved in the prototype.

2) Development of the components involved in the prototype of a business intelligent module for management orders to suppliers.

3) Application in the prototype of appropriate methods for machine learning and preparation of forecast models for calculating the required quantities of consumer goods.

4) Development of interactive reports and dashboards to support the daily tasks of the Sales Department specialists.

5) Proposal for implementation of the module in a retail software system.

In the context of improving the quality of data, their detailed profiling is important. In addition to the built-in methods in the Ataccama Data Quality Analyzer (DQA) software, additional business analysis rules have been developed. They meet the considered business requirements for the data and complement the built-in algorithms for statistical analysis of the software. The presented business rules examine the quality of the data in each of the columns in the input file. They are prepared according to specific business requirements, but it is possible to develop standardized rules with a wider scope.

Based on the problems identified during profiling, a **prototype of a component for improving data quality** was developed in accordance with the presented business rules and the conceptual model of a component for improving the

quality of data. The prototype was developed with the methods of Alteryx Designer software product and is presented in fig. 7.

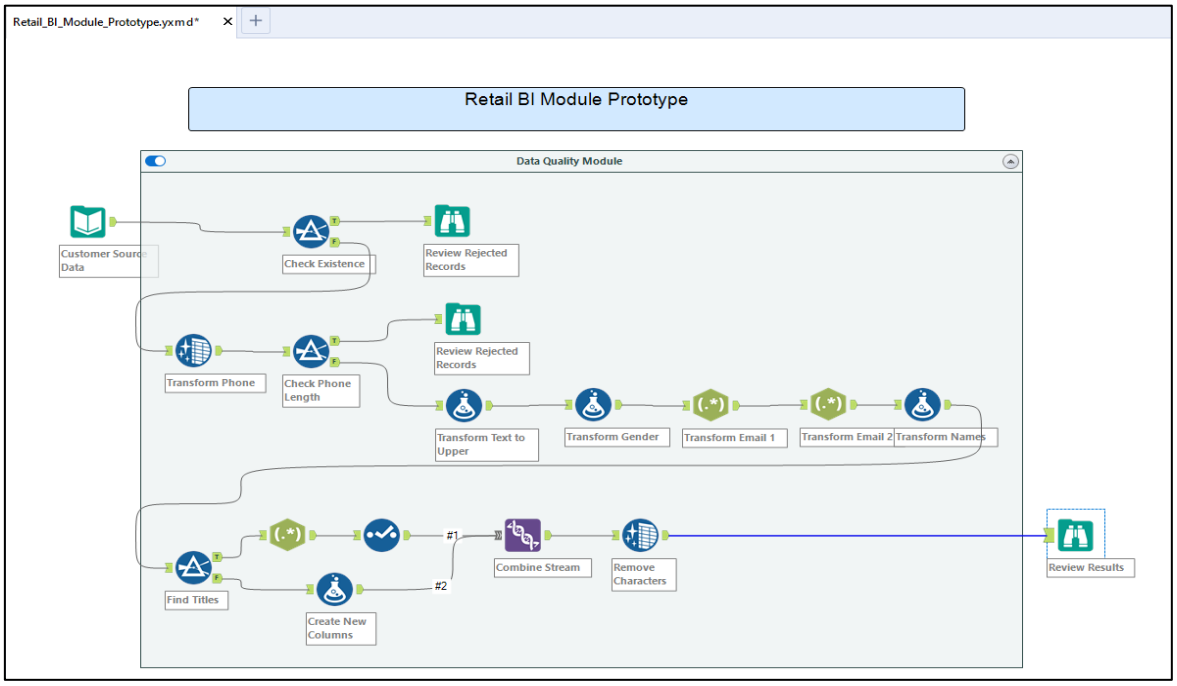

Fig. 7. Prototype of a component for improving data quality

Alteryx Designer is a low code programming development environment and contains predefined data processing algorithms that are presented in the form of a visual object called a step. When starting the process, the steps that are interconnected are performed sequentially and activate the defined settings, checks, formulas, transformations, etc. Their purpose is summarized in Table 3.

Table 3.

Main steps involved in the enhancement component data quality

№	Designation	Purpose	Example
1		Input data reader	Read the input file containing user information
2		Output data visualization	Visualize the data to determine their status at the end of the operations to improve data quality
3		Split data by condition	Apply rules for categorizing data and separating valid records
4		Data standardization	Remove leading spaces, double spaces, numbers, or symbols
5		Use formulas	Replace a character or letter in the data ReplaceChar ([cio_gender], "F", "Ж")
6		Use regular expressions	Extract some of the values in the field (mailto:) ([- a-z0-9 ._] + @ [- a-z0-9 ._] + \. [a-z] +)
7		Rename columns	Introduce a single standard for naming columns in the file
8		Combine data stream	Combine the two data streams into one

After cleansing the data in the data quality component, it is directed to the key data management component. The following operations are applied in it: grouping, finding similarities in the data and merging. Due to the high level of specialization of the development environment, the listed actions are covered by two visual steps, which are presented in fig. 8.

Data grouping is used to check for similarity groups in the data. The technique is applied by highlighting key fields that can uniquely identify a record, excluding the field containing code. The client data can be grouped by name and address to check for records that point to the same client. Based on the created groups, the step for finding matches and merging records checks the values and looks for exact or partial matches based on a set allowable percentage of similarity. In the merge operation, the records marked as identical are transformed into one record.

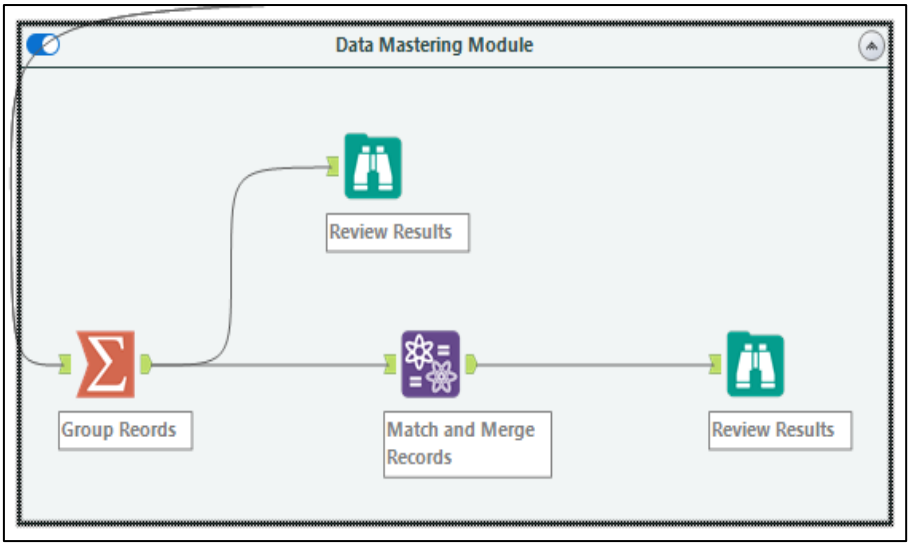

Fig. 8. Prototype of a key data management component

After completing the last operation in the key data management component, the data flow is directed to the **future sales forecasting component**. The main data set is formed from the historical sales data extracted from the POS system. The data for customers, meteorological forecast and calendar holidays have a complementary function. Real historical data for several food products from the group „Sweets" for a period of two years (2015 and 2016) were provided by the retail chain Dar. **Real historical data was provided by the Dar retail chain** for several food products from the group for a period of two years (2015 and 2016), it became possible to carry out experiments on their basis.

Due to the specific requirements of machine learning methods, it is necessary to make previous transformations of the input data, including approval of a single standard for naming columns; check for missing values in the timeline and automatically replenish them with a step in Alteryx; check for missing values from the sales order and replace them with a weighted average value. The first stage of the implementation in Alteryx of the component for forecasting future sales is presented on fig. 9.

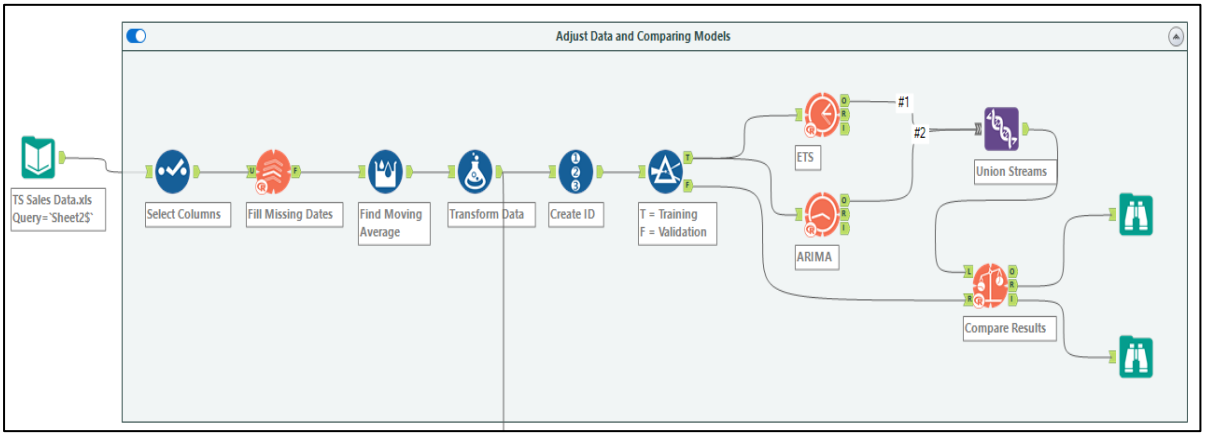

Fig. 9. First stage in the component for forecasting future sales

The current study combines the comparison of ETS and ARIMA methods for time series analysis and the choice of the more accurate of the algorithms for forecasting. Alteryx experts recommend that when comparing two or more methods, two sets of data for model training and verification be identified. This is achieved by dividing the input data into two groups. Data that is closer to the present is included in the verification data and the rest in the training data. After the separation of the two groups, the data flow is directed to the steps for performing the analyzes of time periods ARIMA and ETS. The purpose of the specialized steps in the component is presented in Table 4.

Table 4.

Specialized steps involved in the component for forecasting future sales

№	Designation	Purpose	Example
1		Automated population of missing values	Check the time series based on day and create a new line with a consecutive date, if necessary.
2		Custom operation into single column	Example of calculating weighted average values: ***IF*** *isNull([Y])* ***THEN*** *(([Row-1:Y] + [Row+1:Y]) / 2)* ***ELSE*** *[Y]* ***ENDIF***
3		Create a new column with pre-filled data	Create a unique number type identifier for each row in the dataset.
4		Use of ETS	Preparation of a forecast for weekly sales of a product from the product group “Sweets”
5		Use of ARIMA	Preparation of a forecast for weekly sales of a product from the product group “Sweets”

№	Designation	Purpose	Example
6		Compare results from time series analysis algorithms	Comparing the results of the forecasts with ETS and ARIMA and creating a visual interpretation.
7		Create visualization of time series analysis results	Visualization of results after forecasting with ETS algorithm.
8		Algorithm for analysis of more than one variable	Analyze time series for sales of products from a whole product group.
9		Creating a visual interpretation of the analysis of more than one variable	Visualization of results from the analysis of time series for sales of products from a whole product group.

The forecasting results are combined into a common data stream and a comparison is made through the model comparison step. It is presented visually as a static report and an interactive dashboard. The static report presents the superimposition of the results of the forecasts with ETS and ARIMA and the independent study of the time series. The breakdown shows the dynamics of sales of the researched product over time. The interactive dashboard shows the same information but allows you to understand in detail the results and examine a specific period. The two visual interpretations of the compared data are presented in fig. 10.

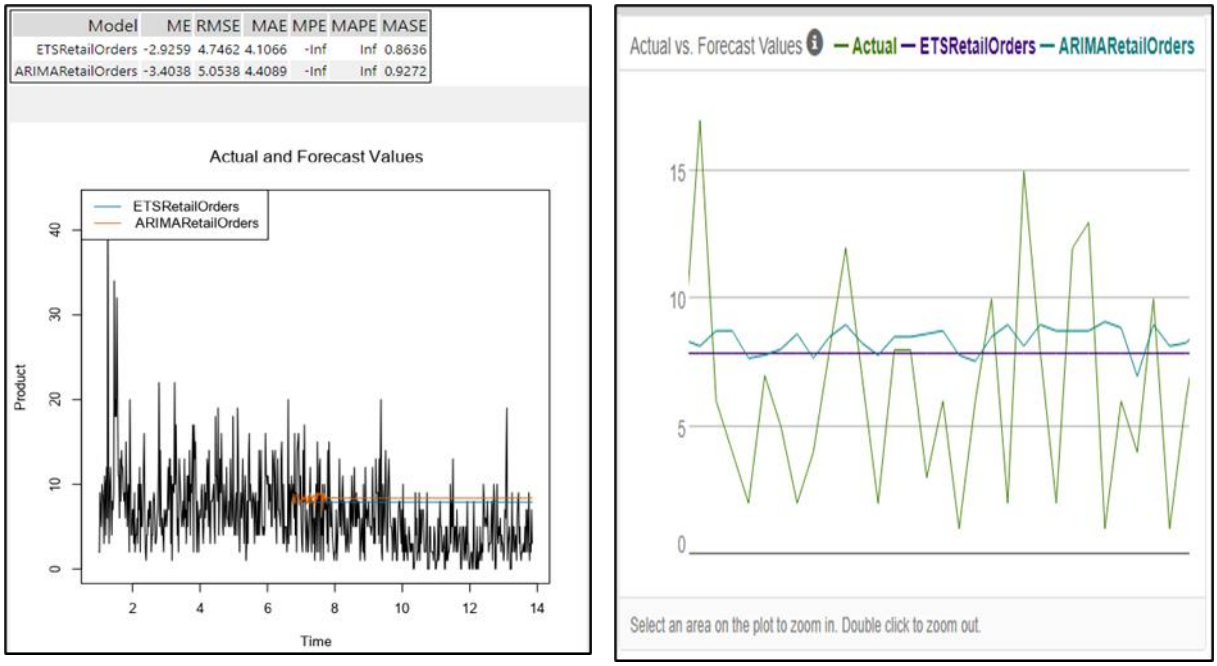

Fig. 10. Comparison of the results of the comparison of forecasts of the ETS and ARIMA algorithms

According to the purpose of the algorithms in Alteryx and their specifics, it is

considered that an algorithm with lower result values is more suitable for the business intelligent module and time series forecasting.

Based on the results obtained, it is considered that the ETS algorithm results more reliable estimate the required quantities of consumer goods. Using it, the second stage in the component for forecasting future sales has been developed (Fig. 11.).

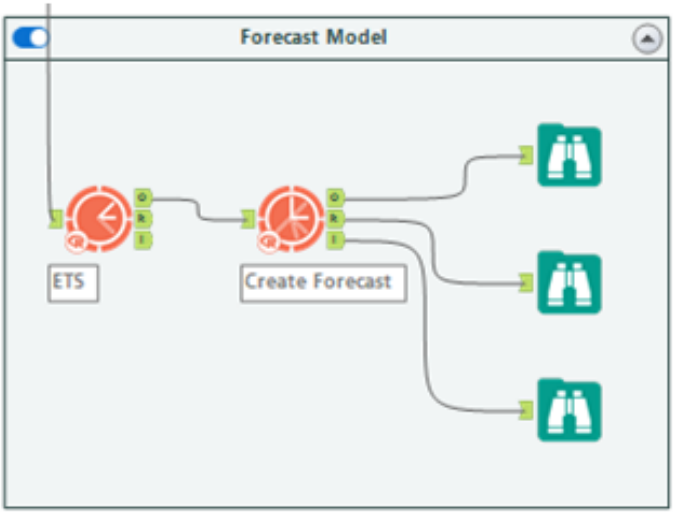

Fig. 11. Second stage of the component for forecasting future sales

The researched product Sweet Cake Balkan Box - Cocoa 38 g has an expiration date of about 1 month and regular deliveries are once a week. Therefore, a summary of the results of the forecast can be set on a weekly basis. Other important parameters of the consumer goods forecast are a study of 6 future periods (6 weeks) with an accuracy of results between 80% and 95%. The visual interpretation of the results is shown in fig. 12.

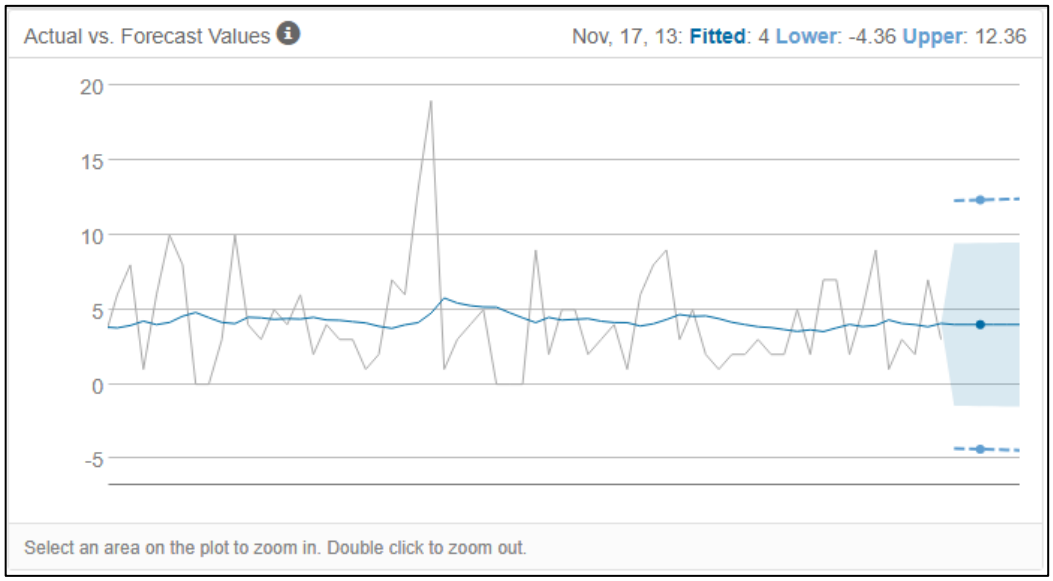

Fig. 12. Results of forecasting future sales

The last date participating in the analyzed data set is 31.10.2016, the forecast covers the period up to 11.12.2016. The forecast shows three possible result values: minimum (-4.36), average (4) and maximum (12.36). Adherence to the maximum value is recommended to ensure sufficient availability of the goods. The expectations are that the quantity sold of the tested product is 12 for a week when it is offered at a regular price. The peaks observed in the historical data are the result of promotional campaigns. When the product participates in a promotion, the number of sales increases. The impact of complementary data sets, namely calendar information, also needs to be considered. In summary, the quantities involved in the improved formula for calculating the required amount of test product can be deduced: **predicted amount of time series analysis; participation in a promotion, considering the amount of the discount; a national or church holiday on which a certain percentage of the store's customers celebrate; minimum inventory; period until the next delivery.**

Alteryx Designer can forecast the sales of many products (analyzed variables), provided that the input data meet the requirements for participation in the time series analysis. After completing the calculation, the value obtained for the required quantity of goods is sent to the existing in the software system of the company Dar component for visualization and editing of new orders.

The component for reporting uses according to set criteria a variety of techniques for getting data the main goal is to keep the data accurate and up to date.

One of the techniques used is Web Scrapping. It gets content from a web page and makes it available in Power BI. In this way, the data is available for further processing. The one discussed in fig. 13. example shows the extraction of the current weather forecast from the site SINOPTIK.bg. It is mandatory to enter the exact link to the web content in the Get Data from Web function available in Power BI. The captured content is displayed in tabular form.

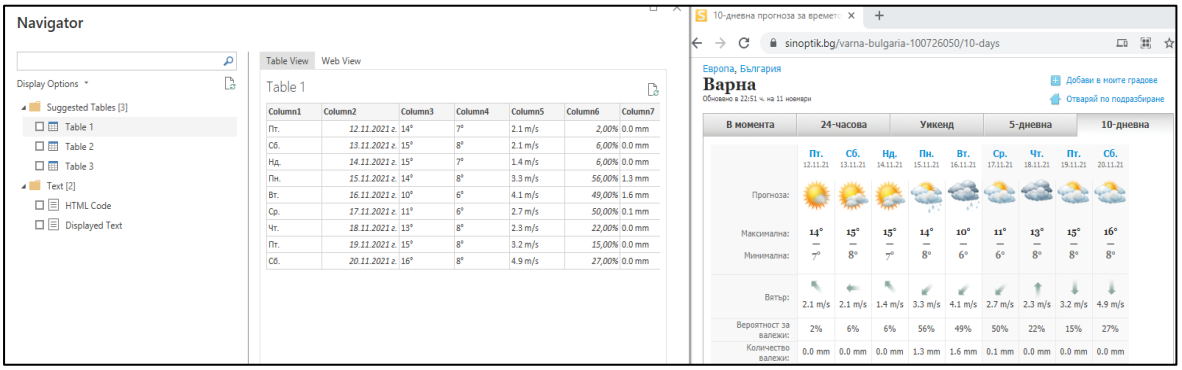

Fig. 13. Extract data from a website with Web Scraping

After implementing data transformations to achieve higher quality, they are included in the overall business data model to add value to dashboards and reports. Daily data updates are also provided through Power BI functionality. A prototype of the visual interpretation of the meteorological data is presented in fig. 14.

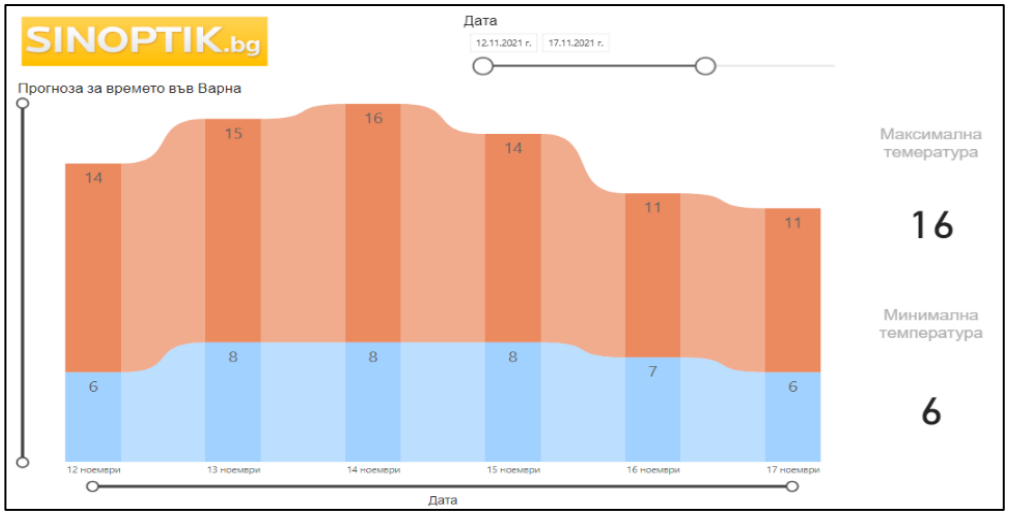

Fig. 14. Visualization of data extracted from the site SINOPTIK.bg

Other reports have been developed, which present the realized sales for products at regular price, promoted products, seasonal products, and key indicators, which monitor the state of the trade promotional campaigns, sales revenues, total profit, etc. The dashboards have dynamic filters that allow the user to display

information only for a specific sample, e.g., by period, supplier, product group, etc. Table 5 presents the key indicators that participate in the interactive tracking board. Due to the specifics of the indicators and the need to use data sets that are subject to confidentiality, test data were used for development.

Table 5.

Key indicators for status reporting of retail operations

Indicator name	Purpose	Example
Sales Cycle Length	Measuring the average time for consumers to make purchases.	Sales Cycle Length = Length of all Deals by the Day/Total Number of Deals
Sales Conversion Rate	Identifying potential new digital customers to attract them.	Sales Conversion Rate = (Total Amount of Sales/Number of Qualified Leads) * 100
Consumer Retention Rate	The percentage of customers who have made more than one purchase for a certain period of time is reported. S - Number of customers at the beginning of the period E - Number of customers at the end of the period N - Number of new customers acquired during the period under review	Customer Retention Rate = [(E-N)/S] * 100
Customer Churn Rate	It measures the outflow of consumers during the study period.	Customer Churn Rate = (Leaving Customers/Total Customers) * 100
Monthly Sales Growth	Determining the profit or loss ratio based on the previous month.	Month Sales Growth = [(Current Month Revenue – Previous Month Revenue) / Previous Month Revenue] * 100

An interactive dashboard has been developed, showing the sales of products at a regular price, with the following attributes: dynamic filter for selecting a period; cards showing weekday and weekend sales; detailed breakdown of sales by day; summary view of sales by week. The visualization provides an opportunity for dynamic change of the period (month, quarter, year, etc.). The developed dashboards are deployed to the Power BI cloud service, where they are available to authorized employees. Additional visualizations and aggregated information are presented on the

dashboard on fig. 15 intended for the management of the company. They show samples on a monthly and weekly basis needed to set medium- and short-term company goals. Profit summaries and KPI values show the trend of current sales.

In our opinion, the presented option for testing the developed prototype can be considered as initial in the context of the infrastructure of the retail chain Dar. The potential opportunities for the development of the prototype in the future include building an infrastructure for communication with cloud technologies, such as using cloud services of Microsoft Azure, Google Cloud or Amazon Web Services, enhancing analytics by developing individual algorithms through software languages such as Python and R and development with Robotic Process Automation methods of a robot that independently creates and executes orders to suppliers.

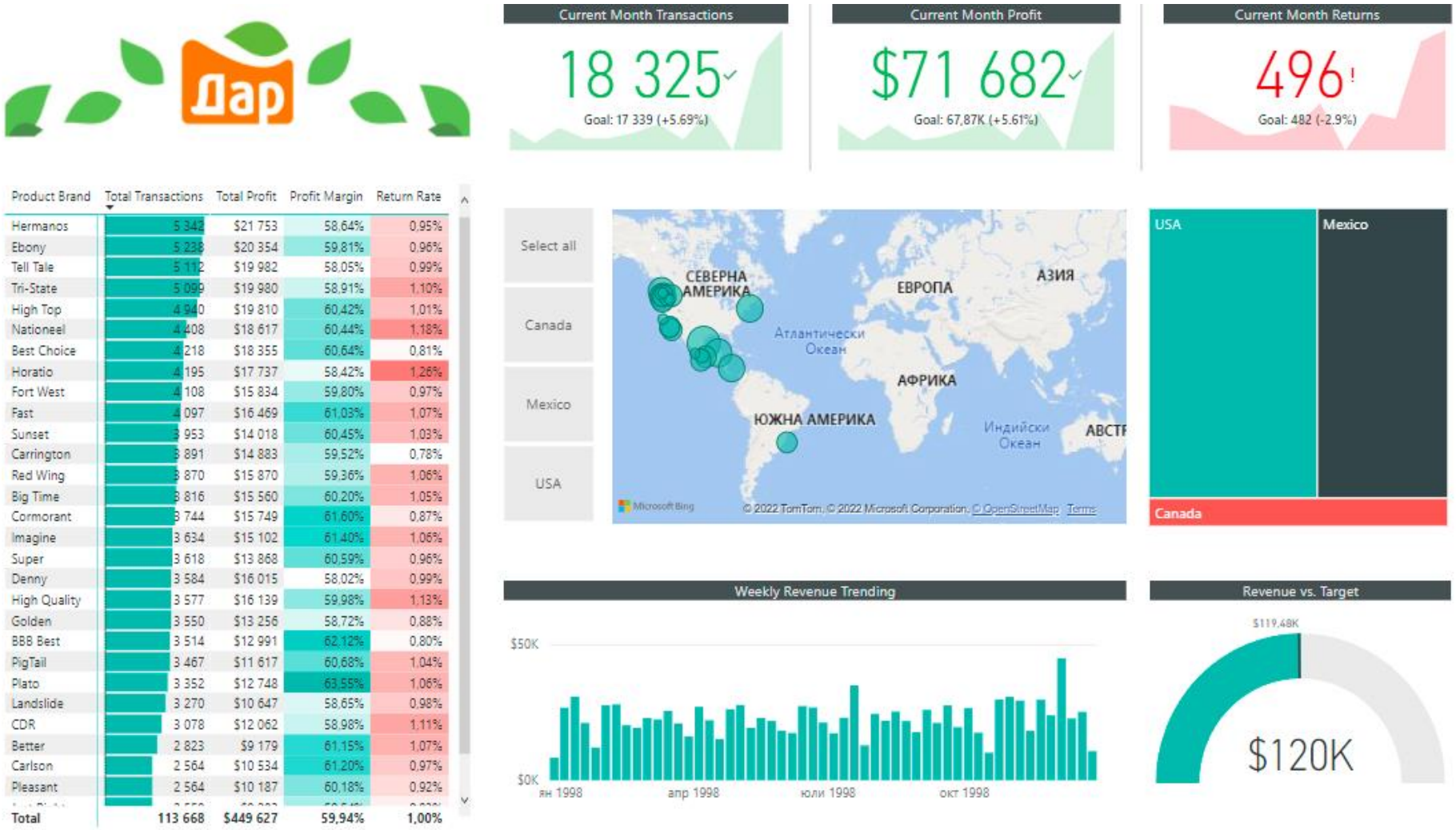

Fig. 15. Dashboard for tracking KPIs on a monthly and weekly basis

IV. REPORT ON THE CONTRIBUTIONS IN THE DISSERTATION

Based on the conducted research in the dissertation the following scientific and applied contributions can be indicated for achievement:

1) The possibilities for improvement of the retail information systems are investigated, as the thesis is presented and developed that using the potential of BI technologies and modeling tools can optimize the processes of managing orders to suppliers, and thus the overall functioning of trade business processes.

2) A conceptual business model of an intelligent module for management orders to suppliers is proposed. It includes a total of five components visually represented using UML and BPML standards.

3) Based on the diverse input data entering the business intelligent module, their profiling and processing is proposed through the created component for improving their quality.

4) A prototype of a business intelligent module for managing orders to suppliers has been developed with the application of a business intelligent approach and the use of statistical and machine learning methods.

5) The prototype of the intelligent module has been tested in an existing retail chain for fast-moving consumer goods based on real sales data. As a result, its capabilities for precise and accurate determination of the required quantities of goods to order from suppliers have been demonstrated.

V. PUBLICATIONS ON THE DISSERTATION

Studies

1. Parusheva, S., Pencheva, D. (2020). Model of a business intelligent system for management orders to suppliers in a retail chain. Annual Book. University of Economics - Varna, Varna: Publishing House Science and Economics, 90, 1, pp. 188 - 227. (In Bulgarian)

Articles

1. Pencheva, D. (2020). Use of Factors Related to the Consumption of Fast Moving Consumer Goods in Business Intelligence System for Managing Orders to Suppliers in Retail Chain. Izvestia Journal of the Union of Scientists - Varna. Economic Sciences Series, Varna: Union of Scientists - Varna, 9, 2, pp. 124 - 135.

2. Pencheva, D. (2021). Fast Moving Consumer Goods Retail: Business Intelligence Approach in Retail Information Systems. Egyptian Computer Science Journal, Cairo, Egypt: Egyptian Computer Society, 45(3), pp. 47-57.

Papers

1. Parusheva, S., Pencheva, D. (2021). Modeling a Business Intelligent System for Managing Orders to Supplier in the Retail Chain with Unified Model Language. In: Magdi, D.A., Helmy, Y.K., Mamdouh, M., Joshi, A. (Eds.) Digital Transformation Technology. Proceedings of ITAF 2020, December 16 – 17, 2020, Springer, Lecture Notes in Networks and Systems, Vol. 224, pp. 375-393. (Indexed in Scopus)