# **BI Big Data Analytics In Marketing**

Vanya Lazarova<sup>1</sup>

**Abstract:** Information is an important business resource just like money, materials, and goods. Over the years, the company was collected thousands of pieces of data for many processes. But this does not mean that automatically these data can be easily converted into information that could be analysed and that company managers, marketing staff, customers, suppliers could use. Transforming the multitude of collected big data into useful information, is a complex process that requires additional skills and knowledge. Using the right tools, marketing analysts in the company could use the collected data to develop effective and competitive strategies. The purpose of this paper is to explore the use of big data in marketing, the sources for its collection, and to present some BI technologies for the practical use of big data by the company's marketers.

**Key words:** BI, Big Data in marketing, tools for analytics, BI Analytics Technologies in Marketing. Impala tool

**JEL:** M31, O00

# Business Intelligence (BI) in Marketing

Information and related knowledge are a powerful decision-making tool for company development. Information only makes sense to accumulate if it can be processed with appropriate tools and provided to those who request it in a form that can best serve them. The accumulation of huge amounts of data must be accompanied by an increase in the company's ability to process this data. This is where business intelligence (BI) comes in. Business intelligence is a term that refers to data from a database, text extracting documents, the Internet, using special software and then analyzing this data and converting it into information that can be used to make informed business decisions and acting. The company's business analyst can get a complete picture of the business, see broader trends based on aggregated data, and get dependencies that he did not expect. The following sample questions are answered:

How do the current year's total sales of all products compare to the previous year's total sales? In which country, in the last five years, have the company's profits been the greatest? How many products were sold in two regions of the country/countries this month compared to the same month last year? For each customer age group, what is the yield distribution by product category? Who are the best sellers, distributors, suppliers, customers, partners?

Answers to these kinds of questions are not easy, nor can they be derived solely from the operational information that finance and accounting offer. For this purpose, something qualitatively new is needed, which only business intelligence can provide. Business intelligence systems (BI systems) can do a huge job: They extract information from various sources - internal company sources and external sources; they centralize, organize, and standardize information while cleaning and supplementing data as needed; provide

<sup>&</sup>lt;sup>1</sup> Assoc. Prof. Dr. Vanya Lazarova

University of National and World Economy – Sofia, Bulgaria Department of Information Technologies and Communications ORCID: <u>https://orcid.org/0000-0003-2849-4818</u> email: <u>vlazarova@unwe.bg</u>

analytical tools that enable a wide range of business and technical professionals to search for the information they need, discover patterns, and solve problems. Accumulation of big data in the company is the first step that will lead to the need for a BI system.

# **Big Data in Marketing**

According to some expert evaluation more exabytes than 2 (2,000,000,000,000,000 bytes) of data are generated in the world every day. The volume of data is growing at about 40% per year and is expected to reach 175 zettabytes (175,000,000,000,000,000,000 bytes) in 2025 (Reinsel D.; Gantz J.; Rydning J. (2018). Big data is characterized by a volume exceeding the capabilities of traditionally used software tools for recording, storing, and processing data in an acceptable time. While traditional tools evolving are and incorporating functionalities that attempt to handle big data, at the same time the scope of big data is also expanding. Big data now includes structured, semi-structured and unstructured data, with the main emphasis increasingly placed on the processing of unstructured data. Big data requires a set of new techniques and technologies capable of processing and analysing in real time hundreds of terabytes of data. Business Intelligence systems allow companies to deal with one of the main challenges of today - the huge amount of data (Мурджева, А.; Цанева, М.; Радоев, М.; Михова, В. (2010)). The term "big data" refers to data that is so large in volume, so complex, or requires such rapid processing that it is difficult or impossible to using traditional process methods (Йорданова, С.; Стефанова, К. (2019)).

The sources of big data in marketing are many and can be grouped into three main categories:

- Data generated by information systems and various transactional systems and websites.

- Data generated by machines and sensor devices.
- Social data from the population through social networks and other media.

Another qualifying feature is the orderliness of big data. According to this feature, big data is divided into:

- Structured created by a computer with human intervention or automatically, by machines and sensor devices, data from web logs, from websites.
- Semi-structured data stored in XML or JSON format
- Unstructured data from social networks, mobile phones, tablets, photos, text messages, etc.

A major source of big data in marketing is contain databases that operational information. This is structured information. This information if is stored, is archived, is updated and is maintained continuously in the database the company will be sure that all the information going back over the years will be available in a timely manner. Transactions are entered into the database and processed usually as soon as they occur, preferably in real time. Example - a software system in a large warehouse where data about goods is loaded as their labels pass through a barcode reader. Data is automatically entered and processed directly by the server. The computer immediately updates the stock information in the warehouse. This allows immediately after processing to be able to generate and issue electronic documents (orders, inquiries, receipts, etc.) with all the details of a given item. Other such systems are, for example, for data processing of customer orders. reservations, etc. \_ this is real-time transaction data that needs to be processed and analysed appropriately so that it can be used to its best advantage.

Machine-generated data is defined as information that is generated by industrial

equipment, sensors that are installed in machines, and even logs that track user behaviour. This type of data is growing exponentially as the Internet of Things (IoT) becomes more prevalent in a variety of fields. Sensors embedded in medical devices, smart meters, traffic cameras, satellites, household appliances and all sorts of other devices are delivering an ever-increasing volume and variety of data.

Social data includes likes, tweets, comments, photos, video, etc. that are uploaded and shared through the world's major social media platforms. This kind of data provides valuable information about consumer behaviour and sentiment and could have a huge impact on marketing analytics. The Internet is a massive source of social data, which greatly improves the volume of big data.

Structured data is data that has a clear, predefined structure that does not change or rarely changes. Structured data usually has a specific length and format. They follow schemas. These schemas outline where each data item is located and what it means. Structured data can be easily organized in tabular form. Most often, structured data is stored in relational databases and is accessed through standardized tools, such as the SOL language. Until recently, structured data was the only data used by businesses. As a result of rapid technological development, structured data has new sources that arise in real time and in large volumes. Structured data in marketing is collected by:

- Smart meters, medical devices, Global Positioning System data. This data is used for supply chain management and inventory control.
- Sales data with Barcode, QR code, RFID, NFC. This data is used for sales monitoring, supply management and inventory optimization.

Semi-structured data does not have as rigid a structure as structured data in a relational database. They do, however, have some

defining and constant characteristics that allow for some form of data organization. In addition to these constant characteristics, each of the objects about which data is stored can have many other properties that distinguish it from other objects. Semi-structured data is often represented as documents stored in Java Script Object Notation (JSON) or XML format. An example would be sending an email - date and time of sending, email addresses to and from, IP address from the device, sender, and other pieces of information are related to the actual content of the email. In this case, the characters that make up the email are not structured, but there are components that allow the data to be grouped based on certain characteristics and to be processed.

Unstructured data is that data that has no fixed structure and cannot be applied to any data model. The lack of structure makes this data much more difficult to store, search, process and analyze, which is why until recently it was rarely used by businesses (Йорданова, Ст.; Стефанова, К. (2017). The wealth of information in unstructured data is already available and today it can be processed with various algorithms and technologies. These technologies have elevated unstructured data to an extremely valuable resource for organizations. Most data stored on computer and other digital systems is unstructured. As unstructured, data from social networks, mobile phones, tablets, photos, text messages, etc. can be considered mostly. But, as stated above, characteristics can still be found that make this data processable, although their content is not structured.

The availability of big data implies the availability of technologies with which it can be processed and analysed to contribute real business benefit.

# Big Data BI Analytics Technologies in Marketing

There are various technologies for analysing big data in marketing, the commonly used of which are:

- Queries and Reports
- Online Analytical Processing (OLAP)
- Data Mining
- Text Mining
- Computer simulations, etc.

The usage of these technologies depends on the data type and subject area (Mihova, Veska. (2015), Stefanov, G. (2015), Marzovanova, M. (2015)). We will consider only the first of the technologies, with the emphasis on the analysis of the accumulated structured big data.

#### **Queries and Reports**

A query returns processed data retrieved from a database or data warehouse, while a report returns a representation of that data, based on a query. The query returns a set of data that usually looks like a spreadsheet rows and columns of data, and the report returns the data prepared for the end user with sums, averages, in a form suitable for viewing on a screen or for printing. Based on the queries, reports are created and made available to the end user in a readable, useful form. Queries should be generated in a manner to match the specific needs of users. There are many tools for querying big data stores – e.g., Google BigQuery, Amazon Redshift... An example of such a query mining technology tool is Apache IMPALA – interactive SQL for queries on big data. The tool is paid, but UNSS has free access so that marketing students could use the latest technologies.

We will look at an example using Apache IMPALA. A company has collected data on sales of goods, in different years, in different and countries regions. The data is accumulated in a flat csv file. Flat files are most often columns and rows written in a table. Flat files are widely used for data exchange between heterogeneous systems, between different operating systems and different database systems, as well as data source in data storage applications. The data collected by the company is approximately 50 million records, and each record contains a lot of data, the most important of which are: region, country, item, date of sale, unit price, quantity, sale price of the product. Fig. 1 shows part of the data - over 1 million records. They cannot be fully visualized due to the limited capabilities of the computer, but this does not mean that they cannot be processed.

Файл	Начало	Bytotta	іны Офа	рылен	10.143	страниците	s sopu	ули Дано	и Преслед	Variet	Tiowo	4			
2.	G	A.	Calibre	,	11	- A' A'	× =	- % E	(\$ fipe-ac		e	Other		-	
12.0	Floctas	3	ч К р	1 - 1	- 5	× - A -	E 8		E 🔛 Obrane	N N WAALDAD	e# =	\$ - %	an 🙀	-1 4000	сярено б азтиране т ко
Onese	Creto	1916 Fa		11pm	e	1	s		Cupaeumann		- 5	104	che :	14	
A10485	76	IXV.	fa Europ	pe											
	A		c	D		E		G	11	1	1	×		м	N
1048554	Europe	Finland	Beverages	Offline	1		7/21/2013	814078605	8/10/2013	2805	47.45	31.79	133097.3	89170.95	43926.3
1048555	Middle Ear	Qatar	Beverages	Online	H	3	1/27/2013	558969628	11/50/2013	389	47.45	11.79	18458.05	12366.31	6091.74
1048556	Midille Ear	Somelia	Vegetable	Online	C	1.1	1/30/2016	525303828	10/1/2016	7931	154,06	90.93	1221850	721165.8	500684
048557	Asia	Myanmar	Housebold	Online	£.		4/7/2015	892817533	4/24/2015	3403	668.27	502.54	2274123	1710144	563979.2
1040558	Sub-Sahar	Swadland	Beverages	Offline	H		1/27/2015	111559569	1/31/2015	1811	47.45	31.79	85931.95	57571.69	28360.26
1048559	Central An	Panama	Meat	Offline	н		9/3/2015	751891267	10/13/2015	189	421.89	364.69	79737.21	68926.41	10810.8
1048560	Sub-Sahar	Democrat	Baby Food	Offline	M	1	2/16/2012	954486453	12/24/2017	2607	255.28	159.42	665515	415607.9	249907
1048561	Europe	Malta	Beyerages	Online	c	8	2/12/2012	280331456	2/20/2012	5162	47.45	31.79	244936.9	164100	80836.92
1048562	Europe	Luxembou	Housebold	Online	H		6/9/2015	160725104	7/5/2015	985	668.27	502.54	658246	495001.9	163244.1
1040563	Middle East	Syria	Office Sup	Offline	M		2/18/2011	747278726	3/1/2011	5236	651.21	524.96	3409736	2748691	661045
1048564	Sub-Sahari	Sudan	Meat	Offline	ι.		2/30/2014	461916148	12/31/2014	108	421.89	364,69	45564,12	39386.52	6177.6
1048565	Middle Ear	Azerboijan	Cereal	Offline	H		8/1/2013	945873700	9/19/2013	9824	205.7	117.11	2020797	1150489	870308.2
1048566	Europe	Bulgaria	Fruits	Offline		- 1	1/13/2013	314700210	7/26/2013	5640	9.33	6.92	52621.2	39028.8	13592.4
1048567	Europe	Portugal	Snacks	Offine	H	100	1/24/2014	798257005	9/29/2014	842	152.58	97.44	128472.4	82044.48	46427.88
1040568	Asia	Tajkistan	Baby Food	Online	11	1.0	1/29/2014	200717914	7/18/2014	5746	235.28	159.42	1466839	916027.3	550811.6
1048569	Sub-Sahari	Chaff	Office Sup	Online	M		5/21/2011	445569264	7/7/2011	2702	651.21	524.96	1759569	1418442	341127.5
1048520	Australia a	Vanuatu	Beverages	Offline	-1		2/21/2016	826609885	3/4/2016	1089	47.45	11.79	51673.05	34619.31	17053.74
1048571	Sub-Saher	Cameroor	Cersul	Online			5/6/2010	257951104	5/27/2010	2319	205.7	117.11	477018.3	271578.1	205440.2
1048572	Sub-Sahar	Senegal	Clothes	Online	M		9/25/2810	424016749	10/12/2010	4834	109.28	35.84	528259.5	173250.6	355009
1048573	Europe	Portugal	Cosmetics	Offine	M		\$/20/2012	829543149	5/29/2012	9107	437.2	263.33	3981580	2398146	1583434
1048574	Sub-Sahari	South Alvi	Beverages	Coline	M		3/24/2015	149579131	5/5/2015	2644	47.45	31.79	125457.8	84052.76	41405.04
1048575	Europe	Denmark	Office Sup	Online	c		9/4/2012	279486024	9/29/2012	4980	651.21	524.96	3243026	2614301	628725
ULERS TO	Europe	San Marin	Darmana	Challers	144		nininaa	484033055	010/1010	1087	47.45	24.70	112100 4	40,565,077	43045 70

Figure 1: Flat data file - 1,048,567 records.

Based on these data, a variety of queries can be made:

• how much are the sales amount of a given product in a given country, in a given region for the entire period or for year;

• how sales change over the years in different regions;

• which are the best-selling items in different countries/regions, etc.

Using interactive SQL for queries on big data query can be made, for example, for Japan, in which to display all goods by sales amount.

SELECT item_type,	
count(DISTINCT order_id),	
round(SUM(total_cost),o) AS	to-
tal_cost	
FROM sales_1000000	
WHERE (country = 'Japan')	
GROUP BY item_type	
ORDER BY total_cost DESC;	

The illustration (Fig. 2) shows an example of all goods by sales amount and by product in Japan for the entire period.

≡ H\Ue		And a Second descendent.	John B 🕾 Barren Jackery
Co Co Electronic     Concentration     Concentration     Concentration     Concentration     Concentration     Concentration     Concentration     Concentration     Concentration	(109) + 2	1000 Database of white Type last + 0 1 interface of the type last + 0 1 interface of	Takes     File.     File.     File.     restart
<ul> <li>Marce, L., London</li> <li>Marce, M., Landon</li> <li>Marce, M., Landon</li> <li>Marce, M., Landon</li> <li>Marce, M., Marce, M., Marcee, M., Mar</li></ul>			deb. 4 Oper deb. 300 Operation articly 301 Operation articly 302 Operation articly 302 Operation 1010 (2000) 1010

#### Figure 2: Using Apache IMPALA technology tool to extract data

According to the needs of the end user, various reports can be displayed by making sections and illustrating them with suitable diagrams. Charts can be presented through different views. Summary graphic reports are best perceived by end user.

# Conclusion

Analysing data in marketing uses huge amount of collected information - internal company data and data from external sources. There are many tools for analysing and processing marketing information. The company must select the most suitable technological tool, considering the analysis the necessary reports, and the needs. preparedness of the staff. BI software should become an essential part of the company's software, and of course there must be trained specialists who can perform analytical analysis. Only then it can be said that the purchase and use of technological tools for BI analysis of big data in marketing is justified and contributes to the implementation of the company strategy.

# References

- Reinsel David; Gantz John; Rydning John (2018). The Digitization of the World From Edge to Core. IDC White Paper. <u>https://www.seagate.com/files/wwwcontent/our-story/trends/files/idcseagate-dataage-whitepaper.pdf</u>
- 2) Clissa, Luca (2021) Survey of Big Data sizes in 2021. <u>https://doi.org/10.48550/arXiv.2202.076</u> 59
- Йорданова, С.; Стефанова, К. (2019) Предизвикателствата на големите данни - същност, характеристики и технологии. София, УИ "Стопанство". Научни трудове на УНСС. 2019, т. 1, с.13-31
- 4) Йорданова, Ст.; Стефанова, К. (2017)
   Извличане на знания от неструктурирани данни чрез анализ на

мнението на потребители. София, УИ "Стопанство". Икономически и социални алтернативи. 2017, брой 1, с.13-27

- 5) Мурджева, А.; Цанева, М.; Радоев, М.; Михова, В. (2010) Бизнес интелигентни системи за оптимизиране и управление на приложни системи: Комуникации, информационни технологии и статистика. Актуални проблеми на теорията и практиката. София,
- 6) Marzovanova, M. (2015) Intelligent Tagging and Search as a Fully Automated System. Economic alternatives. 2015, Issue 2, pp. 95-100
- 7) Mihova, Veska. (2015) Methods of Using Business Intelligence Technologies for Dynamic Database Performance Administration. Economic Alternatives. 2015, issue 3, pp. 105-116
- 8) Stefanov, G. (2015) Methods for Heterogeneity Detection During Multi-Dimensional Dat