World Journal of Social Sciences
Vol. 3. No. 6. November 2013 Issue. Pp. 105 – 114

Data Quality Awareness as an Optimal Marketing Strategy: A Case Study of a Saudi Manufacturing Company

Mohammad Almotairi*

The ultimate objective of all marketing strategies is maximum profit and a good market reputation. Among the many aspects to be considered while concentrating on this ultimate objective, data quality awareness is one of the essential dimensions with the potential to support the objective at a larger scale. The ignorance of or lack of adequate knowledge about the importance of data quality results in discrepancies, deficiencies, and financial and social losses to business organizations as well as the development of a bad company image. This study aims to investigate data quality awareness as an optimal marketing strategy in a manufacturing company (Aluminum Product Company) in Riyadh, Saudi Arabia. The primary data of selected company have been collected through a detailed questionnaire. The company employees were divided into three clusters representing lower, middle and upper level workers. Data were collected from each cluster randomly to avoid bias and increase the reliability of the results. After checking for reliability and consistency of the responses, a dichotomous logistic regression model was used to investigate the determinants of overall data quality awareness level in the Saudi manufacturing company. The study reveals a lack of data quality awareness at the upper management level, which may cause financial losses on one hand and "deficient marketing" of the product and the company on the other hand. The study will enhance data quality awareness in order to cure socioeconomic losses and to yield efficiency, good reputation and financial benefits to Saudi manufacturing companies.

Keywords: Data quality awareness, Case study, Optimal Marketing Strategy

1. Introduction

In the last few decades, data quality has attracted the attention of many researchers due to the extended use of data warehouse systems (Heinrich et al. 2007). Because the key factor for the success of an organization is the quality of data provided, it is essential to understand the importance of data quality. Numerous industry surveys indicate that business organizations are increasingly experiencing problems that adversely affect their socio-economic reputations (Giannoccaro et al. 1999). Effective business decision making depends on good quality data; poor data quality can be costly and sometimes disastrous. Thus, business organizations need to understand data quality and establish procedures to ensure the quality of data in a data warehouse (Giannoccaro et al. 1999). This research aims to investigate data quality awareness as an optimal marketing strategy in a manufacturing company (Aluminum Product Company) in Riyadh, Saudi Arabia. Practitioners know that creating awareness of the problems and their impact is a significant first step in the resolution of data quality problems (Redman 1998).

*Ph.D, Department of Marketing, College of Business Administration, King Saud University, Saudi Arabia, Email: mohamed104@hotmail.com

Awareness is the process of updating knowledge, developing skills, enacting attitudinal and behavioral changes, and improving a person's ability to perform his/her task efficiently and effectively. The basic responsibility of an organization is to educate and train employees in data quality (Talib et al. 2011). Many executives and managers ignore the incredible damage caused by poor data quality, as well as its impact on the business organization. This is due to the lack of understanding of how to solve data quality problems, which leads management to disregard this critical issue (Whitehead, 2006). Data quality management is important because organizational data are used for many purposes (transactions, deliveries, invoices, marketing, sales forecasting, financial forecasting, supply chain management, etc.) (Whitehead 2006). Therefore, employee awareness concerning data quality is necessary for the achievement of high-quality data.

2. Material and Methods

2.1 Research Approach – The Case Study

A case study approach has been adopted for this study. The main purpose of the case study is to help the researcher to understand the phenomena that are studied. The case study allows the researcher to attain maximum information from small units. Case studies are used to closely examine the data within a specific context. In many cases, a small geographic area or a minimum number of individuals are selected as subjects. Case studies are normally used to explore and investigate contemporary real-life phenomena with a contextual analysis of a minimum number of events or conditions, as well as their relationships (Zainal, Jun 2007).

2.2 Research Tools

The different data collection methods available for quantitative study include observation, document analysis, and questionnaire study. This research uses a self-administered questionnaire survey, defined as "a data collection technique in which the respondent read the survey questions and record his / her responses without the assistance of a trained interviewer" (Shiang 2012). This method was chosen for the following reasons. First, it is an effective tool to reach out to the respondents, who are distributed in different departments of the selected manufacturer, thereby producing generalized results. Specially, the questionnaire study allows data to be collected from data source users from various departments of the selected manufacturer, in order to obtain representative and heterogeneous data. Second, this approach allows the researchers to collect data from a large number of respondents within a reasonable time frame utilizing limited resources and human power (Shiang 2012).

2.3 Operationalization of Variables

In order to gather information from the samples, a set of tools is needed to assess the variables in this research study. Therefore, appropriate indicators (i.e., measurement items) need to be identified to assess these variables. The items used in this study have been developed from a review of the relevant literature; sources are shown in Table 1. A total of

37 measurement items are used to assess the variables. The table shows the summary of the numbers and sources of items used to assess each variable.

Table 1: Sources of Questionnaire Items

Variables	No. of Items	Sources
Dimensions (Accuracy, Completeness)	12	(Lee and Strong, 2004), (Batini et al., 2009), (Haug et al., 2009)
Factors Affecting the Data Quality (Training, Software Friendliness)	9	(Idris and Ahmad, 2011), (Singh and Singh, 2011)
Impacts	11	(Loshin, 2011)
Improvement Strategy	5	(Redman, 1995)

All the variables are operationalized by using 5-point Likert scales, ranging from 1 (strongly disagree) to 5 (strongly agree). The Likert scale was selected because it takes little time and is easy to answer (Shiang 2012).

2.4 Questionnaire Design

After the variables are operationalized, they are ready to be included in the questionnaire. The questionnaire includes a cover page presenting the non-disclosure statement to ensure the confidentiality of the company information. Table 2 shows the structure of the questionnaire and the nature of the questions in respect to particular variables. The questionnaire has been divided into two parts: part 1 collects profile and categorical information about the respondents through multiple choice questions, and part 2 consists of the measurement items and uses Likert scales to collect the data and allow for certain analytical operations to produce statistical results.

Table 2: Structure of Questionnaire

Part	Variables	Scale Type
1	Personal Profile	
	• Gender	Categorical
	• Highest Level of Education	Categorical
	• Department	Categorical
	• Job Title	Categorical
	• Working Experience	Categorical
2	• Dimensions (Accuracy, Completeness)	Likert Scale
	• Factors Affecting Data Quality (Training, Software Friendliness)	Likert Scale
	• Impacts of Poor Data Quality	Likert Scale
	• Improvement Strategy	Likert Scale

No negative questions are used in the questionnaire. The questions are arranged in an ordinal scale from strongly disagree to strongly agree in order to identify the data quality awareness in the Saudi manufacturer.

2.5 Population and Sampling

The population for this study consists of the Information System Group (ISG)/data warehouse users at the top management level of the selected manufacturer. It was necessary to use this population because the purpose of this study is to determine the awareness about data quality from the user's experience and interaction with system data. This study uses the non-probability purposive sampling technique to determine how the samples are selected, because it would be difficult to conduct a study including all employees in the selected company. Purposive sampling technique is limited to samples that can provide the desired information or conform to the criteria of the study (Shiang 2012). The sample criteria for this research are: 1) The respondent must be from the data warehouse/ISG, and 2) the respondent must be working in the data warehouse of manufacturing. The purposive sampling technique was chosen because it provides consistent and robust data and allowed the study to balance out the respondents who are dispersed in different departments of the selected company (Shiang 2012).

Collecting data from the whole population would be impossible due to the limited time and resources for the individual research project. Thus, the most common method adopted for this study is random sampling within the target group. A 30% sample size has been selected to collect the required data. The purpose of sampling is to select a sufficient number of samples from the population, in order to understand the properties of the samples and to generalize the findings to the population.

2.6 Data Analysis

Data collected regarding the variables of interest have been analyzed through descriptive statistics. As the data have been collected on an ordinal scale, Median and Mode as a central tendency will be used to reveal the average behaviour of the data. The respondents' responses regarding the data quality are collected on an ordinal scale ranging from strongly agree to strongly disagree. These five points (Strongly agree, agree, neutral, disagree and strongly disagree) are then summarized to draw conclusions about data quality awareness in the area under study.

The following analytical techniques are used to identify the data quality awareness.
First, the SPSS statistical package (version 19) has been used to check the consistency of the items in the survey questionnaire. The reliability and validity of the variables have been identified by the Cronbach's alpha test, which ensures that responses collected for a given item are strongly correlated.

Cronbach's alpha can be defined as:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K}\sigma_{Y_i}^2}{\sigma_X^2}\right)$$

Where K represent the number of components (*K-items* or *testlets*), σ^2_X the variance of the observed total test scores, and $\sigma^2_{Y_i}$ the variance of component *i* for the current sample of persons.

Second, as the data are collected on the ordinal scale, Median and mode, as the most commonly used descriptive statistics, are the best measures of calculating the central tendency of the data. Median and mode have been computed as the average value of the responses to all of the items in the survey instrument.

Third, we use the logistic model to show how the probability of overall data quality awareness changes given a change in the selected items. Data quality is peroxide by 2/3 of the median score of accuracy and completeness dimensions. If a respondent chooses at least 8 of 12 agree or strongly agree options from the accuracy and completeness dimensions, then the situation will represent the overall data quality awareness of the respondent.

The logistic model is thus specified as:

$$\mathrm{Xi} = [\frac{\alpha \mathrm{i}}{(1-\alpha \mathrm{i})}] = \beta 0 + \beta 1\mathrm{X}1 + \beta 2\mathrm{X}2 + \beta 3\mathrm{X}3 + \beta 4\mathrm{X}4 + \beta 5\mathrm{X}5 + \mu \mathrm{i}$$

Where;

$\mathbf{Xi} = [\frac{\boldsymbol{\alpha}\mathbf{i}}{(\mathbf{1}-\boldsymbol{\alpha}\mathbf{i})}]$ is odd ratios in favor of an increase in individual respondents' overall data quality awareness to the probability that an individual overall data quality awareness has not increased.

X_1= Respondent's gender
X_2 = Respondent's department
X_3 = Respondent's job title
X_4 = Respondent's education level
X_5 = Respondent's experience

3. Results and Discussion

The respondents who participated in the survey belong to different departments of the company and have different positions and education levels.

Table 3 presents individual attributes of the respondents. Out of all respondents, 38.5% are operational staff, 12.8% are middle managers, followed by administrative staff and technicians, who constitute, respectively, 10.3% and 7.7% of all respondents. The table also shows the education level of the respondents; 48.7% have a bachelor's degree, 33.3% have a diploma, and the remaining 7.7% and 10.3% hold certificates and high school certificates, respectively. Moreover, 56.4% of the respondents have between one and five years of experience, and 43.6% have more than five years of experience. These figures show that the respondents suitably represent the samples for the study.

Table 3: Individual Attributes of Respondents

Characteristics	Categories	Frequency	Percent
Sex	Male	22	56.4
	Female	17	43.6
	Total	39	100.0
Department	Production Planning	3	7.7
	Engineering/Product Development	12	30.8
	Quality Control	13	33.3
	Production Floor	7	17.9
	IT Department	2	5.1
	Other	2	5.1
	Total	39	100.0
Job Title	Middle Manager	5	12.8
	Operational Staff	15	38.5
	Administrative Support	4	10.3
	Technician	3	7.7
	Other	12	30.8
	Total	39	100.0
Education Level	High School	4	10.3
	Certificate	3	7.7
	Diploma	13	33.3
	Bachelor's Degree	19	48.7
	Total	39	100.0
Working Experience	Below 5 years	22	56.4
	6 to 10 years	7	17.9
	11 to 15 years	6	15.4
	16 to 20 years	3	7.7
	21 to 25 years	1	2.6
	Total	39	100.0

3.1 Reliability Result

A commonly used method for reliability analysis is Cronbach's alpha, which identifies the correlation between items. With this approach, we construct the reliability of the first section using Cronbach's alpha. Table 3.1 shows the Cronbach's α result obtained from the two dimensions (accuracy and completeness) with multiple indicators to check the scale of reliability of the indicators. The value for Cronbach's α is 0.906, which shows a high level of internal scale of consistency. The reliability of the dimensions is internally consistent and reliable for further statistical estimation. As a rule of thumb, the observed value of Cronbach's alpha should be at least 0.70.

Table 3.1: Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	No. of Items
0.906	0.911	12

3.2 Reliability Result

Table 4.7 shows the Cronbach's α result obtained from section two. In this section, we identify factors affecting the data quality/awareness (training and software friendliness) and their impacts, with multiple indicators to check the scale of reliability of the indicators. The value for Cronbach's α is 0.914, which shows a high level of internal scale of consistency.

Table 3.2: Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	No. of Items
0.914	0.918	20

3.3 Reliability Result for Section Three

Table 3.3 shows the Cronbach's α result obtained from the improvement strategy, with multiple indicators to check the scale of reliability of the indicators. The value for Cronbach's α is 0.865, which also shows a high level of internal scale of consistency. The reliability of these indicators is internally consistent and reliable for further statistical estimation.

Table 3.3: Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	No. of Items
0.865	0.693	5

3.4 Logistic Regression

Finally, we use the logistic model to show how the probability of overall data quality awareness changes given a change in the selected items. Data quality is proxied by 2/3 of the median score of the accuracy and completeness dimensions. If a respondent chooses at least 8 of 12 agree or strongly agree options from the accuracy and completeness dimensions, then the situation will represent the respondent's overall data quality awareness.

Thus, the logistic model is specified as:

$$\mathrm{Xi} = \left[\frac{\alpha\mathrm{i}}{(1-\alpha\mathrm{i})}\right] = \beta 0 + \beta 1\mathrm{X}1 + \beta 2\mathrm{X}2 + \beta 3\mathrm{X}3 + \beta 4\mathrm{X}4 + \beta 5\mathrm{X}5 + \mu\mathrm{i} \ldots\ldots\ldots\ldots\ldots\ldots \ldots (1)$$

Where;

$\mathbf{Xi} = [\frac{\alpha i}{(1-\alpha i)}]$ is odd ratio in favor of an increase in individual respondents' overall data quality awareness to the probability that an individual overall data quality awareness has not increased.

X_1= Respondent's gender
X_2 = Respondent's department
X_3 = Respondent's job title
X_4 = Respondent's education level
X_5 = Respondent's experience
μi = Stochastic error term

The following is the output of the logistic regression model:

Table 3.4: Logistic Regression Estimates

Demographics	Wald	df	Sig.	Exp(B)
Gender	0.238	1	.626	.211
Department	2.600	1	.332	.417
Job	5.866	1	.032	.668
Education	4.014	1	.035	.950
Experience	4.297	1	.055	1.741
Constant	2.101	1	.147	141.33

Finally, the study applies the logistic regression model to investigate how demographic characteristics influence the overall level of data quality awareness of the employees in the Saudi manufacturing company. The logistic regression model reports odd ratios along with other important statistics. The odd ratio for each demographic characteristic has been reported to show changes in the probability of employees' overall level of data quality awareness given a change in any demographic characteristic. Two variables – sex and departmental affiliation of the employee – are statistically insignificant, indicating no significant change in overall data quality awareness. The remaining three variables – job, education and experience – are statistically significant, representing a significant change in the employee's overall level of data quality awareness given a change in any of the mentioned variables. Experience is the most influential variable to change overall data quality awareness. If an employee's experience increases by one unit (one year), then probability in favor of overall data quality awareness will increase 1.74 times.

4. Conclusion

In order to answer the research questions, the descriptive data from 39 respondents were summarized and divided into three sections. A total of 37 questions were asked in the survey questionnaire based on the three objectives mentioned in chapter one. The questionnaire was based on an ordinal scale, in which respondents answered using a Likert ranging from 1 (strongly disagree) to 5 (strongly agree). For the reliability test and descriptive analysis, we divided the questionnaire into three sections. The first section includes 12 questions related to two dimensions by which to identify the level of awareness regarding data quality. The second section includes 20 questions to identify the factors that affect the data quality and its impact, and the third section contains 5 questions about the improvement of strategy. These three sections have been tested individually for the reliability and validity analysis. The results shows that all three sections of the questionnaire are internally consistent, indicating the relevance of the questions for each section and the internal scale of consistency. The information thus collected about the data quality awareness is reliable and statistically significant.

This study aims to investigate data quality awareness as an optimal marketing strategy in a manufacturing company (Aluminum Product Company) in Riyadh, Saudi Arabia. The primary data of the selected company were collected through a detailed questionnaire. The company employees were divided into three clusters representing lower, middle and upper level workers. Data were collected from each cluster randomly to avoid bias and increase reliability of the results. After checking the responses for reliability and consistency, a dichotomous logistic regression model was used to investigate the determinants of overall data quality awareness level in the Saudi manufacturing company. The study reveals that there is a lack of data quality awareness at the upper management level, which may cause financial losses on one hand and “deficient marketing” of the product and the company on the other hand. The study will enhance data quality awareness to compensate socioeconomic losses and to yield efficiency, good reputation and financial benefits for the Saudi manufacturing companies.

4.1 Improvement Strategies

Question three: How can the data quality awareness in Saudi manufacturers be improved? Our findings indicate that relevant qualification and experience are the best approaches to improve data quality and awareness. We also find identify that training is the best solution for the improvement of data quality awareness. For the improvement strategy, we first identify the factors affecting data quality and awareness, then classify their impacts on the organization by how much these impacts influence on the organization. Highlighting the issues and impacts related to data quality in the organization will help employees to understand the importance of data quality for the organization. In this context, Redman (1998) argues that creating awareness of the problems and their impact is a significant first step in the resolution of data quality problems. As discussed earlier, awareness is the process of updating knowledge, developing skills, creating attitudinal and behavioral changes, and improving a person’s to perform his/her task efficiently and effectively. Thus, with the help of training related to data quality, the employees will be able to understand and thus improve data quality. Talibe et al. (2011) argue that the basic responsibility of an organization is to educate and train their employees in data quality.

The demographic results show that job title, experience, and education level significantly influence data quality. These demographic characteristics are more important for the improvement of data quality and awareness.

For an organization to achieve high-quality data, they need to conduct refresher courses and workshops, as well as train their employees regarding data quality. Organizations also need to institute reward procedures on the basis of best performance and to provide suitable and peaceful environment to utilize their skills for the improvement of data quality.

References

Haug A, StentoftArlbjorn, J & Pedersen, A 2009, '*A classification model of ERP system data quality'*, *Industrial Management & Data Systems,* vol. 109, no. 8, pp. 1053-1068.

Cleven, A & Felix 2010, 'Uncovering four strategies to approach master data management'*, Proceedings of the 43rd Hawaii International Conference on System Sciences.* pp.1-10.

Giannoccaro, A, Shanks G, & PetaDarke 1999, 'Stakeholder Perceptions of Data Quality in a Data Warehouse Environment'*, 10th Australian Conference on Information Systems*, pp. 344-355.

Batini, C, Cappiello, C, Francalanci, C & Maurino, A (2009*),* 'Methodologies for data quality assessment and improvement', *ACM Computing Surveys*, vol. 41, no. 3

Loshin, D 2011, '*The practitioner's Guide to Data Quality Improvement',* pp.1-16.

Strong, DM, Lee, YW & Wang RY 1997, Data Quality in Context, *Communications of the ACM* ,vol. 40, no. 5, pp. 103-110.

Talib F, Rahman, Z & Qureshi MN 2011*,* 'Assessing the awareness of total quality management in Indian service industries: An empirical investigation'*, Asian Journal on Quality, v*ol. 12, no. 3, pp. 228 – 243.

Shanks, G & Corbitt, B 1999, 'Understanding Data Quality: Social and Cultural Aspects', *10th Australasian Conference on Information Systems,* pp. 785-797.

Haug, A, Zachariassen, F, & Van Liempd, D 2011, 'The cost of poor data quality', *Journal of Industrial Engineering and Management*, vol. 4, no. 2, pp. 168-193.

Whitehead, JC 2006, 'Solving the Data Quality Problems'*, Trituns Innovation, LLC,www.tritunsinnovation.com,* pp. 1-3.

NorizamIdris & Ahmad, K 2011, 'Managing Data Source Quality for Data Warehouse in manufacturing Services', *IEEE 2011 International Conference on Electrical Engineering and Informatics.*

Singh, R & Singh, K (2010), 'A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing'*, IJCSI International Journal of Computer Science Issues,* vol.7, no. 2, pp. 41-50.

Redman TC 1998, 'The Impact of Poor Data Quality on the Typical enterprise', *Communications of the ACM,* vol. 41 no.2, pp.72-82.

RistoSilvola, Jaaskelainen, O, Kropsu-Vehkapera, H & HarriHaapasalo 2011, 'Managing one master data – challenges and preconditions', *Industrial Management & Data Systems,* vol. 111, no.1, pp.146-162*.*

Zainal, Z 2007, *'Case Study as a Research Method', JurnalKemanusiaanbil,* pp. 1-6.