INFO- 2121 UDK : 349.22:3:654.07
Primljeno / Received: 2014-03-25 Izvorni znanstveni rad / Original Scientific Paper

BUSINESS INTELLIGENCE AND ANALYSIS OF SELLING IN RETAIL

POSLOVNA INTELIGENCIJA I ANALIZA PRODAJE U MALOPRODAJI

Sanja Bijakšić, Brano Markić, Arnela Bevanda
Faculty of Economy, University of Mostar, Mostar, Bosnia and Herzegovina
Ekonomski fakultet, Sveučilište u Mostaru, Mostar, Bosna i Hercegovina

Abstract
Business intelligence system is a set of software solutions, among which can be singled out three subsets: queries and reports, decision support systems and executive information systems. Research and analysis of huge amounts of data using appropriate techniques and methods can in organization diagnose essential processes, identify and anticipate the direction of change, interpret financial results, classify and cluster data to model the behavior of the system, aggregate data, detect changes and deviations to the objectives, determine the correlation between variables, generate association rules. This paper presents a business intelligence system for the analysis of sales in retail as an institutional form of an exchange process. Shown are all elements of the system: transaction data stored in tables of relational database, their extraction, transformation and loading into data warehouse and the use of appropriate data mining methods for the analysis. Planning of marketing activities uses the results of the analysis with the aim of increasing the sale and profitability.

Sažetak
Sustav poslovne inteligencije je, s informatičkog aspekta, skup softverskih rješenja među kojima se mogu izdvojiti tri podskupa: upiti i izvješća, sustavi potpore odlučivanju i izvršni informacijski sustavi. Istraživanja i analize velikIh količina podataka, uporabom odgovarajućih tehnika i metoda, mogu u organizaciji dijagnosticirati ključne procese, identificirati i anticipirati smjer promjena, interpretirati financijske rezultate, klasificirati i klasterirati podatke, oblikovati model ponašanja sustava, agregirali podaci, detektirati promjene i devijacije u odonosu na ciljeve, odrediti korelaciju među varijablama, generirati asocijativna pravila. Rad prikazuje sustav poslovne inteligencije za analizu prodaja u trgovoni kao institucionalnom obliku procesa razmjene. Prikazani su svi elementi sustava poslovne inteligencije: transakcijski podaci pohranjeni u tablicama relacijske baze, njihova ekstrakcija, transformacija i punjenje u skladište podataka i uporaba metoda data mining u analizi. Planiranje marketing aktivnosti koristi rezultate analize s ciljem povećanja prodaja i profitabilnosti.

1. INTRODUCTION

Business intelligence is both a process and a system in which outputs and knowledge are intended for decision-makers in organisational systems. In that process different data sources are used whereat some sources ensure structured and unstructured information /**1**/. Data can be qualitative and quantitative in different ways. Most frequently they are formatted and stored in data warehouse. Information for business intelligence can be in a form of structured reports (in advance known form), control dashboards, scorecards etc. Business intelligence is a result of an information process creation of useful information and knowledge that are necessary in decision-making and organisational systems managing. Besides, business intelligence is observed as a process of creation, analysis and distribution of knowledge and information. Therefore business intelligence uses appropriate tools that

software industry offers but also the solutions that are developed and adapted to the needs and requirements in organisation systems. Certain software solutions use business intelligence tools from data in relation databases, data warehouses, and structured and unstructured data (text) from internal and external sources. They create information and knowledge for easier and faster decision making. The term business intelligence is often equalised with or does not differ from the term knowledge management, data mining, and discovering knowledge in data. Although those terms are mutually correlated the sign of equality cannot be put between them. Business intelligence uses appropriate tools to provide decision makers with correct, timely and concise information necessary for decision making. Within the organisational system there can be and develop more different support solutions to decision making as well as solutions of business intelligence systems. Each solution is acceptable if it, with minimal expenses, provides decision makers with necessary information. Knowledge management is related to the all necessary activities that will, as much as possible, contribute to knowledge resources utilisation in organisational system. Those activities are creation, dissemination and use of knowledge. Artificial intelligence is focused on presentation of knowledge and its use. It is critical for knowledge management. Artificial intelligence is not only focused on explicit knowledge that can be relatively easy to formalise with some form of its presentation (manufacturing rules, semantic networks, triplet object of attribute value, frameworks, predicates of the first row). Relationships between business intelligence and other technologies that are directly related to business intelligence can be observed through five characteristics: inputs, nature of inputs, outputs, components, and users.

	Business intelligence	Knowledge management	Data warehouse	Data mining	DSS[1] and (ADS)
Inputs	Data, information	Data, information, knowledge	Data (from more systems)	Data	Data, information, knowledge
Nature of inputs	Internal or external, structured or unstructured	Internal or external, structured or unstructured	Internal, structured	Internal, structured	Internal or external, structured
Outputs	Information and explicit knowledge	Tacit knowledge and explicit knowledge	Data (in one logical warehouse)	Information	Decision proposals (in case of DSS) or automatic decision (in case of ADS)
Components	Information technologies	Information technologies, social mechanisms, structural arrangement	Information technologies	Information technologies	Informational technologies
Users	Within the whole organisation	Within the whole organisation	IT personnel	IT personnel and others trained for IT	Special, targeted users

Table 1: Differences between business intelligence and other related technologies /**2**/

[1] DSS is abbrivation from *Decision support systems* and ADS from *Automatic decision system*.

Business intelligence differs from other information technologies such as data warehouses, data mining, and decision support systems. Data warehouse and data mining are concentrated on the data. Data warehouse is subject oriented repository of data that is loaded from transactional databases such as relation tables that keep data from sales place using tools for data extraction and transformation. Data are from different sources and in order to be consistent they have to be transformed after extraction. Data mining also proceeds from data and using algorithms of their transformations such as cluster analysis, inductive rules, logistic regression associative rules, and others they intend to discover regularities and legalities that are within the data. Decision support systems use data and models to create information and knowledge at their exit (exit from the system). In that way data mining can be a part of a system to support decision making as well as data warehouse. All these technologies are mutually connected and compatible, and they have common aim to create information and knowledge in organisational system to improve quality of decision making, business functions' managing and whole organisational system managing. Difficulties in development and application of business intelligence solutions are numerous. First of all, training for application of business intelligence tools is expensive, business intelligence tools are complex to use, needs for business intelligence cannot be easily identified, it is necessary to customise business intelligence solutions to a user; benefits of business intelligence cannot be recognised in economic sense (speed of return of invested).

2. Business intelligence and data warehouse

Data for business intelligence algorithm application and key indicators of sales calculation are found in a relation database or data warehouse. Data warehouse are subject oriented, integrated, unchanged in time and only for reading stored data in function management support /**3**/. Inmon firstly defines four characteristics to which special attention should be paid with those databases.

Data in database are not oriented to applications as operational data which main task is to reflect everyday transactions.

"The data in the database are dynamic. Change over time, unlike the data warehouse, which is time–dependent, but fixed and subject-oriented information. " /**4**/ There is a chain of interrelated and dependent activities. The first is the selection of data, and then their filtering and cleaning (a process known as extraction (E), transformation (T) and filling (L)). The result of selecting the data often must be further transformed and processed by simply aggregating. Model of integrating unsupervised and supervised learning is clearly illustrated at the next figure:

The first activity is called selection (choice) data. The results of selecting are the target data that were selected from the data warehouse or database. These data are usually in the form of tables of relational model. The selection focuses on a subset of data of a database, a subset that is essential to the process of generating information. Of course this is not a simple activity" /**5**/. It proposes clear and defined information requirements. Filtration, followed by the treatment of data, aims to remove potential inconsistency of data, and to apply the algorithm for data mining.

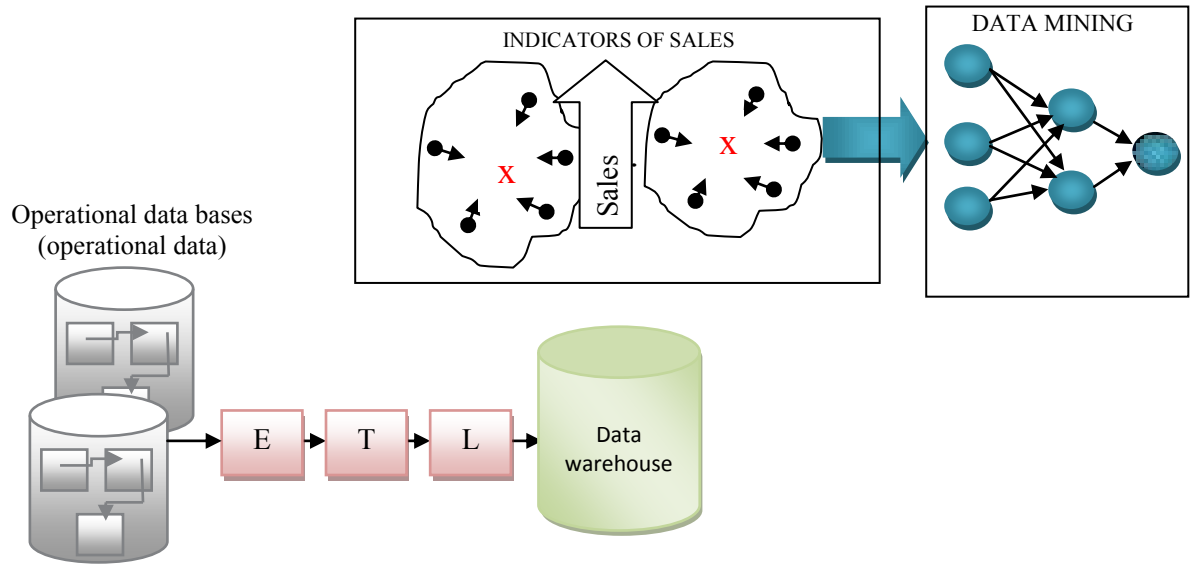


Figure 1: Business intelligence and data warehouse /**6**/

Data in data warehouse are related to different objects such as customers, products, regions. Those data that have operational character and serve only to run business process and do not serve to support decision making, they do not have application in data warehouses.

Data come from different applications. They have to be consolidated to be overtaken into data warehouse. That is firstly related to names (names of attributes), measures, and coding. The aim is only consistent and integrated condition of data on which different procedures and algorithms can be applied. Time is necessary dimension in data warehouse (unlike operational data). It enables data comparison in a company according to one additional dimension – time dimension. Data in a warehouse are kept longer (from 5 to 30 years) unlike operational data which are often deleted after processing. Therefore there are high requirements towards data warehouse with regards to data access, data storage techniques, and volume of stored data.

To build a business intelligence system that will be able to analyse sales rapidly and simply and then to provide marketing with information about changes and sales trends the global plan of data warehouse must be firstly set. Global plan is scenically shown by bus model (Figure 2).

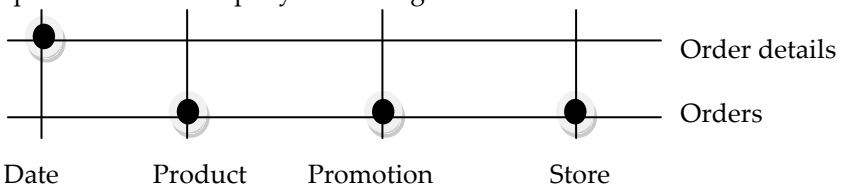


Figure 2: Bus model to show plan to develop data warehouse

The columns show all dimensions in the warehouse, while the rows show all factual (value) tables. Such global plan can then transform architecture of the data warehouse shown in the following entity relation diagram made for SQL server database:

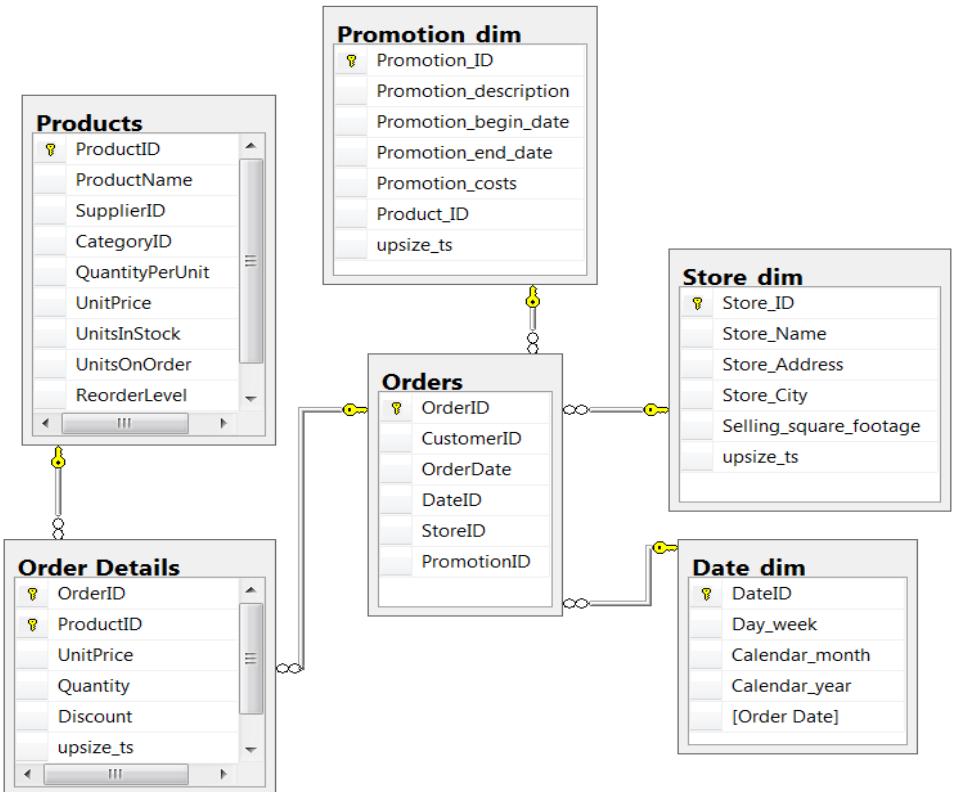


Figure 3: Data warehouse for sales analysis

Data warehouse keeps data about products, transactions (orders), time when transaction is made, promotion activities (start of promotion, duration of promotion, expenses of certain types of promotion): Value table is Order_Details in which data about product code, unit prices, quantities of sales at discount price. Dimension tables are Promotion_dim, Products, Orders, Date, Store_dim.

3. BUSINESS INTELLIGENCE USING INDICATORS OF RETAIL SALES

Based on dimensions in data warehouse and values (quantitative data) it is possible to shape more indicators that serve to analyse sales per products, shops, days, weeks, months, effects of promotional activities (for example discounts, prize competitions, advertising). Business intelligence shows information to decision makers in a simple form so decision maker uses information in order to bring decisions using minimal "efforts". Therefore it is necessary to visualise analysis results. Decision makers in marketing must constantly analyse the results of business activities based on information that get a form of an indicator of doing business results. Integrating business intelligence tools with data in data warehouse it is possible to form great number of such indicators. However, function marketing extracts one special set of indicators (metrics) to analyse sales and it consists of the following:

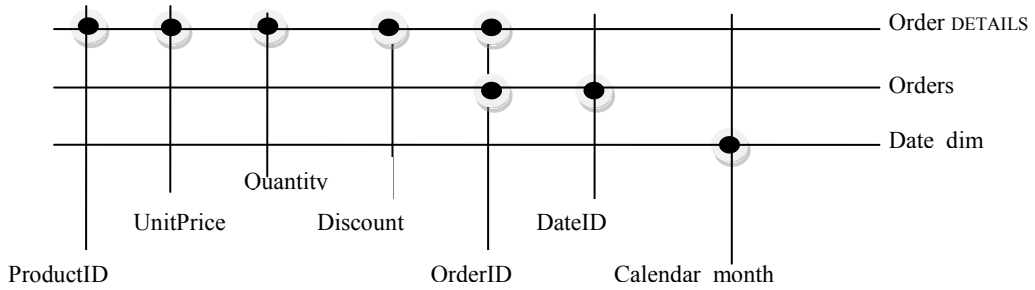
1. Total product sales (in currency) per days in a week for one month.
2. Product sales per days in a week (amount) for certain period of time (for example in January)
3. Sales of all products (value) per days in a week for certain period of time (for example the first quarter).
4. Sales per organisational units (for example shops).
5. Changes in sales quantities with regard to promotion expenses in certain period of time (for examples per weeks for certain quarter).
6. Changes in sales (value) with regard to promotion expenses in certain period of time (for example per weeks for certain quarter).
7. Identify connectivity (association) between certain products (algorithm of associative rules).

8. Number of orders per days in a week etc.

Business intelligence is therefore not only a system but also both set of tools and programmes that enable data warehouse customisation and data analysis by generating abovementioned indicators. Number of indicators is surely higher than eight. In this paper we will be limited to certain part of indicators but the logic of their customisation is similar. Data are stored in SQL server database, while business intelligence system uses R language and its packages that are applicable to business intelligence.

3.1 Agile business intelligence and sales indicators

The first indicator is the sales of all products, expressed in money, per days in a week for month of January. Such information that sales indicator has, business intelligence system can generate form appropriate data recorded in data warehouse using appropriate number of orders in R language. Necessary dimensions and their attributes are shown in the following bus model:



It is possible to ensure information from sales business process by integrating different software tools in business intelligence system. Besides data warehouse tables in SQL server in this paper we also use R language. Package sqldf() enables necessary questions over the database.

```
library(RODBC)
myconSQL<- odbcConnect("Bintell")[2]
library(sqldf)
library(tcltk)
qrODs<-sqlQuery(myconSQL, "Select * FROM
[Order Details]")

qrOs<-sqlQuery(myconSQL, "Select * FROM
Orders")
qrPRs<-sqlQuery(myconSQL, "Select * FROM
Products")
qrPMs<-sqlQuery(myconSQL, "Select * FROM
Promotion_dim")
qrDAs<-sqlQuery(myconSQL, "Select * FROM
Date_dim")
join_stringSQL<- 'SELECT qrDAs.Day_week,
sum(qrODs.UnitPrice * qrODs.Quantity -
qrODs.UnitPrice * qrODs.Quantity *
qrODs.Discount) AS Total,
qrDAs.Calendar_month
FROM  qrOs INNER JOIN qrODs ON
qrOs.OrderID = qrODs.OrderID INNER JOIN
qrDAs ON qrOs.DateID = qrDAs.DateID
WHERE qrDAs.Calendar_month="January"
GROUP BY qrDAs.Day_week'
salAmountSQL<-sqldf(join_stringSQL)
```

[2] The connection of R language with database SQL server requires the setup a user DSN data source pointing at our SQL Server using ODBC. The data source will be called from R using the package "RODBC". First has to be open *Control Panel->System Security-> Administrative Tools* and "ODBC Data Sources (32 bit)" and now is relatively easy to follow the next steps. Finally, after selection the default database „Bintell" it is possible to connect to database from R language by statement
mycon <- odbcConnect("Bintell").

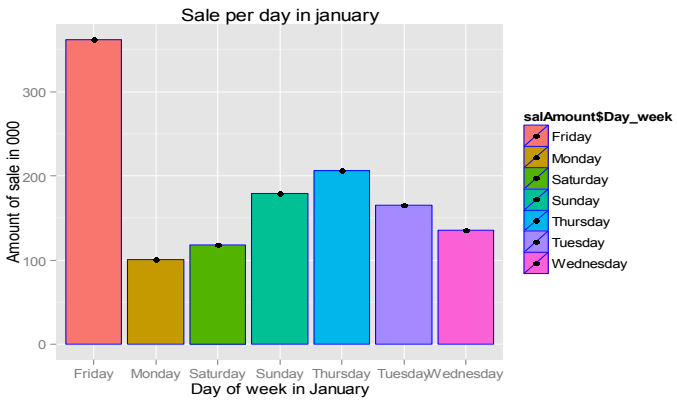


Figure 4: Sales per day of all products

Data frame ***salAmountSQL*** is a result of order Select SQL language application. It consists of three columns: days in a week (attribute Day_week), sales value of all products for every day in a week[3] expressed (in money) (attribute Total), and months in a year (attribute Calendar_month).

```
> salAmountSQL
  Day_week    Total Calendar_month
1 Friday 361509.2       January
2 Monday 100405.4       January
3 Saturday 117571.9      January
4 Sunday 178773.8       January
5 Thursday 206290.5      January
6 Tuesday 165136.0       January
```

Business intelligence always attempts to show results in a visual form to decision makers. Therefore it uses appropriate graphical presentations. Data can be visualised using function ggplot() of R language and appropriate arguments that are completely intuitive and are understood without additional explanations. Then demand for function ggplot() follows:

```
bar_graphSQL<-
ggplot(data=salAmountSQL,aes(x=salAmount
SQL$Day_week,y=salAmountSQL$Total/1000,
fill= salAmountSQL$Day_week)) + ge-
om_bar(colour="blue",stat="identity") +
xlab("Day of week in January") +
ylab("Amount of sale in 000")+ ggtitle("Sale per
day in january")
bar_graphSQL+geom_point()
```

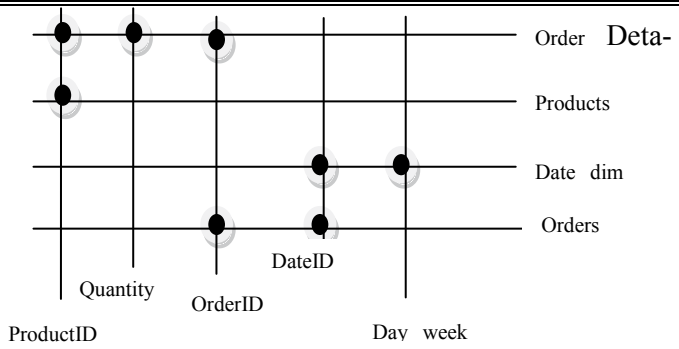
The result is two dimensional presentations of sales per days in a week in a form of bar graph.

in January It is simple to observe that product sales at sales spots are highest on Friday and the lowest on Monday. That information is important to marketing since it is the best to plan promotion activities for days of the biggest sales since the concentration of customers is highest then. However, marketing activities planning are never focused on one indicator. Marketing experts are also interested in other indicators. They are also interested in the amount of sales for every product per days in a week. In other words, they want to know how much in quantities per days in January are products sold to be able to compare them with the same month in previous year or other previous periods.

Business intelligence system must have an answer for such requirements and questions from marketing experts. Again, possible solution is application of a script made in R language. Appropriate SQL question that selects appropriate attributes of dimensions in data warehouse "Bintell" is reshaped. The answer of agile business intelligence on such question of marketing business function will be therefore shown in R language. Necessary dimensions and their attributes are shown in a new model of bus that has the following appearance[4]:

[3] Data are sorted in ascending order of days in a week.

[4] Bus columns are dimensions and value tables, and rows are attributes of those tables. In cross section there are attributes of appropriate SQL question that enables necessary operations of relation algebra to generate information about sales.



Based on visualisation of appropriate SQL question in a form of a bus it is necessary to write a small programme, script in R language that will approach to data warehouse (Bintell), create appropriate data set and then visualise dataset using function of the package ggplot()

```
>join_stringSQL1<-
'SELECT qrODs.ProductID, qrDAs.Day_week, sum(qrODs.Quantity) AS Total,
        qrDAs.Calendar_month
FROM  qrOs INNER JOIN qrODs ON qrOs.OrderID = qrODs.OrderID INNER JOIN  qrDAs
        ON qrOs.DateID = qrDAs.DateID INNER JOIN  qrPRs ON
       qrODs.ProductID=qrPRs.ProductID
WHERE qrDAs.Calendar_month="January"
GROUP BY qrDAs.Day_week, qrPRs.ProductID
ORDER BY qrDAs.Day_week'
```

```
>salAmountSQL1<-sqldf(join_stringSQL1)
> salAmountSQL1
```

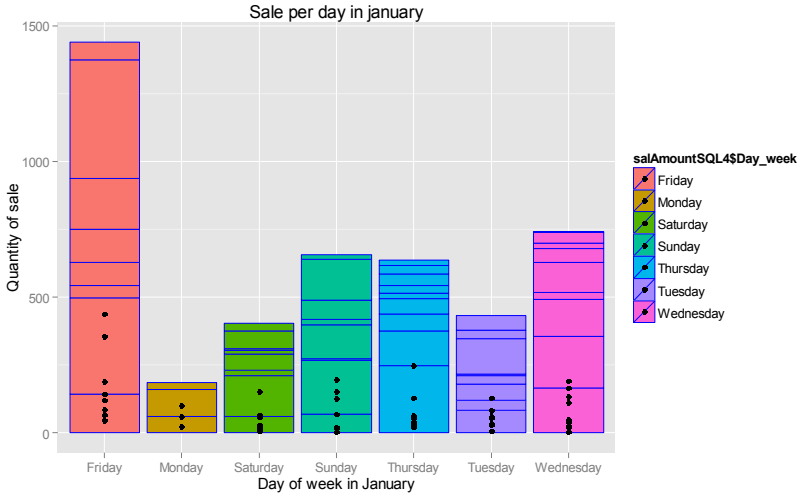
```
ProductID  Day_week  Total   Calendar_month
1      1          Friday   142    January
2      2          Friday   356    January
.....................................................................
76     1          Monday   60   January
77     4          Monday   100   January
78     8          Monday    24   January
.....................................................
139    1          Saturday    60   January
140    2          Saturday   150   January
```

Again, data in a form of table are not sufficient for business intelligence. Sales quantities are graphically shown per days in a week for January whereat those are product whose ProductID <10 and they belong to group 1 (category 1). It is sufficient to create new SQL question and demand ggplot() function of R language with appropriate parameters:

```
join_stringSQL4<-
'SELECT qrODs.ProductID, qrDAs.Day_week, sum(qrODs.Quantity) AS Total,
       qrDAs.Calendar_month
FROM  qrOs INNER JOIN qrODs ON qrOs.OrderID = qrODs.OrderID INNER JOIN  qrDAs
        ON qrOs.DateID = qrDAs.DateID INNER JOIN  qrPRs ON
        qrODs.ProductID=qrPRs.ProductID
WHERE qrDAs.Calendar_month="January"
GROUP BY qrDAs.Day_week, qrPRs.ProductID
HAVING (qrPRs.ProductID) < 10'
```

The orders of R language that follow enable visualisation of product sales which code (ProductID) is less than 10. Those are the products that belong to group 1 (the same group).

```
bar_graphSQL4<-
ggplot(data=salAmountSQL4,aes(x=salAmountSQL4$Day_week,y=salAmountSQL4$Total,
fill=salAmountSQL4$Day_week))+ geom_bar(colour="blue",stat="identity")+ xlab("Day of week in January") + ylab("Quantity of sale")+ ggtitle("Sale per day in january")
bar_graphSQL4+geom_point()
```



Again, "domination" of the products from the group one is visible on Friday and those sales surpasses all other days in a week. However, sales of products from the group 1 do not follow the same trend as the sales of all products since sales on Tuesday and Wednesday relatively differ with regards to expressed value of whole sales. Promotion activities and certain forms of integrated marketing communication should be focused and intensified on Wednesday and Friday during a week for the products from the group one (Product ID<10). Nowadays agile business intelligence in marketing is sine qua non of successful decision making, managing and making of business plans. Business intelligence can also include sophisticated algorithms to discover hidden relationships between variables. Business intelligence is exactly one of the key areas on which data mining, process of revealing and correlation explaining, patterns and trend analysis (sifting through large amounts of data). That process is based on CRISP methodology (Cross-Industry Standard Process of Data Mining) that includes six phases: business understanding, data understanding, data preparation, model building, testing and evaluation, and deployment. Data mining is focused on the operations of revealing anomalies in data (outlier detection), clustering, classification, associative rules, and regression analysis etc. Application of appropriate data mining algorithms is possible only if data are prepared in an appropriate way in data warehouses and then in form of datasets. That part of data preparation activity and business context understanding is the most complex and lasts the longest period of time.

CONCLUSION

One approach of agile business intelligence applicable in sales analysis has been shown in this paper. Business intelligence differs from knowledge management, data warehouses, data mining and knowledge support system according to the form of inputs, outputs, components and users. Agile business intelligence attempts to generate relevant information in fast and timely manner by using a concise form that is simple, acceptable, and understandable for a user. A part of business intelligence system was shown in this paper. It is used to analyse the sales in a retail shop as an institutional form or goods and services exchange. Business intelligence system uses technology and software to prepare data in a form of data warehouse ("Bintell" data warehouse in SQL server

database) and appropriate dataset that is extracted from it. After that such dataset is visualised using R language, and a user has simple presentation and sales process "diagnosis" and signals to define form and intensity of different types of communication with a market. SQL server base and packages of R language have shown simple and satisfactory development power for agile business intelligence.

Notes

/1/ Rajiv Sabherwal, (2007*), How do Knowledge Management Announcements Affect Firm Value?* A Study of Firms Pursuing Different Business Strategies, *IEEE Transactions on Engineering Management*, 54 (3): 409-422

/2/ Rajiv Sabharval, Irma-Becerra-Fernandez, (2010), *Business Intelligence, Practices, Technologies, and Management*, John Wiley & Sons.

/3/ Inmon, W.H., *Building the Data Warehouse*, John Wiley &Sons, 4th edition 2005.

/4/ John Maindonald and John Braun, (2003), *Data Analysis and Graphics Using R - An Example-Based Approach,* Cambridge University Press.

/5/ Ibid

/6/ Markić, B. (2011), Customer *segmentation by integrating unsupervised and supervised learning*, Proceedings from International Conference Economic Theory and Practice: Meeting the New Challenges, Faculty of Economics, University of Mostar, Mostar.