IMPACT OF BIG DATA ON THE RETAIL INDUSTRY

A. Seetharaman*, Indu Niranjan*, Varun Tandon*, A. S. Saravanan**

*S P Jain School of Global Management, Singapore-Sydney-Dubai-Mumbai
**Taylors University, Malaysia

Abstract

With the recent emergence of Big Data with its Volume, Variety and Velocity (3V's), data analysis has emerged as a crucial area of study for both practitioners and researchers, reflecting the magnitude and impact of data-related problems to be resolved in business organizations, including the retail industry. This study has methodically identified and analysed four factors, namely, data source, data analysis tools, financial and economic outcomes and data security and data privacy, to gauge their influence on the impact of Big Data in the retail industry. This research analyses the impact of big data analysis on retail firms that use data and business analytics to make decisions, termed a data-driven decision-making (DDD) approach. The new finding is arrived that financial and economic outcome showed a strong support and have direct relationship with data analysis tools of retail industry. Data for the study were collected using a survey of various business practices and investments in information technology by retail organizations. The data analysis showed that retail organizations which use DDD have higher output and productivity. Using SMART PLS data analysis methods with solid support of review from ISI Journals, the relationship between DDD and performance is also evident in aspects of organization such as the utilization of inventory, customer engagement and market value in the retail industry.

Keywords: Data Source, Data Analysis Tools, Data Security, Financial and Economic Outcomes, Retail Industry

1. INTRODUCTION

The world is inundated with data and this is increasing exponentially day by day. Computer systems store vast amounts of data. Researchers at the University of California, Berkeley, recently estimated that approximately 1 Exabyte (1 million terabytes) of data is generated annually worldwide, 99.97% of which is available only in digital form (Keim, 2009). The marketing industry is teeming with data captured by companies and the rise of social media, multimedia and the Internet will add exponential growth in the near future (Manyika et al., 2011).

The retail industry is one of the largest sectors in the world. This industry is expected to grow as the middle classes are increasing substantially in size and in buying power. Retail purchases via e-commerce and m-commerce are growing at a high rate due to the advent of high-speed internet connections, advancements in Smartphone technology and online-related technology, improvements in the product lines of e-commerce firms, a selection of delivery options and better payment options (Keim, 2009; Wixom, B.H. and Watson, H.J., 2001). It is estimated that consumers and large organizations generate 2.5 billion GB of data yearly and this is increasing at the rate of 40% year on year (Manyika, J. *et al.*, 2011).

This growth in data is possible with the advent of high-speed Internet access and the availability of new data types for data analysis. The introduction of these data types has become possible with the introduction of Smartphones, tablet computers and other electronic devices. The data are collected because retail companies – including those engaged in some kinds of e-commerce – view them as a source of potentially valuable information, which, as a strategic asset, could provide competitive advantage (Keim, 2001). These retail data from Big Data are a powerful means of creating a way forward for marketers to accomplish their objectives in an effective manner.

Business intelligence and analytics (BI&A) and the related field of big data analytics have become increasingly important in both the academic and business communities over the past two decades (Chen *et al.*, 2012; Watson, H.J. *et al.*, 2007). In the rising tide of retail business transaction data, these tools help distinguish what are strategic assets and what are not worth collecting in the first place (Keim, 2009). The analysis of these new data types can make the decision-making process more effective in marketing. Until recent times, appropriate software tools and algorithms were scarce in marketing research (Baier, D. and Daniel, I., 2012). However, with the advancement in technology, software tools and algorithms coupled with velocity of Big Data are now available to analyse content uploaded at different locations in the form of images/snaps, music, or video. For innovation, growth and to excel in competition, analysis of these large data sets (also known as big data) plays an important role (Manyika et al., 2011). Big data not only have an impact on data-oriented managers, data analysts, etc., but also on the entire retail sector which will increasingly have to cope with the Volume, Variety and Velocity (3V's) of big data.

Business analytics was identified as a technological trend in the 2010s by the IBM Tech Trends Report (2011). In a survey of the state of business analytics by Bloomberg Businessweek (2011), 97% of companies with revenues exceeding $100 million were found to use some form of business analytics. Such analysis has been facilitated by the advent of advanced tools for data analysis and techniques such as visual data exploration, which allows the visualization of data to gain insights and come up with new hypotheses. In addition to granting the user direct involvement, visual data exploration has several key advantages over automatic data-mining techniques in statistics and machine learning (Keim, 2009). Other powerful data analysis tools, such as IBM's SPSS, SAS, Hadoop and Big Data, can also enable users to undertake analysis using a combination of statistical tools (Russom, P., 2011), technology and a process of strategic thinking in marketing.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

A literature review was undertaken encompassing academic and research papers. This section reviews Big Data as undertaken by organizations and then addresses several key aspects identified from the literature review, leading to the development of the research hypotheses.

Data analysis in organizations employs a qualitative approach that includes statistical procedures. The analysis comprises an ongoing iterative process in which data are continuously collected and analysed in real time. To analyse the data, patterns in the data are sought during the data collection phase (Savenye W.C. and Robinson R.S., 2004). The qualitative approach adopted determines the form of the analysis to be performed. For example, the types of data that can be analysed using content analysis include field notes, documents, audiotapes, videotapes, etc.

Maintaining data integrity is an import aspect of Data analysis; similarly, accurate and appropriate analysis of the data is also essential in maintaining data integrity to derive the research findings. Data integrity issues that include statistical and non-statistical data are relevant in data analysis. Incorrect data analysis will have a negative impact on the scientific findings and the public perception of research (Shepard, 2002).

The evolution of data analysis methods employed in organizations are listed below (ISACA, 2011)

- Ad-hoc – This data analysis technique is used in the initial investigation, mainly to support a specific project. This type of analysis technique is rarely applied directly to live systems or production systems. The technique is highly dependent on the skills of the individual.
- Repeatable – This data analysis technique is predefined and scripted to perform the same tests on similar data. Data access tools may be used to import data directly from production systems. The technique is less dependent on the individual and the acquisition process is automated to improve the output of data analysis.
- Centralized analytics – For development, operation and data storage purposes, a centralized approach should be developed. Standards for the development of data analysis from Big Data are documented. Batch jobs are created for the applications to run the data analysis against the centralized storage location. Data can either be pushed or pulled from different sources.
- Continuous monitoring – Data analytics, referenced to the centralized storage of Big Data, is a continuous process using automated jobs. These jobs are monitored and maintained by operations teams with the help of technical teams.

Turning to the key aspects of Big Data analysis in the retail context identified in the literature, four focal areas are discussed in turn: i) data source; ii) data analysis tools; iii) data security and data privacy; iv) financial and economic outcomes.

2.1. Data source

For many years, data collection and analysis have proven helpful for marketing purposes (Ashley, C. and Noble, S.M., 2013). Marketing researchers and practitioners collect data and analyse them (Baier *et al.*, 2012). Data are collected because companies, including those engaged in some kind of e-commerce, view them as a source of potentially valuable information which, as a strategic asset, could provide competitive advantage (Keim, 2009; Wixom, B.H. and Watson, H.J., 2001). The data that are collected by businesses about their customers are one of the greatest assets of the business (Ahmed, 2004).

The quality of data from a source largely depends on the degree to which they are governed by schema and on integrity constraints controlling permissible data values (Van Till, S., 2013; Wu, J., 2002). Vast quantities from Big Data are used as raw material to enable searches using data analysis tools and group the data according to desired criteria that could be useful for future targeted marketing (Ahmed, 2004).

New data types are becoming available for data analysis and classification in marketing with advancements in Smartphones, tablet computers and other equipment (Baier *et al.*, 2012; Chiang et al., 2012). In today's world of social media, customer perceptions can very easily be accessed across the Internet in the form of blogs and on-line forums. These customer data are mostly sought by retail organizations to gauge the extent of positive or negative perceptions (Cerchiello and Giudici, 2012) and also for quick decision support (Lohr, S., 2012; Graen, M., 1999). The analysis of these data makes a considerable difference to companies. The valuable information obtained could make significant difference for an organization in running its business, the way it interacts with current and prospective customers and enabling it to gain a competitive edge on their customers (Ahmed, 2004).

Based on the above, the following hypotheses are proposed:

H1. The data source, used as the basis for performing data analysis and comprising data from various sources, has a positive impact on the use of data analysis tools.

H2. The data source, used as the basis for performing data analysis, has a positive impact on data analysis.

2.2. Data analysis tools

Data analysis tools are used to extract buried or previously unknown information from large databases using different criteria, allowing the discovery of patterns and relationships (Ahmed, 2004). Data analysis tools can be divided into data-profiling and data-mining tools (Written, I. H. *et al.*, 2011). A large number of commercial tools support the extraction, transformation and loading (ETL) process for data warehouses in a comprehensive way (Turban E. *et al.*, 2008). In short they solidly support Volume, Variety and Velocity (3V's) of Big Data.

In the 1980s, factor analysis became one of the more widely used procedures in the arsenal of analytic tools for market research (Stewart D.W., 1981). Core variables were selected from the collected data and then analysed. When the amount of data is so large as to be beyond comprehension, factor analysis can be used to search data for qualitative and quantitative distinctions (Stewart D.W., 1981).

Companies use multiple data sources to undertake analyses. Drawing data from multiple sources and employing various types of analysis can provide robust findings and overcome the risk of method bias (Davis *et al.*, 2011). Advancements in technology have helped researchers analyse and group respondent data (Lohr, S., 2012) available in the form of videos, images or audio files by using different algorithms and software (Berry, M.J.A. and Linoff, G. (1997); Baier *et al.*, 2012). Text categorization has become one of the key techniques for handling and organizing data in textual format with the rapid growth of online information (Cerchiello and Giudici, 2012).

The information that is derived from such analyses can be used for decision support, prediction, forecasting and estimation to make important business decisions (Sallam, R.L. *et al.*, 2011). Indeed, they can help a business to gain a competitive edge (Ahmed, 2004). In recent years, business intelligence tools and technology are improving and all business are taking advantage of this situation (Chaudhuri S. *et al.*, 2011). Big data analytics has come to be considered the most advanced data analysis technology, it helps in.

Based on the above, the following hypothesis is proposed:

H3. Data analysis tools, used to analyse data collected from different sources, have a positive impact on the data analysis performed.

2.3. Data security & data privacy

The data that the companies collect regarding their customers are one of the greatest assets of the company. The security and privacy of these data are important concerns from the companies' point of view as well as from the customers' point of view (Van Till, S., 2013; Wu, J., 2002). Data security and data privacy are also important aspects of electronic commerce (Acquisti, 2004). A PWC study in 2000 stated that nearly two thirds of the consumers surveyed would shop more online if they knew retail sites would not do anything with their personal information.

As the Internet develops and matures, its success will depend in large part on gaining and maintaining the trust of visitors. This will be of paramount importance to sites that depend on consumer commerce (McKnight D.H. *et. al.*, 2002). The development of trust not only affects the intention to buy, but it also directly affects the effective purchasing behaviour in terms of preference, cost and frequency of visits and therefore the level of profitability provided by each consumer. In addition, analyses show that trust in the Internet is particularly influenced by the level of security perceived by consumers regarding the handling of their private data (Wu, J., 2002; Flavián and Guinalíu, 2006).

Thus, the following hypotheses are proposed:

H4. Ensuring data security and data privacy has a positive impact on financial and economic outcomes.

H5. Ensuring data security and data privacy has a positive impact on the data source.

H6. Ensuring data security and data privacy has a positive impact on the effect of data analysis in the retail industry.

2.4. Financial & economic outcomes

The data derived from consumer transactions is increasing by 40% per year (Johnsen, 2013). Making sense of these data makes a considerable difference for businesses. Such analysis will potentially give companies a competitive advantage (Smith, 2014; Lohr, S., 2012). To develop better decision making, retail companies are using new technology, such as big data analytics, not to build massive databases or develop costly technological products, but to help in identifying five to ten combinations of existing and new data sources that can drive better decision making when combined with sophisticated real-time analytics (Johnsen, 2013; Sallam, R.L. *et al.*, 2011). As noted by Andrew Appeal, IRI's new CEO, at the company summit in March 2014 in Las Vegas, data analysis technologies, including big data analytics and other services, will reach \$17 billion by 2015, from a base of just \$3 billion in 2010 (Johnsen, 2013). This is a modest estimate.

Companies leveraging average analytic capabilities are 20% more likely to provide higher returns for their stakeholders than their non-analytic-orientated competitors; companies that use advanced analytic capabilities, such as those using big data, are 50% more likely to provide higher returns (Johnsen, 2013; Savitz, E., 2012). According to IRI, the retailing industry could potentially see more than $10 billion in annual value created as a result of the improved application of advanced analytics to support brands and channels.

Thus, the following hypotheses are proposed:

H7. Financial and economic outcomes have a positive impact on the use of data analysis tools.

H8. Financial and economic outcomes have a positive influence on the impact of Big data analysis in the retail industry.

The findings from the literature review are summarized in terms of the flow of related research over the years as illustrated in Table 1 below. This table captures how retail organizations have used data analysis since advancements in technology for data capture and data analysis and it shows the impact of big data analysis on organizations' growth. Moreover, it illustrates the advancements in the technology and tools available for data analysis, as well as the importance that organizations accord data analysis in taking operational and financial decisions. This table gives the chronological contribution of various experts in the field with a view to identify relevant independent variables influencing the Data Analysis through Big data. At the end the Table further illustrates that the findings from this paper strengthen and enrich the earlier findings further it also introduces new variable on Financial & Economic Outcome, which is distinctly different from earlier contributions.

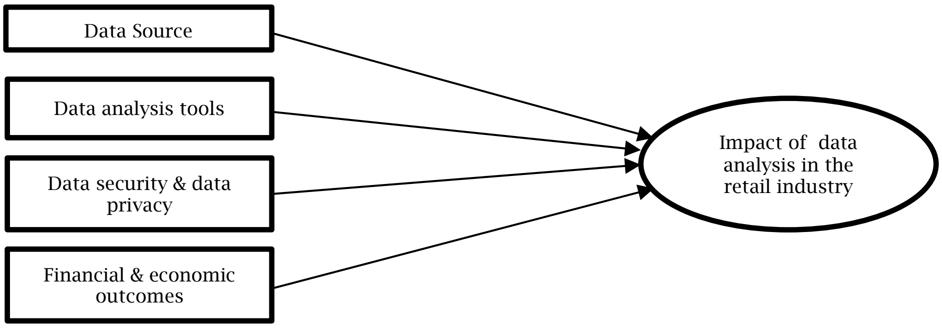
Table 1. Comparison of the impact of big data analysis in the retail industry

Study	Data Source	Data Analysis Tools	Data Security & Data Privacy	Financial & Economic Outcome	Impact of Big Data Analysis on Retail Industry
Donthu and Yoo (1998)	Data from various stores related to customers, collected and used for analysis	Data envelopment analysis (DEA)	No	Helps to measure productivity in the retail industry	Organizations should use data analysis techniques to measure retail productivity
Ahmed (2004)	Businesses collect data on their customers	Data mining tools	Confidentiality and individuality of the common man should be preserved	Data mining can help businesses plan for the peak periods of consumption, irregular transactions, etc.	Improvement in the market. Organizations should be able to react quickly to changes
Brynjolfsson *et al.* (2011)	Survey data on 179 firms	Instrumental variable testing	No	Firms adopting DDD have 5–6% higher output and productivity	Particularly for large firms, DDD is worth adopting
Baier *et al.* (2012)	New data types available in market with the advancement of Smartphones, tablet computers and other equipment	Analysis of image data is possible with the help of the SPSS-like software package IMADAC	No	Organizations exploring advanced data analysis technology	Grouping of content is possible with the use of proper analytic tools
Wyner (2013)	N/A	Discussion of customer relationship management (CRM)	No	Data can help in predictive analysis, customer segmentation and also have operational and financial benefits	Used by companies to improve customer experience and increase profitability
Falcioni (2013)	N/A	Discussion of big data	No	Data analysis can predict purchasing trends, predictive analysis and operational benefits	Helps companies to plan for production, reduce inventory and improve margins
This research paper	With advancements in technology, new data sources are available for analysis	Advanced data analysis tools like big data analytics are helping in the analysis of complex data sources	Technology is helping companies to protect the security and privacy of data	Data analysis has a positive impact on the financial and economic outcomes of retail companies	Users have demonstrated a strong inclination for research variables for the impact of the Data Analysis on the Retail industry

3. RESEARCH METHODOLOGY

The research methodology was developed after the completion of the literature survey largely from ISI Thomson listed journals. The research methodology is centred on the variables identified that appeared repeatedly during the literature review. A meaningful direct and indirect relationship between independent variable and dependent variables was identified and a research model created to understand their dominance, as illustrated in Figure 1.

Figure 1. Research framework



3.1. Data collection

Secondary data were obtained by following the ground research (Glaser & Strauss, 1967) from the literature survey, resulting in four independent variable. Then, a survey questionnaire was developed to gather primary data on the impact of big data analysis in the retail sector from various individuals working for global medium-sized and large retail organizations. The questionnaire was piloted through personal interviews with 20 respondents and selected experts from medium-sized and large retail organizations (Departmental stores, super markets and online retailing etc.) to obtain feedback. Based on the responses from these interviews, the questions were restructured and the final survey questionnaires were sent.

Table 2. Demographic characteristics of respondents

Survey Participants (n = 238)		
No. of employees in analysis team		
Less than 10	68	28.6%
10–49	43	18.1%
50–99	28	11.8%
100–300	31	13.0%
More than 300	68	28.6%
Geographic region for data analysis		
Asia	128	53.8%
Australia	19	8.0%
Europe	34	14.3%
US	45	18.9%
UAE	12	5.0%
Annual revenue		
Less than USD 100,000	24	10.1%
USD 100,001 to USD 1 million	17	7.1%
USD 1 million to USD 10 million	26	10.9%
USD 10 million to USD 100 million	31	13.0%
USD 100 million to USD 1 billion	10	4.2%
More than USD 1 billion	130	54.6%
Market share of the organization		
Less than 10%	69	29.0%
10–29%	90	37.8%
30–50%	41	17.2%
More than 50%	38	16.0%
Type of data analysis performed		
Customer data analysis	28	11.8%
Report generation	93	39.1%
Financial report data analysis	12	5.0%
Transaction data analysis	97	40.8%
All of the above	8	3.4%
Frequency of data analysis		
Daily	65	27.3%
Weekly	41	17.2%
Monthly	83	34.9%
Quarterly	44	18.5%
Half yearly	2	0.8%
Yearly	3	1.3%
Organizations' spend on data analysis compared to marketing spend		
Less than 5%	107	45.0%
5–19%	79	33.2%
20–40%	31	13.0%
More than 40%	21	8.8%
Organizational growth experienced with data analysis initiative		
Less than 5%	77	32.4%
5–9%	71	29.8%
10–14%	26	10.9%
15–20%	37	15.5%
More than 20%	27	11.3%

The survey was undertaken using a questionnaire based on the four independent variables identified from the literature survey, covering both quantitative and qualitative aspects. The final survey was sent only to those aware of data analysis and within the retail industry. At the start of the questionnaire a brief description was provided, highlighting the purpose of the research together with an assurance regarding the confidentiality of the data collected. The questionnaire was divided into various sections, each containing questions to obtain information related to the variables identified. For each variable, three to five questions were formulated using a five-point Likert scale to capture the use and adoption of data analysis by the companies in taking decisions. Each question is solidly supported by literature mostly from ISI Journals.

The final online survey was then sent to a vast number participants using personal contacts via email. The profile of the respondents were people who belong to retail industry particularly Super Markets, Departmental Stores and online retailing (e-Commerce). A total of 284 responses were received, of which only 238 were found useful for analysis due to missing data. All the Likert-scale questions were mandatory and there were other optional items that were optional, such as comments. Table 2 summarizes the demographic characteristics of the 238 respondents. The respondents in this research were from developed countries around the globe.

3.2. Data analysis

SmartPLS (Wende *et al.*, 2005; Ringle *et al.*, 2005) software was used for the following:

- To analyse the model
- To test the hypotheses developed
- For path modelling with latent variables
- To measure the validity and reliability of the constructs

SmartPLS uses the partial least squares (PLS) technique, a component-based approach for examining and testing theory without imposing any normality condition on the data (Hulland, 1999). PLS is also useful for relatively small amounts of data and when the data are skewed (Wong, 2011). PLS makes no assumption about data distribution (Vinzi *et al.*, 2010) and removes the problem of undesirable solutions (Löhmoller, 1989; Wold, 1989). Structural equation modelling (SEM) is a component-based estimation method (Tenenhaus, 2008) that allows testing of both built theories and concepts (Rigdon, 1998; Haenlein & Kaplan, 2004). The structural modes should be compatible with experimental designs (Bagozzi, 1980). Hoyle (1995) suggested using a sample size of 100–200 for path modelling analysis. The sample size used for analysing this model is 238.

The analysis was done in stages: during the first stage the structural model was estimated to assess the quality of the measures, followed by hypothesis validation using the structural model (Jöreskog and Sorbom, 1993). Browne *et al.* (2002) suggested validating the model in the first stage before examining the hypotheses. Reliability (consistency of measures) and validity (measure of concept) are the two criteria for testing the measures (Sekaran and Bougie, 2010).

3.2.1. Reliability

To evaluate the reliability and consistency of the model, composite reliability and Cronbach's alpha were used. Composite reliability is a comprehensive estimate of reliability (Chin and Gopal, 1995). For adequacy, the constraints are a Cronbach's score of 0.6 and above (Hair J.F. *et al.*, 2012) and a value greater than 0.7 for composite reliability is recommended (Gefen *et al.*, 2000). Table 3 shows the composite reliability values are greater than 0.7 and the Cronbach's alpha values are above 0.6. This shows that the model is robust and reliable.

Table 3. Reliability validation for latent constructs

Overview	*AVE*	*Cronbach's alpha*	R^2
Data analysis tools	0.5814	0.6399	0.6671
Data security & data privacy	0.7494	0.8326	0
Data source	0.6951	0.9505	0.6763
Financial & economic outcomes	0.591	0.9458	0.6115
Impact of Big data analysis	0.684	0.9575	0.7163

3.2.2. Convergent validity analysis

Convergent validity is the measure of variable indicators. This measures the extent of conformity between scores. Construct validity is also tested with the help of convergent validity (Straub et al., 2004; Fornell and Larcker, 1981). A value above 0.7 is considered the ideal value for the convergent validity of each item (Chin, Marcolin, & Newsted, 2003) and the average variance extracted (AVE) for each construct should be above 0.5 (Barclay *et al.*, 1995).

As shown in Table 3, the minimum AVE value is 0.58, which is above 0.5 as required. Also, as shown in Table 4 the loading reliability indicator is more than 0.7 (Chin *et al.*, 2003). Moreover, the loading constructs for all items are above 0.5, so the measurement model satisfies the requirements for convergent validity.

Table 4. Results for reflective outer models

Construct	*Loadings (indicator reliability) (min–max)*
Data analysis tools	0.7304–0.7908
Data security & data privacy	0.8321–0.8913
Data source	0.7197–0.9052
Financial & economic outcomes	0.7104–0.8729
Impact of Big data analysis	0.7356–0.9364

Table 5 compares the item-to-construct correlation against the correlations with other constructs.

3.2.3. Discriminant validity

Discriminant validity is used to assess the degree of discrimination between variables. Smart PLS was employed to validate discriminant validity by comparing the measured value for each variable with other constructs; if a weak correlation is resulted, discriminant validity is established (Hulland, 1999). Also, following Compeau *et al.* (1999), the average variance for an indicator should be greater than the variance of other variables. To calculate discriminant validity, the square root of the AVE value for each indicator is calculated and then compared to the AVE of other variables. The value of the square root of the AVE should be greater than the AVE of other variables (Fornell and Larcker, 1981). Table 6 shows the results of the discriminant validity testing, with the square root of AVE in bold.

Table 5. Comparison of item-to-construct correlation against correlations with other constructs

Construct	Item definition	Data analysis tools	Data security & data privacy	Data source	Financial & economic outcomes	Impact of big data analysis
Data analysis tools	Cost analysis	0.7908	0.4664	0.4865	0.7701	0.4863
	Advanced	0.7304	0.5461	0.555	0.5261	0.7839
	Specific	0.765	0.3506	0.6051	0.5695	0.4921
Data security & data privacy	Data security important	0.4703	0.8321	0.7346	0.6162	0.5341
	Data privacy important	0.5211	0.8727	0.6641	0.6274	0.5976
	Latest technology	0.567	0.8913	0.7356	0.7772	0.587
Data source	Research	0.6649	0.6149	0.7535	0.6791	0.7981
	Mobile devices	0.5178	0.5071	0.7197	0.546	0.5353
	Internet	0.5375	0.8315	0.8671	0.6993	0.5898
	Other internal & external sources	0.5821	0.67	0.8614	0.6647	0.7093
	Promotional data	0.5488	0.6136	0.8487	0.7456	0.6047
	Consumer buying patterns	0.6107	0.7491	0.9052	0.752	0.6863
	Store video data	0.6199	0.8455	0.8634	0.7983	0.6204
	Customer feedback	0.6244	0.5961	0.7571	0.6057	0.5835
	Inputs from employees	0.6896	0.6331	0.8379	0.8111	0.6993
Financial & economic outcomes	Impact on cash flow	0.5847	0.7405	0.9005	0.7762	0.7158
	Reduce human intervention to improve accuracy	0.6138	0.4829	0.587	0.7062	0.7491
	Ability in data analysis	0.7908	0.4664	0.4865	0.7701	0.4863
	Elevate the quality of people	0.4523	0.5235	0.4212	0.7184	0.2131
	Reducing human error	0.6972	0.5997	0.6597	0.8556	0.6348
	Optimization of inventory	0.41	0.4061	0.3992	0.7154	0.2436
	Prevent local and corporate losses	0.7086	0.6648	0.7312	0.8377	0.6432
	Help in product mix	0.705	0.6498	0.7991	0.8561	0.6947
	Customer purchase behaviour	0.6414	0.7394	0.7609	0.8729	0.6422
	Impact on customer behaviour	0.3952	0.6494	0.6526	0.7244	0.4919
Impact of big data analysis	Understand customer needs	0.5488	0.6136	0.8487	0.7456	0.6047
	Identify customer behaviour	0.641	0.8151	0.811	0.8256	0.6723
	Retain customers	0.6481	0.6921	0.7456	0.7883	0.7871
	Attract new customers	0.7908	0.4664	0.4865	0.7701	0.4863
	Identify customer satisfaction levels	0.6139	0.5444	0.622	0.7104	0.691
	Increase customer engagement	0.6665	0.4626	0.5557	0.5194	0.8216
	Improve customer spending patterns	0.6	0.4879	0.583	0.5734	0.863
	Improve demand	0.6918	0.5275	0.7091	0.7415	0.8908
	Real-time customer purchase patterning for better decision making	0.7315	0.5101	0.6596	0.6229	0.9055
	Re-marketing	0.4586	0.3662	0.4829	0.5016	0.7466
	Impact on the client relationship	0.5951	0.4356	0.5491	0.5394	0.8078

Note: Construct item loadings are highlighted in grey.

Table 6. Discriminant validity

	Data analysis tools	***Data security & data privacy***	***Data source***	***Financial & economic outcomes***	***Impact of Big data analysis***
Data analysis tools	**0.7625**				
Data security & data privacy	0.6016	**0.8657**			
Data source	0.7197	0.8224	**0.8337**		
Financial & economic outcomes	0.7255	0.752	0.7541	**0.7688**	
Impact of big data analysis	0.759	0.6619	0.7894	0.7543	**0.8270**

As can be seen from Table 6, discriminant validity is proven and supported as all the square root AVE values are greater than the AVE of the other variables.

3.2.4. Structural equation modelling

To test the hypotheses, SmartPLS software was used for path analysis and the 238 samples were bootstrapped using the re-sampling procedure to establish confidence intervals (Mooney & Duval, 1993; Manski, 1996). To model the unknown population, bootstrapping results can be used (Hesterberg *et al.*, 2003). The t-statistic is used as the basis for checking the level of significance. The different significance levels (p-values) and corresponding t-values are given in Table 7 (Cowles & Davis, 1982; Neyman and Pearson, 1933).

Table 7. Significance levels

	Significance	*t-value*
Significance values	$p < 0.1$	1.650
	$p < 0.05$	1.968
	$p < 0.01$	2.592

4. RESULTS

Table 8 shows the results of the hypothesis testing. A total of eight hypotheses were created and of these, seven hypotheses are supported. As can be observed, H1 ($\beta = 0.0856$, $p > 0.1$) is not supported because the path from data source to data analysis tools is not significant. This might be expected, as a good and advanced data analysis tool can make sense of a very basic data source. With the advancements in Smartphone, tablet, computer and other technology, an increasing number of data types are available on the market for analysis (Baier *et al.*, 2012). This is contrary to our findings in the sense that though more Data Sources are available they has less impact on Data Analysis.

Table 8. Results of hypothesis testing

		Path coefficient (β)	*Mean*	*St. Dev.*	*St. Error*	*t-value*	*Supported*
H1	Data Source > Data Analysis Tools	0.0856	0.0902	0.0804	0.0804	1.0647	No
H2	Data Source > Impact of Big Data Analysis	0.4667**	0.4736	0.1837	0.1837	2.5404	Yes
H3	Data Analysis Tools > Impact of Big Data Analysis	0.412***	0.407	0.083	0.083	4.9636	Yes
H4	Data Security & Data Privacy > Financial & Economic Outcome	0.782***	0.7854	0.0433	0.0433	18.055	Yes
H5	Data Security & Data Privacy > Data Source	0.8224***	0.8256	0.0263	0.0263	31.2101	Yes
H6	Data Security & Data Privacy > Impact of Big Data Analysis	0.6822***	0.6862	0.0478	0.0478	14.265	Yes
H7	Financial & Economic Outcome > Data Analysis Tools	0.7424***	0.7393	0.0814	0.0814	9.1212	Yes
H8	Financial & Economic Outcome > Impact of Big Data Analysis	0.3663*	0.3556	0.1982	0.1982	1.8486	Yes

Notes: *, ** and *** denote significance at the 10%, 5% and 1% levels respectively (one-tailed).

H2 is supported ($\beta = 0.4667$, $p < 0.05$) showing a significant link between data source and the impact of big data analysis. Using the wrong data source will provide spurious information and this will potentially have a negative impact on the organization. The incorrect analysis of data has a significant impact on the organization's revenue (Chartered Institute of Management Accountants, 2013). This is in conformity with our findings.

H3 is supported ($\beta = 0.412$, $p < 0.01$), with a significant path from data analysis tools to impact on data analysis. The use of different tools will affect the data analysis as different tools use different algorithms. Information in blogs and forums, for example, can be analysed by combining different analysis trees. Good results can be obtained using the Kruskal–Wallis and Brunner–Dette–Munk tests (Cerchiello and Giudici, 2012). This is in conformity with our findings.

The paths from data security and data privacy to financial & economic outcomes ($\beta = 0.782$, $p < 0.01$) and to data source ($\beta = 0.8224$, $p < 0.01$) are significant, providing support for H4 and H5 respectively. This indicates that ensuring data security and data privacy and using appropriate and diverse data sources will have a positive impact on financial and economic outcomes. This is in conformity with earlier findings by Ahmad, 2004 that confidentiality and individuality of the common man should be preserved.

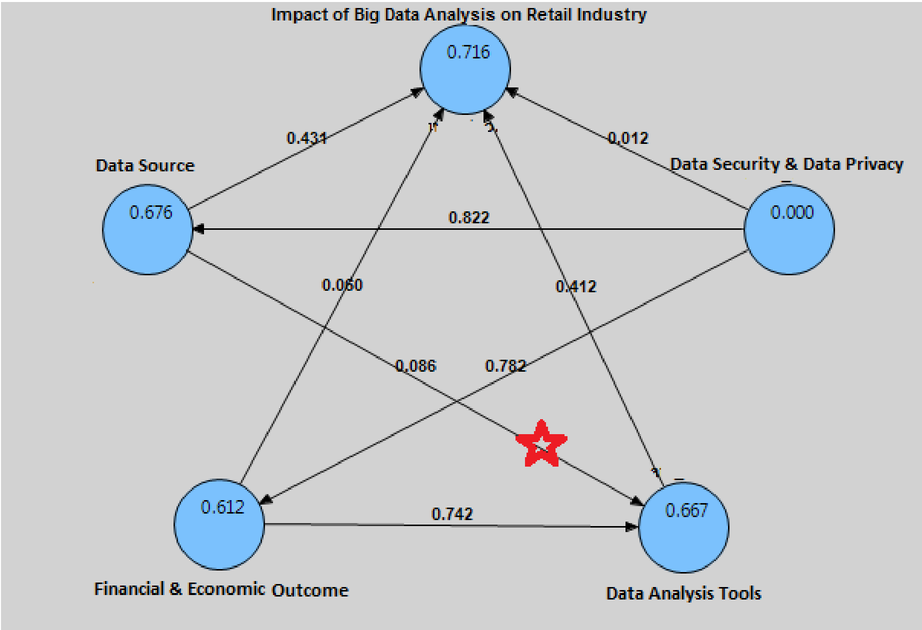
H6 is also supported ($\beta = 0.6822$, $p < 0.01$), demonstrating the link from data security and data privacy to the impact of big data analysis. This is in conformity with earlier findings that security issues are a major concern for most organizations (Chen *et al.*, 2012) and will have an effect on data analysis.

H7 ($\beta = 0.7424$, $p < 0.01$) is strongly supported. Financial and economic outcomes show a direct relationship with data analysis tools of retail industry. This is our new finding contributing to the literature of Big Data.

There is also support for H8 ($\beta = 0.3663$, $p < 0.1$), showing that the path from financial & economic outcomes to the impact of big data analysis is significant. In a survey of the state of business analytics by Bloomberg Businessweek (2011), our findings are validated by the statement that 97% of companies with revenues exceeding $100 million were found to use some form of business analytics (Chen *et al.*, 2012).

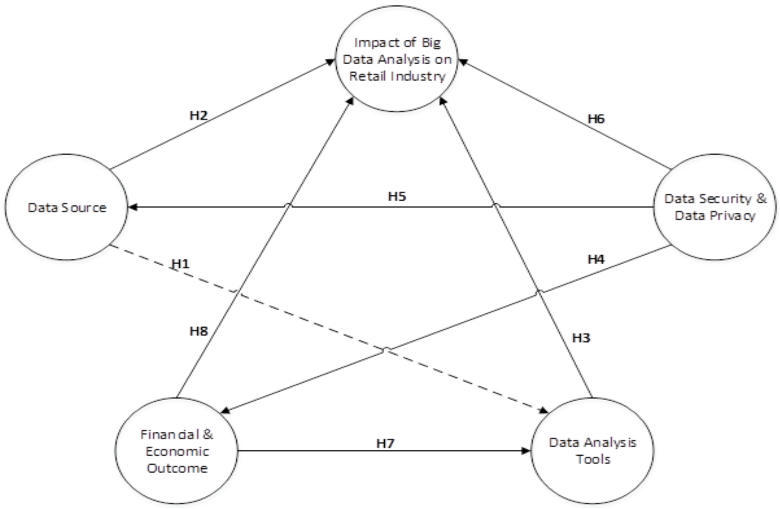
The results of the PLS structural modelling are shown in Figures 2 and 3 below.

Figure 2. Results of PLS analysis (extracted from SmartPLS) showing direction of path and beta coefficients



Note: The line in red represents the non-supported hypothesis.

Figure 3. Results of PLS structural model analysis showing paths and hypotheses



Note: Significant relationship (→), insignificant relationship (--->).

4.1. Goodness of Model Fit

ADANCO (Wende *et al.*, 2005; Ringle *et al.*, 2005) software was used to analyse the overall goodness of fit of the model. If model does not fit the data, then it means that the data contain more information than the model conveys. For frame of reference it is needed to determine the model fit both for the estimated model and for the saturated model. (Hensler et al., 2014)

The SRMR value of a perfect model fit is 0, but a value less than 0.05 indicates an acceptable fit (Byrne, 2013). Hensler et al., 2014, suggest that SRMR value above 0.06 is also acceptable. Hu and Bentler (1999) proposed 0.08 as a cut-off value for SRMR. Also a value between 0.05 and 0.08 suggest reasonable error of approximation (Browne and Cudeck, 1993).

As can be seen from Table 8 & 9, SRMR value is less than 0.08 for saturated and estimated model.

Table 8. Results of Goodness-of-model-fit (saturated model)

	Value	*HI95*	*HI99*
SRMR	0.0712	0.0452	0.0470
dULS	0.6793	0.1999	0.2167
dG	1.6020	0.5666	0.6393

Table 9. Results of Goodness-of-model-fit (estimated model)

	Value	*HI95*	*HI99*
SRMR	0.0785	0.0539	0.0599
dULS	0.7520	0.2849	0.3511
dG	1.6203	0.5552	0.6580

5. IMPLICATIONS FOR THE RETAIL INDUSTRY

The retail industry contributes to 6–7% of the world economy and covers a large ecosystem. The retail industry includes specialty groceries, consumer products, goods, e-commerce, department stores, apparel, discount drugstores, home improvement, discount retailers, electronics, and specialty retailers. Retail is increasing day by day with the invention of new technology. The squeeze on industry incumbents is coming from e-commerce and new "point, scan and analyse" technologies that give shoppers decision-making tools – powerful pricing, promotion and product information, often in real time. Applications in iPhones and Android, such as Red Laser, can scan barcodes and provide immediate price, product and cross-retailer comparisons. They can even point customers to the nearest retailer providing free shipping (total cost of purchase optimization). This leads to further margin erosion for retailers that compete based on price.

Thus, it is important for businesses, whether large or small, to ensure they are competitive. The competition intensifies as online retailers interact with their customers in real time. Data analysis through Big Data can help retailers in the following ways coupled with Volume, Varity and Velocity (3V's of Big Data):

- Improved customer service – Data analysis can help an organization to improve its customer service to attract more customers and retain existing ones. For example, when customers complain online or through social media, data analysis can provide background to the issue, helping customer care to address it and provide a better service. This will result in good customer handling, quicker resolution of problems and the customer feeling privileged and important.
- Greater customisation – Consumers shop with the same retailer in different ways, for example online, using mobile apps, etc. When data are collected in real time for analysis from multiple sources, companies can provide a customised experience for customers. For example, data analysis helps in segmenting customers, identifying those who are loyal customers and those who are new. This helps organizations to reward loyal customers, whilst also appealing to and attracting new customers.
- Product status and availability – With the increase in technology, customers like to know the real-time availability, status and location of their orders. This is challenging when many parties are involved in the product delivery. To keep customers happy, it is important for them to be able to know the exact status of the product. To communicate the real-time status of the product, all the relevant parties, including third parties involved in the transaction, should communicate with each other. To implement this functionality, companies should start early and improve their services over time.
- Managing fraud – The availability of larger data sets helps improve fraud detection rates, but to achieve this, companies require huge infrastructure. This infrastructure leads to a safer environment in which to run businesses and also helps improve profitability. For example, to detect online fraud, companies need to process their transactions against pre-defined fraud patterns in real time, otherwise it is not possible to detect fraudulent behaviour.
- Predictive analytics – Irrespective of a firm's size, analytics is crucial for all online retailers. Without analytics it is difficult to sustain a business in this competitive environment. Predictive analysis helps organizations to identify events before they occur. This can be achieved using data analysis and many businesses are now using predictive analytics to plan for the future. For example, retailers can use this analysis to plan inventory, helping them to save on inventory costs and avoid out-of-stock issues.
- Dynamic pricing – With the increase in competition, dynamic pricing is very important to compete on pricing with other sites. To achieve this, companies collect data from multiple sources, such as product sales, competitor pricing and customer actions, to determine the right price for the sale of products. Large online retail giants such as Amazon already support this functionality. This analysis gives large businesses a huge competitive advantage over small and medium competitors.

6. LIMITATIONS AND SCOPE FOR FURTHER RESEARCH

The variables considered in this study are based on the current state of data analysis in the retail sector. Future research could incorporate other critical variables, such as geographic location and socio-economic data, to examine social implications. This will help provide a more social and global perspective on data analysis in the retail industry. The inclusion of companies and respondents from other geographies and from SMEs would also help in making comparisons of the impact of big data analysis in the retail industry around the globe and in retail organizations of different sizes.

With the increased adoption of Big Data, another recommendation would be to conduct further research into factors that influence the choice of data analysis tools in the retail industry (for example, revenue, macro and micro factors influencing the success of data analysis, etc.).

The final recommendation is to research and analyse the company's performance in terms of the financial and operating benefits that companies can achieve with data analysis.

7. CONCLUSIONS

Data analysis through Big Data has been acclaimed as a tool that can revolutionize the retail industry. There are various advantages of data analysis technology and an increasing number of firms are in the process of implementing the data analysis through Big Data to provide insights into and improve their revenues. Research has considered many businesses and the operational challenges companies face in the adoption and implementation of data analysis in the retail industry. Despite these challenges, data analysis has become a significant strategy for companies in gaining competitive advantage and accelerating growth (Davenport, T.H., 2006). James, L. (2010) on data analysis in the retail industry suggests that it is an integral part of business, revealing the customer data that should be analysed and the benefits that retail organizations can obtain.

This study has methodically analysed four factors, namely, data source, data analysis tools, financial and economic outcomes and data security and data privacy, to gauge their influence on the impact of Big Data in the retail industry. Based on this analysis, there is a remarkable change in the factors identified that affect data analysis in the retail industry. The data source and data analysis tools are now perceived as given factors or "must haves" in terms of their impact and these are not considered to be differentiators for companies. Rather, it is observed that data security and data privacy are major considerations for companies in adopting data analysis, followed by the financial and economic impact.

Although the data source and data analysis tools are not the most significant influences for companies to adopt data analysis, both still have a positive impact on data analysis in the retail industry. This observation is supported by Wixom and Watson (2001) and Park and Kim (2003). The research also discovered that the ability to perform data analysis is one of the most important factors driving an organization's success in the retail industry. This observation is supported by Davenport, T.H. (2006).

Based on this research, it can be predicted that data analysis will continue to exert an impact of the retail industry and extended its use could also potentially benefit other sectors, such as healthcare, etc.

REFERENCES

1. Acquisti, A. (2004), Privacy in electronic commerce and the economics of immediate gratification, in EC '04 Proceedings of the 5th ACM conference on Electronic commerce, pp. 21–29.
2. Ahmed, S.R. (2004), "Applications of data mining in retail business", IEEE Information Technology: Coding and Computing, Proceedings. ITCC 2004, International Conference, Vol. 2
3. Ashley, C. and Noble, S.M. (2013), "It's closing time: territorial behaviours from customers in response to front line employees", Journal of Retailing, Vol. 90 No. 1, pp. 74–92.
4. Bagozzi, R.P. (1980), "Causal Models in Marketing", Wiley, New York.
5. Baier, D. and Decker, R. (2012), "Special issue on data analysis and classification in marketing – preface by the guest editors", Advances in Data Analysis and Classification, Vol. 6 No. 4, pp. 249–251.
6. Baier, D., Daniel, I., Frost, S. and Naundorf, R. (2012), "Image data analysis and classification in marketing", Advances in Data Analysis and Classification, Vol. 6 No. 4, pp. 253–276.
7. Barclay, D.W., Thompson, R. and Higgins, C. (1995), "The partial least squares (PLS) approach to causal modelling: personal computer adoption and use an illustration", Technology Studies, Vol. 2 No. 2, pp. 285–309.
8. Bentler, P.M. and Hu, L. (1999), "Cutoff criteria for fit indexes in covariance structure analysis:
9. Conventional criteria versus new alternatives", Structural Equation Modelling: A Multidisciplinary Journal, Vol. 6, pp.1-55.
10. Berry, M.J.A. and Linoff, G. (1997), Data Mining Techniques for Marketing, Sales, and Customer Support, Wiley, New York.
11. Bloomberg Businessweek (2011), "The Current State of Business Analytics: Where Do We Go from Here?", Bloomberg Businessweek Research Services, available at: http://www.sas.com/resources/asset/busanalytics study_wp_08232011.pdf
12. Browne, M.W., MacCallum, R.C., Kim, C.-T., Andersen, B.L. and Glaser, R. (2002), "When fit indices and residuals are incompatible", Psychological Methods, Vol. 7, No. 4, pp. 403–421.
13. Browne, M.W. and Robert C. (1993), "Alternative ways of assessing model fit", Sage Focus Editions 15, pp. 136-136.
14. Brynjolfsson, E., Hitt, L. and Kim, H. (2011), "Strength in numbers: how does data-driven decision making affect firm performance?", available at: http://dx.doi.org/10.2139/ssrn.1819486
15. Cerchiello, P. and Giudici, P. (2012), "Non parametric statistical models for on-line text classification", Advances in Data Analysis and Classification, Vol. 6 No. 4, pp. 277–288.
16. Chaudhuri, S., Dayal, U. and Narasayya, V. (2011), "An overview of business intelligence technology", Communications of the ACM, Vol. 54 No. 8, pp. 88–98.
17. Chen, H., Chiang, R.H.L. and Storey, V.C. (2012), "Business intelligence and analytics: from big data to big impact", MIS Quarterly, Vol. 36 No. 4, pp. 1165–1188.
18. Chiang, R.H.L., Goes, P. and Stohr, E.A. (2012), "Business intelligence and analytics education and program development: a unique opportunity for the information systems discipline", ACM Transactions on Management Information Systems, Vol. 3 No. 3, doi > 10.1145/2361256.2361257.
19. Chin, W.W. and Gopal, A. (1995), "Adoption intention in GSS: relative importance of beliefs", ACM SIGMIS Database, Vol. 26, No. 2-3, pp.42–64.
20. Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003), "A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and voice mail emotion/adoption study", Information Systems Research, Vol. 14 No. 2, pp. 189-217.
21. Chartered Institute of Management Accountants (2013), "Incorrect analysis of big data has significant impact on revenue, say third of finance professionals", available at: http://www.cimaglobal.com/About-us/Press-office/Press-releases/2013/Incorrect-analysis-of-big-data-has-significant-impact-on-revenue-say-third-of-finance-professionals/
22. Compeau, D.R., Higgins, C.A. and Huff, S. (1999), "Social cognitive theory and individual reactions to computing technology: a longitudinal-study", MIS Q, Vol. 23 No. 2, pp. 145–158.

23. Cowles, M., and Davis, C. (1982), "On the origins of the .05 level of statistical significance", American Psychologist, Vol. 37 No. 5, pp. 553-558.
24. Davenport, T.H. (2006), "Competing on analytics", Harvard Business Review, Vol. 84 No. 1, pp. 98–107.
25. Davis, D.F., Golicic, S.L. and Boerstler, C.N. (2011), "Benefits and challenges of conducting multiple methods research in marketing", Journal of the Academy of Marketing Science, Vol. 39 No. 3, pp. 467-479.
26. Donthu, N. and Yoo, B. (1998), "Retail productivity assessment using data envelopment analysis", Journal of Retailing, Vol. 71 No. 1, pp. 89–105.
27. Esposito Vinzi, V., Trinchera, L. and Amato, S. (2010), "PLS path modelling: from foundations to recent developments and open issues for model assessment and improvement", in Esposito Vinzi, V., Chin, W.W., Henseler, J. and Wang, H. (Eds.), Handbook of Partial Least Squares: Concepts, Methods and Applications, Springer Berlin Heidelberg, Berlin, pp.47–82.
28. Falcioni, J.G. (2013), "Smart factories, smarter engineers", Marketing Engineering, available at: http://memagazineblog.org/2013/10/07/smarter-factories-smarter-engineers/
29. Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996), "From data mining to knowledge discovery: an overview", in: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and Data Mining, MIT Press, Massachusetts, pp. 2-18.
30. Flavian, C. and Guinaliu, M. (2006), "Consumer trust, perceived security and privacy policy: Three basic elements of loyalty to a web site", Industrial Management & Data Systems, Vol. 106 No. 5, pp. 601-620.
31. Fornell, C. and Larcker, D. (1981), "Evaluating structural equitation models with unobservable variables and measurement errors", Journal of Marketing Research, Vol. 18 No. 1, pp. 39–50.
32. Gefen, D., Straub, D. and Boudreau, M. (2000), "Structural equation modelling techniques and regression: guidelines for research practice", Communications of the Association for Information Systems, Vol. 7, No. 7, pp.1–78.
33. Glaser, B.G. and Strauss, A.L. (1967), The Discovery of Grounded Theory: Strategies for Qualitative Research, Transaction Publishers, Chicago.
34. Graen, M. (1999), "Technology in Manufacturer/Retailer Integration between Wal-Mart and Procter & Gamble, available at: http://citebm.business.illinois.edu/IT_cases/PG-Graen.htm
35. Haenlein, M. and Kaplan, A.M. (2004), "A beginner's guide to partial least squares analysis", Understanding Statistics, Vol. 3 No. 4, pp. 283–297.
36. Hair, J.F., Sarstedt, M., Pieper, T.M. and Ringle, C.M. (2012), "The use of partial least squares structural equation modeling in strategic management research: a review of past practices and recommendations for future applications", Long Range Planning, Vol. 45 No. 5/6, pp. 320–340.
37. Hair, J.F., Sarstedt, M., Ringle, C.M. and Mena, J.A. (2012), "An assessment of the use of partial least squares structural equation modeling in marketing research", Journal of the Academy of Marketing Science, Vol. 40 No. 3, pp. 414–433.
38. Henseler, J. and Sarstedt, M. (2013), "Goodness-of-fit indices for partial least squares path modeling", Computational Statistics, Vol. 28 No. 2, pp. 565–580.
39. Hesterberg, T., Moore, D.S., Monaghan, S., Clipson, A. and Epstein, R. (2003), "Bootstrap methods and permutation tests", in Moore, D.S. and McCabe, G.P. (Eds.) Introduction to the Practice of Statistics (5th edition), available at: http://bcs.whfreeman.com/pbs/cat_160/PBS18.pdf
40. Hoyle, R.H. (Ed.), (1995) Structural Equation Modeling, SAGE Publications, Inc., Thousand Oaks, CA.
41. Hulland, J. (1999), "Use of partial least squares (PLS). Strategic management research: a review of four recent studies", Strategic Management Journal, Vol. 20, No. 4, pp.195–204.
42. IBM Tech Trends Report (2011), available at: http://www/ibm.com/developerworks/techtrends report
43. ISACA (2011), "Data Analysis – A Practical Approach", available at: http://www.algaonline.org/DocumentCenter/View/2847
44. Johnsen, M. (2013), "Using big data to drive retail personalization", Drugstorenews.com, available at: http://www.drugstorenews.com/article/using-big-data-drive-retail-personalization
45. Jöreskog, K. and Sorbom, D. (1993) LISREL. VIII Scientific Software, Chicago, IL.
46. Keim, D. A., Mansmann, F., Schneidewind, J. and Ziegler, H. (2006), "Challenges in Visual Data Analysis", IEEE Information Visualization.
47. Keim, D.A. (2009), "Visualization techniques for mining large databases: a comparison", IEEE Transactions on Knowledge and Data Engineering, Vol. 8 No. 6, pp. 923–937.
48. James, L. (2010), "Data Analysis for the Retail Industry", 24 Oct, available at: http://www.yellowfinbi.com/YFCommunityNews-Data-Analysis-for-the-Retail-Industry-Part-1-100068
49. Lohmöller, J.B. (1989), "Latent Variable Path Modeling with Partial Least Squares", Physica-Verlag, Heidelberg.
50. Lohr, S. (2012), "The Age of Big Data", The New York Times, available at: http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=0
51. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H. (2011), "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute Report, available at: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
52. Manski, C.F. (1996), "Book review: an introduction to the bootstrap, by Efron, B. and Tibshirani, R.J.", Journal of Economic Literature, Vol. 34 No. 3, pp. 1340–1342.
53. McKnight, D.H., Choudhury, V. and Kacmar, C. (2002), "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology", Information Systems Research, Vol. 13 No. 3, pp. 334-359.
54. Mooney, C.Z. and Duval, R.D. (1993), Bootstrapping: A Nonparametric Approach to Statistical Inference, Sage, Newbury Park, CA.
55. Neyman, J. and Pearson, E.S. (1933), "The testing of statistical hypotheses in relation to probabilities a priori", Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 29, pp. 492–510.
56. Park, C-H. and Kim, Y-G. (2003), "Identifying key factors affecting consumer purchase behavior in an online shopping context", International Journal

of Retail & Distribution Management, Vol. 31 No. 1, pp. 16–29.
57. Ringle, C.M., Wende, S. and Will, A. (2005), "SmartPLS 2.0 M3 (beta)", University of Hamburg, available at: www.smartpls.de (accessed 22 April, 2014).
58. Rigdon, E.E. (1998), "Structural equation modeling" in Marcoulides, G.A. (Ed.), Modern Methods for Business Research, Erlbaum, Mahwah, pp.251–294.
59. Russom, P. (2011), "Big data analytics", TDWI Best Practices Report, available at: https://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx.
60. Sallam, R.L., Richardson, J., Hagerty, J. and Hostmann, B. (2011), Magic Quadrant for Business Intelligence Platforms, Gartner Group, Stamford, CT.
61. Savitz, E. (2012), "Why Big Data Is All Retailers Want for Christmas", CIO Network available at http://www.forbes.com/sites/ciocentral/2012/12/12/why-big-data-is-all-retailers-want-for-christmas/
62. Savenye, W.C. and Robinson, R.S. (2005), "Using qualitative research methods in higher eduction", Journal of Computing in Higher Education, Vol. 16 No. 2, pp. 65-95.
63. Sekaran, U. and Bougie, R. (2010), "Research Methods for Business: A Skill Building Approach", Wiley, UK.
64. Shepard, R.J. (2002), "Ethics in exercise science research", Sports Medicine, Vol. 32 No. 3, pp. 169-183.
65. Smith, B. (2014), "Embrace Big Data", available at: cabinet-maker.co.uk
66. Stewart, D.W., (1981), "The Application and Misapplication of Factor Analysis in Marketing Research", Journal of Marketing Research, Vol. 18 No. 1, pp. 51-62.
67. Straub, D., Boudreau, M. C., & Gefen, D. 2004, "Validation guidelines for IS positivist
68. Research", Communications of the Association for Information Systems, Vol. 14, pp. 380-426.
69. Tenenhaus, M. (2008), "Component-based structural equation modelling", Total Quality Management & Business Excellence, Vol. 19, pp.871–886.
70. Tenenhaus, M., Amato, S. and Vinzi, V.E. (2004), "A global goodness-of-fit index for PLS structural equation modeling", in Proceedings of the XLII SIS Scientific Meeting, CLEUP, Padova, pp. 739–742.
71. Turban, E., Sharda, R., Aronson, J.E. and King, D. (2008), Business Intelligence: A Managerial Approach (2nd Edition), Pearson Prentice Hall, Boston.
72. Van Till, S. (2013), "The Predictive Value of Big Data in Retail, Public Venues and Property Management", available at http://www.securitymagazine.com/articles/84979-the-future-of-big-data-for-retail-and-property
73. Watson, H.J. and Wixom, B.H. (2007), "The current state of business intelligence", IEEE Computer, Vol. 40 No. 9, pp. 96–99.
74. Wetzels, M., Odekerken-Schröder, G. and van Oppen, C. (2009), "Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration", MIS Quarterly, Vol. 33 No. 1, pp. 177–195.
75. Wixom, B.H. and Watson, H.J. (2001), "An empirical investigation of the factors affecting data warehousing success", MIS Quarterly, Vol. 25 No. 1, pp. 17–41.
76. Wold, H. (1989), "Introduction to the second generation of multivariate analysis" in Wold, H. (Ed.), Theoretical Empiricism, Paragon House, New York, NY, pp.vii–xl.
77. Wong, K.K. (2011), "Book review: handbook of partial least squares: concepts, methods and applications, by V. Esposito Vinzi, W.W. Chin, J. Henseler & H. Hwang (Eds)", International Journal of Business Science & Applied Management, Vol. 6 No. 2, pp. 52–54.
78. Written, I. H., Frank, E. and Hall, M. (2011), Data Mining: Practical Machine Learning Tools and Techniques (3rd edition), Morgan Kaufmann, San Francisco.
79. Wu, J. (2002), "Business Intelligence: The Value in Mining Data", DM Review, available at: http://www.information-management.com/news/4618-1.html
80. Wyner, G. (2013), "Data, Data Everywhere", Marketing Management, Vol. 47 No. 3, p. 18.