إقـــــرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

"The Impact Of Data Quality On Business Processes And Decision Making – Case Study Insurance Companies In Gaza Strip"

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وإن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل لنيل درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

DECLARATION

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification

Student's name: Hanan Hasan AL-Talla — اسم الطالب: حنان حسن الطلاع

Signature: — التوقيع:

Date: 09/09/2015 — التاريخ: 2015/09/09

The Islamic University- Gaza
Deanship of Graduates Studies
Faculty of Commerce
Business Administration Department

الجامعة الإسلاميـة–غزة
عمادة الدراسات العليـا
كليـة التجـارة
قسـم إدارة الأعمـال

MBA Thesis

THE IMPACT OF DATA QUALITY ON BUSINESS PROCESSES AND DECISION MAKING – CASE STUDY INSURANCE COMPANIES IN GAZA STRIP

تأثير جودة البيانات على العمليات الادارية وعملية اتخاذ القرار – دراسة حالة شركات التأمين في قطاع غزة

Submitted by:

Hanan Hasan Al-Talla

220110034

Supervised by:

Prof. Dr. Yousif Hussain Ashour

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master in Business Administration

1436 - 2015

بسم الله الرحمن الرحيم

الجامعة الإسلامية – غزة
The Islamic University - Gaza

مكتب نائب الرئيس للبحث العلمي والدراسات العليا هاتف داخلي 1150

الرقم. ج. س. غ/35/..... Ref

التاريخ 2015/07/11م Date

نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحثة/ **حنان حسن سلامة الطلاع** لنيل درجة الماجستير في **كلية *التجارة*/ قسم إدارة الأعمال** وموضوعها:

تأثير جودة البيانات على العمليات الإدارية وعملية اتخاذ القرار
دراسة حالة شركات التأمين في قطاع غزة

The Impact of Data Quality on Business Processes and Decision Making Case Study Insurance Companies in Gaza Strip

وبعد المناقشة التي تمت اليوم السبت 24 رمضان 1436هـ، الموافق 2015/07/11م الساعة الثانية عشرة ظهرًا، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

أ.د. يوسف حسين عاشور	مشرفاً و رئيساً
د. وسيم إسماعيل الهابيل	مناقشاً داخلياً
د. يوسف العبد حمودة	مناقشاً خارجياً

وبعد المداولة أوصت اللجنة بمنح الباحثة درجة الماجستير في **كلية *التجارة*/قسم إدارة الأعمال**.

واللجنة إذ تمنحها هذه الدرجة فإنها توصيها بتقوى الله ولزوم طاعته وأن تسخر علمها في خدمة دينها ووطنها.

والله ولي التوفيق ،،،

مساعد نائب الرئيس للبحث العلمي والدراسات العليا

أ.د. فؤاد علي العاجز

P.O. Box 108, Rimal, Gaza, Palestine fax: +970 (8) 286 0800 فاكس Tel: +970 (8) 286 0700 ص.ب. 108 الرمال، غزة، فلسطين هاتف
public@iugaza.edu.ps www.iugaza.edu.ps

بسم الله الرحمن الرحيم

"وما توفيقي الا بالله عليه توكلت واليه انيبُ"(هود : 88)

Abstract

The impact of data quality on business processes and decision making – case study Insurance companies in Gaza Strip

Data quality is a very important aspect that affects organization's business operations. Data quality means "Fitness for use". Now a day's organizations are paying more attention to data quality and focusing on how to improve it to enhance the overall performance of the organization.

In this research, the main objective is to measure the effect of data quality on business processes and decision making. Data quality has many dimensions and five of them were selected to be evaluated by this research. These dimensions are data completeness, accuracy, consistency, validity of data and relevance. The research target the insurance sector in Gaza Strip which contains six companies that where under the study of this research.

In this research, a questionnaire was designed as a data collection tool for the purpose of evaluating the impact of data quality on the business processes and decision making in the insurance companies working in Gaza Strip. 56 questionnaires were distributed and 50 of them were collected successfully.

The results show that there is a significant impact of data quality on daily business processes and decision making which indicates the important effect of the data quality dimensions on the organizational successes. Also it shows that insurance companies in Gaza Strip already knows the importance of data quality and is working to maintain it. The main recommendation of this research are to develop a data quality assessment methodology and developing a tool to measure the costs of poor data quality.

الملخص

تأثير جودة البيانات على العمليات الادارية وعملية اتخاذ القرار – دراسة حالة شركات التأمين في قطاع غزة

تعتبر جودة البيانات جانب هام جدا ومؤثر على العمليات الادراية في المؤسسات. يمكن تعريف مفهوم جودة البيانات على انه "ملاءمة البيانات للغرض التي ستستخدم له". ومنذ فترة غير بعيدة بدات المؤسسات بإيلاء المزيد من الاهتمام لنوعية البيانات المستخدمة في عملياتها والتركيز على كيفية تحسينها لتحسين الأداء الكلي للمؤسسة.

إن الهدف الرئيسي من هذا البحث هو قياس تأثير جودة البيانات على العمليات االدارية واتخاذ القرار في المؤسسة. تتكون جودة البيانات من العديد من المعايير وقد تم اختيار خمسة منها ليتم تقييمها من قبل هذا البحث. هذه المعايير هي اكتمال البيانات، دقة البيانات، تناسق البيانات، صحة البيانات وأن تكون البيانات ذات الصلة بالموضوع الذي ستسنخدم له. استهدف البحث قطاع التأمين في قطاع غزة والذي يحتوي على ست شركات عاملة في وقت اجراء البحث.

في هذا البحث، تم تصميم استبانة كأداة لجمع المعلومات لغرض تقييم أثر جودة البيانات على العمليات الادارية واتخاذ القرار في شركات التأمين العاملة في قطاع غزة. وقد تم توزيع 56 استبانة وجمعت 50 منها بنجاح.

وقد أظهرت النتائج أن هناك تأثيرا كبيرا لجودة البيانات على العمليات الادارية في المؤسسة واتخاذ القرار مما يدل على وجود تأثير مهم لمعايير جودة البيانات على نجاح المؤسسة. وأيضا فإنه يدل على أن شركات التأمين في قطاع غزة على معرفة مسبقة بأهمية جودة البيانات وعلى أنها تعمل على الوصول لأعلى درجات الجودة.

في نهاية هذا البحث, فانه يوصي بتطوير منهجية تقييم لجودة البيانات وتطوير أداة لقياس تكاليف انخفاض جودة البيانات على المؤسسات.

Dedication

To the soul of my parents who rose me up

To my sisters "Alaa, Tasneem and Weam" and brothers "Mohammed and Ibrahim", who stood beside me and give all the support they can

To my lovable companion "Walid" who supported me in all the steps of my thesis

To all my dearest friends who gave support through my Master degree

To all who encouraged, supported, and helped me

I dedicate this research

Acknowledgement

I would like to express my great gratitude to Prof. Dr. Yusuf H. Ashour who was a great mentor, who guided me through my thesis and provided all the help and advice he can to support me until the end.

Also my thanks to my teachers at Faculty of Commerce for the knowledge they gave.

I would like also to send my thanks to insurance companies' staff and especially to Mr. Yunis Abu Marshoud Gaza regional manager of Trust Insurance Company who was a great help to explore the insurance sector in Gaza Strip.

Finally, my appreciation is presented to all who provided help, advice or information in order to complete my master thesis.

Table of content

List of Tables

List of Figures

Chapter 1

The Research General Framework

1.1 Introduction

1.2 The Research Problem Statement

1.3 The Research Variables

1.4 The Research Hypotheses

1.5 The Research Objectives

1.6 The Research Importance

1.1 Introduction

Data is the foundation element in organization's operations. Data and information are becoming more and more important in our lives. In business, having valid and accurate data will lead to more mature decisions, running business processes more effectively and in developing the organization's overall performance.

The beginning of the 80's witnessed the focus of the researchers on the subject of data quality in the management field specifically in the manufacturing systems to detect and eliminate quality problems (Scannapieco, Missier, & Batini, 2005).

Data quality has become a critical concern when dealing with daily business operations (Stvilia, Gasser, Twidale, & Smith, 2007). Operation's data are in continuous access by managers and information users where this access increases the need for high quality data and more accurate recourses to collect data from.

Data quality is recognized as a relevant performance issue and a bottle-neck problem that plays a critical role in all business and governmental applications. It is important to business processes, to decision making activities, and to inter-organizational cooperation requirements (Batini, Cappiello, Francalanci, & Maurino, 2009)

It is estimated that poor data costs United States of America businesses over than $600 billion a year (Fan, Geerts, Ma, Tang, & Yu, 2013). With this comes the need for improved data quality, and to add accuracy and value to business processes (Fan, Geerts, Ma, Tang, & Yu, 2013). Poor data quality appears to be a major aspect in the failure of business projects (Piprani & Ernst, 2008). Which also may lead to bad decisions whish in the first place was based on such poor data. Also it can imply negative consequences in an organization that is not identified and corrected which have significantly negative economic and social impacts on the organization (Haug, Zachariassen, & van Liempd, 2011). Therefore, high data quality is very crucial to organization's success (Haug, Zachariassen, & van Liempd, 2011).

Organizations need to improve business process efficiency to stay competitive and adapt quickly to changing market environments. High data quality is one of the critical factors for the smooth run of business processes and trustful management decisions.

Data quality is very important in any financial environment. Because of the immediate strategic impact of the financial operations, data quality is more strictly monitored than in other business units (Moges, Dejaeger, Lemahieu, & Baesens, 2013). Since the financial data is highly redundant, timely and nonlinear, data quality has significant impacts on financial decision making (Du & Zhou, 2012).

Poor data quality can make it hard for insurance companies to attain an accurate estimate of its insurance-in-force. This may lead to miscalculating both its premium income and

the amount of its loss reserve required for future insurance claims (Herzog, Scheuren, & Winkler, 2007).

In this research, insurance companies in Gaza will be the subject for the research as case study. The focus will be on the effect of the quality of data collected in decisions related to the organizational procedures and that related to insurance decisions.

1.2 Problem Statement

Lack of quality in data collected from various resources to support business processes may lead to non-efficient business operations and less effective decisions which reduce the overall organization's performance, the ways that the business processes are intended to run or more dramatically, to business failure (Haug, Zachariassen, & van Liempd, 2011). Such low quality data negatively impacts the achievement of business objectives by presenting financial impacts, productivity impacts and increasing the risk organization's face while running its operations (Loshin, 2011).

Organizations continuously seek to improve its business processes and making efficient decisions. To reach such goals, the data used in its operations have to be identified correctly. This data can be contained and evaluated to be corrected from whether the sources it was collected from, the tools used in collecting such data and the storing techniques used by the organization.

On the other hand, poor data quality is a recognized bottle neck for business process improvements, so for organizations to develop and expand their operations the need for evaluating its core data is presented.

In this research, an examination of the effect of data quality on business processes and decision making in insurance companies in Gaza Strip will be presented where the question will be:

"To what degree does data quality affect business processes and decision making in insurance companies in Gaza Strip?"

1.3 Research Variables

The variables in this research are divided to dependent and independent variables as follow:

The dependent variables are: Business processes and decision making

The independent variables are: Data Quality, which is sub divided to five major variables:

1- Data completeness

2- Data accuracy

3- Data consistency

4- Validity of data

5- Data relevance

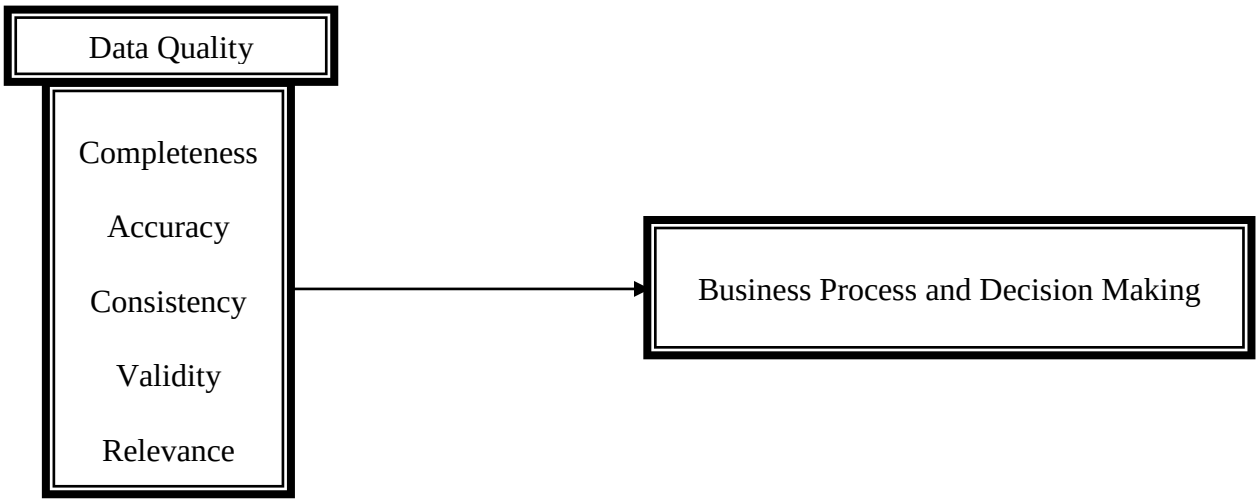

Figure 1.1: Conceptual map, source Researcher 2014

1.4 Research Hypotheses

The research hypotheses are as follow:

H1: There is significant impact between data completeness and business processes and decision making at ($\alpha \leq 0.05$) level

H2: There is significant impact between data accuracy and business processes and decision making in insurance companies in Gaza Strip at ($\alpha \leq 0.05$) level

H3: There is significant impact between data consistency and business processes and decision making in insurance companies in Gaza Strip at ($\alpha \leq 0.05$) level

H4: There is significant impact between validity of data and business processes and decision making in insurance companies in Gaza Strip at ($\alpha \leq 0.05$) level

H5: There is significant impact between data relevance and business processes and decision making in insurance companies in Gaza Strip at ($\alpha \leq 0.05$) level

H6: There is a statistically significant differences attributed to the demographic variables (gender, age, academic qualification, employment level, work

experience, former knowledge of data quality and whether the respondent is a decision maker) in insurance companies in Gaza Strip at $\alpha \leq 0.05$ level

1.5 Research Objectives

This research primary objective is to evaluate the effect of data quality on business processes and decision making. Also the research aimed to achieve the following objectives:

a) To put insight to the area of data quality in insurance companies in Gaza Strip.
b) To identify data quality importance and its effect on insurance companies in Gaza Strip.
c) To identify data quality dimensions that affect insurance companies in Gaza Strip.
d) To investigate if insurance companies in Gaza Strip put interest in data quality issues.
e) To measure the effect of data quality on business processes and decision making in insurance companies in Gaza Strip.

1.6 Research Importance

This research is considered important because it is focusing on the importance of data quality to insurance companies in Gaza Strip. The research importance came from:

a) Drawing organization's attention to data quality importance to business success.
b) Helping organizations valuing the data used in their operations and to better identify sources for data collection and entry procedures to enhance organization overall performance.
c) Offering new knowledge to the researchers in the field of data quality.
d) Attracting more attention to the effect of data quality on organizations.
e) Giving a new topic to the library of IUG with new dimensions for more researches in data quality area while it's a very important and essential field in business success as data becoming a more valuable asset in 21th century.

1.7 Research Operational Definitions

In this research, operational definitions were presented by the researcher to facilitate the understanding of data quality dimensions meaning. See Table 1.1 Data quality dimensions' definition by the researcher.

Table1.1: Data quality dimensions definition's by the researcher

Data quality dimension	Definition
Completeness	is that all the necessary data is available to achieve the task
Accuracy	the degree to which the value in hand matches the real world value or the value that is accepted to be true
Consistency	having the same values representing the data in all its used versions over time
Validity of data	data is considered valid if its correct when compared to the business rules implemented in the organization
Relevance	data is considered relevant if it is what the user needs to achieve a certain task in high efficiency

1.8 Research Contribution

This research is considered to the first on the field of data quality in Palestine and specially in Gaza Strip. This field is considered very important to the business sectors while data quality is essential for all life aspects.

This will open the door for the researchers and put the corner stone to start working on this field of science since it has its significance all aspects.

Chapter 2

The Research Theoretical Framework

2.1 Introduction

2.2 Data Quality

2.3 Data Quality and Business Processes

2.4 Data Quality and Decision Making

2.1 Introduction

A web search for the term “data quality” through the search engine Google, returned about 10,600,000 of results, which is an indicator that data quality issues are real and increasingly important (Google, 2015).

Since organizations are using more and more data to complete its tasks and more sources were used to attain such data, more chances to face poor data quality is presented (Moges, Dejaeger, Lemahieu, & Baesens, 2013). The rapid increase in the exchange and availability of data has made an important need for users to require more and more details on the quality of data used.

Poor data quality has a powerful impact to organization’s failure as much as the good data quality has on organization success (Haug, Zachariassen, & van Liempd, 2011). The more the organization is seeking for its success the more it depends on larger data which use more complicated sources that results to business processes failure and ineffective decisions regarding the operating of its tasks (Haug, Zachariassen, & van Liempd, 2011). Such impacts are customer satisfaction, productivity, employees’ satisfaction, financial operations and increases risks (Haug, Zachariassen, & van Liempd, 2011).

Although these important effects, less organizations are thinking of data quality neither as an asset nor as a liability (Loshin, 2011). And because most organizations are not measuring data quality in objective or quantitative way, the less chance they have to improve their data quality to support its operations (Friedman & Smith, 2011).

Data quality has many dimensions that affect organizations. Among these dimensions there are completeness, accuracy, flexibility, consistency, validity, reference, reliability, and many other dimensions (Haug, Zachariassen, & van Liempd, 2011).

This chapter will define data quality, surf the literature for more clearance on data quality, define data quality dimensions and specifically those used as the variables of this research. And finally will spot some light on the importance of data quality on business processes and decision making.

2.2 Data Quality

2.2.1 Data Quality Definition

To understand the concept of 'data quality', a distinction between data, quality and data quality is presented. And also a clarification is made between data and information.

The term data in the language was defined in oxford dictionary as "Facts and statistics collected together for reference or analysis" (Oxford Dictionaries, 2015). Data can also be defined as "Discrete, objective facts about events" (Haug, Zachariassen, & van Liempd, 2011, p171).

From more business perspective, data can be defined as "Information in raw or unorganized form (such as alphabets, numbers, or symbols) that refer to, or represent, conditions, ideas, or objects. Data is limitless and present everywhere in the universe" (Business Dictionary, 2015)

Data and information have different definitions that distinguish them from each other. Information is defined in (Oxford Dictionaries, 2015) as "Facts provided or learned about something or someone". (IAIDQ, 2015) defines information as "1) Data in context, i.e., the meaning given to data or the interpretation of data based on its context; 2) the finished product as a result of processing, presentation and interpretation of data. (Larry English)" and as "Information can be defined as all inputs that p eople process to gain understanding. It is a difference (a distinction) that makes a difference, an answer to a question. A set of related data that form a message. (Martin Eppler)". Information is also defined as "data transformed by the value-adding processes of contextualization, categorization, calculation, correction and condensation" (Haug, Zachariassen, & van Liempd, 2011, p171).

Information in business is defined as "Data that is (1) accurate and timely, (2) specific and organized for a purpose, (3) presented within a context that gives it meaning and relevance, and (4) can lead to an increase in understanding and decrease in uncertainty. Information is valuable because it can affect behavior, a decision, or an outcome." (Business Dictionary, 2015).

From these definitions, the differences between data and information are that data is the fundamental input for information that is processed and have meaning. In this research, the word data and information will have same meaning.

Quality, on the other hand, was defined in English language as "The standard of something as measured against other things of a similar kind; the degree of excellence of something" (Oxford Dictionaries, 2015).

The meaning of 'quality' depends on the context in which it is applied. In business it's defined as "In manufacturing, a measure of excellence or a state of being free from defects, deficiencies and significant variations. It is brought about by strict and consistent commitment to certain standards that achieve uniformity of a product in order to satisfy specific customer or user requirements." (Business Dictionary, 2015). The International Organization for Standardization (ISO) defines quality in the ISO 8402-

1986 slandered as “the totality of features and characteristics of a product or service that bears its ability to satisfy stated or implied needs” (Business Dictionary, 2015).

Quality is intangible asset that is hardly to be measured. Its conditional and subjective term that may be understood differently by different people. In product/service environment, consumers may focus on the specification or on how to compare the competitors in the marketplace. So quality can be recognized as the degree of excellence in a product, performance or service.

The concept of data quality is vast, including different definitions and interpretations (Oliveira, Rodrigues, & Henriques, 2005). Data quality was defined by different researchers in different but similar meaning definitions. J. M. Juran defined data quality as “data is in high quality if they are fit for their intended uses in operations, decision making and planning" (Wikipedia, 2015).

The International Association for Information and Data Quality (IAIDQ, 2015) uses the phrase data quality in the same meaning of information quality. IAIDQ defines data quality in two different meaning. The first definition is “Information quality: (1) Consistently meeting all knowledge worker and end-customer expectations in all quality characteristics of the information products and services required to accomplish the enterprise mission (internal knowledge worker) or personal objectives (end customer). (2) The degree to which information consistently meets the requirements and expectations of all knowledge workers who require it to perform their processes. (Larry English)”. The second definition is “Information Quality: the fitness for use of information; information that meets the requirements of its authors, users, and administrators. (Martin Eppler)” (IAIDQ, 2015).

J. M. Juran was the first who used the term fitness for use that is widely used in data quality research area. According to (Haug, Zachariassen, & van Liempd, 2011, pp171) “data quality is often defined as 'fitness for use', i.e. an evaluation of to which extent some data serve the purposes of the user” (Shankaranarayanan & Cai, 2006). It has been generally accepted in both academic and industrial settings to use this term as data quality definition where to notice that it may not perfectly conforms to its intended design (Neely, 2005). So to consider data of high quality they have to be Fit for Use in their intended operational, decision-making and other procedures in the organization (Herzog, Scheuren, & Winkler, 2007).

2.2.2 Data Quality Overview

It has been two decades since Wang et al (1995) proposed a framework for data quality based on a product manufacturing approach. In his fundamental first paper, data quality was identified and analyzed according to seven functions: management responsibilities, operation and assurance costs, research and development, production, distribution, personnel management, and the legal function (Neely, 2005).

Data quality was addressed in different contexts including statistics, management and computer science. Statisticians were the first to study some of the problems related to data quality by proposing some mathematical theories for duplicates in statistical data sets, this was in the late 60's. Then management researchers followed them, who, at the beginning of the 80's, focused on how to control data manufacturing systems in order to find and remove any data quality problems and focused in the aspects or dimensions of data quality like accessibility, accuracy, relevancy and completeness. Only at the beginning of the 90's computer scientists began considering the problem of data quality from more technical point of view as in electronic data (databases, data warehouses) (Scannapieco, Missier, & Batini, 2005) (Oliveira, Rodrigues, & Henriques, 2005) (Batini & Scannapieca, 2006). In this context, the researcher follows the management perspective.

When considering data quality, we must keep in mind the distinct differences between data and other goods (Neely, 2005). Data is much more dynamic in its nature, created and used from different users, used across different operational and analytical applications (Loshin, 2011). A product or service is exhausted in its use. Data, on the other hand, is not. Many users with different needs and working environments can use the same data at the same time and the data will still available for all of them. This characteristic of data is important when evaluating the quality of data in business (Neely, 2005). Also, data quality for one use may not be applicable for other uses. For example, the level of data completeness for accounting tasks may not be considered required for sales prediction tasks (Moges, Dejaeger, Lemahieu, & Baesens, 2013)

Data quality can be classified into three major categories: master data, transactional data, and historical data (Haug, Zachariassen, & van Liempd, 2011) (McGilvray & Thomas, 2008). Master data is defined as the basic characteristics illustrations of business entities such as customers, products, employees, and suppliers. Typically, master data are created once, used more frequently, and doesn't change often over time. Transactional data describes relevant events in an organization, e.g. invoices, orders, payments, storage records, etc. Transactional data uses master data so if there is any slight error in the master data, errors will appear in transactional data and then affecting the organizations revenue. For instance, if the address of a customer was wrong, this may result to deliver the product to the wrong address (Haug & Arlbjørn, 2011) (McGilvray & Thomas, 2008). Historical data are significant facts, of a certain point in time, that should not be changed except to correct an error (McGilvray & Thomas, 2008). For example, historical price, price/earnings ratio, revenues and revenue growth, earnings and earnings growth. Most organizations use data in two directions: transactional/operational use (running the business), and analytical use (improving the business) (Loshin, 2011).

Data quality is a critical aspect for reaching strategic and operational business goals (Ofner, Otto, & Osterle, 2012) (Oliveira, Rodrigues, & Henriques, 2005). Where data is the foundation for operational, tactical and strategic decisions, it's critically important for organization's data to be in a suitable level of quality (Haug & Arlbjørn, 2011). All organizations make decisions and service customers based on the data they have at their hand (Geiger, 2004). Improved decision making, effective customer relationship

management (CRM), adherence to regulatory requirements and supply chain management excellence all depends on such high quality data (Ofner, Otto, & Osterle, 2012).

Most literature considers data as an asset that is generated through multiple processes, with multiple feeds of raw data that are combined, processed, and fed out to multiple customers both inside and outside the organization (Loshin, 2011). Data quality involves multiple stakeholders, with possibility of conflicts of interest among them (Storey, Dewan, & Freimer, 2012). In spite of that, it's still premature to think that organizations list data as a significantly important item, either as an asset or as a liability on its balance sheet (Loshin, 2011).

Poor data is a critical problem for organizations leading to inefficient daily operations, incorrect decision making and eventually wasting business resources (Chiang & Miller, 2008). Data quality has serious consequences, reaching the significance, efficiency and effectiveness of organizations (Batini & Scannapieca, 2006). For a piece of data to considered poor, its value must violate an explicit statement of its correct value (Chiang & Miller, 2008). Although there are much literature claims the importance of addressing poor data quality, only very few studies show how to identify, categorize, measure and establish fundamental relations between poor data quality and their monetary effects (Haug, Zachariassen, & van Liempd, 2011) (Eppler & Helfert, 2004). (Herzog, Scheuren, & Winkler, 2007, p7) said that "discovering whether data are of acceptable quality is a measurement task, and not a very easy one". Poor data quality can be a major cause for damages on business processes while there are various associated costs linked with its collection, storage, security and proper quality level maintenance (Storey, Dewan, & Freimer, 2012). In spite of the need to repair these damages, there are many challenges associated with establishing procedures to assess risk related to data failures as well as to monitor conformance to business user expectations (Loshin, 2011).

Data quality issues are a perfect example of the business saying, "You can't manage what you don't measure" (Friedman & Smith, 2011, p10). Most organizations fail at building a business case for data quality improvement efforts because they don't make the required effort to measure the quality of their data in objective or quantitative way (Friedman & Smith, 2011) (Batini & Scannapieca, 2006).

Many industry surveys indicated that data quality is an area to which many organizations seem not giving enough attention to or even know how to deal with it efficiently (Haug, Zachariassen, & van Liempd, 2011). Organizations with time are discovering that poor data quality have a significant impact on their strategic business initiatives that holding them from achieving the growth they seek and the competitiveness that they desire (Friedman & Smith, 2011). In spite of this importance, there seems to be a general agreement in literature that poor data quality is still a problem in many organizations (Haug, Zachariassen, & van Liempd, 2011).

To be able to link data quality issues to business impacts, organizations must be able to both classify data quality expectations as well as business impact criteria (Loshin, 2006). Loshin (2006) categorized poor data quality impacts that affect organizations into four dimensions (see Figure 2.1 impacts of poor data quality):

1- Increased Costs
2- Decreased Revenues
3- Decreased confidence
4- Increased Risk

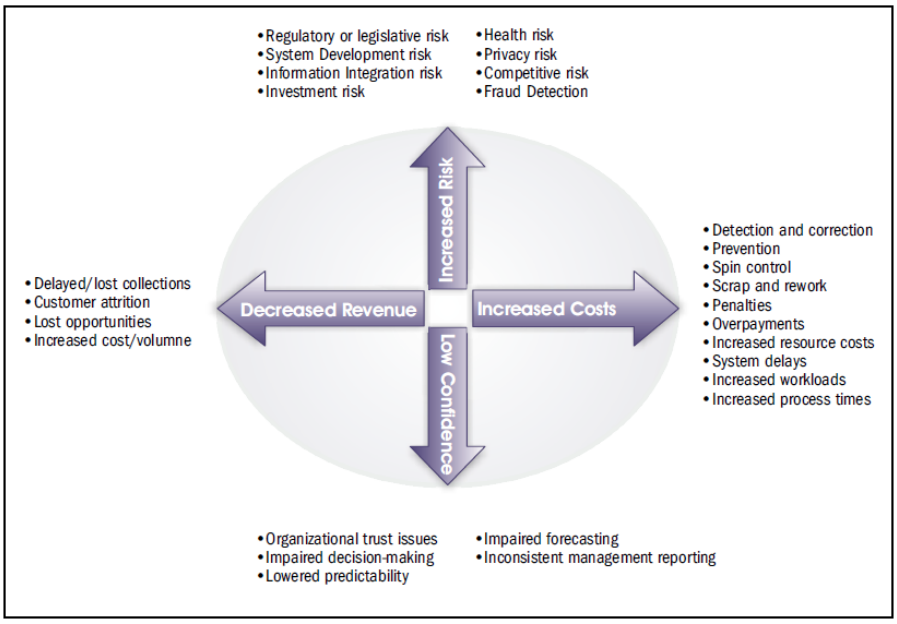

Figure 2.1: Impacts of poor data quality, source (Loshin, 2006, p8)

Calculating costs caused by poor data quality is in-deed difficult because many of it are indirect costs, where there is no clear link between poor data quality and the negative monetary effects, noting that data quality costs are context-dependent (i.e. the types of damage caused by poor data quality is dependent on the type of the managed data, its uses and responses) (Eppler & Helfert, 2004).

Poor data quality costs are divided into two categories, direct and indirect costs. Direct costs are the negative monetary effects that are directly associated with poor data quality (costs of re-entering incomplete data, costs of compensation for damages to others based on poor data). Indirect costs are those negative monetary effects that associated with intermediate effects from poor data quality (low quality decisions) (Eppler & Helfert, 2004) (see Figure 2.2 data quality costs).

Data quality costs	Costs caused by low data quality	Direct costs	Verification costs
			Re-entry costs
			Compensation costs
		Indirect costs	Costs based on lower reputation
			Costs based on wrong decisions or actions
			Sunk investment costs
	Costs of improving or assuring data quality	Prevention costs	Training costs
			Monitoring costs
			Standard development and deployment costs
		Detection costs	Analysis costs
			Reporting costs
		Repair costs	Repair planning costs
			Repair implementation costs

Figure 2.2: Data quality costs, source (Haug, Zachariassen, & van Liempd, 2011, p177)

Poor data quality effects organizations in several ways and can imply many negative consequences that if it's not identified and corrected can have significantly negative social and economic impacts on the organization (Haug, Zachariassen, & van Liempd, 2011). The quality of data strongly effects the quality of the results (garbage in, garbage out principle) (Oliveira, Rodrigues, & Henriques, 2005). At the operational level, poor data quality has an impact on customer and employee satisfaction, increased running costs, lower organizational performance and increases operational costs since resources are spent identifying and correcting errors (Moges, Dejaeger, Lemahieu, & Baesens, 2013) (Haug, Zachariassen, & van Liempd, 2011) (Chiang & Miller, 2008). At the strategic level, poor data quality affects the quality of the decision making process (Moges, Dejaeger, Lemahieu, & Baesens, 2013) (Chiang & Miller, 2008). Since data are created and used in all organizational daily operations and are critical inputs to all decisions, it makes a significant contribution to organizational culture and may create negative effects on the organizational culture (Haug, Zachariassen, & van Liempd, 2011).

Data is used as a shared resource in multiple business processes, such as procurement, manufacturing, or marketing. When using this data, each business process follows its own needs and compiling data according to its own procedures. As a result, data are often used across business processes in an uncoordinated way, which raises negative effects on data quality and produces higher costs (Ofner, Otto, & Osterle, 2012).

Poor data quality can reduce customer satisfaction, can lower employee job satisfaction leading to extreme turnover and then losing of key process knowledge and can also raise organizational mistrust and make it hard to support efforts needed to achieve required improvements (Herzog, Scheuren, & Winkler, 2007) (Chiang & Miller, 2008).

In the same time, it's difficult to set organizational data quality objectives and auditing their implementation. This difficulty came from the fact that data quality depends on the whole business process and all of its affected users (Storey, Dewan, & Freimer, 2012).

Achieving high data quality is not an easy process to do. According to (Haug & Arlbjørn, 2011) data quality have some barriers that stands in the way of organizations

to obtain high data quality. In (Figure 2.3 data quality barriers), possible data quality barriers are listed according to (Haug & Arlbjørn, 2011):

(1) lack of delegation of responsibilities for maintenance of data;
(2) lack of rewards for ensuring valid data;
(3) lack of data control routines;
(4) lack of employee competencies; and
(5) lack of user-friendliness of the software used to manage data.

Figure 2.3: Data quality barriers, source (Haug & Arlbjørn, 2011, p296)

The set of potential data sources has dramatically increased in size and scope (Batini, Cappiello, Francalanci, & Maurino, 2009). Also the vast development of information technology has enabled organizations to collect and store huge amounts of data leading to more complexity in managing them (Haug, Zachariassen, & van Liempd, 2011) & (Haug & Arlbjørn, 2011).

Poor data quality risk increases as larger and more complex data resources are collected and managed in organizations today (Moges, Dejaeger, Lemahieu, & Baesens, 2013) (Haug & Arlbjørn, 2011).

When having poor data quality, it means that it becomes difficult to build trust in the organization's data, which may indicate a lack of user acceptance of any initiatives based on such data (Haug, Zachariassen, & van Liempd, 2011). Therefore, it's no longer accepted to ignore such defects in data, and organizations must demonstrate the accuracy of information that they report to auditors, regulators and the public (Friedman & Smith, 2011).

Many industry experts such as Gartner Group, Price Waterhouse Coopers and The Data Warehousing Institute, have done researches that claims to identify a crisis in data quality management (Haug, Zachariassen, & van Liempd, 2011) (Haug & Arlbjørn, 2011). (Haug, Zachariassen, & van Liempd, 2011, p174) & (Haug & Arlbjørn, 2011, p294) summarizes the findings from such surveys into the following points (quoted):

1- 88% of all data integration projects either fail completely or significantly over-run their budgets.
2- 75% of organizations have identified costs stemming from dirty data.
3- $611bn per year is lost in the US in poorly targeted mailings and staff overheads alone.
4- According to Gartner, bad data is the number one cause of CRM system failure.
5- Less than 50% of companies claim to be very confident in the quality of their data.

6- Only 15% of companies are very confident in the quality of external data supplied to them.
7- Organizations typically overestimate the quality of their data and underestimate the cost of errors.
8- Business processes, customer expectations, source systems and compliance rules are constantly changing. Data quality management systems must reflect this.

So how high data quality can be obtained? This question was answered by (Herzog, Scheuren, & Winkler, 2007) in three different procedures to use:

1- Prevention: keep poor data out of the organization's system: this could be done by allowing the system to edit the data before using it.
2- Detection: proactively look for poor data already entered in the system: fix the poor data by using different measurement methods related to the type of data.
3- Repair: let the poor data find you and then fix things.

As illustrated, approaches differ in the timing, i.e. when to address poor data. In the real life, all the three procedures are used simultaneously with percentage of prevent: 10%, detect: 30% and repair: 60% (Herzog, Scheuren, & Winkler, 2007).

2.2.3 Data quality management

Data Quality Management (DQM) is defined as "quality oriented management of data as an asset, that is, the planning, provisioning, organization, usage, and disposal of data that supports both decision-making and operational business processes, as well as the design of the appropriate context, with the aim to improve data quality on a sustained basis" (Storey, Dewan, & Freimer, 2012, p434).

Data quality management deals with the creation and implementation of roles, responsibilities, policies, and procedures concerning the acquisition, repair, distribution, and disposition of data (Geiger, 2004). It involves multiple stakeholders, with possibility of conflict of interest among them (Storey, Dewan, & Freimer, 2012).

Data quality management involves ensuring that data is accurate, reliable, consistent and complete. It requires a strategy and it involves continuous monitoring and maintenance of data quality. It also provides a focus on ongoing, measurable improvements in data quality (O'Neal, 2014).

Deploying a data quality management group is not easy; there are significant challenges that must be firstly overcome. Some of these challenges are responsibility, cross functionality, problem recognition, discipline, investment, on-going effort and return on investment (see Figure 2.4 Data quality management challenges) (Geiger, 2004):

- No business unit or department feels it is responsible for the problem.
- It requires cross-functional cooperation.
- It requires the organization to recognize that it has significant problems.
- It requires discipline.
- It requires an investment of financial and human resources.
- It is perceived to be extremely manpower-intensive.
- The return on investment is often difficult to quantify.

Figure 2.4: Data quality management challenges, source (Geiger, 2004, p3)

2.2.4 Data quality dimensions

One way to understand the concept of data quality is by dividing it into smaller dimensions (Haug, Zachariassen, & van Liempd, 2011). However, there is no single, precisely defined set of data quality dimensions among literature and that's because data quality is a complex, context dependent and multi-dimensional concept (Moges, Dejaeger, Lemahieu, & Baesens, 2013) (Batini & Scannapieca, 2006). This situation has led to much confusion within the data quality community and more importantly to business stakeholders (Askham, et al., 2013).

A data quality dimension was defined by (Askham, et al., 2013, p3) "a recognized term used by data management professionals to describe a feature of data that can be measured or assessed against defined standards in order to determine the quality of data".

Data quality dimensions are indicators helping to measure and communicate the quality of data (Askham, et al., 2013). A data quality dimension is a feature of data and a way to classify data quality needs. Dimensions are used to define, measure, and manage the quality of the data (McGilvray, 2011).

These dimensions represent the views, criteria and measurement characteristics for data quality problems that can be assessed, interpreted and possibly improved individually by assigning scores to these dimensions. The overall data quality value can be calculated as an accumulated value of single dimensions relevant to its given context (Sattler, 2009).

Data quality dimensions are dependent, correlations exist among them and effects each other (see Figure 2.5 data quality dimensions have relation among each other) (Moges, Dejaeger, Lemahieu, & Baesens, 2013) & (Scannapieco, Missier, & Batini, 2005). Organizations select data quality dimensions based on their business context, requirements, level of risk, etc. Noting that organizations determine which dimensions to use and the weight of each dimension to contribute to the overall data quality measurement (Askham, et al., 2013). However, favoring one dimension on the others may negatively affect other dimensions (Moges, Dejaeger, Lemahieu, & Baesens, 2013)

& (Scannapieco, Missier, & Batini, 2005). Compromises may need to be made between different dimensions according to the context. For example, if considering an e-banking system, completeness, accuracy and consistency dimensions are more important than timeliness dimension, and therefore delays are mostly understood in favor of correctness (Scannapieco, Missier, & Batini, 2005).

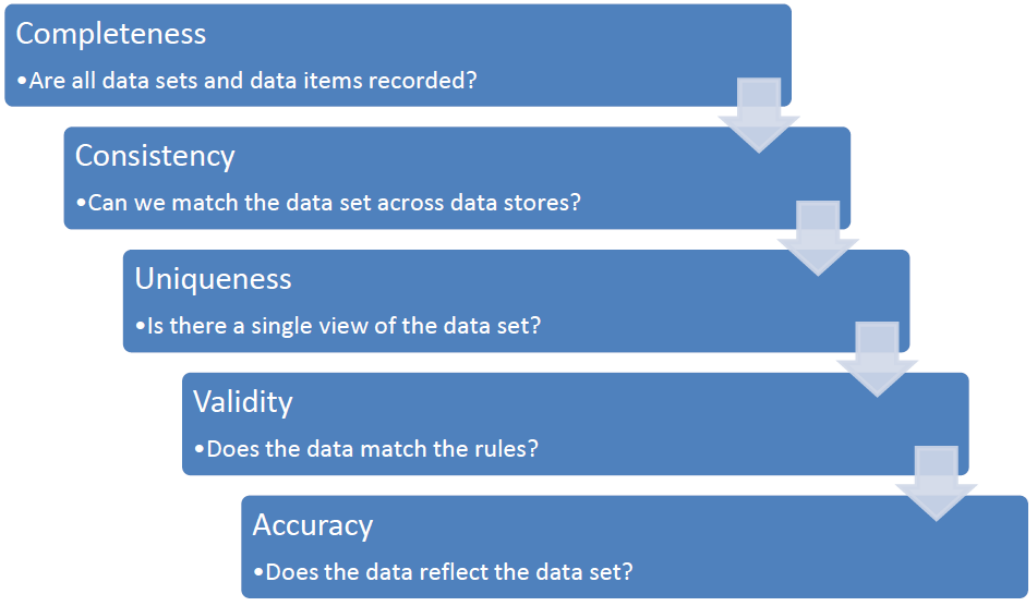

Figure 2.5: Data quality dimensions have relation among each other, source (Askham, et al., 2013, p6)

There are three main approaches for introducing comprehensive sets of data quality dimensions' definitions, theoretical, empirical and intuitive. The theoretical approach implements a formal model in order to justify or describe the dimensions. The empirical approach uses a set of experiments, interviews, and questionnaires to define dimensions. And the intuitive approach uses common sense and practical experience to define them (Batini & Scannapieca, 2006).

Data quality has many dimensions that differs between literature but have many similarities. (Haug, Zachariassen, & van Liempd, 2011) in (Figure 2.6 Data quality dimensions) has listed what they consider a data quality dimensions.

Other set of dimensions was defined by (Chen, Hailey, Wang, & Yu, 2014) by dividing what dimension to be considered of high data quality and what to be of poor data quality. (Figure 2.7 data quality dimensions) illustrates that.

Accuracy	25	Flexibility	5	Sufficiency	3	Informativeness	2
Reliability	22	Precision	5	Usableness	3	Level of detail	2
Timeliness	19	Format	4	Usefulness	3	Quantitativeness	2
Relevance	16	Interpretability	4	Clarity	2	Scope	2
Completeness	15	Content	3	Comparability	2	Understandability	2
Currency	9	Efficiency	3	Conciseness	2		
Consistency	8	Importance	3	Freedom from bias	2		

Figure 2.6: Data quality dimensions, source (Haug, Zachariassen, & van Liempd, 2011, p172)

Item	Attribute
High data quality (38)	Completeness, accuracy or positional accuracy, timeliness or up-datedness or currency, validity, periodicity, relevance, reliability, precision, integrity, confidentiality or data security, comparability, consistency or internal consistency or external consistency, concordance, granularity, repeatability, readily useableness or usability or utility, objectivity, ease with understanding, importance, reflecting actual sample, meeting data standards, use of standards, accessibility, transparency, representativeness, disaggregation, data collection method or adjustment methods or data management process or data management
Poor data quality (11)	Missing data, under-reporting, inconsistencies, data errors or calculation errors or errors in report forms or errors resulted from data entry, invalid data, illegible hand writing, non-standardization of vocabulary, and inappropriate fields

Figure 2.7: Data quality dimensions, source (Chen, Hailey, Wang, & Yu, 2014, p 5177)

As presented above, each data quality researcher or data quality leading corporations has its different primary set of data quality dimensions but not rejecting others dimensions. As for (Askham, et al., 2013) the core data quality dimensions are Completeness, Uniqueness, Timeliness, Validity, Accuracy and Consistency (see Figure 2.8 core data quality dimensions). Herzog, Scheuren, & Winkler (2007) sets the most important dimensions to Relevance, Accuracy, Timeliness, Comparability and Completeness. For (Batini & Scannapieca, 2006) the set include Accuracy, Completeness, Time-Related Dimensions (Currency, Timeliness, and Volatility) and Consistency. (Haug, Zachariassen, & van Liempd, 2011, p171) states "Ballou and Pazer (1985), has divide data quality into four dimensions: accuracy, timeliness, completeness, and consistency". Data quality dimensions according to (Friedman & Smith, 2011) are validity, completeness, consistency and accuracy. (Moges, Dejaeger, Lemahieu, & Baesens, 2013) have summarized the most repeated data quality dimensions in the literature along with their definitions (see Figure 2.9 Most recurring data quality dimensions in literature

and their definitions). Completeness, Accuracy and Timeliness were the most used dimensions amongst a total of 49 dimension of data quality (Chen, Hailey, Wang, & Yu, 2014). The majority of researchers define a basic set of data quality dimensions, including Accuracy, Completeness, Consistency, and Timeliness (Batini, Cappiello, Francalanci, & Maurino, 2009).

Figure 2.8: Core data quality dimensions, source (Askham, et al., 2013, p7)

Cat.	DQ dimensions	Definitions
Intrinsic	Accuracy (AC)	The extent to which data are certified, error-free, correct, flawless and reliable
	Objectivity (OBJ)	The extent to which data are unbiased, unprejudiced, based on facts and impartial
	Reputation (REP)	The extent to which data are highly regarded in terms of its sources or content
Contextual	Completeness (COM)	The extent to which data are not missing and covers the needs of the tasks and is of suffcient breadth and depth to the task at hand
	Appropriate amount (APM)	The extent to which the volume of information is appropriate for the task at hand
	Value-added (VAD)	The extent to which data are beneficial and provides advantages from its use
	Relevance (REL)	The extent to which data are applicable and helpful for the task at hand
	Timeliness (TIM)	The extent to which data are suffciently up-to-date for the task at hand
	Actionable (ACT)	The extent to which data is ready for use
Representation	Interpretable (INT)	The extent to which data are in appropriate languages, symbols, and the definitions are clear
	Easily understandable (EU)	The extent to which data are easily comprehended
	Representational consistent (RC)	The extent to which data are continuously presented in same format
	Concisely represented (CR)	The extent to which data is compactly represented, well presented, well-organized, and well-formatted
	Alignment (AL)	The extent to which data is reconcilable (compatible)
Access	Accessibility (ACC)	The extent to which data is available, or easily and swiftly retrievable
	Security (SEC)	The extent to which access to data is restricted appropriately to maintain its security
	Traceability(TRA)	The extent to which data is traceable to the source

Figure 2.9: Most recurring data quality dimensions in literature and their definitions, source (Moges, Dejaeger, Lemahieu, & Baesens, 2013, p46)

In all these data quality dimension sets, the researchers have agreed on common dimensions like accuracy, completeness and consistency. In this research a set of five data quality dimensions is used to evaluate the effect of data quality on insurance companies in Gaza strip. This set was selected after reviewing literature and selecting the most important dimensions to achieve this task. These dimensions are Completeness, Accuracy, Consistency, Validity of data and Relevance.

Understanding data quality dimensions is the first step to maintain data quality in the business. Being able to isolate data flaws by dimension allows managers to apply techniques to improve both the data and the processes that create and manipulate that data. In the following section, a comprehensive view on the selected data quality dimensions is presented.

2.2.4.1 Completeness

Completeness is an important data quality dimension that all literature agreed on. Completeness is defined in (Oxford Dictionaries, 2015) "The state or condition of having all the necessary or appropriate parts". (Moges, Dejaeger, Lemahieu, & Baesens, 2013, p46) defined completeness as "The extent to which data are not missing and covers the needs of the tasks and is of sufficient breadth and depth to the task at hand". Completeness is defined by (Batini, Cappiello, Francalanci, & Maurino, 2009, p16:7) as "the degree to which a given data collection includes data describing the corresponding set of real-world objects". Another definition by (Herzog, Scheuren, & Winkler, 2007, p10) "no records are missing and that no records have missing data elements". Also it was defined by (Scannapieco, Missier, & Batini, 2005, p6) "the extent to which data are of sufficient breadth, depth, and scope for the task at hand". (Shankaranarayanan & Cai, 2006, p304) defines completeness as "the extent to which the value is present for that specific data element".

Definitions differ in the context to which they refer. But in all the above mentioned definitions, researchers have defined completeness in similar way with substantial agreement on the abstract definition that is no missing data and is a fit for the task in hand.

A proper definition for completeness would be "completeness is that all the necessary data is available to achieve the task" (The Researcher, 2015). It means that data is complete so it doesn't leave any room for questions. So to say this data is complete or not, it should have no missing data or blanks. In some cases, missing data is irrelevant and do not have to be 100% complete, but when the data that is missing is critical to a specific business process, completeness becomes an issue. Completeness is measured to the degree to which it matches user's expectations and data availability.

In order to understand completeness, it is important to know why the value is missing. A value can be missing either because it exists, but is not known, or because it does not exist (Batini, Cappiello, Francalanci, & Maurino, 2009).

The evaluation of completeness is quite straightforward, as long as the focus is on whether the data is complete or not rather than the degree to which the data is complete (Haug & Arlbjørn, 2011).

Missing data (incomplete data) lead to serious implications affecting business processes. In context such as financial, missing entire data entry can have catastrophic consequences. In survey and administrative context, missing data can have serious consequences if they are associated with large organizations or with a large amount of employees (Herzog, Scheuren, & Winkler, 2007).

2.2.4.2 Accuracy

Accuracy is confirmed to be the most important data quality dimension (Moges, Dejaeger, Lemahieu, & Baesens, 2013). Accuracy is defined in English language as "The quality or state of being correct or precise" (Oxford Dictionaries, 2015). Several definitions were provided for the term accuracy, (IAIDQ, 2015) defines accuracy as "the

degree of conformity of a measure to a standard or a true value. Level of precision or detail". (Askham, et al., 2013, p12) defines accuracy as "The degree to which data correctly describes the real world object or event being described". (McGilvray, 2011, p1) defined accuracy as "A measure of the correctness of the content of the data (which requires an authoritative source of reference to be identified and accessible)". (Stvilia, Gasser, Twidale, & Smith, 2007, p32) defined accuracy "The degree to which an information object correctly represents another information object, process, or phenomenon in the context of a particular activity or culture". Another definition by (Batini & Scannapieca, 2006) is the extent to which data are correct, reliable and certified free of error. (Scannapieco, Missier, & Batini, 2005, p4) also defined accuracy as "accuracy measures the distance between a value v and a value v' which is considered correct". Some researchers give accuracy the same meaning as validity (Chen, Hailey, Wang, & Yu, 2014).

Although the different definition, they all have the same meaning. Accuracy is defined as "the degree to which the value in hand matches the real world value or the value that is accepted to be true" (The Researcher, 2015).

Accuracy answers the questions: Do data precisely represent the real-world values, without incorrect spellings of product or person names, addresses? Dose the data have the ability to draw correct conclusions? (Loshin, 2006). Accuracy is a dimension of data quality that deals with the data being exact when describing the physical characteristics of products. It can be measured as the minimum gab between data and reality (lower is better). Researchers argue that accuracy is the easiest to evaluate from other dimensions, as it is only a matter of analyzing the difference between the correct value and the real value (Haug, Zachariassen, & van Liempd, 2011)

Not giving enough attention to accuracy can impact operational and analytical applications (Loshin, 2006). In some cases, accuracy can be easily detected (e.g. misspellings) but it is more difficult in other cases (Scannapieco, Missier, & Batini, 2005). Accuracy is considered one of the dimensions that are independent from the context in which the data are used in (Moges, Dejaeger, Lemahieu, & Baesens, 2013).

2.2.4.3 Consistency

Consistency is one of the data quality dimensions that most literature agreed on. It's defined in English language by (Oxford Dictionaries, 2015) as "The quality of achieving a level of performance which does not vary greatly in quality over time". (IAIDQ, 2015) defines consistency "A measure of information quality expressed as the degree to which a set of data is equivalent in redundant or distributed databases. (Larry English)". Consistency is defined by (Gharib & Giorgini, 2015, p9) "consistency means that multiple recordings of the same information should be the same across time and space". (Askham, et al., 2013, p13) defined consistency as "The absence of difference, when comparing two or more representations of a thing against a definition". (Alsaidi & Zbar, 2013, p23) defined consistency as "applying the same policies, rules and procedures

from time to time, where the absence of this dimension prevents the interpretation of the organization results across periods of time". (Du & Zhou, 2012, p79) defines consistency as "two or more representations of the same concept (or instance) exist". Another definition is "A measure of the equivalence of information stored or used in various data stores, applications, and systems, and the processes for making data equivalent" (McGilvray, 2011, p1). (Batini & Scannapieca, 2006, p38) defines consistency as "the extent to which data are always presented in the same format and are compatible with the previous data".

After reviewing the definitions in the literature, consistency can be defined as "having the same values representing the data in all its used versions over time" (The Researcher, 2015).

Consistency is a data quality dimension that detects the violation of rules defined over a set of data items (Scannapieco, Missier, & Batini, 2005). Consistency answers the questions: Do the values in multiple units match? Is the data the same, regardless of where it exists? Is data identical for all processes and departments? Is the data in harmony across the organization? Do distinct data provide conflicting values about the same original data object? (Loshin, 2006).

The evaluation of consistency dimension is more complex than the other dimensions (accuracy or completeness), that is because this evaluation requires two or more representation outlines for comparison (Haug & Arlbjørn, 2011) (Haug, Zachariassen, & van Liempd, 2011). Data consistency means that data across the entire organization should be in synchronization with each other (Execution-MiH, 2015).

Data can be accurate as representing the real world, but still inconsistent. It is possible to have consistency without having validity or accuracy (Askham, et al., 2013). Data is inconsistent, when it is unified in one department of an organization, but not across all the other departments of the organization. Also data can be complete, but inconsistent (Execution-MiH, 2015). Consistency can be achieved by making the data representation with the same value in all cases (Tejay, Dhillon, & Chin, 2006). That is data to be clear, unambiguous, consistent and have the same structure and format across the organization (Chapman, 2005).

2.2.4.4 Validity of data

Validity of data is defined in (Oxford Dictionaries, 2015) as "The quality of being logically or factually sound; soundness or cogency" and "The state of being legally or officially binding or acceptable". (IAIDQ, 2015) defines validity "A characteristic of information quality measuring the degree to which the data conforms to define business rules (Larry English)". (Askham, et al., 2013, p11) has defined validity as "Data are valid if it conforms to the syntax (format, type, range) of its definition". According to (Piprani & Ernst, 2008, p756) validity is "Conformance of data values that are edited for acceptability reducing the probability of error". (Stvilia, Gasser, Twidale, & Smith, 2007, p31) describes validity as "The extent to which information is legitimate or valid

according to some stable reference source such as a dictionary or set of domain constraints and norms".

Summarizing all the above definitions, validity is "data is considered valid if it's correct when compared to the business rules implemented in the organization" (The Researcher, 2015).

Validity and accuracy are related dimensions because in order to be accurate values must be valid and in the correct representation (Askham, et al., 2013). In spite of this relation, validity is not accuracy. A value may be valid, but still be incorrect. For example, a customer date of first service can be a valid date (within the correct range) and yet not be an accurate date (IAIDQ, 2015). Validity answers the question Does the data match the rules? (Askham, et al., 2013). Organizations must define data validity constraints to measure data quality in their systems (Loshin, 2011), noting that all data can be measured for validity (Askham, et al., 2013). Data is considered valid when it can be verified as being true and satisfying appropriate standards related to business rules and other dimensions such as accuracy, timeliness, completeness and security. Even if validity dimension is high, other dimensions may be low. Business rules are expressed as data quality expectations for measuring aspects of data validity. Examples for such rules are: (Loshin, 2011) & (Loshin, 2006)

- What data is missing or unusable?
- Which data values are in conflict?
- Which records are duplicated?
- What linkages are missing?

2.2.4.5 Relevance

Relevance is one of the data quality dimensions that are context dependent. Relevance definition in English language is "relevant: Closely connected or appropriate to what is being done or considered" and "Appropriate to the current time, period, or circumstances; of contemporary interest" (Oxford Dictionaries, 2015). (Chen, Hailey, Wang, & Yu, 2014, p5193) defines relevance as "Relevance: comparing the data categories with those in upper level report to evaluate whether the data collected satisfied management information needs". Another definition by (McGilvray, 2011, p1) is "A measure of the perception of and confidence in the quality of the data; the importance, value, and relevance of the data to business needs". (Stvilia, Gasser, Twidale, & Smith, 2007, p32) defines relevance as "The extent to which information is applicable in a given activity". Relevance is the degree to which statistics meet current and potential user needs. (Ehling, et al., 2007, p9) "Is the degree to which statistics meet current and potential user needs. It refers to whether all statistics that are needed are produced and the extent to which concepts (definitions, classifications etc.) reflect user needs".

A proper definition for relevance that go with this research context is ”data is considered relevant if it is what the user needs to achieve a certain task in high efficiency” (The Researcher, 2015).

Relevance is one of the data quality dimensions that cannot be measured objectively. For example, data relevance typically depends on the task, because data that are relevant for one task doesn’t have to be relevant to another task (Moges, Dejaeger, Lemahieu, & Baesens, 2013).

Relevance deals with the use of data by its users. It is concerned with the relation between data, information and behavior, in a given context (Tejay, Dhillon, & Chin, 2006). This does not mean that irrelevant information to one user is of poor quality for the rest of the users or the organization.

Relevance answers the questions: Is the data helping the users to know what they want? Do the data meet the basic needs for which they were collected and used? Is it useful to the business in its pursuit of objectives? Can the data be used for additional purposes? (Herzog, Scheuren, & Winkler, 2007).

2.3 Data quality and Business Processes

Data is considered with a value to the organization only if it supports a business process or organizational decision making (Askham, et al., 2013). Data are collected, stored, and exchanged in all departments of the organization to provide services to business processes (Batini & Scannapieca, 2006).

A business process consists of a group of activities that are performed in coordination to achieve organizational success. A classic business process begins with a stakeholder (inside or outside the organization), the process goes through a number of different business departments that may use, update, or enlarge the data gathered from previous steps in the process to finally achieve the task trigger initially by the stakeholder (Storey, Dewan, & Freimer, 2012). An example of a regular business process is the purchase process illustrated in (Figure 2.10 typical purchase process) (Storey, Dewan, & Freimer, 2012).

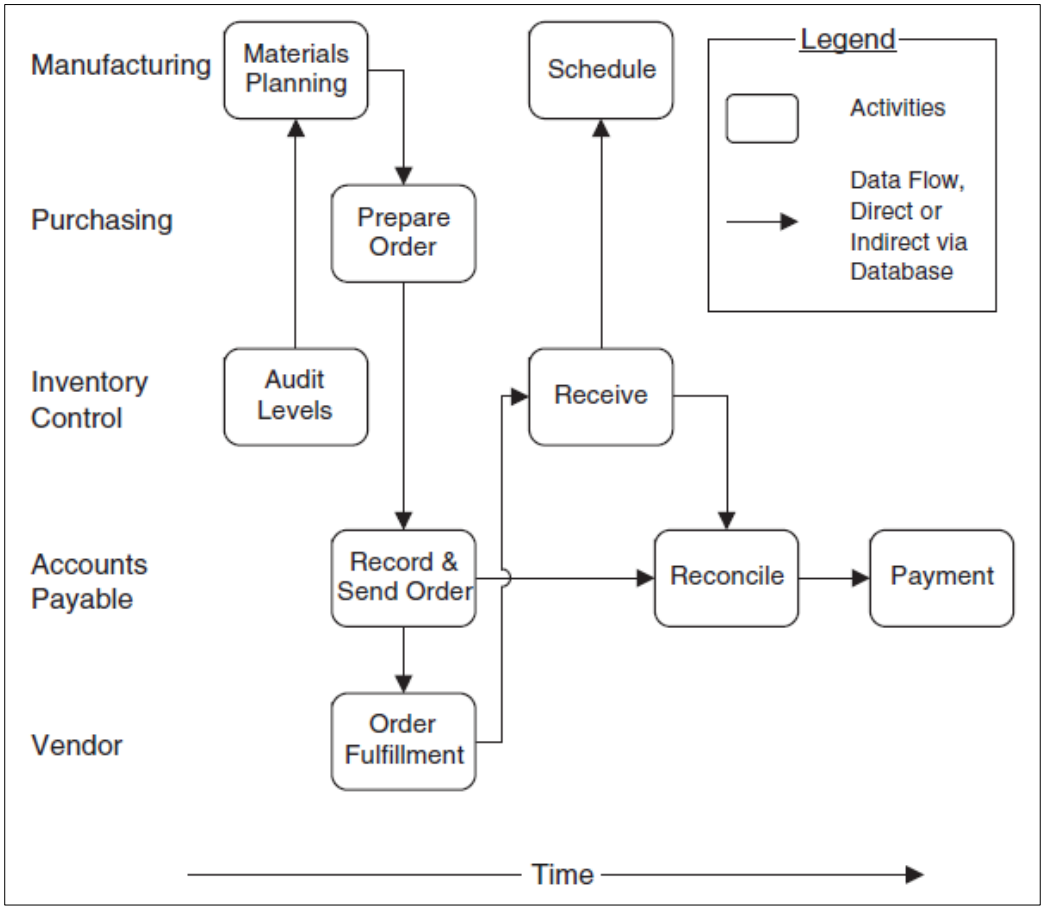

Figure 2.10: Typical purchase process, source (Storey, Dewan, & Freimer, 2012, p 435)

Data quality is a significant issue for operational processes in organizations. The Data Warehousing Institute report on data quality states that there is a significant gap between awareness and reality regarding data quality in many organizations, and this problem costs U.S. businesses more than 600 billion dollars a year (Batini & Scannapieca, 2006). To focus in the data quality problem, organizations tend to set data quality goals. Setting reasonable goals and auditing its implementation is a difficult procedure because data quality depends on the whole business process with all its users with possibly of conflicts of interest amongst them. (Storey, Dewan, & Freimer, 2012).

The failure to obtain high data quality leads to a significant impact on business processes. Some examples of these impacts are (Askham, et al., 2013, p4) (quoted):

- Incorrect or missing email addresses would have a significant impact on any marketing campaigns.
- Inaccurate personal details may lead to missed sales opportunities or a rise in customer complaints.
- Goods can get shipped to the wrong locations.
- Incorrect product measurements can lead to significant transportation issues.

These impacts can be classified into categories related to organizational main processes. Since every business industry is different, there may be several ways for poor data quality to affect organizational operations or for achieving its goals (Loshin, 2011) (Friedman & Smith, 2011). According to (Loshin, 2011), there are four categories (but not limited to) to identify the impacts of poor data quality to the organization. These impacts categories are (Loshin, 2011):

1- Financial impacts: are associated with expectations related to costs, financial management, and revenues. Some examples are: increased operating costs, decreased revenues, missed opportunities, reduction or delays in cash flow, or increased penalties, fines, or other charges.

2- Confidence and Satisfaction-based impacts: are associated with expectations related to meeting consumer marketplace expectations or satisfying the internal ability to execute against strategy. Some examples are: customer satisfaction, employee satisfaction, supplier management, decreased organizational trust and incorrect forecasting, reporting and decisions.

3- Productivity impacts: are associated with measurements of operational efficiency. Some examples are: increased workloads, decreased outputs increased processing time, decreased end-product quality and supply chain management issues. These impacts are considered direct costs of poor data quality.

4- Risk and Compliance impacts: are associated with the methods that data problems can increase exposure to various risks. Some examples are: credit assessment, investment risks, competitive risk, capital investment, fraud, and

compliance with government regulations, industry expectations, or self-imposed policies (such as privacy policies).

The monetary value of these impacts is divided to direct and indirect costs that are associated with it. In (Figure 2.11 four types of costs incurred by poor quality data) an example of these costs are presented and categorized by direct and indirect costs (Haug, Zachariassen, & van Liempd, 2011):

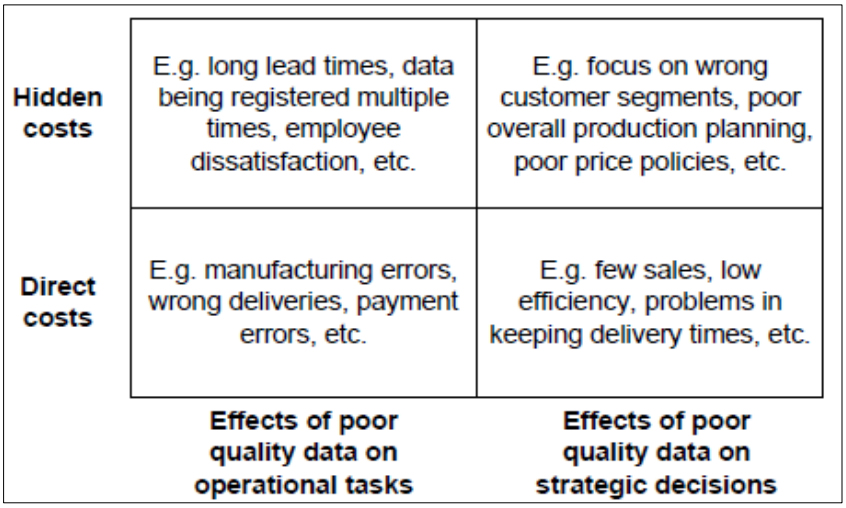

Figure 2.11: Four types of costs incurred by poor quality data, source (Haug, Zachariassen, & van Liempd, 2011, p181)

Many literatures introduced approaches to asses' data quality in organizations. One of these approaches is introduced in Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information by Danette McGilvray (McGilvray, 2008). The Ten Steps process is the approach for measuring, improving, and creating quality. These ten steps are illustrated in the following (Figure 2.12 The ten steps process to data quality) (McGilvray, 2008):

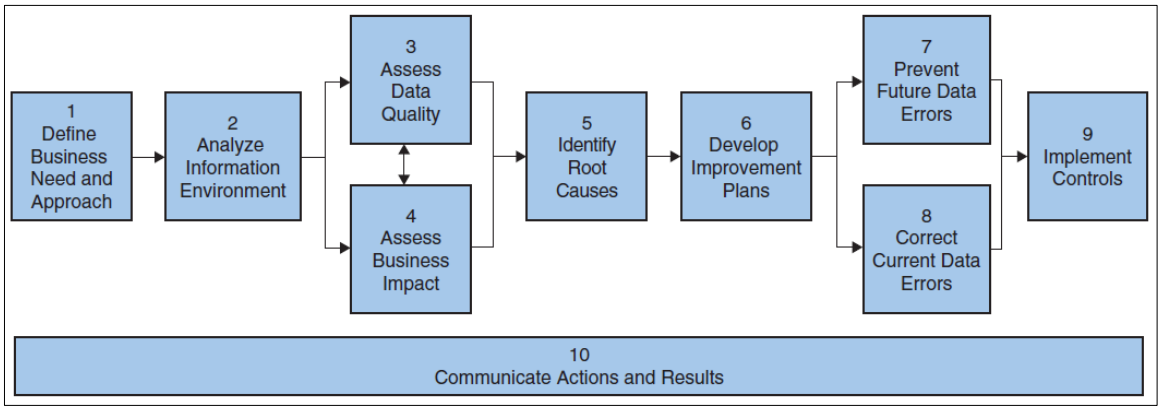

Figure 2.12: The ten steps process to data quality, source (McGilvray, 2008, p1)

2.4 Data Quality and Decision Making

Data are used in almost all the activities performed by the organization and establishes the basis for operational and strategic decisions. Poor data quality has significant negative impacts on the organization's performance, while high data quality is often essential to the organization's success (Haug, Zachariassen, & van Liempd, 2011). Therefore, data is vital to decision making and data quality is a key factor for decision performance (Du & Zhou, 2012).

Data are created and used in all business daily operations and it is a critical input to almost all organizational decisions (Haug, Zachariassen, & van Liempd, 2011). The decision making process consist of four phases: intelligence, design, choice and implementation (see Figure 2.13 the decision making process) (Du & Zhou, 2012).

1- The intelligence phase: in this phase, data are collected to support the next phases. This phase outlines the goals and outcomes of the decision.

2- The design phase: using brain storming, a set of possible solutions are presented, then a discussion about the advantages and disadvantages of each choice is outlined.

3- The choice phase: the agreed upon solution is selected.

4- The implementation phase: in this phase, the selected solution (decision) is implemented.

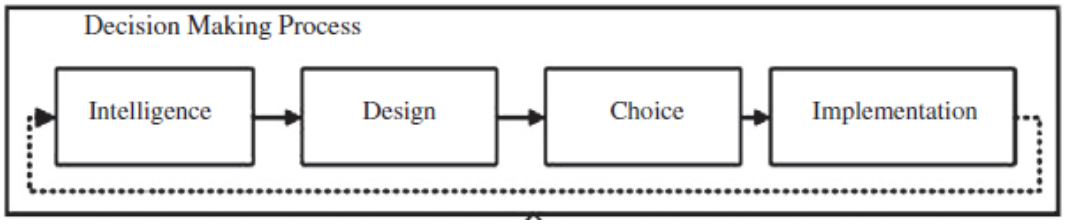

Figure 2.13: The decision making process, source (Du & Zhou, 2012, p80)

With the enormous amount of data being collected within organizations and used for achieving competitiveness, maintaining a high data quality for supporting business decisions continues to be important, but very difficult (Storey, Dewan, & Freimer, 2012). (Storey, Dewan, & Freimer, 2012) agreed that data quality should be accurate, timely, and consistent for the organization to succeed and make powerful decisions. The consequences of poor data quality have negative effects to the business and lead to inefficient decision making (Haug, Zachariassen, & van Liempd, 2011).

The importance of data quality in decision making process is clearly stated in the quality declaration of the European Statistical System, where its mission is: "We provide the European Union and the world with high quality information on the economy and society at the European, national, and regional levels and make the information available

to everyone for decision making purposes, research, and debate." (Batini & Scannapieca, 2006).

For maximum support to decision making process, it is important to address all aspects of data quality (Stvilia, Gasser, Twidale, & Smith, 2007). Data quality is a multidimensional concept that concerns both objective and contextual aspects. These aspects intrinsic to the data and vary across tasks and users (Du & Zhou, 2012).

Using poor data effect the decision making process in negative way. Capturing this poor data values is crucial to avoid making poor and costly decisions (Chiang & Miller, 2008). The costs based on wrong decisions are considered indirect costs of poor data quality (Haug, Zachariassen, & van Liempd, 2011).

In today's decision environment, there is no restriction by any business-unit or organizational boundaries to access organizational data. Decision making in the business environment uses large amount of data and include a wide variety of decision tasks. Decision makers have access to data anywhere and at any time which enforces them to become more responsive in the business environment. To do so, it is important to assure data quality to decision makers (Shankaranarayanan & Cai, 2006).

Quality of the data is dependent on the task where its influenced by the decision-task and the decision-maker (Shankaranarayanan & Cai, 2006). Decision makers ignore the value created for others to achieve their own value (Storey, Dewan, & Freimer, 2012). Allowing decision makers to assign weights to data quality related dimensions may introduce a large bias in the overall decision process. The assignment of weights is based on how a decision maker sees the importance and relevance of the data in the context of the decision-task. And since each decision-maker has the ability to assess the data quality based on his needs, personal gains may interfere (Shankaranarayanan & Cai, 2006).

Chapter 3

Previous studies

3.1 Introduction

3.2 Previous studies

3.3 Comments on previous studies

3.1 Introduction

In this chapter, the researcher has reviewed the previous literature on the subject of data quality and its relation with business processes and decision making. The previous literatures were collected from electronic libraries and online websites where it was selected to be published between 2004-2015.

This research has reviewed seventeen foreign studies and one Arabic study on data quality and its relation to business processes and decision making and on the dimensions of data quality. The researches were listed by Arabic studies then the foreign studies which were listed from the newest to the oldest and then in alphabetical order.

3.2 Previous studies

3.2.1 Arabic studies:

3.2.1.1 (Alsaidi & Zbar, 2013), "Quality of Information and its Impact on Strategic Decision – a Field Study on a Sample of Iraqi Banks"

This research aims to detect the information quality, identify its reliability and suitability to the decision maker and measure its impact on strategic decisions taken by senior and middle management. This research was applied on a sample of 40 employees representing the senior and middle management in the sample of Iraqi Banks.

The research concludes that strategic decisions depends on the accuracy of information, its efficiency and its forecasting and recommends to better identify data collection sources that affect the strategic decisions and getting information suitable with the challenges facing banking industry.

3.2.2 Foreign studies:

3.2.2.1 (Chen, Hailey, Wang, & Yu, 2014), "A Review of Data Quality Assessment Methods for Public Health Information Systems"

This research reviewed the current data quality assessment methods used in public health information systems. High data quality assessment is required for accurately evaluating the impact of public health interventions and measuring public health outcomes and to decision-making.
The research reviewed the literature of 39 studies from 2001 to 2013 extracted from several electronic bibliographic databases like Scopus, IEEE Xplore, Web of Science, ScienceDirect, PubMed, Cochrane Library and ProQuest. And some well-known institutions' websites were also reviewed.

The research found that the dimensions of data quality that was most frequently assessed are completeness, accuracy, and timeliness among a total of 49 of data quality dimensions and for those dimensions there was no standard definition.

As recommended for further research, a development of consistent data quality definitions and attributes should be done and more research efforts should be given to assess the data quality use and the data collection process.

3.2.2.2 (Moges, Dejaeger, Lemahieu, & Baesens, 2013), "A Multidimensional Analysis of Data Quality for Credit Risk Management: New Insights and Challenges"

This research explores data quality dimensions, assesses the data quality level using a scorecard index and identifies challenges to data quality and their causes. The research performs an empirical study using a questioners distributed to150 financial institutions worldwide. In the conclusion, it's confirmed that most of the financial institutions participated in the research are unaware of their data quality problems, which stops them from taking severe corrections to address these problems. Future study suggests applying sensitivity analysis to understand the possible impact of data quality as well as the relative importance of individual data quality dimensions.

3.2.2.3 (Storey, Dewan, & Freimer, 2012), "Data Quality: Setting Organizational Policies"

This research is intended to analyze data quality from an organizational perspective, where the improvements in decisions made were a key measure of data quality value. Poor data quality can cause a major damages or losses of business processes.

The research investigates that individual employees in absence of any managerial intervention might not invest enough in data quality. This research looks on this problem and proposes a solution.

An analytical model was presented for predicting quality choices made by individuals and organizations. It suggests that different users have different data quality needs. The benefits of data quality are presented to the whole organization and to the employees working in it. To solve this problem, a system to be implemented to focus on employee empowerment, decentralization, and create mechanisms to measure and reward individuals for their data quality efforts.

The research recommends that a future research is needed to implement the model into a tool to be tested in the real word.

3.2.2.4 (Du & Zhou, 2012), "Improving Financial Data Quality Using Ontologies"

This research is asking the following question: "How should one address the quality problems of financial data so as to improve the performance of financial decision-making?". Nowadays financial problems highlighted the importance of financial decision-making to individuals, businesses, and organizations. The research used both qualitative and quantitative approaches to address various types of data quality problems. The research was conducted on the financial data of S&P 500 companies from Standard& Poor's Compust at North America dataset.

An ontology-based framework was proposed in this research to improve financial data quality and address various quality problems which indicated a positive impact on the performance of financial decision-making.

3.2.2.5 (Ofner, Otto, & Osterle, 2012), "Integrating a Data Quality Perspective into Business Process Management"

This research presents a model to improve business process modeling and decision making. This modeling method conceptualizes data quality in the context of business process modeling and introduces data quality as further criteria in a multi-criteria decision making process during business process re-designs.

A case study on Siemens enterprise communications was used to evaluate the presented model.

As a conclusion, data quality is considered as a successes factor in business modeling and the proposed model facilitates and improves decision making process at managerial level. More research is needed to generalize the results.

This research supports decision makers in taking a data quality perspective in business process remodeling.

3.2.2.6 (Haug & Arlbjørn, 2011), "Barriers to Master Data Quality"

This research attempts to identify the most important barriers preventing organizations from achieving high data quality and presents an improved understanding of how to achieve higher level of data quality in organizations. By identifying and focusing on these barriers, organizations will have better chances of increasing their data quality.

A set of data quality barriers is identified and a questionnaire was presented on 90 Danish companies and the results indicated that the main data quality barrier is the lack of delegation of responsibilities in these companies.

Further investigation by other research studies must be presented to confirm the result of this research.

3.2.2.7 (Loshin, 2011), "Evaluating the Business Impacts of Poor Data Quality"

The research works on the impacts of poor data quality in different business dimensions like financial, trust issues, productivity and risk. It considers types of risks linked to poor data quality as well as identifying an approach to relating business impacts to data flaws.

The research was conducted in USA. It describes the value of the data as "the data is worth what we pay for it or what others may pay for it" but in the same time it's difficult to give monetary value for the data, however it's possible to have estimated value as how much this data doesn't meet business expectations. The value of the data must be linked to meet business expectations and reaching its goals. This involves looking in the root cause such poor data.

The research concludes that classifying business impacts into smaller categories makes determining how poor data quality impacts business processes a much more manageable task. Further segmentation to the categories in this research need to be investigated as the research recommends.

3.2.2.8 (Mendes & Rodrigues, 2011), "Main Barriers for Quality Data Collection in EHR"

This research aims to review the main barriers that may face data collection in health sector and to better understand what is possible to achieve high data quality. A qualitative review study has been introduced. The database research was held on Google, Google Scholar, PubMed, Scopus, ISI Web of Knowledge and ScienceDirect using the following key-words: barriers, high quality data, data

collection, EHR, Electronic Health Records.

The research concluded that the main barriers could be summarized in data sources, data format, data accuracy and data accessibility. As a result, the research recommends that data quality, data quality context, data quality representation and data quality accessibility were identified as major data quality characteristics. The solution to such problems should be identified from the beginning and a well-structured forms and designs should be presented. Also a periodic accuracy monitoring and feedback reports maybe very helpful.

3.2.2.9 (Friedman & Smith, 2011), "Measuring the Business Value of Data Quality"

The research looks in the anticipated value of data quality that affects the success of new business initiatives. Both in planning and execution phases, poor data quality lead to operational inefficacy and risk mitigation that effect decisions made in each of these areas.

The research found that poor data quality is a primary cause for new business initiatives to fail in achieving its targeted benefits. Using Six Sigma and Harry and Schroeder estimations, the research calculated the average monetary value that business processes lose and the productivity costs when using poor data quality.

At decision making level, depending on the data available assumptions and projections regarding business is made. If the data presented is week then the decision built on it is also weak. It indicates that most organizations never try to estimate the cost of poor data which lead to continuous fails and more money spent on repairing what is caused by this poor data.

The research recommends businesses to measure business value achieved of improving data quality used on business processes, overall productivity cycle and making business decisions.

3.2.2.10 (Haug, Zachariassen, & van Liempd, 2011), "The Costs of Poor Data Quality"

The research argues that high quality data should not be the ultimate goal of business. It provides a clarification of what are the costs of poor data quality and defines the relation to data quality maintenance effort. A case study was introduced to investigate the optimal data maintenance effort and to classify the cost of poor data quality.

The research was inducted in the manufacturing sector using the qualitative and exploratory method.

In the end, the research concluded that data effect could be different depending on the characteristics of the business. The research recommends more similar case studies with large number of questioners.

3.2.2.11 (Batini, Cappiello, Francalanci, & Maurino, 2009), "Methodologies for Data Quality Assessment and Improvement"

In this research, 13 data quality methodologies were described and compared. The whole data quality research field is currently evolving, and cannot be considered mature. Methodologies for data quality measurement and improvement are still evolving.

The comparison of methodologies clearly indicates that methodologies tend to focus on a subset of data quality issues. These differences can be recognized by classifying methodologies into four categories: complete methodologies, audit methodologies, operational methodologies and economic methodologies.

As a conclusion, the researchers found that limited number of research contributions focused on semi-structured and unstructured data in the data quality domain. As future topics, the researchers suggest more empirical validation of the models and to expand the analysis to a wider set of dimensions and to more specific types of business processes.

3.2.2.12 (Chiang & Miller, 2008), "Discovering Data Quality Rules"

Poor data is a serious problem for organizations leading to ineffective decision making, inefficient daily operations, low performance and ultimately wasting both time and money.

This research proposes a new data-driven tool that can be used by organization's data quality management process to suggest possible rules, and to identify dirty data records.

Many organizations employ consultants to develop a data quality management process while not taking into consideration that users who are involved in the consultation processes have specific knowledge of business policies and may not be aware of the existing rules that can be useful towards enforcing semantic data consistency.

The research used seven datasets (six real and one synthetic) to evaluate the proposed tool. Real datasets used are UCI Machine Learning Repository (Adult), Census-Income (KDD), Auto, Mushroom, Statlog German Credit and Insurance Company Benchmark datasets. Synthetic dataset is consisting of individual US tax records.

The research concludes that the tool presented can discover reasonable number of rules but not a complete set of cleaning rules which needs to be considered in future researches.

3.2.2.13 (Shankaranarayanan & Cai, 2006), "Supporting Data Quality Management in Decision-Making"

This research proposed a decision-support framework that permits decision makers to measure quality in both objective and context-dependent manner. The research emphasizes that data quality is dependent on the decision-task and that the same data may be viewed with two or more different quality perspectives according to the decision-task it is used for.

The proposed framework was presented to evaluate completeness as a data quality dimension. The sample test was on a supply chain decision-environment.

As a conclusion, the research showed that giving the decision makers the right to assign weight to different data quality dimensions may increase the bias upon everyone's one needs.

As a future research, more frameworks have to be presented to systematically address the data quality problem in the networked environment.

3.2.2.14 (Oliveira, Rodrigues, & Henriques, 2005), "A Formal Definition of Data Quality Problems"

This research presents taxonomy of data quality problems and a definition for each one. These problems are an obstacle to effective data usage and have negative effects on the results extracted from this data affecting its usefulness and correctness in supporting decision making. Data is a key asset to improve efficiency in today's dynamic and competitive business environment.

The research was conducted using real-world data from the retail sector to identify data quality problems of the usual data organization model and followed a bottom-up approach. This research was the first step to develop a tool to automatically detect data quality problems.

3.2.2.15 (Scannapieco, Missier, & Batini, 2005), "Data Quality at a Glance"

In this research, an introduction to data quality dimensions and several examples on how they can be measured were presented. Accuracy, completeness, time-related dimensions (currency, timeliness and volatility) and consistency.

Data quality dimensions are dependent on each other and correlations exist among them. Also it may imply negative consequences if one dimension is favored on the others.

The research was conducted on Internet Movie Database (http://www.imdb.com) where the techniques used to evaluate data quality dimensions also suggest a method for error correction.

As a conclusion, the research presented a definition to data quality definitions that are applied to many contexts, including e-Business and e-Government but for some data quality dimensions, there is still no general agreement on their definitions. More researches must be done to have more agreed-on definitions.

3.2.2.16 (Neely, 2005), "The Product Approach to Data Quality and Fitness for Use: A Framework for Analysis"

This research proposes a theory of data quality based on the five principles defined by J. M. Juran for product and service quality and Wang et al's framework for data quality research.

Different users have different needs, and this lead to the same data is used differently by different users. Also the same user may use the same data differently in different situations.

The research used a 76 journal articles from 1996 through 2004 to evaluate the literature that have been conducted on the two main principles of Juran and Wang. The research found that significant researches have been completed regarding the production and distribution of data and information and to data quality dimensions but not to how to measure these dimensions.

The research concludes that data quality as a field of research is beginning to mature where more and more researches are focusing on the term "Data Quality". Recommendations on more empirical studies to be conducted to support frameworks presented in different researches.

3.2.2.17 (Eppler & Helfert, 2004), "A Classification and Analysis of Data Quality Costs"

This research reviews and categorizes the potential costs associated with low data quality, potential negative monetary effects and to identify cost saving potentials.

There was no validated economic theory of data quality costs that could be used as a basis for data quality cost analysis.

Through exploring the data quality literature, the research identified specific cost examples and then clustered these examples into cost groups based on shared criteria. A useful management framework was implemented through converting the findings and been evaluated but not in a testable model which is recommended for future research.

The results show that data quality costs can be categorized in terms of several informative dimensions and that it can be grouped in terms of their measurability or effect.

The research recommends validating the model in additional researches through real-life observations.

3.3 Comments on Previous Studies

There were general agreement among literature about the importance of data quality to organization's and its effect on the ongoing daily business processes and decision making.

Arabic and foreign studies were reviewed and the following are the general comments on them:

a) Organizations have given little attention to data quality issues.
b) There is no unified tool used to address poor data quality.
c) In academic field, there wasn't a unified definition or specification for data quality but a similarity on its foundation.
d) Also there wasn't a general agreement on the key dimensions neither of data quality nor to its definition.
e) There is no quantified tool that can precisely measure the business value of data quality.
f) Less empirical studies have been done on the field of data quality and many assessment methodologies have been proposed theoretically.
g) Little literature has established a concrete relation between poor data quality and its monetary value.

This research has many common characteristics with the research viewed in this chapter. All the research has proved that the data quality in its pure concept is one of the main elements for organization's success. The dimensions used in this research are the most agreed on among the research.

The main deference between this research and the previous studies is that it highlighted the role of data quality dimensions and specifically studies the insurance sector in Gaza Strip. This sector is very vulnerable to the changes in the Palestinian economy and is affected directly by poor data quality.

This research is distinguished from the previous studies that it has put insight on the insurance companies and the knowledge about data quality in this sector. This research is considered to be the first one in Palestine and in Gaza Strip that addresses the problem of data quality in organizations.

This research has faced some limitations where there was little Arabic literature on the subject of data quality (one related paper prepared on 2013 and used in this chapter).

Chapter 4

Population Profile

Insurance companies in Gaza Strip

4.1 Introduction

4.2 National Insurance Company

4.3 Trust Insurance Company

4.4 Palestine Insurance Company

4.5 Ahleia Insurance Group Company

4.6 Global United Insurance Company

4.7 AlMultazem for Insurance and Investment

4.1 Introduction

The research population consists of the insurance companies based in Gaza Strip. According to the Palestinian Insurance Federation (PIF), there are ten registered insurance companies in Palestine. But in Gaza Strip, there are only seven companies with branches. Where only six of them are really operating.

The companies registered in the PIF are divided in two categories:

- Have branch in Gaza Strip:
    1- National Insurance Company
    2- Trust Insurance Company
    3- Palestine Insurance Company
    4- Ahleia Insurance Group Company
    5- Global United Insurance Company
    6- Al-Takaful Insurance Company
    7- Al-Mashriq Insurance Company (one employee)

- Doesn't have branch in Gaza Strip:
    1- Arab Life & Accidents Insurance Company
    2- Palestine Mortgage & Housing Corporation
    3- American Life Insurance Company (ALICO)

Also there is one company that is not registered in the PIF but operating widely in Gaza Strip, AlMultazem for Insurance and Investment will be added to the research population.

In the following section, a brief overview will be held on each of the six following companies:

1- National Insurance Company
2- Trust Insurance Company
3- Palestine Insurance Company
4- Ahleia Insurance Group Company
5- Global United Insurance Company
6- AlMultazem for Insurance and Investment

4.2 National Insurance Company (NIC)

About the company:

National Insurance Company Ltd. (NIC) was founded in 1992 and registered on 15/3/1992 as a public shareholding company with the Registrar of Companies in the city of Ramallah, and got a license to work on 15/2/1993, and started its functioning on the date of 1/3/1993. Main branch is in Al-Beireh. NIC has established branches in Bethlehem, Ramallah, Hebron, Nablus, Jenin, Tulkarem, Al-Ezareyeh, and Gaza (National Insurance Company Ltd., 2015).

NIC was founded by a group of experienced Palestinian professionals with paid capital in the amount of JD 3,500,000 (the equivalent of 5 million U.S. Dollars) to be raised in 2010 to capital of USD 10,000,000 divided by 10,000,000 shares with a par value of USD 1 share. Its establishment occurred in the midst of an uncertain eco-political climate, which included the withdrawal of Israeli insurance companies from the Palestinian market. Since then, NIC has achieved numerous successes due to its unparalleled strengths within Palestine (National Insurance Company Ltd., 2015).

NIC is the largest insurance company licensed under the applicable laws of the Palestinian Authority. Guided by the vision of the board of directors, its primary function is to increase shareholder value by providing a distinctive service to all its customers while maintaining risk based policies and practices (National Insurance Company Ltd., 2015).

Mission (Natioanl Insurance Company Ltd., 2014):

1- Granting sophisticated and comprehensive insurance coverage.

2- Raising public awareness about insurance.

3- Fair competition based on improving service levels and performance.

4- Honesty in dealing with the public.

5- Maintaining a balance amongst all stakeholders' interests; shareholders, customers, and employees.

6- Contributing to the development of the national economy.

7- Providing job opportunities to the largest number of people through the company national investments.

Vision (Natioanl Insurance Company Ltd., 2014):

1- To maximize shareholders value by providing best in class insurance services thorough product innovation, acting with transparency and integrity with all stakeholders, and provide support and attention to the local community and economy.

2- Be the first in the insurance sector through providing distinctive insurance services for all economic sectors in Palestine, and work to provide the best coverage with the easiest way to customers.

Services (National Insurance Company Ltd., 2015):

1- Motor Insurance
2- Property Insurance
3- Marine Insurance
4- Engineering Insurance
5- Specialized Insurance
6- Personal Insurance

4.3 Trust Insurance Company

About the company:

Trust International Insurance Company is a Palestinian public shareholding company founded by a group of Palestinian and Arab businessmen with impressive and outstanding experience in the field of insurance in Arab and international markets. The company began its work in the Gaza Strip in the middle of 1994 and then extended its activities to include all parts of the West Bank by middle of 1995 (Trust International Insurance Palestine, 2015).

It is an associate company of the Group of Trust International Insurance Company, one of the largest insurance groups and reinsurance in the Middle East. The capital of Trust Group in its various branches is more than two hundred million dollars. The share capital of the company's is ten million US dollars. This company has achieved a famous Arab and international cherish and proved that the Palestinian competencies are capable of building and giving wherever he exists and under the most difficult circumstances (Trust International Insurance Palestine, 2015).

Mission (Trust International Insurance Palestine, 2015):

1- Continuing to strengthen the relationship with customers by providing high-quality services and provide the best placement to meet their demands and conditions of services.
2- Develop and expand the network of branches and agents and offices to meet existing customers and reach all possible future customers.

3- Investment in human resources and enhance their ability to negotiate and communicate to enable the company to achieve its goals and accomplish its vision.
4- To make the name “Trust International Insurance Palestine” the best in the Palestinian insurance market and have the echo of success as a model that the community is looking forward.

Vision (Trust International Insurance Palestine, 2015):

To become the first choice for any consumer, as the company is the result of accumulated experience over decades which has established solid background of knowledge and experience that prepared all the possibilities to achieve this vision in practice during the next phase. Trust International Insurance Palestine is working to take leadership in the Palestinian insurance market to be the first choice for any consumer through establishing power, reputation, service quality, speed response to customer requirements, and quality goods and performance to take the company to the first class in production, profits, expansion, shareholders' equity and customer satisfaction.

Services (Trust International Insurance Palestine, 2015):

1- Vehicle Insurance
2- Fire Insurance
3- Engineering Insurance
4- Civil Liability Insurance
5- Marine Insurance
6- Health and Life Insurance
7- Personal Insurance

4.4 Palestine Insurance Company

About the company

Palestine Insurance Company (PIC) is a public shareholding company founded 1994 by a group of Palestinian businessmen, with a capital of $5 million. Palestine Insurance Company puts all its expertise and potentials in the Palestinian national economy to service and support industrial, commercial and financial institutions.

Palestine Insurance Company provide protection and security to property of citizens and to provide Palestine with expert and talented manpower to contribute effectively in the quality of the Palestinian economy (Palestine Insurance Company (PIC), 2013).

Mission and values (Palestine Insurance Company (PIC), 2013)

Palestine Insurance Company believes in the importance of its role and contribution in building the Palestinian economy and is committed to its leadership role in economic and social development, through its responsibility towards society and high credibility in dealing with the insured and responsibility towards them.

Palestine Insurance Company is also committed to the moral and social role towards the company and the community, where the ethical value, principles and standards of governance and practices of administrative governance. And as a national institution it's committed to a code of ethics in all its practices and activities, and is keen to disseminate good citizenship standards, devoted to work and take into account the improvement of the internal work environment, and also strengthen its commitment to support initiatives and community development projects.

Services (Palestine Insurance Company (PIC), 2013)

1- Professional insurance
2- Money Insurance
3- Personal Accident Insurance
4- Insurance for workers
5- Comprehensive home insurance
6- Insurance on fire
7- Health insurance
8- Vehicle insurance

4.5 Ahlia Insurance Group Company

About the company:

Ahlia Insurance Group (AIG) is a limited public shareholding company, established under the name "Gaza Ahleia Insurance Company". Ahleia Insurance Group was founded in 1994 by a group of Palestinian professionals to be the first public shareholding company in the Palestinian territories founded from a national capital and operated by national experts. At AIG, we strive to lead in the Palestinian insurance market while contributing to the development of the Palestinian community (Ahlia Insurance Group, 2013) (Ahlia Insurance Group, 2014).

Mission (Ahlia Insurance Group, 2013):

1- We embody the aspirations of the Palestinian people, seeking security, and dependability, while furthering our own imperative for prosperity.
2- We operate under a banner of integrity, employing nothing but the highest standards of ethics and professionalism in everything we do.
3- Our focus is to contribute to the growth of the Palestinian economy, while solidifying our position as the premier risk manager and superior risk underwriter in Palestine.

Objectives (Ahlia Insurance Group, 2014):

1- Provide insurance coverage to attain financial and manpower treasures of Palestine and fulfill current customers' needs and future expectations through qualified national technical and administrative staff.
2- Improve the performance level of the insurance services to simulate the international insurance service.
3- Rehabilitation and training of local staff to enrich the Palestinian insurance market and meet its needs.
4- Provide insurance awareness and focus on the importance of the insurance role in contributing to the economic rise and servicing Palestine.
5- Providing consultancy and insurance services, taking into account the common interest of the insured, the company and Palestine.
6- Activation of the vital role of insurance in maintaining money, property and lives against all threats to ensure the continuation of production and service projects. As well as the safety of the human element. And focus on the importance of providing all precautions and safety regulations to reduce and prevent accidents.

Services (Ahlia Insurance Group, 2013):

1- Motor Insurance
2- Property Insurance & Business Interruption
3- Engineering Insurance (Car/EAR/Plant & Machinery)
4- Liability Insurance: Pubic, Products, Professional, and Medical Malpractice (includes worker's compensation)
5- Personal Accidents (group scheme, individual, and travel)
6- Fidelity Insurance (Blanket & Individual)
7- Money Insurance
8- Medical Insurance
9- Marine Cargo Insurance

10- Credit and Mortgage Insurance
11- Life Insurance
12- Workmen's Compensation Insurance
13- Household Comprehensive
14- Bankers Blanket Bond Insurance
15- Fidelity Guarantee

4.6 Global United Insurance Company

About the company (Global United Insurance Company, 2013):

Global United Insurance Company, a public shareholding Palestinian company was founded in 2010 by a group of businessmen with vast expertise in the insurance industry with a capital of $5 million.

In the context of development and keep pace with economic and political changes experienced by the region, the company has developed a long-term and clear strategic plan. Global United Insurance Company is running at full effort and determination to constitute to the Palestinian insurance market, and for that Global United Insurance provide its services at best performance.

Mission (Global United Insurance Company, 2013):

1- Global United Insurance Company strives to be the market leader through a combination of professional skills and human resources with a focus on customer service to achieve service excellence.
2- Global United Insurance Company seeks to provide a full range of products with a high degree of service.
3- Global United Insurance Company seeks to provide optimal performance in customer service by providing regional and international insurance services and alternative solutions.
4- Honesty in dealing with the public.
5- Support and contribute to the development and construction of the Palestinian economy.

Vision (Global United Insurance Company, 2013):

The vision of Global United Insurance Company is to meet the growing demand for insurance in the region and also devote our energies to provide excellent services and protection from the risk as widely as possible.

We at Global United Insurance Company strive to be real partners and actors for our customers to manage the risks surrounding their work, and we want to be the most effective in building a leading insurance in our country, this commitment and loyalty to our country and our citizens, we've assembled the most prestigious experiences and studied the experiences of others because we are seeking to excellence and superiority.

The needs of the people are always changing and to cope with these changes, we found it necessary for us to adapt our services accordingly by creating the necessary technical solutions and provide insurance that go along with those changes. Because we care about you, let's plan together to protect your families, your homes, and your belongings.

4.7 AlMultazem for Insurance and Investment

About the company

AlMultazem Insurance & Investment Company was founded in 2008 by a group of investors, businessmen and well-known companies with distinctive, efficient and powerful experience in management and investment (AlMultazem Insurance & Investment Company, 2012). AlMultazem Insurance & Investment Company is not registered in the Palestinian Insurance Federation.

Mission and vision (AlMultazem Insurance & Investment Company, 2012):

1- Distinctive insurance coverage
2- Provide special customers service
3- Separation between stakeholders' and shareholders' funds
4- The distribution of the insurance surplus to subscribers (insured)
5- Highly qualified staff

Services (AlMultazem Insurance & Investment Company, 2012):

1- Vehicle insurance
2- Fire insurance
3- Insurance against theft
4- Insurance against all contractor's risks
5- Insurance against the dangers of installation
6- Work injury insurance
7- Insurance on machinery and equipment

8- Insurance glass panels
9- Insurance breach of trust
10- Marine insurance
11- Yacht and boat insurance
12- Money Insurance
13- Total home insurance
14- Personal insurance
15- Civil liability insurance
16- Travel insurance
17- Health insurance

Chapter 5

The Research Methodology

5.1 Methodology and Procedures

5.2 Testing of Research Tools

5.3 Analysis of Sample Characteristics

5.4 Data Analysis and Hypotheses Testing

5.1 Methodology and Procedures

5.1.1 Introduction

This chapter describes the methodology used in this research. The adopted methodology to accomplish this research used the following techniques: the information about the research design, research population and sample, questionnaire design, content validity, and statistical data analysis.

5.1.2 The research methodology

The analytical descriptive approach was used by this research to compare, explain and evaluate the research hypotheses in order to attain meaningful results to enrich knowledge about data quality. This research consists of two parts, the descriptive part of the research, which attempt to illustrate the concepts of data quality, data quality dimensions, business processes, and decision making. The second one is the analytical part where the questionnaire has presented and distributed to the research sample and used the Statistical Package for the Social Sciences (SPSS) to analyze the collected data.

5.1.3 Data collection sources

This research used two sources for data collection primary and secondary.

The primary data collection source is the questionnaire.

The secondary data collection sources are:

a) Journal articles.
b) Books.
c) Conference preceding.
d) Websites
e) Reports.

5.1.4 Population and Sampling

The population of this research includes employees in insurance companies in Gaza Strip, which is a universal sample. The questionnaires were distributed to 56 employees from the research population and only 50 questionnaires were retrieved successfully (see Table 5.1 Population and sample of the research).

Table 5.1: Population and sample of the research, source The Researcher, 2015

Company	Number of distributed questionnaire	Number of collected questionnaire
National Insurance Company	10	10
Trust Insurance Company	15	12
Palestine Insurance Company	9	6
Ahleia Insurance Group Company	5	5
Global United Insurance Company	7	7
AlMultazem for Insurance and Investment	10	10
Total	**56**	**50**

5.1.5 Research Design

a. The first phase: it includes developing the research proposal, which contains identifying, and defining the research problem, the establishment of the research objectives and the development of the research plan.

b. The second phase: included a comprehensive literature review on data quality in business processes and decision-making.

c. The third phase: the development of the research questioner, which measures the research variables.

d. The fourth phase: focused on the modification of the questionnaire design.

e. The fifth phase: it focused on distributing the questionnaire on the insurance companies operating in Gaza city. This questionnaire was used to collect the required data in order to achieve the research objective. 56 questionnaires were distributed where 50 of them were retrieved successfully.

f. The sixth phase: data analysis and discussion. The Statistical Package for the Social Sciences (SPSS) was used to perform the required analysis.

g. The seventh and final phase: includes the research conclusions and recommendations.

Figure (5.1 Research design) shows the methodology flowchart, which leads to achieve the research objectives.

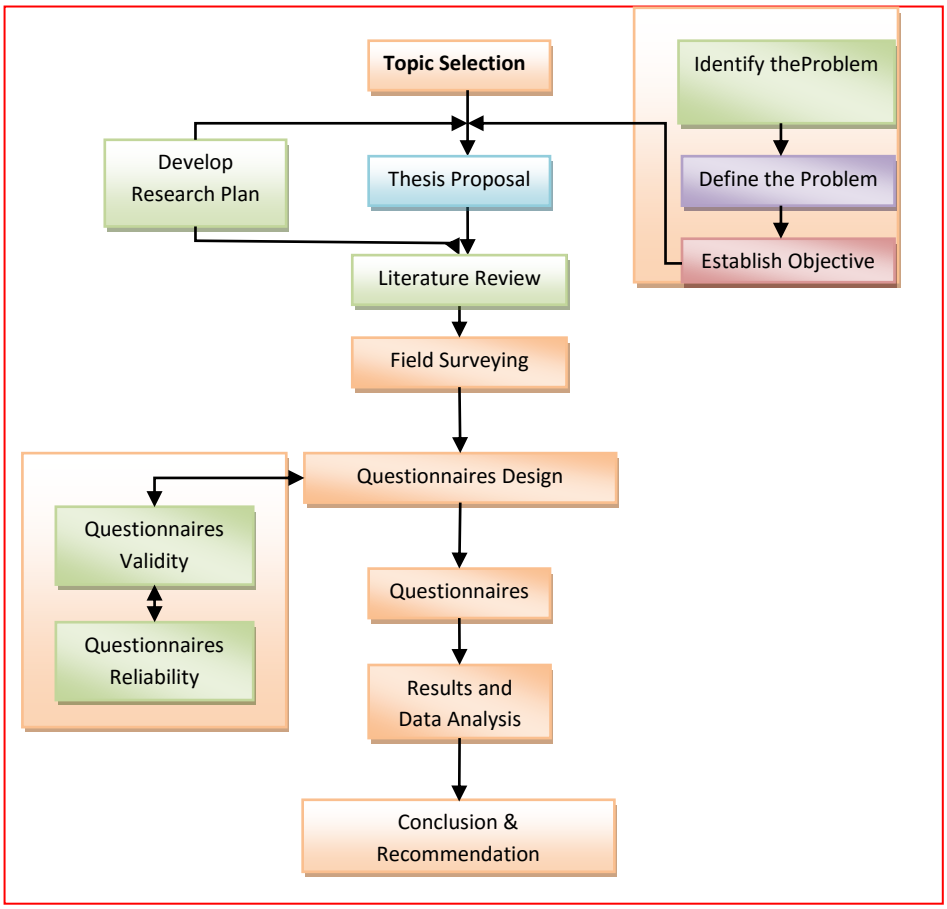

Figure 5.1: Research design, source the Researcher 2015

5.1.6 Questionnaire Design

The questionnaire was carefully design in order to facilitate the collection process, in addition to maximize the reliability and validity of data gathered from respondents. The questionnaire is composed of the following sections:

- Section one: It focused on the personal and organizational characteristics of the respondents including (gender, age, educational level, years of work in the company, employment level, and decision making responsibilities).
- Section two: focused on the impact of data quality on business processes which consist of 48 statements divided as follows:
    - Data completeness include 10 questions
    - Data accuracy include 9 questions
    - Data consistency include 9 questions

- Validity of data include 11 questions
- Data relevance include 9 questions

- Section three: focused on the impact of data quality on decision making which consist of 33 statements divided as follows:
  - Data completeness include 7 questions
  - Data accuracy include 7 questions
  - Data consistency include 6 questions
  - Validity of data include 6 questions
  - Data relevance include 7 questions

5.1.7 Data Measurement

In order to be able to select the appropriate method of analysis, the level of measurement must be understood. For each type of measurement, there is/are an appropriate method/s that can be applied and not others. In this research, all questions are answered on a scale from 1-10, which "1" represents completely disagree; while "10" represents completely agree.

Level	Strongly Disagree	…….	Completely agree
Scale	1	…….	10

5.1.8 Statistical methods are as follows:

1- Frequencies and Percentile
2- Alpha-Cronbach Test for measuring reliability of the items of the questionnaires
3- Person correlation coefficients for measuring validity of the items of the questionnaires.
4- Spearman–Brown Coefficient
5- One sample t test
6- Independent samples T test for the difference between means of two independent samples
7- One-way ANOVA test for the difference between means

5.2 Testing of Research Tools

In this section a validity and reliability of the used instruments to conduct this research is done.

5.2.1 Validity of the Questionnaire

We can define the validity of an instrument as a determination of the extent to which the instrument actually reflects the abstract construct being examined. "Validity refers to the degree to which an instrument measures what it is supposed to be measuring". High validity is the absence of systematic errors in the measuring instrument. When an instrument is valid, it truly reflects the concept it is supposed to measure. Achieving good validity required the care in the research design and sample selection.

5.2.2 Content Validity of the Questionnaire

The content validity of the questionnaire was conducted through the supervisor's review. Also nine academics from the Islamic University of Gaza, AlAzhar University, and AlAqsa University reviewed the questionnaire. The provided notes were valuable and used to improve its validity in order to ensure that the content of the questionnaire is consistent with the research objectives and is reflecting the research problem.

5.2.3 Statistical Validity of the Questionnaire

To insure the validity of the questionnaire, two statistical tests should be applied. The first test is Criterion-related validity test (Pearson test) which measures the correlation coefficient between each item in the field and the whole field. The second test is structure validity test (Pearson test) that used to test the validity of the questionnaire structure by testing the validity of each field and the validity of the whole questionnaire. It measures the correlation coefficient between one filed and all the fields of the questionnaire that have the same level of similar scale.

5.2.4 Criterion Related Validity

Internal consistency:

Internal consistency of the questionnaire is measured by a selected sample, through measuring the correlation coefficients between each question in one field and the whole filed. Tables (5.2 The correlation coefficient between each question in the field and the whole field and 5.3 The correlation coefficient between each question in the field and

the whole field) below shows the correlation coefficient and p-value for each field items. As show in the Table the p-values are less than 0.05 or 0.01, so the correlation coefficients of this field are significant at $\alpha \leq 0.01$ or $\alpha \leq 0.05$, so it can be said that the paragraphs of this field are consistent and valid to be measure what it was set for.

Table 5.2: The correlation coefficient between each question in the field and the whole field, (Section one: Business Processes)

No.	Question	Pearson coefficient	p-value
	Data completeness		
1	Data used in the daily business processes is complete for its purpose of use	0.502	0.005
2	When there is a problem in data completeness, you look for its source to ensure its completeness	0.457	0.011
3	When there is a problem in data completeness, you report the problem	0.442	0.014
4	When there is a problem in data completeness, you continue working without considering the problem	0.446	0.013
5	The organization works on improving quality of data used for all organizational purposes	0.417	0.022
6	The lack of data completeness causes problems to daily business processes	0.393	0.032
7	Where there is a problem in data completeness, you stop working till the problem is solved	0.597	0.000
8	The lack of data completeness affects the daily business processes	0.532	0.002
9	Data completeness is important for daily business process in the organization	0.433	0.017
10	The more used data is complete, the more the quality of business process.	0.372	0.043
	Data accuracy		
11	The data used in the daily business processes are accurate for its purpose of use	0.619	0.000
12	When there is a problem in the data accuracy you look for its source to ensure its accuracy	0.588	0.001
13	When there is a problem in data accuracy, you report the problem	0.482	0.007
14	When there is a problem in the data accuracy you continue working without considering the problem	0.496	0.005
15	The lack in data accuracy causes problems to daily business processes	0.446	0.014
16	When there is a problem in the data accuracy, you stop	0.756	0.000

	working until the problem is solved		
17	The lack of data accuracy affects the daily business processes	0.516	0.004
18	Data accuracy is important for daily business process in the organization	0.557	0.001
19	The more used data is accurate, the more the quality of business process.	0.628	0.000
	Data consistency		
20	The data used in daily business processes are consistent for its purpose of use	0.570	0.001
21	When there is a problem in the data consistency, you look for its source to ensure its consistency	0.788	0.000
22	When there is a problem in the data consistency, you report the problem	0.565	0.001
23	When there is a problem in the data consistency, you continue working without considering the problem	0.554	0.002
24	The lack of data consistency causes problems to daily business processes	0.673	0.000
25	When there is a problem in the data consistency, you stop working until the problem is solved	0.756	0.000
26	The lack of data consistency affects the daily business processes	0.735	0.000
27	Data consistency is important for daily business process in the organization	0.558	0.001
28	The more used data is consistent, the more the quality of business process.	0.624	0.000
	Validity of data		
29	The data used in the daily business processes are valid for its purpose of use	0.748	0.000
30	When there is a problem in the data validity , you look for its source to ensure its validity	0.737	0.000
31	When there is a problem in the data validity, you report the problem	0.739	0.000
32	Poor data quality affects business processes decision making	0.519	0.003
33	When there is a problem in the data validity, you continue working without considering the problem	0.594	0.001
34	The lack of data validity causes problems to daily business processes	0.387	0.035
35	When there is a problem in the data validity, you stop working until the problem is solved	0.774	0.000
36	Obtaining high data quality is not important to the organization’s success	0.502	0.005
37	The lack of data validity affects the daily business processes	0.711	0.000

38	The data validity is important for daily business process in the organization	0.608	0.000
39	The more used data is valid, the more the quality of business process.	0.659	0.000
	Data relevance		
40	The data used in the daily business processes are relevance for its purpose of use	0.706	0.000
41	When there is a problem in the data relevancy, you look for its source to ensure its relevancy	0.554	0.001
42	When there is a problem in the data relevancy, you report the problem	0.829	0.000
43	When there is a problem in the data relevancy, you continue working without considering the problem	0.653	0.000
44	The lack of data relevancy causes problems to daily business processes	0.671	0.000
45	When there is a problem in the data relevancy, you stop working until the problem is solved	0.730	0.000
46	The lack of data relevancy affects the daily business processes	0.593	0.001
47	The data relevancy is important for daily business process in the organization	0.596	0.001
48	The more used data is relevant, the more the quality of business process.	0.737	0.000

Table 5.3: The correlation coefficient between each question in the field and the whole field, (Section two: Decision making)

No.	Question	Pearson coefficient	p-value
	Data completeness		
49	Data completeness affects the decision making process in the organization	0.502	0.005
50	It is difficult to reach a proper decision when the data is incomplete	0.457	0.011
51	When there is missing data, you try to collect the missing data to facilitate the process of decision making	0.442	0.014
52	The organization works on improving quality of data used for all organizational purposes	0.446	0.013
53	Having complete data is a group effort that affect all staff in the organization	0.417	0.022
54	A data that is complete for making one decision is not necessarily complete for another	0.493	0.006
55	The organization works on achieving the highest level of	0.597	0.000

	data completeness for all its employees in there different jobs		
	Data accuracy		
56	Data accuracy affects the decision making process in the organization	0.433	0.017
57	It is difficult to reach a proper decision when the data is inaccurate	0.372	0.043
58	When there is inaccurate data, you try to improve data to facilitate the process of decision making	0.482	0.007
59	Having accurate data is a group effort that affect all staff in the organization	0.619	0.000
60	Improving data quality leads to raising the value of the whole organization	0.588	0.001
61	A data that is accurate for making one decision is not necessarily accurate for another	0.532	0.002
62	The organization works on achieving the highest level of data accuracy for all its employees in there different jobs	0.496	0.005
	Data consistency		
63	Data consistency affects the decision making process in the organization	0.756	0.000
64	It is difficult to reach a proper decision when the data is inconsistence	0.516	0.004
65	When there is inconsistent data, you try to improve data to facilitate the process of decision making	0.557	0.001
66	Having consistent data is a group effort that affect all staff in the organization	0.628	0.000
67	A data that is consistent for making one decision is not necessarily consistent for another	0.393	0.032
68	The organization works on achieving the highest level of data consistency for all its employees in there different jobs	0.570	0.001
	Validity of data		
69	Data validity affects the decision making process in the organization	0.493	0.006
70	It is difficult to reach a proper decision when the data is not valid	0.667	0.000
71	When there is invalid data, you try to improve data to facilitate the process of decision making	0.473	0.008
72	Having valid data is a group effort that affect all staff in the organization	0.618	0.000
73	A data that is valid for making one decision is not necessarily valid for another	0.618	0.000
74	The organization works on achieving the highest level of data validity for all its employees in there different jobs	0.424	0.020
	Data relevance		

75	Data relevancy affects the decision making process in the organization	0.803	0.000
76	It is difficult to reach a proper decision when the data is not relevant	0.714	0.000
77	When there is irrelevant data, you try to improve data to facilitate the process of decision making	0.712	0.000
78	Having relevant data is a group effort that affect all staff in the organization	0.765	0.000
79	A data that is relevant for making one decision is not necessarily valid for another	0.568	0.001
80	The organization works on achieving the highest level of data relevancy for all its employees in there different jobs	0.654	0.000
81	Obtaining high data quality is not important to the organization's success	0.746	0.000

5.2.5 Structure Validity of the Questionnaire

Structure validity is the second statistical test that is used to test the validity of the questionnaire structure by testing the validity of each field and the validity of the whole questionnaire. It measures the correlation coefficient between one field and all the fields of the questionnaire that have the same level of liker scale.

As shown in Table (5.4), the significance values are less than 0.01, so the correlation coefficients of all the fields are significant at α=0.01, so it can be said that the fields are valid to measure what it was set for to achieve the main aim of the research.

Table 5.4: Structure Validity of the Questionnaire

Section	Sub-section	Pearson correlation coefficient	p-value
Section one (Business Processes)	Data completeness	0.792	0.000
	Data accuracy	0.863	0.000
	Data consistency	0.920	0.000
	Validity of data	0.608	0.000
	Data relevance	0.716	0.000
Section two (Decision making)	Data completeness	0.828	0.000
	Data accuracy	0.905	0.000
	Data consistency	0.896	0.000
	Validity of data	0.650	0.000
	Data relevance	0.759	0.000

5.2.6 Reliability of the Research

Reliability of an instrument is the degree of consistency with which it measures the attribute it is supposed to be measuring. The test is repeated to the same sample of people on two occasions and then compares the scores obtained by computing a reliability coefficient. For the most purposes reliability coefficient above 0.70 are considered satisfactory. Period of two weeks to a month is recommended between two tests due to complicated conditions that the research sample is facing at the time being, it was too difficult to ask them to responds to our questionnaire twice within short period. The statistician's explained that, overcoming the distribution of the questionnaire twice to measure the reliability can be achieved by using Cronbach Alpha coefficient and Half Split Method through the SPSS software.

5.2.6.1 Cronbach's Coefficient Alpha

This method is used to measure the reliability of the questionnaire between each field and the mean of the whole fields of the questionnaire. The normal range of Cronbach's coefficient alpha value between 0.0 and + 1.0, and the higher values reflects a higher degree of internal consistency. As shown in Table (5.5 Reliability Cronbach's Alpha) the Cronbach's coefficient alpha was calculated. The general reliability for all items equal 0.9045. This range is considered high; the result ensures the reliability of the questionnaire.

Table 5.5: Reliability Cronbach's Alpha

Section	Sub-section	Cronbach's Alpha
Section one (Business Processes)	Data completeness	0.8678
	Data accuracy	0.8896
	Data consistency	0.8391
	Validity of data	0.9157
	Data relevance	0.8924
Section two (Decision making)	Data completeness	0.8721
	Data accuracy	0.9045
	Data consistency	0.8957
	Validity of data	0.9242
	Data relevance	0.8721
	All Items	**0.9045**

5.2.6.2 Half Split Method

This method depends on finding Pearson correlation coefficient between the means of odd rank questions and even rank questions of each field of the questionnaire. Then, correcting the Pearson correlation coefficients can be done by using Spearman Brown correlation coefficient of correction. The corrected correlation coefficient (consistency coefficient) is computed according to the following equation:

Consistency coefficient = 2r/(r+1), where r is the Pearson correlation coefficient. The normal range of corrected correlation coefficient 2r/(r+1) is between 0.0 and + 1.0 as shown in Table (5.6 Split-Half Coefficient method), and the general reliability for all items equal 0.8965, and the significant (α) is less than 0.05 so all the corrected correlation coefficients are significance at α = 0.05. It can be said that according to the Half Split method, the dispute causes group are reliable.

Table 5.6: Split-Half Coefficient method

Section	Sub-section	person-correlation	Spearman-Brown Coefficient	Sig. (2-Tailed)
Section one (Business Processes)	Data completeness	0.7296	0.8436	0.000
	Data accuracy	0.7525	0.8588	0.000
	Data consistency	0.6924	0.8182	0.000
	Validity of data	0.7895	0.8824	0.000
	Data relevance	0.7725	0.8717	0.0000
Section two (Decision making)	Data completeness	0.7568	0.8616	0.0000
	Data accuracy	0.8145	0.8978	0.0000
	Data consistency	0.7758	0.8737	0.000
	Validity of data	0.7924	0.8842	0.0000
	Data relevance	0.7525	0.8588	0.000
	All Items	**0.8124**	**0.8965**	**0.0000**

5.3 Analysis of Sample Characteristics

Personal information characteristics

In this section, the personal characteristics of the research sample will be discussed.

5.3.1 Gender

Table (5.7 Gender) shows that 24.0% from the sample are “Female”, and 76.0 % from the sample are "Male". The contribution of the females in the workforce in insurance companies is limited and less than male contribution by three times. Referring to the Palestinian central bureau of statistics PCBS 2013, the females’ contributions in the workforce are limited to 17.3%.

Table 5.7: Gender

Gender	Frequency	Percentages
Female	12	24.0
Male	38	76.0
Total	**50**	**100.0**

5.3.2 Age group

Table (5.8 Age group) shows that 14.0% of the sample of ages less than 25 years, 46.0% from the sample of age between 25 and 34 years old, 28.0 % from the sample of age between 35 to 44 year old, and 12.0 % from the sample of age more than 45 years old. This means that the majority of insurance sector employees in Gaza Strip considered young where around 60% of them are less than 35 years. This reflects that insurance companies are relaying on employing fresh graduates.

Table 5.8: Age group

Age group	Frequency	Percentages
Less than 25 years	7	14.0
From 25 to 34 year	23	46.0
From 35 to 44 year	14	28.0
More than 45 year	6	12.0
Total	**50**	**100.0**

5.3.3 Academic qualification

Table (5.9 Educational Level) shows that 14.0% of the sample holds diploma degree, 74.0% of the sample holds bachelor degree, 12.0% of the sample holds master degree. It is noticed that there is 0.0% of PhD holders in insurance companies sector in Gaza strip. The fact that more than 85.0% of the employees are educated (bachelor and master) is reflecting the interest of insurance companies in educated employees. This refers that insurance companies are more interested with educated employees.

Table 5.9: Educational Level

Academic qualification	Frequency	Percentages
Diploma	7	14.0
Bachelor	37	74.0
Master	6	12.0
PhD	0	0.0
Total	**50**	**100.0**

5.3.4 Employment level

Table (5.10 Employment level) shows that 18.0% from the sample of employment level are Department heads, 2.0% from the sample of employment level are vice branch manager, 6.0% from the sample of employment level are branch manager, and 74.0% from the sample of employment level are employee. The low number of senior management employees is because the small numbers of insurance companies in Gaza strip where each company needs one employee for senior management.

Table 5.10: Employment level

Employment level	Frequency	Percentages
Department head	9	18.0
Vice branch manager	1	2.0
Branch manager	3	6.0
Employee	37	74.0
Total	**50**	**100.0**

5.3.5 Work experience

Table (5.11 Work experience) shows that 36.0% of the sample has less than 5 years' experience, 28.0% of the sample is having experience from 5 to 9 years, 16.0% of the sample is having experience from 10 to 14 year, and only 20.0% of the sample is having more than 15 years of experience. From the researcher point of view this is because those insurance companies were established after 1994.

Table 5.11: Work experience

Work experience	Frequency	Percentages
Less than 5 years	18	36.0
From 5 to 9 years	14	28.0
From 10 to 14 year	8	16.0
More than 15 years	10	20.0
Total	**50**	**100.0**

5.3.6 Former knowledge of "data quality" concept

Table (5.11 Former knowledge of "data quality"?) shows that 80.0% of the sample has answered "Yes" to the question and 20.0% of them have answered "No". This indicates that most of insurance companies' employees knows about the concept of data quality, which gives a good perception that they will understand the purpose of the questionnaire, and answers questions without any ambiguity.

Table 5.12: Former knowledge of "data quality"?

Former knowledge of "data quality"?	Frequency	Percentages
Yes	40	80.0
No	10	20.0
Total	**50**	**100.0**

5.3.7 Decision making responsibilities

Table (5.12 Are you a decision maker?) shows that 66.0% of the sample has decision making responsibilities, but 34.0% are not. This indicates that more than two-third of the population sample answered the decision making section in reliable manner.

Table 5.13: Are you a decision maker?

Are you a decision maker?	Frequency	Percentages
Yes	33	66.0
No	17	34.0
Total	**50**	**100.0**

5.4 Data Analysis and Hypotheses Testing

5.4.1 Tests of Normality

1-Sample K-S test will be used to identify if the data follow normal distribution or not, this test is considered necessary in case testing hypotheses as most parametric Test stipulate data to be normally distributed and this test is used when the size of the sample are greater than or equal 50.

Test results shown in Table (5.14 1-Sample k-s), clarifies that the calculated p-value is greater than the significant level which is equal 0.05 (p-value > 0.05), this in turn denotes that data follows normal distribution, and so parametric Tests must be used.

Table 5.14: 1-Sample k-s

Section	Sub-section	Statistic test	P-value
Section one (Business Processes)	Data completeness	0.727	0.751
	Data accuracy	0.849	0.316
	Data consistency	0.960	0.150
	Validity of data	1.045	0.224
	Data relevance	1.323	0.091
Section two (Decision making)	Data completeness	0.475	0.978
	Data accuracy	0.650	0.792
	Data consistency	0.858	0.453
	Validity of data	0.886	0.412
	Data relevance	0.994	0.276
	All Items	**1.045**	**0.224**

5.4.2 Discussion and Hypotheses test

In the following Tables, a one sample t test is used to test if the opinion of the respondent in the content of the sentences are positive (weighted mean is greater than "60.0%", the p-value is less than 0.05 and t-value is positive) or the opinion of the respondent in the content of the sentences are neutral (p-value is greater than 0.05 and regardless the sign of the t-value) or the opinions of the respondents in the content of the sentences are negative (weighted mean is less than "60.0%", the p-value is less than 0.05 and t-value is negative).

5.4.2.1 Dependent variable 1: Business Processes

The dependent variable "Business processes" is studied against five independent variables which are data completeness, data accuracy, data consistency, validity of data and data relevance.

- **Data completeness**

The opinions of the respondents about data completeness in business processes are shown in Table (5.15 Data completeness in business processes).

Table 5.15: Data completeness in business processes

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	Data used in the daily business processes is complete for its purpose of use	8.24	2.592	82.40	6.112	0.000
2	When there is a problem in data completeness, you look for its source to ensure its completeness	8.08	2.820	80.80	5.215	0.000
3	When there is a problem in data completeness, you report the problem	8.30	2.401	83.00	6.773	0.000
4	When there is a problem in data completeness, you continue working without considering the problem	6.52	3.598	65.20	1.022	0.312
5	The organization works on improving quality of data used for all organizational purposes	7.92	2.641	79.20	5.141	0.000
6	The lack of data completeness causes problems to daily business processes	7.82	2.723	78.20	4.726	0.000
7	Where there is a problem in data completeness, you stop working till the problem is solved	5.72	3.207	57.20	-0.617	0.540
8	The lack of data completeness affects the daily business processes	7.58	2.822	75.80	3.959	0.000
9	Data completeness is important for daily business process in the organization	8.34	2.592	83.40	6.383	0.000
10	The more used data is complete, the more the quality of business process.	8.64	2.422	86.40	7.706	0.000
	All items	**7.72**	**2.080**	**77.16**	**5.833**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

The results shown in Table (5.15 Data completeness in business processes) donate the following:

1- The highest two items according to the weighted mean are:
    a. In item no. (10) the weighted mean equal "86.40%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means higher data completeness results in higher quality of business processes.
    b. In item no. (9) the weighted mean equal "83.40%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that indicates the high importance of data completeness to business processes.

2- The lowest two items according to the weighted mean are:
    a. In item no. (4) the weighted mean equal "65.20%" and p-value equal "0.312" which is greater than 0.05 with positive t-value, this means when facing a problem in data completeness, no general agreement nor disagreement among respondents about what to do in such situation was shown by the results.
    b. In item no. (7) the weighted mean equal "57.20%" and p-value equal "0.540" which is greater than 0.05 with negative t-value, this means when facing a problem in data completeness, no general agreement nor disagreement among respondents about what to do in such situation was shown by the results.

For general the results for all items of the field show that the average mean equal 7.72 and the weighted mean equal 77.16% which is greater than “60%", and the value of t test equal 5.833 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means insurance companies in Gaza strip gives data completeness a great attention at significance level $\alpha \leq 0.05$.

- **Data accuracy**

The opinions of the respondents about data accuracy in business processes are shown in Table (5.16 Data accuracy in business processes).

Table 5.16: Data accuracy in business processes

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	The data used in the daily business processes are accurate for its purpose of use	8.24	2.544	82.40	6.226	0.000
2	When there is a problem in the data accuracy you look for its source to ensure its accuracy	8.00	2.523	80.00	5.604	0.000
3	When there is a problem in data accuracy, you report the problem	7.76	2.707	77.60	4.597	0.000
4	When there is a problem in the data accuracy you continue working without considering the problem	4.68	3.329	46.80	-2.804	0.007
5	The lack in data accuracy causes problems to daily business processes	7.32	2.567	73.20	3.636	0.001
6	When there is a problem in the data accuracy, you stop working until the problem is solved	5.94	3.026	59.40	-0.140	0.889
7	The lack of data accuracy affects the daily business processes	7.30	3.025	73.00	3.038	0.004
8	Data accuracy is important for daily business process in the organization	8.24	2.576	82.40	6.149	0.000
9	The more used data is accurate, the more the quality of business process.	8.52	2.435	85.20	7.319	0.000
	All items	**7.33**	**1.795**	**73.33**	**5.253**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

The results shown in Table (5.16 Data accuracy in business processes) donate the following:

1- The highest two items according to the weighted mean are:

   a. In item no. (9) the weighted mean equal "85.20%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means a general agreement is shown about the importance of accurate data to the quality of business processes.

   b. In item no. (1) the weighted mean equal "82.40%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means insurance companies assure the accuracy of data used in daily business processes.

2- The lowest two items according to the weighted mean are:

   a. In item no. (6) the weighted mean equal "59.40%" and p-value equal "0.889" which is greater than 0.05 with negative t-value, that means when facing a problem in data accuracy, no general agreement nor disagreement about what to do in such situation was shown by the results.

   b. In item no. (4) the weighted mean equal "46.80%" and p-value equal "0.007" which is less than 0.05 with negative t-value, that means respondents stop working if they face a problem in data accuracy.

For general the results for all items of the field show that the average mean equal 7.33 and the weighted mean equal 73.33% which is greater than "60%" and the value of t test equal 5.253 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means data accuracy in business processes is considered important to insurance companies at significance level $\alpha \leq 0.05$.

- **Data consistency**

The opinions of the respondents about data consistency in business processes are shown in Table (5.17 Data consistency in business processes).

Table 5.17: Data consistency in business processes

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	The data used in daily business processes are consistent for its purpose of use	8.14	2.474	81.40	6.115	0.000
2	When there is a problem in the data consistency, you look for its source to ensure its consistency	7.96	2.356	79.60	5.883	0.000
3	When there is a problem in the data consistency, you report the problem	8.06	2.606	80.60	5.589	0.000
4	When there is a problem in the data consistency, you continue working without considering the problem	5.06	3.139	50.60	-2.117	0.039
5	The lack of data consistency causes problems to daily business processes	7.24	3.027	72.40	2.896	0.006
6	When there is a problem in the data consistency, you stop working until	5.44	3.176	54.40	-1.247	0.218

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
	the problem is solved					
7	The lack of data consistency affects the daily business processes	7.22	3.125	72.20	2.760	0.008
8	Data consistency is important for daily business process in the organization	7.62	2.934	76.20	3.904	0.000
9	The more used data is consistent, the more the quality of business process.	7.84	2.881	78.40	4.516	0.000
	All items	**7.18**	**2.038**	**71.76**	**4.079**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

The results shown in Table (5.17 Data consistency in business processes) donate the following:

1- The highest two items according to the weighted mean are:

   a. In item no. (1) the weighted mean equal "81.40%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means insurance companies gives good attention to ensure data consistency in business processes.

   b. In item no. (3) the weighted mean equal "80.60%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means if there is a problem in data consistency, respondents report the problem to try to fix it.

2- The lowest two items according to the weighted mean are:

   a. In item no. (6) the weighted mean equal "54.40%" and p-value equal "0.218" which is greater than 0.05 with negative t-value, that means no general agreement nor disagreement among respondents response to the problem.

   b. In item no. (4) the weighted mean equal "50.60%" and p-value equal "0.039" which is less than 0.05 with negative t-value, that means when there is a problem in the data consistency, employees stop working until the problem is resolved.

For general the results for all items of the field show that the average mean equal 7.18 and the weighted mean equal 71.76% which is greater than " 60%" and the value of t test equal 4.079 which is greater than the critical value which is equal 2.01 and the p-value

equal 0.000 which is less than 0.05, that means insurance companies put interest on having consistent data at significance level $\alpha \leq 0.05$.

- **Validity of data**

The opinions of the respondents about validity of data in business processes are shown in Table (5.18 Validity of data in business processes).

Table 5.18: Validity of data in business processes

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	The data used in the daily business processes are valid for its purpose of use	8.12	2.662	81.20	5.631	0.000
2	When there is a problem in the data validity, you look for its source to ensure its validity	7.88	2.647	78.80	5.022	0.000
3	When there is a problem in the data validity, you report the problem	7.76	2.638	77.60	4.717	0.000
4	Poor data quality affects business processes decision making	7.42	2.949	74.20	3.405	0.001
5	When there is a problem in the data validity, you continue working without considering the problem	4.56	3.302	45.60	-3.084	0.003
6	The lack of data validity causes problems to daily business processes	7.44	2.991	74.40	3.404	0.001
7	When there is a problem in the data validity, you stop working until the problem is solved	6.32	3.133	63.20	0.722	0.474
8	Obtaining high data quality is not important to the organization's success*	4.76	3.589	47.60	-2.443	0.018
9	The lack of data validity affects the daily business processes	7.90	2.697	79.00	4.981	0.000
10	The data validity is important for daily business process in the organization	8.02	2.714	80.20	5.263	0.000
11	The more used data is valid, the more the quality of business	8.10	2.735	81.00	5.430	0.000

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
	process					
	All items	**7.25**	**1.917**	**72.51**	**15.682**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01
*q8 is negative

The results shown in Table (5.18 Validity of data in business processes) donate the following:

1- The highest two items according to the weighted mean are:
   a. In item no. (1) the weighted mean equal "81.20%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means insurance companies are keen in obtaining valid data to be used in its daily business processes.
   b. In item no. (11) the weighted mean equal "81.00%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means most respondents agreed on the importance of valid data to high quality business processes.

2- The lowest two items according to the weighted mean are:
   a. In item no. (8) the weighted mean equal "47.60%" and p-value equal "0.018" which is less than 0.05 with negative t-value, that means respondents agreed on the importance of high data quality to the success of organization.
   b. In item no. (5) the weighted mean equal "45.60%" and p-value equal "0.003" which is less than 0.05 with negative t-value, that means respondents stop working if they face a problem in data validity.

For general the results for all items of the field show that the average mean equal 7.25 and the weighted mean equal 72.51% which is greater than "60%" and the value of t test equal 15.682 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means validity of data is considered important for business processes in insurance companies at significance level $\alpha \leq 0.05$.

- **Data relevance**

The opinions of the respondents about data relevance in business processes are shown in Table (5.19 Data relevance in business processes).

Table 5.19: Data relevance in business processes

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	The data used in the daily business processes are relevance for its purpose of use	7.84	2.979	78.40	4.368	0.000
2	When there is a problem in the data relevancy, you look for its source to ensure its relevancy	7.92	2.813	79.20	4.827	0.000
3	When there is a problem in the data relevancy, you report the problem	7.96	2.610	79.60	5.309	0.000
4	When there is a problem in the data relevancy, you continue working without considering the problem	5.02	3.473	50.20	-1.995	0.052
5	The lack of data relevancy causes problems to daily business processes	7.68	2.699	76.80	4.402	0.000
6	When there is a problem in the data relevancy, you stop working until the problem is solved	7.82	2.238	78.20	5.751	0.000
7	The lack of data relevancy affects the daily business processes	7.60	2.900	76.00	3.902	0.000
8	The data relevancy is important for daily business process in the organization	7.84	2.874	78.40	4.527	0.000
9	The more used data is relevant, the more the quality of business process	8.16	2.637	81.60	5.792	0.000
	All items	**7.54**	**2.158**	**75.38**	**5.039**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

The results shown in Table (5.19 Data relevance in business processes) donate the following:

1- The highest two items according to the weighted mean are:

   a. In item no. (9) the weighted mean equal "81.60%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means that there is general agreement about the importance of data relevance to high quality business processes.

   b. In item no. (3) the weighted mean equal "79.60%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means any problem data relevance will be reported to be solved.

2- The lowest two items according to the weighted mean are:

a. In item no. (7) the weighted mean equal "76.00%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means there is general agreement among respondents about the importance of data relevance to business processes and any lack in it rises problems.

b. In item No. (4) the weighted mean equal "50.20%" and p-value equal "0.052" which is greater than 0.05 with negative t-value, that means no general agreement nor disagreement among respondents about what to do if there is a problem in data relevancy.

For general the results for all items of the field show that the average mean equal 7.54 and the weighted mean equal 75.38% which is greater than "60%" and the value of t test equal 5.039 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means data relevance in business processes have great effect on business processes in insurance companies in Gaza Strip at significance level $\alpha \leq 0.05$.

- **Business Processes (all sub-sections)**

A one sample t test was used for the opinion of the respondents about all sub-sections of business processes and the results are shown in Table (5.20 Business processes). It shows that the average mean for all sub-sections equal 7.40 and the weighted mean equal 74.03% which is greater than "60%" and the value of t test equal 16.248 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means data quality is important to business processes in insurance companies in Gaza Strip at significance level $\alpha \leq 0.05$.

Table 5.20: Business processes

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	Data completeness	7.72	2.080	77.16	16.030	0.000
2	Data accuracy	7.33	1.795	73.33	17.072	0.000
3	Data consistency	7.18	2.038	71.76	14.487	0.000
4	Validity of data	7.25	1.917	72.51	15.682	0.000
5	Data relevance	7.54	2.158	75.38	14.869	0.000
	All items	**7.40**	**1.916**	**74.03**	**16.248**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

Dependent variable 2: Decision making

The dependent variable "Decision making" is studied against five independent variables which are data completeness, data accuracy, data consistency, validity of data and data relevance.

- **Data completeness**

The opinions of the respondents about data completeness in decision making are shown in Table (5.21 Data completeness in decision making).

Table 5.21: Data completeness in decision making

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	Data completeness affects the decision making process in the organization	8.10	2.581	81.00	5.753	0.000
2	It is difficult to reach a proper decision when the data is incomplete	8.02	2.622	80.20	5.447	0.000
3	When there is missing data, you try to collect the missing data to facilitate the process of decision making	7.98	2.714	79.80	5.158	0.000
4	The organization works on improving quality of data used for all organizational purposes	7.70	2.794	77.00	4.302	0.000
5	Having complete data is a group effort that affect all staff in the organization	7.94	2.551	79.40	5.378	0.000
6	A data that is complete for making one decision is not necessarily complete for another	7.98	2.638	79.80	5.307	0.000
7	The organization works on achieving the highest level of data completeness for all its employees in there different jobs	7.88	2.488	78.80	5.343	0.000
	All items	**7.94**	**2.507**	**79.43**	**5.479**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

The results shown in Table (5.21 Data completeness in decision making) donate the following:

1- The highest two items according to the weighted mean are:
    a. In item no. (1) the weighted mean equal "81.00%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means respondents agrees that data completeness affect decision making in the organization.
    b. In item no. (2) the weighted mean equal "80.20%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means if the data used is not complete, the decision may be wrong or in low quality.

2- The lowest two items according to the weighted mean are:
    a. In item no. (7) the weighted mean equal "78.80%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means insurance companies in Gaza strip works hard to archive the highest level of data quality to its employees.
    b. In item no. (4) the weighted mean equal "77.00%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means insurance companies works on improving quality of data used for all organizational purposes.

For general the results for all items of the field show that the average mean equal 7.94 and the weighted mean equal 79.43% which is greater than " 60%" and the value of t test equal 5.479 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means data completeness in decision making have high importance in insurance companies in Gaza Strip at significance level $\alpha \leq 0.05$.

- **Data accuracy**

The opinions of the respondents about data accuracy in decision making are shown in Table (5.22 Data accuracy in decision making).

Table 5.22: Data accuracy in decision making

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	Data accuracy affects the decision making process in the organization	7.76	2.832	77.60	4.394	0.000
2	It is difficult to reach a proper	7.76	2.759	77.60	4.510	0.000

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
	decision when the data is inaccurate					
3	When there is inaccurate data, you try to improve data to facilitate the process of decision making	7.58	2.807	75.80	3.980	0.000
4	Having accurate data is a group effort that affect all staff in the organization	7.94	2.445	79.40	5.612	0.000
5	Improving data quality leads to raising the value of the whole organization	8.10	2.525	81.00	5.880	0.000
6	A data that is accurate for making one decision is not necessarily accurate for another	7.58	2.492	75.80	4.484	0.000
7	The organization works on achieving the highest level of data accuracy for all its employees in there different jobs	8.26	2.165	82.60	7.382	0.000
	All items	**7.85**	**2.348**	**78.54**	**5.584**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

The results shown in Table (5.22 Data accuracy in decision making) donate the following:

1- The highest two items according to the weighted mean are:

   a. In item no. (7) the weighted mean equal "82.60%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means insurance companies tries to achieve the highest data accuracy level to its employees at different levels.

   b. In item no. (5) the weighted mean equal "81.00%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means respondents agrees that improving data quality leads to raising the value of the whole organization.

2- The lowest two items according to the weighted mean are:

   a. In item no. (3) the weighted mean equal "75.80%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means respondents are aware of data accuracy and if they face any problem they try to resolve the problem and improve the data to facilitate the process of decision making.

b. In item no. (6) the weighted mean equal "75.80%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means respondents are aware that accurate data is subjective to the situation and the decision to be made.

For general the results for all items of the field show that the average mean equal 7.85 and the weighted mean equal 78.54% which is greater than "60%" and the value of t test equal 5.584 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means data accuracy in decision making have high importance to insurance companies in Gaza Strip at significance level $\alpha \leq 0.05$.

- **Data consistency**

The opinions of the respondents about data consistency in decision making are shown in Table (5.23 Data consistency in decision making).

Table 5.23: Data consistency in decision making

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	Data consistency affects the decision making process in the organization	7.86	2.755	78.60	4.773	0.000
2	It is difficult to reach a proper decision when the data is inconsistence	7.60	2.814	76.00	4.021	0.000
3	When there is inconsistent data, you try to improve data to facilitate the process of decision making	7.72	2.703	77.20	4.499	0.000
4	Having consistent data is a group effort that affect all staff in the organization	7.46	2.712	74.60	3.807	0.000
5	A data that inconsistent for making one decision is not necessarily consistent for another	7.58	2.627	75.80	4.253	0.000
6	The organization works on achieving the highest level of data consistency for all its employees in there different jobs	7.96	2.579	79.60	5.374	0.000
	All items	**7.70**	**2.573**	**76.97**	**4.663**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

The results shown in Table (5.23 Data consistency in decision making) donate the following:

1- The highest two items according to the weighted mean are:
   a. In item no. (6) the weighted mean equal "79.60%" and p-value equal 0.000" which is less than 0.05 with positive t-value, that means respondents agreed that the organization works on achieving the highest level of data consistency for all its employees in their different jobs.
   b. In item no. (1) the weighted mean equal "78.60%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means data consistency affects the decision making process in the organization.

2- The lowest two items according to the weighted mean are:
   a. In item no. (5) the weighted mean equal "75.80%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means respondents are aware that data consistent to make one decision is not necessarily consistent for making another.
   b. In item no. (4) the weighted mean equal "74.60%" and p-value equal "0.000 " which is less than 0.05 with positive t-value, that means having consistent data is a group effort that affect all staff in the organization.

For general the results for all items of the field show that the average mean equal 7.70 and the weighted mean equal 76.97% which is greater than “60%" and the value of t test equal 4.663 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means data consistency in decision making is important to insurance companies in Gaza Strip at significance level $\alpha \leq 0.05$.

- **Validity of data**

The opinions of the respondents about validity of data in decision making are shown in Table (5.24 Validity of data in decision making).

Table 5.24: Validity of data in decision making

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	Data validity affects the decision making process in the organization	7.98	2.503	79.80	5.594	0.000
2	It is difficult to reach a proper decision when the data is not valid	8.08	2.602	80.80	5.653	0.000

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
3	When there is invalid data, you try to improve data to facilitate the process of decision making	7.92	2.709	79.20	5.011	0.000
4	Having valid data is a group effort that affect all staff in the organization	7.90	2.667	79.00	5.038	0.000
5	A data that is valid for making one decision is not necessarily valid for another	7.52	2.901	75.20	3.704	0.001
6	The organization works on achieving the highest level of data validity for all its employees in there different jobs	8.02	2.875	80.20	4.968	0.000
	All items	**7.90**	**2.470**	**79.03**	**5.449**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

The results shown in Table (5.24 Validity of data in decision making) donate the following:

1- The highest two items according to the weighted mean are:

   a. In item no. (2) the weighted mean equal "80.80%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means it is difficult to reach a proper decision when the data is not valid.

   b. In item no. (6) the weighted mean equal "80.20%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means insurance companies works on achieving the highest level of data validity for all its employees in their different jobs.

2- The lowest two items according to the weighted mean are:

   a. In item no. (4) the weighted mean equal "79.00%" and p-value equal "0.000 " which is less than 0.05 with positive t-value, that means having valid data is a group effort that affect all staff in the organization.

   b. In item no. (5) the weighted mean equal "75.20%" and p-value equal "0.001" which is less than 0.05 with positive t-value, that means respondents agrees that a data that is valid for making one decision is not necessarily valid for making another.

For general the results for all items of the field show that the average mean equal 7.90 and the weighted mean equal 79.03% which is greater than "60%" and the value of t test equal 5.449 which is greater than the critical value which is equal 2.01 and the p- value equal 0.000 which is less than 0.05, that means validity of data in decision making is considered important to insurance companies in Gaza Strip at significance level $\alpha \leq 0.05$.

- **Data relevance**

The opinions of the respondents about data relevance in decision making are shown in Table (5.25 Data relevance in decision making).

Table 5.25: Data relevance in decision making

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	Data relevancy affects the decision making process in the organization	7.82	2.862	78.20	4.497	0.000
2	It is difficult to reach a proper decision when the data is not relevant	7.78	2.916	77.80	4.317	0.000
3	When there is irrelevant data, you try to improve data to facilitate the process of decision making	7.68	2.931	76.80	4.053	0.000
4	Having relevant data is a group effort that affect all staff in the organization	7.64	2.812	76.40	4.124	0.000
5	A data that is relevant for making one decision is not necessarily relevant for another	7.46	2.823	74.60	3.657	0.001
6	The organization works on achieving the highest level of data relevancy for all its employees in there different jobs	7.72	2.778	77.20	4.378	0.000
7	Obtaining high data quality is not important to the organization's success*	6.30	3.576	63.00	0.593	0.556
	All items	**7.26**	**2.205**	**72.57**	**13.654**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01
*q7 is negative

The results shown in Table (5.25 Data relevance in decision making) donate the following:

1- The highest two items according to the weighted mean are:
    a. In item no. (1) the weighted mean equal "78.20%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means data relevancy affects the decision making process in the organization.
    b. In item no. (2) the weighted mean equal "77.80%" and p-value equal "0.000" which is less than 0.05 with positive t-value, that means respondents agreed that it is difficult to reach a proper decision when the data is not relevant.

2- The lowest two items according to the weighted mean are:
    a. In item no. (5) the weighted mean equal "74.60%" and p-value equal "0.001" which is less than 0.05 with positive t-value, that means a data that is relevant for making one decision is not necessarily relevant for making another.
    b. In item no. (7) the weighted mean equal "63.00%" and p-value equal "0.556" which is greater than 0.05 with positive t-value, that means no general agreement nor disagreement among respondents on the importance of data quality to the organization's success.

For general the results for all items of the field show that the average mean equal 7.26 and the weighted mean equal 72.57% which is greater than "60%" and the value of t test equal 13.654 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means data relevance in decision making is considered important to insurance companies in Gaza Strip at significance level $\alpha \leq 0.05$.

- **Decision making (all sub-sections)**

A one sample t test was used for the opinion of the respondents about all sub-sections of decision making and the results shown in Table (5.26 Decision making) which show that the average mean for all sub-sections equal 7.73 and the weighted mean equal 77.27% which is greater than "60%" and the value of t test equal 14.542 which is greater than the critical value which is equal 2.01 and the p-value equal 0.000 which is less than 0.05, that means data quality is important to decision making process at insurance companies in Gaza Strip at significance level $\alpha \leq 0.05$.

Table 5.26: Decision making

No.	Items	Mean	Standard deviation	Weight mean	t-value	P-value
1	Data completeness	7.94	2.507	79.43	13.940	0.000
2	Data accuracy	7.85	2.348	78.54	14.619	0.000
3	Data consistency	7.70	2.573	76.97	12.908	0.000
4	Validity of data	7.90	2.470	79.03	14.037	0.000
5	Data relevance	7.26	2.205	72.57	13.654	0.000
	All items	**7.73**	**2.298**	**77.27**	**14.542**	**0.000**

Critical value of t at df "49" and significance level 0.05 equal 2.01

5.4.2.2 Research Hypotheses

The research hypotheses are as follow:

H1: There is significant impact between data completeness and business processes and decision making at ($\alpha \leq 0.05$) level

A Pearson correlation test was used to test the impact between data completeness on business processes and decision making at significance level $\alpha \leq 0.05$ and the results shown in Table (5.27 A correlation between data completeness and business processes and decision making at significance level $\alpha \leq 0.05$) show that the correlation coefficient equal 0.957 which is greater than critical value 0.271 and p-value equal 0.000 which is less than 0.05, that mean there is a positive correlation between data completeness and business processes and decision making at significance level $\alpha \leq 0.05$.

This result shows the importance of data completeness to the successes of the daily business processes and decision making. This results are consistent with most of the literature like (Chen, Hailey, Wang, & Yu, 2014), (Herzog, Scheuren, & Winkler, 2007), (Shankaranarayanan & Cai, 2006) and (Scannapieco, Missier, & Batini, 2005) which agreed that completeness in as important dimension of data quality that should be addresses by the organizations to attain organizational success.

Table 5.27: A correlation between data completeness and business processes and decision making at significance level α ≤ 0.05

Section	Statistic	Data completeness	Data accuracy	Data consistency	Validity of data	Data relevance	Business processes and decision making(all sub-section)
Data completeness in business processes and decision making	Pearson Correlation	0.897	0.872	0.801	0.833	0.818	0.957
	P-value	0.000	0.000	0.000	0.000	0.000	0.000
	N	50	50	50	50	50	50

*Correlation is significant at the 0.01 level (2-tailed)

H2: There is significant impact between data accuracy and business processes and decision making at (α ≤ 0.05) level

A Pearson correlation test was used to test the impact between data accuracy and business processes and decision making at significance level α ≤ 0.05 and the results shown in Table (5.28 A correlation between data accuracy and business processes and decision making at significance level α ≤ 0.05) shows that the correlation coefficient equal 0.966 which is greater than critical value 0.271 and p-value equal 0.000 which is less than 0.05, that mean there is a positive correlation between data accuracy and business processes and decision making at significance level α = 0.05.

The results show that respondents has agreed on the importance of data accuracy to the daily business processes and decision making and these results are consistent with most literature like (Moges, Dejaeger, Lemahieu, & Baesens, 2013), (Alsaidi & Zbar, 2013), (Batini, Cappiello, Francalanci, & Maurino, 2009) and (Herzog, Scheuren, & Winkler, 2007) which all agrees on the significant effect of data accuracy on the organization overall successes.

Table 5.28: A correlation between data accuracy and business processes and decision making at significance level α ≤ 0.05

Section	Statistic	Data completeness	Data accuracy	Data consistency	Validity of data	Data relevance	Business processes and decision making(all sub-section)
Data accuracy in business processes and decision making	Pearson Correlation	0.839	0.857	0.834	0.806	0.824	0.966
	P-value	0.000	0.000	0.000	0.000	0.000	0.000
	N	50	50	50	50	50	50

*Correlation is significant at the 0.01 level (2-tailed)

H3: There is significant impact between data consistency and business processes and decision making at (α ≤ 0.05) level

A Pearson correlation test was used to test the impact between data consistency and business processes and decision making at significance level α ≤ 0.05 and the results shown in Table (5.29 A correlation between data consistency and business processes and decision making at significance level α ≤ 0.05) show that the correlation coefficient equal 0.962 which is greater than critical value 0.271 and p-value equal 0.000 which is less than 0.05, that mean there is a positive correlation between data consistency and business processes and decision making at significance level α ≤ 0.05.

The results show that data consistency has an effect on the organization's business processes and decision making which is consistent with most literature like (Askham, et al., 2013), (Haug, Zachariassen, & van Liempd, 2011) and (Batini & Scannapieca, 2006) where they all agree to the importance of consistency dimension.

Table 5.29: A correlation between data consistency and business processes and decision making at significance level α ≤ 0.05

Section	Statistic	Data completeness	Data accuracy	Data consistency	Validity of data	Data relevance	Business processes and decision making(all sub-section)
Data consistency in business processes and decision making	Pearson Correlation	0.819	0.828	0.856	0.788	0.772	0.962
	P-value	0.000	0.000	0.000	0.000	0.000	0.000
	N	50	50	50	50	50	50

*Correlation is significant at the 0.01 level (2-tailed)

H4: There is significant impact between validity of data and business processes and decision making at (α ≤ 0.05) level

A Pearson correlation test was used to test the impact between validity of data and business processes and decision making at significance level α ≤ 0.05 and the results shown in Table (5.30 A correlation between validity of data and business processes and decision making at significance level α ≤ 0.05) show that the correlation coefficient equal 0.950 which is greater than critical value 0.271 and p-value equal 0.000 which is less than 0.05, that mean there is a positive correlation between validity of data and business processes and decision making at significance level α ≤ 0.05.

The respondents have agreed on the importance of validity of data on the daily business processes and decision making where this result are consistent with many literatures like (Chen, Hailey, Wang, & Yu, 2014), (Askham, et al., 2013), (Loshin, 2011) and (Stvilia, Gasser, Twidale, & Smith, 2007).

Table 5.30: A correlation between validity of data and business processes and decision making at significance level α ≤ 0.05

Section	Statistic	Data completeness	Data accuracy	Data consistency	Validity of data	Data relevance	Business processes and decision making(all sub-section)
Validity of data in business processes and decision making	Pearson Correlation	0.871	0.848	0.816	0.754	0.757	0.950
	P-value	0.000	0.000	0.000	0.000	0.000	0.000
	N	50	50	50	50	50	50

*Correlation is significant at the 0.01 level (2-tailed)

H5: There is significant impact between data relevance and business processes and decision making at (α ≤ 0.05) level

A Pearson correlation test was used to test the impact between data relevance and business processes and decision making at significance level α ≤ 0.05 and the results shown in Table (5.31 A correlation between data relevance and business processes and decision making at significance level α ≤ 0.05) show that the correlation coefficient equal 0.967 which is greater than critical value 0.271 and p-value equal 0.000 which is less than 0.05, that mean there is a positive correlation between data relevance and business processes and decision making at significance level α ≤ 0.05.

The results show that the respondents have agreed on the important effect of data relevance on business processes and decision making to attain organizational successes. This result is consistent with most literature like (Chen, Hailey, Wang, & Yu, 2014), (Moges, Dejaeger, Lemahieu, & Baesens, 2013) and (Herzog, Scheuren, & Winkler, 2007).

Table 5. 31: A correlation between data relevance and business processes and decision making at significance level α ≤ 0.05

Section	Statistic	Data completeness	Data accuracy	Data consistency	Validity of data	Data relevance	Business processes and decision making(all sub-section)
Data relevance in business processes and decision making	Pearson Correlation	0.922	0.867	0.797	0.812	0.773	0.967
	P-value	0.000	0.000	0.000	0.000	0.000	0.000
	N	50	50	50	50	50	50

*Correlation is significant at the 0.01 level (2-tailed)

5.4.2.3 Demographic hypothesis analysis

There is a statistically significant differences attributed to the personal information of the respondents at the level (α ≤ 0.05) about the impact of data quality on business processes and decision making

And this hypothesis is divided into sub-hypotheses as follows:

H1-There is a statistically significant difference at the level (α ≤ 0.05) about the impact of data quality on business processes and decision making refer to gender?

The Independent Samples Test is used to test the hypothesis and the results is illustrated in Table (5.32 Independent Samples Test for differences about the impact of data quality on business processes and decision making refer to gender) show that the p-value equal 0.499 which is greater than 0.05 and the absolute value of T test equal 0.681 which is less than the critical value which is equal 2.01. This means that gender has no effect on the results of the questions regarding the impact of data quality on business processes and decision making in insurance companies in Gaza Strip.

Table 5.32: Independent Samples Test for differences about the impact of data quality on business processes and decision making refer to gender

Field	Gender	N	Mean	Std. Deviation	T	P-value
Business processes	**Female**	12	7.153	1.885	-0.515	0.609
	Male	38	7.482	1.944		
Decision making	**Female**	12	7.235	2.292	-0.848	0.401
	Male	38	7.882	2.309		
All items	**Female**	12	7.186	2.018	-0.681	0.499
	Male	38	7.645	2.039		

Critical value of t at df "48" and significance level 0.05 equal 2.01

H2-There is a statistically significant difference at the level ($\alpha \leq 0.05$) about the impact of data quality on business processes and decision making refer to having former knowledge of "data quality"

The Independent Samples Test is used to test the hypothesis and the results is illustrated in Table (5.33 Independent Samples Test for differences about the impact of data quality on business processes and decision making refer to having former knowledge of "data quality") show that the p-value equal 0.731 which is greater than 0.05 and the absolute value of T test equal 0.370 which is less than the critical value which is equal 2.01. This means that having former knowledge of "data quality" has no effect on the results of the questions regarding the impact of data quality on business processes and decision making in insurance companies in Gaza Strip.

Table 5.33: Independent Samples Test for differences about the impact of data quality on business processes and decision making refer to having former knowledge of "data quality"

Field	Former knowledge of "data quality"	N	Mean	Std. Deviation	T	P-value
Business processes	**Yes**	40	7.463	1.997	0.440	0.662
	No	10	7.163	1.619		
Decision making	**Yes**	40	7.770	2.350	0.267	0.791
	No	10	7.552	2.190		
All items	**Yes**	40	7.588	2.085	0.370	0.713
	No	10	7.321	1.838		

Critical value of t at df "48" and significance level 0.05 equal 2.01

H3-There is a statistically significant difference at the level ($\alpha \leq 0.05$) about the impact of data quality on business processes and decision making refer to have a decision making role

The Independent Samples Test is used to test the hypothesis and the results is illustrated in Table (5.34 Independent Samples Test for differences about the impact of data quality on business processes and decision making refer to have a decision making role) show that the p-value equal 0.038 which is less than 0.05 and the absolute value of T test equal 2.135 which is greater than the critical value which is equal 2.01. This means that having a decision making role in the organization has an impact on the results of the questions regarding the impact of data quality on business processes and decision making at insurance companies in Gaza Strip.

Table 5.34: Independent Samples Test for differences about the impact of data quality on business processes and decision making refer to have a decision making role

Field	Having a decision making role	N	Mean	Std. Deviation	T	P-value
Business processes	**Yes**	33	7.818	1.677	2.215	0.032
	No	17	6.598	2.139		
Decision making	**Yes**	33	8.163	1.950	1.919	0.061
	No	17	6.881	2.724		
All items	**Yes**	33	7.958	1.714	2.135	0.038
	No	17	6.713	2.361		

Critical value of t at df "48" and significance level 0.05 equal 2.01

H4- There is a statistically significant difference at the level ($\alpha \leq 0.05$) about the impact of data quality on business processes and decision refer to age group

The one-way ANOVA test was used and the results is illustrated in Table (5.35 One-way ANOVA test for differences about the impact of data quality on business processes and decision making refer to age group) show that the p-value equal 0.055 which is greater than 0.05, and the value of F_{stat}=2.721 which is less than $F_{critical}$=2.81. This means age has no effect on the results of the questions regarding the impact of data quality on business processes and decision making in insurance companies in Gaza Strip.

Table 5. 35: One way ANOVA test for differences about the impact of data quality on business processes and decision making refer to age group

Field	Source	Sum of Squares	df	Mean Square	F value	Sig.(P-Value)
Business processes	**Between Groups**	26.450	3	8.817	2.643	0.060
	Within Groups	153.451	46	3.336		
	Total	179.901	49			
Decision making	**Between Groups**	39.240	3	13.080	2.740	0.054
	Within Groups	219.611	46	4.774		
	Total	258.851	49			
All items	**Between Groups**	30.231	3	10.077	2.721	0.055
	Within Groups	170.344	46	3.703		
	Total	200.575	49			

Critical value of F at df "3,46" and significance level 0.05 equal 2.81

H5- There is a statistically significant difference at the level ($\alpha \leq 0.05$) about the impact of data quality on business processes and decision refer to academic qualification

The one-way ANOVA test was used and the results is illustrated in Table (5.36 One-way ANOVA test for differences about the impact of data quality on business processes and decision making refer to academic qualification) show that the p-value equal 0.297 which is greater than 0.05, and the value of F_{stat} =1.246 which is less than $F_{critical}$ =3.20. This means that academic qualification has no effect on the results of the questions regarding the impact of data quality on business processes and decision making in insurance companies in Gaza Strip.

Table 5.36: One way ANOVA test for differences about the impact of data quality on business processes and decision making refer to academic qualification

Field	Source	Sum of Squares	df	Mean Square	F value	Sig.(P-Value)
Business processes	**Between Groups**	9.299	2	4.650	1.281	0.287

	Within Groups	170.602	47	3.630		
	Total	179.901	49			
Decision making	**Between Groups**	11.330	2	5.665	1.076	0.349
	Within Groups	247.522	47	5.266		
	Total	258.851	49			
All items	**Between Groups**	10.099	2	5.049	1.246	0.297
	Within Groups	190.476	47	4.053		
	Total	200.575	49			

Critical value of F at df "2,47" and significance level 0.05 equal 3.20

H6- There is a statistically significant difference at the level ($\alpha \leq 0.05$) about the impact of data quality on business processes and decision refer to employment level

The one-way ANOVA test was used and the results is illustrated in Table (5.37 One way ANOVA test for differences about the impact of data quality on business processes and decision making refer to employment level) show that the p-value equal 0.682 which is greater than 0.05, and the value of F_{stat} =0.503 which is less than $F_{critical}$ =2.81. This means that employment has no effect on the results of the questions regarding the impact of data quality on business processes and decision making in insurance companies in Gaza Strip.

Table 5.37: One way ANOVA test for differences about the impact of data quality on business processes and decision making refer to employment level

Field	Source	Sum of Squares	df	Mean Square	F value	Sig.(P-Value)
Business processes	**Between Groups**	3.833	3	1.278	0.334	0.801
	Within Groups	176.068	46	3.828		
	Total	179.901	49			
Decision making	**Between Groups**	11.544	3	3.848	0.716	0.548
	Within Groups	247.307	46	5.376		
	Total	258.851	49			

All items	**Between Groups**	6.368	3	2.123	0.503	0.682
	Within Groups	194.207	46	4.222		
	Total	200.575	49			

Critical value of F at df "3,46" and significance level 0.05 equal 2.81

H7- There is a statistically significant difference at the level ($\alpha \leq 0.05$) about the impact of data quality on business processes and decision making refer to work experience

The one-way ANOVA test was used and the results is illustrated in Table (5.38 One-way ANOVA test for differences about the impact of data quality on business processes and decision making refer to work experience) show that the p-value equal 0.545 which is greater than 0.05, and the value of F_{stat}=0.889 which is less than $F_{critical}$=2.81. This means that work experience has an impact on the results of the questions regarding the impact of data quality on business processes and decision making at insurance companies in Gaza Strip because high experienced employees know the importance of data quality and are more committed to it.

Table 5.38: One way ANOVA test for differences about the impact of data quality on business processes and decision making refer to work experience

Field	Source	Sum of Squares	df	Mean Square	F value	Sig.(P-Value)
Business processes	**Between Groups**	13.417	3	4.472	1.236	0.308
	Within Groups	166.484	46	3.619		
	Total	179.901	49			
Decision making	**Between Groups**	14.184	3	4.728	0.889	0.454
	Within Groups	244.667	46	5.319		
	Total	258.851	49			
All items	**Between Groups**	13.527	3	4.509	1.109	0.355
	Within Groups	187.047	46	4.066		
	Total	200.575	49			

Critical value of F at df "3,46" and significance level 0.05 equal 2.81

5.4.2.4 Multiple linear regression

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable *x* is associated with a value of the dependent variable *y*. The population regression line for *p* explanatory variables x_1, x_2,..., x_p is defined to be $\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$. This line describes how the mean response μ_y changes with the explanatory variables. The observed values for *y* vary about their means μ_y and are assumed to have the same standard deviation σ. The fitted values b_0, b_1, ..., b_p estimate the parameters β_0, β_1, ..., β_p of the population regression line.

Since the observed values for *y* vary about their means μ_y, the multiple regression model includes a term for this variation. In words, the model is expressed as DATA = FIT + RESIDUAL, where the "FIT" term represents the expression $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_p x_p$. The "RESIDUAL" term represents the deviations of the observed values *y* from their means μ_y, which are normally distributed with mean 0 and variance σ. The notation for the model deviations is ε.

Formally, the model for multiple linear regression, given *n* observations, is $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... \beta_p x_{ip} + \varepsilon_i$ **for *i* = 1,2, ... *n*.**

In the least-squares model, the best-fitting line for the observed data is calculated by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the deviations are first squared, then summed, there are no cancellations between positive and negative values. The least-squares estimates b_0, b_1, ... b_p are usually computed by statistical software such EViews.

Model no. (1) the dependent variable Business Processes:

Our model is multiple linear regression, consider Y: = **Business Processes** as the response variable, and x1:= Data completeness, x2:= Data accuracy, x3:= Data consistency, x4:= Validity of data, x5:= Data relevance as explanatory variables and the results shown below by using **EViews program**.

The multiple linear regression is

Business Processes = 0.471 - 0.2963 * Data completeness + 0.675 * Data accuracy - 0.288 * Data consistency + 0.569 * Validity of data + 0.192 * Data relevance

Table 5. 39: Regression model for dependent variable Business Processes

Dependent Variable: **Business Processes**				
Method: Least Squares				
Sample: 1 50				
Included observations: 50				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.471786	0.799378	0.590192	0.5581
Data completeness	-0.296907	0.370299	-0.801802	0.4270
Data accuracy	0.675115	0.417717	1.616204	0.1132
Data consistency	-0.288771	0.330289	-0.874298	0.3867
Validity of data	0.569689	0.240094	2.372776	0.0221
Data relevance	0.192274	0.302241	0.636163	0.5280
R-squared	0.629748	Mean dependent var		6.699800
Adjusted R-squared	0.587674	S.D. dependent var		2.027727
S.E. of regression	1.302056	Akaike info criterion		3.477933
Sum squared resid	74.59544	Schwarz criterion		3.707376
Log likelihood	-80.94834	Hannan-Quinn criter.		3.565307
F-statistic	14.96760	Durbin-Watson stat		1.927391
Prob(F-statistic)	0.000000			

The output shows that F=14.96 (P-value= 0.000 < 0.05), indicating that the null hypothesis that the explanatory variables have an effect on response variable ($\beta_1 = \beta_2 = \ldots + \beta_5 = 0$) at the 5% level of significance should be rejected. The results also show that the variable x4:= Validity of data is significant since T=2.372, (P-Value =0.0221< 0.05), but the other explanatory variables are not significant since the P-Value for each greater than 0.05. In addition, the output also shows that Adjusted R-squared=0.5876, and the scatter plot (Figure 5.2: Business processes scatter plot) below show the simple linear regression of explanatory variables to each response variable.

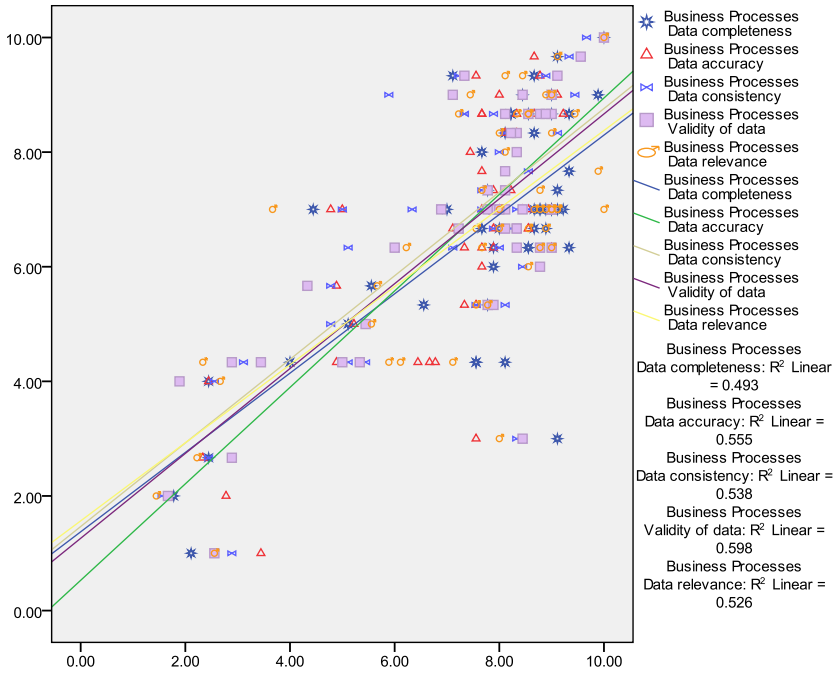

Figure 5.2: Business processes scatter plot

Model no. (2) the dependent variable Decision making:

Our model is multiple linear regression, consider Y:=**Decision making** as the response variable, and x1:=Data completeness, x2:=Data accuracy, x3:=Data consistency, x4:=Validity of data, x5:=Data relevance as explanatory variables and the results shown below by using Eviews program.

The multiple linear regression is

Decision making = 0.106 + 0.2973* Data completeness + 0.3631* Data accuracy + 0.033* Data consistency + 0.288* Validity of data - 0.063* Data relevance

Table 5. 40: regression model for dependent variable Decision making

Dependent Variable: **Decision making**				
Method: Least Squares				
Sample: 1 50				
Included observations: 50				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.106468	0.636225	0.167344	0.8679
Data completeness	0.297603	0.204634	3.454317	0.0041
Data accuracy	0.363469	0.308345	1.178774	0.2448

Data consistency	0.033181	0.217177	0.152785	0.8793
Validity of data	0.288367	0.253815	1.136134	0.2621
Data relevance	-0.063721	0.223846	-0.284663	0.7772
R-squared	0.793159	Mean dependent var		7.367000
Adjusted R-squared	0.769654	S.D. dependent var		2.390217
S.E. of regression	1.147169	Akaike info criterion		3.224639
Sum squared resid	57.90389	Schwarz criterion		3.454081
Log likelihood	-74.61597	Hannan-Quinn criter.		3.312012
F-statistic	33.74471	Durbin-Watson stat		1.247382
Prob(F-statistic)	0.000000			

The output shows that F=33.74471 (P-value = 0.000 < 0.05), indicating that the null hypothesis that the explanatory variables have an effect on response variable($\beta_1 = \beta_2 = \ldots + \beta_5 = 0$) at the 5% level of significance should be rejected. The results also show that the variable x1: = Data completeness is significant since T=3.454, (P-Value = 0.0041< 0.05), but the other explanatory variables are not significant since the P-Value for each greater than 0.05) In addition, the output also shows that Adjusted R-squared=0.769, and the scatter plot (Figure 5.3: Decision making scatter plot) below show the simple linear regression of explanatory variables to each response variable.

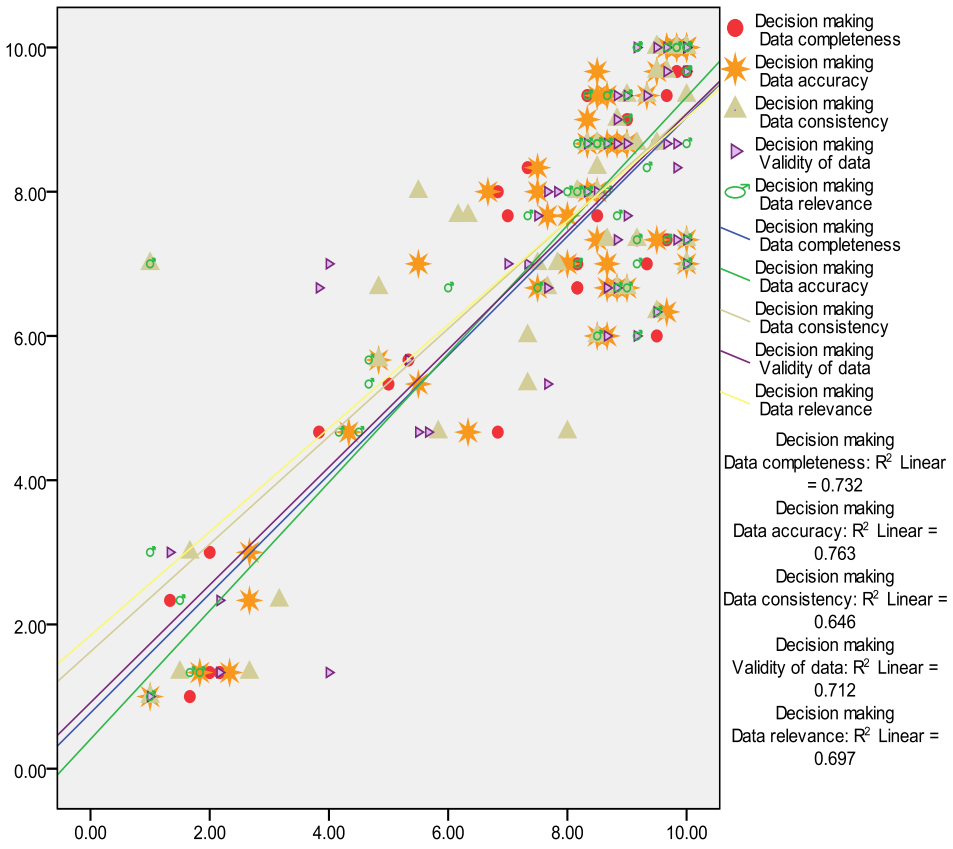

Figure 5.3: Decision making scatter plot

Chapter 6

The Research Conclusions and Recommendations

6.1 Introduction

6.2 Conclusion

6.3 Recommendations

6.4 Suggested Topics for Future Studies

6.1 Introduction

This research aimed to evaluate the impact of data quality and to highlight its importance to the organizational successes. In this chapter the conclusion will be illustrated based on the previous results of the research. Then a recommendation list is presented and in the end, some suggested future studies are presented.

6.2 Conclusion

This research aimed to answer the question "to what degree does data quality affect business processes and decision making in insurance companies in Gaza Strip?" which highlight the important impact of data quality to the organization's processes in general and to business processes and decision making in specific. As the results have showed from the quantitative part of this research, data quality has a significant impact on the insurance companies in Gaza Strip.

And from the results of this research, the objectives of the research were effectively achieved through introducing to the first time the concept of data quality in business and by identifying data quality importance and its impact on the insurance companies in Gaza Strip. Also by elaborating on the data quality dimensions that affect insurance companies. In addition, it has shown that insurance companies in Gaza Strip has already put some effort on obtaining high data quality to assure its operations running smoothly and decisions are made efficiently. And finally insuring the effect of data quality on business processes and decision making which has been proved by the quantitative part of this research.

The research has shown and proved the importance of data quality and its dimensions to successes of the daily business processes and decision making. The quantitative tool used (the questionnaire) has presented five variables which are actually five dimensions of data quality. These dimensions are completeness, accuracy, consistency, data validity and relevance. The questionnaire has proved the importance of these variables to the success of the organization and showed the level of knowledge that insurance companies have on data quality and how these companies are interested in more effective decision making and more efficient business processes.

1- Business processes

Data quality is important to business processes in insurance companies in Gaza Strip which is proved by the quantitative part of this research.

- Data completeness: insurance companies make sure data used in its daily operations are complete and when there is a problem in data completeness employees tries to solve the problem by looking to its sources and balance the effect between completing the work or stop working. The quality of business processes increases when the data used is complete.

- Data accuracy: the results show that insurance companies know the importance of data accuracy and the fact that that the more the data is accurate the more the quality of business processes and work to obtain this high data accuracy.
- Data consistency: the results shows that data consistency is important to the organization business processes and that insurance companies tries to obtain high level of consistency to all the purposes of its operations. Employees responded to that when they face a problem data consistency they report the problem and tries to find its source to assure its consistency.
- Validity of data: the results show that insurance companies work to obtain high data validity and the employees are aware of the importance of data validity. The lack of data validity does affect the quality of daily business processes.
- Data relevance: the results show that insurance companies know the importance of data relevance and work to attain this relevancy to all its business operations. When employees face a problem in data relevancy they look for its source and report the problem.

2- Decision making

Data quality is important to decision making process at insurance companies in Gaza Strip which is proved by the quantitative part of this research.

- Data completeness: the results show that data completeness affect decision making in organizations and that organizations work for achieving high data completeness for all staff. Respondents have agreed that reaching a proper decision is difficult when the data is incomplete and the data to be complete to one decision is not necessarily important for making other decisions.
- Data accuracy: the results show that organizations work on achieving the highest level of data accuracy for all its employees in their different jobs and that it's a group effort to attain high data accuracy. Respondents have agreed that when there is inaccurate data they try to improve it to facilitate the process of decision making.
- Data consistency: the results show that data consistency affects the decision making process in the organization and organizations work on achieving the highest level of data consistency for all its employees in their different jobs. Also having consistent data is a group effort that affects all staff in the organization.
- Validity of data: the results show that reaching proper decision needs valid data and that this processes is a group effort. And organizations work on achieving the highest level of data validity for all its employees in their different jobs.

- Data relevance: results show that data relevance affect decision making on organizations and that organizations do work to achieve high data relevancy to its employees. Respondents agreed that when there is irrelevant data they try to improve data to facilitate the process of decision making.

6.3 Recommendations

Based on the findings, the following recommendations are suggested:

1- Form these results organizations should give more attention to the data used in its processes and decisions.

2- Organizations should have clear model on how to improve data quality to be used in its daily business processes and decision making.

3- Sessions should be made to the employees to increase the awareness of the importance of high data quality and to raise the commitment to achieving such high data.

4- It's necessary for organizations to facilitate the processes of attaining high data quality to employees and to give more privileges to seek this data.

5- Organizations should put guidelines to reach high data quality and develop procedures to help in rising the quality of its data. The guidelines should clear the minimum level of data quality that should be presented to proceed in its operations.

6- Organizations should focus on the feedback of the employees and the clients on the quality of data provided and the quality of its operations and decisions which is directly affected by the level of data quality.

7- Organizations should measure the costs of data quality and its consequences to the business.

8- Organizations should create a job position or a division (based on its size) to work on monitoring and improving the data quality.

9- The use of new technologies and IT to measure the quality of data.

6.4 Suggested Topics for Future Studies

Some suggested topics for future studies in the field of data quality are:

1- Data quality assessment and evaluation methodology
2- An empirical study on some of data quality existing assessment methodologies
3- Creating a model for measuring data quality costs
4- A study on the public sector on the level of data quality and how to improve it
5- The role of IT on attaining high data quality
6- A framework for data quality assessment
7- The barriers to have high data quality

References

Arabic references:

- Ahlia Insurance Group. (2014). *Final Annual Report.* Ramallah: Ahlia Insurance Group.

- Alsaidi, M., & Zbar, S. A. (2013). Quality of information and its impact on strategic decision: A field study in a sample of Iraqi banks. *AL-Qadisiya Journal For Administrative and Economic sciences, Vo.15, issue 3*, pp. 7-41.

- Palestine Insurance Company (PIC). (2013). *Final Annual Report.* Ramallah: Palestine Insurance Company (PIC) .

Foreign references:

- Askham, N., Cook, D., Doyle, M., Fereday, H., Gibson, M., Landbeck, U., . . . Schwarzenbach, J. (2013). *The six primary dimensions for data quality assessment - defining data quality dimensions.* DAMA UK Working Group.

- Batini, C., & Scannapieca, M. (2006). Data Quality Dimensions. In *Data Quality Concepts, Methodologies and Techniques* (pp. 19-49). Springer.

- Batini, C., & Scannapieca, M. (2006). Introduction to Data Quality. In *Data Quality Concepts, Methodologies and Techniques* (pp. 1-18). Springer.

- Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *Journal of ACM Computing Surveys, Vol. 41, No. 3, Article 16.*

- Chapman, A. D. (2005). *Principles of Data Quality.* Global Biodiversity Information Facility.

- Chen, H., Hailey, D., Wang, N., & Yu, P. (2014). A Review of Data Quality Assessment Methods for Public Health Information Systems. *International Journal of Environmental Research and Public Health, 11(5)*, pp. 5170-5207.

- Chiang, F., & Miller, R. J. (2008). Discovering Data Quality Rules. *Journal Proceedings of the VLDB Endowment, Vol 1 Iss 1*, pp. 1166-1177.

- Du, J., & Zhou, L. (2012). Improving financial data quality using ontologies. *Elsevier B.V., Decision Support Systems 54 (2012)*, pp. 76–86.

- Ehling, M., Körner, T., Bergdahl, M., Elvers, E., Földesi, E., Kron, A., . . . Zilhão, M. J. (2007). *Handbook on Data Quality Assessment Methods and Tools.* European Commission, EUROSTAT.

- Eppler, M. J., & Helfert, M. (2004). A classification and analysis of data quality costs. *International Conference on Information Quality (ICIQ-04)*, (pp. 311-325).

- Fan, W., Geerts, F., Ma, S., Tang, N., & Yu, W. (2013). Data Quality Problems beyond Consistency and Deduplication. In *In Search of Elegance in the Theory and Practice of Computation* (pp. 237-249). Springer-Verlag Berlin Heidelberg.

- Friedman, T., & Smith, M. (2011). *Measuring the Business Value of Data Quality.* Gartner, Inc., G00218962.

- Geiger, J. G. (2004). Data Quality Management The Most Critical Initiative You Can Implement. *Proceedings of the Twenty-Ninth Annual SAS® Users Group International Conference.* SAS Institute Inc.

- Gharib, M., & Giorgini, P. (2015). Modeling and Reasoning About Information Quality Requirements. In *Requirements Engineering: Foundation for Software Quality* (pp. 49-64). Springer International Publishing.

- Haug, A., & Arlbjørn, J. S. (2011). Barriers to master data quality. *Emerald Group Publishing Limited, Journal of Enterprise Information Management, Vol. 24 No. 3, 2011*, pp. 288 - 303.

- Haug, A., Zachariassen, F., & van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management, 4(2)*, pp. 168-193.

- Herzog, T. N., Scheuren, F. J., & Winkler, W. E. (2007). What is Data Quality and Why Should We Care? In T. N. Herzog, F. J. Scheuren, & W. E. Winkler, *Data Quality and Record Linkage Techniques* (pp. 7-15).

- Loshin, D. (2006). *The data quality business Case: projecting return on investment.* Informatica Corporation.

- Loshin, D. (2011). *Evaluating the Business Impacts of Poor Data Quality.* Knowledge Integrity Incorporated. IAIDQ Publications.

- McGilvray, D. (2008). *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information, Overview of The Ten Steps Process.* Elsevier Inc.

- McGilvray, D. (2011). *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information, Data Quality Dimensions.* Elsevier Inc.

- McGilvray, D., & Thomas, G. (2008). *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information, Definitions of Data Categories.* Elsevier Inc.

- Mendes, R., & Rodrigues, P. P. (2011). Main barriers for quality data collection in EHR. *Vicente Traver; Ana L. N. Fred; Joaquim Filipe & Hugo Gamboa, ed., 'HEALTHINF' , SciTePress,*, pp. 451-454

- Moges, H.-T., Dejaeger, K., Lemahieu, W., & Baesens, B. (2013). A multidimensional analysis of data quality for credit risk management: New insights and challenges. *Elsevier B.V., Information & Management 50(2013)*, pp. 43–58.

- Neely, M. (2005). The product approach to data quality and fitness for use: framework for analysis. *Proceedings of the Tenth International Conference on Information Quality (ICIQ-05).*

- Ofner, M. H., Otto, B., & Osterle, H. (2012). Integrating a data quality perspective into business process management. *Emerald Group Publishing Limited, Business Process Management Journal, Vol. 18 No. 6, 2012*, pp. 1036-1067.

- Oliveira, P., Rodrigues, F., & Henriques, P. (2005). *A Formal definition of data quality problems.* In F. Naumann, M. Gertz & S. E. Madnick (eds.), IQ, : MIT.

- Piprani, B., & Ernst, D. (2008). A Model for Data Quality Assessment. In *On the Move to Meaningful Internet Systems: OTM 2008 Workshops* (pp. 750-759). Springer Berlin Heidelberg.

- Sattler, K.-U. (2009). Data Quality Dimensions. In *Encyclopedia of Database Systems* (pp. 612-615). Springer US.

- Scannapieco, M., Missier, P., & Batini, C. (2005). Data Quality at a Glance. *Datenbank-Spektrum (14)*.

- Shankaranarayanan, G., & Cai, Y. (2006). Supporting data quality management in decision-making. *Elsevier B.V., Decision Support Systems 42 (2006)*, pp. 302– 317.

- Storey, V. C., Dewan, R. M., & Freimer, M. (2012). Data quality: Setting organizational policies. *Elsevier B.V., Decision Support Systems 54 (2012)*, pp. 434–442.

- Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A Framework for Information Quality Assessment. *JASIST, 58(12)*, pp. 1720-1733.

- Tejay, G., Dhillon, G., & Chin, A. G. (2006). Data Quality Dimensions for Information Systems Security: A Theoretical Exposition. In *Security Management, Integrity, and Internal Control in Information Systems* (pp. 21-39). Springer US.

Websites:

- Ahlia Insurance Group. (2013). *Ahlia Insurance Group*. Retrieved April 18, 2015, from Ahlia Insurance Group website: http://www.ahlia.ps/

- AlMultazem Insurance & Investment Company. (2012). *AlMultazem Insurance & Investment Company*. Retrieved May 2, 2015, from AlMultazem Insurance & Investment Company website: http://mii.ps/index.php

- *Business Dictionary*. (2015). Retrieved March 9, 2015, from Business Dictionary website: http://www.businessdictionary.com/

- Execution-MiH. (2015). *Execution-MiH*. Retrieved April 26, 2015, from Execution-MiH, enterprise development and growth engine website: http://www.executionmih.com/data-quality/accuracy-consistency-audit.php

- Global United Insurance Company. (2013). *Global United Insurance Company*. Retrieved April 29, 2015, from Global United Insurance Company website: http://gui.ps/index/portal

- *Google*. (2015, April 14). Retrieved April 14, 2015, from Google web search engine: https://www.google.com/webhp?sourceid=chrome-instant&rlz=1C1FDUM_enPS606PS606&ion=1&espv=2&ie=UTF-8#q=%22data+quality%22

- *IAIDQ*. (2015). Retrieved March 20, 2015, from International Association for Information and Data Quality: http://iaidq.org/main/glossary.shtml#I

- Natioanl Insurance Company Ltd. (2014). *Final Annual Report.* Ramallah: Natioanl Insurance Company Ltd. Retrieved April 23, 2015, from Natioanl Insurance Company website: http://www.nic-pal.com/index.aspx

- Natioanl Insurance Company Ltd. (2014). *Final Annual Report.* Retrieved April 23, 2015, from Natioanl Insurance Company website: http://www.nic-pal.com/index.aspx

- National Insurance Company Ltd. (2015). *National Insurance Company Ltd.* Retrieved April 23, 2015, from National Insurance Company website: http://www.nic-pal.com/edefault1.aspx?id=145&catid=

- O'Neal, E. (2014). *The Importance of Data Quality Management to UDI Regulation and Compliance*. Retrieved April 27, 2015, from Information Management website: http://www.information-management.com/news/the-importance-of-data-quality-management-to-udi-regulation-and-compliance-10025309-1.html

- *Oxford Dictionaries*. (2015). Retrieved February 16, 2015, from Oxford dictionaries website: http://www.oxforddictionaries.com/definition/english/data

- Trust International Insurance Palestine. (2015). *Trust International Insurance Palestine*. Retrieved April 23, 2015, from Trust International Insurance Palestine website: http://www.trustpalestine.com/index.php

- *Wikipedia*. (2015). Retrieved March 20, 2015, from Wikipedia: http://en.wikipedia.org/wiki/Data_quality

Annexes

Annex (1): List of Referees and their Job Title

S/No.	Referee name	Job title
1	Dr. Wasim Ismail Al-Habil	Associate Professor at the Faculty of Commerce – The Islamic University of Gaza and Head of Business Administration Department
2	Dr. Khalid Abd Dahleez	Assistant Professor at the Faculty of Commerce – The Islamic University of Gaza
3	Dr. Akram Sammour	Assistant Professor at the Faculty of Commerce – The Islamic University of Gaza
4	Dr. Yousif Abd Baher	Associate Professor at the Faculty of Commerce – The Islamic University of Gaza
5	Dr. Nafiz Mohd Barakat	Assistant Professorat the Faculty of Commerce – The Islamic University of Gaza
6	Dr. Mohammed Fares	Assistant Professor at the Faculty of Economics and Administrative Sciences– AlAzhar University
8	Dr. Nehaya Telbani	Associate Professor at the Faculty of Economics and Administrative Sciences– AlAzhar University
7	Dr. Samy Abu Naser	Professor at the Faculty of Engineering and Information Technology - AlAzhar University
9	Dr. Nedal Abdulla	Assistant Professor at the Faculty Administration and Economic – AlAqsa University

Annex (2): English Questionnaire

Islamic University - Gaza Higher Studies Deanship

Commerce Faculty – Business Administration Department

Dear Employees; Greetings;

This questionnaire is a tool for a study to complete master degree requirements in business administration titled "The impact of data quality on business processes and decision making – case study insurance companies in Gaza Strip". The study aims at measuring the impact of data quality on business processes and its success in achieving better levels in accuracy and professionalism in decision making process.

Please answer the following questions using (X) mark in the designated spaces that represent your choice keeping in mind that "1" represents completely disagree, while "10" represents completely agree. Please note that all data collected from you will be used for scientific research purposes only.

Thanks for your cooperation

Researcher

Hasan AL-Talla

Part one:

Personal information

Gender	☐	Male	☐	Female			
Age group	☐	Less than 25	☐	From 25 to 34 year	☐	From 35 to 44 year	☐	More than 45 year
Academic	☐	Diploma	☐	Bachelor	☐	Master	☐	PhD
Employment level	☐	Department head	☐	Vice branch manager	☐	Branch manager	☐	Employee
Work experience	☐	Less than 5 years	☐	From 5 to 9 years	☐	From 10 to 14 year	☐	More than 15 years
Do you have former knowledge of "data quality"?	☐	Yes	☐	No				
Are you a decision maker?	☐	Yes	☐	No				

Part two: Questionnaire:

#	Statement	1	2	3	4	5	6	7	8	9	10
	First: Business Processes										
	Data completeness										
1.	Data used in the daily business processes is complete for its purpose of use										
2.	When there is a problem in data completeness, you look for its source to ensure its completeness										
3.	When there is a problem in data completeness, you report the problem										
4.	When there is a problem in data completeness, you continue working without considering the problem										
5.	The organization works on improving quality of data used for all organizational purposes										
6.	The lack of data completeness causes problems to daily business processes										
7.	Where there is a problem in data completeness, you stop working till the problem is solved										
8.	The lack of data completeness affects the daily business processes										
9.	Data completeness is important for daily business process in the organization										
10.	The more used data is complete, the more the quality of business process.										
	Data accuracy										
11.	The data used in the daily business processes are accurate for its purpose of use										
12.	When there is a problem in the data accuracy you look for its source to ensure its accuracy										
13.	When there is a problem in data accuracy, you report the problem										
14.	When there is a problem in the data accuracy you continue working without considering the problem										
15.	The lack in data accuracy causes problems to daily business processes										
16.	When there is a problem in the data accuracy, you stop working until the problem is solved										
17.	The lack of data accuracy affects the daily business										
18.	Data accuracy is important for daily business process in the organization										
19.	The more used data is accurate, the more the quality of business process.										
	Data consistency										

20.	The data used in daily business processes are consistent for its purpose of use										
21.	When there is a problem in the data consistency, you look for its source to ensure its consistency										
22.	When there is a problem in the data consistency, you report the problem										
23.	When there is a problem in the data consistency, you continue working without considering the problem										
#	Statement	1	2	3	4	5	6	7	8	9	10
24.	The lack of data consistency causes problems to daily business processes										
25.	When there is a problem in the data consistency, you stop working until the problem is solved										
26.	The lack of data consistency affects the daily business processes										
27.	Data consistency is important for daily business process in the organization										
28.	The more used data is consistent, the more the quality of business process.										
	Validity of data										
29.	The data used in the daily business processes are valid for its purpose of use										
30.	When there is a problem in the data validity , you look for its source to ensure its validity										
31.	When there is a problem in the data validity, you report the problem										
32.	Poor data quality affects business processes decision										
33.	When there is a problem in the data validity, you continue working without considering the problem										
34.	The lack of data validity causes problems to daily business processes										
35.	When there is a problem in the data validity, you stop working until the problem is solved										
36.	Obtaining high data quality is not important to the organization's success										
37.	The lack of data validity affects the daily business										
38.	The data validity is important for daily business process in the organization										
39.	The more used data is valid, the more the quality of business process.										
	Data relevance										
40.	The data used in the daily business processes are relevance for its purpose of use										

41.	When there is a problem in the data relevancy, you look for its source to ensure its relevancy										
42.	When there is a problem in the data relevancy, you report the problem										
43.	When there is a problem in the data relevancy, you continue working without considering the problem										
44.	The lack of data relevancy causes problems to daily business processes										
45.	When there is a problem in the data relevancy, you stop working until the problem is solved										
46.	The lack of data relevancy affects the daily business										
47.	The data relevancy is important for daily business process in the organization										
48.	The more used data is relevant, the more the quality of business process.										
#	Statement	1	2	3	4	5	6	7	8	9	10
	Second: Decision making										
	Data completeness										
49.	Data completeness affects the decision making process in the organization										
50.	It is difficult to reach a proper decision when the data is incomplete										
51.	When there is missing data, you try to collect the missing data to facilitate the process of decision										
52.	The organization works on improving quality of data used for all organizational purposes										
53.	Having complete data is a group effort that affect all staff in the organization										
54.	A data that is complete for making one decision is not necessarily complete for another										
55.	The organization works on achieving the highest level of data completeness for all its employees in there										
	Data accuracy										
56.	Data accuracy affects the decision making process in the organization										
57.	It is difficult to reach a proper decision when the data is inaccurate										
58.	When there is inaccurate data, you try to improve data to facilitate the process of decision making										
59.	Having accurate data is a group effort that affect all staff in the organization										
60.	Improving data quality leads to raising the value of the whole organization										

61.	A data that is accurate for making one decision is not necessarily accurate for another										
62.	The organization works on achieving the highest level of data accuracy for all its employees in there										
	Data consistency										
63.	Data consistency affects the decision making process in the organization										
64.	It is difficult to reach a proper decision when the data is inconsistence										
65.	When there is inconsistent data, you try to improve data to facilitate the process of decision making										
66.	Having consistent data is a group effort that affect all staff in the organization										
67.	A data that is consistent for making one decision is not necessarily consistent for another										
68.	The organization works on achieving the highest level of data consistency for all its employees in there										
	Validity of data										
69.	Data validity affects the decision making process in the organization										
70.	It is difficult to reach a proper decision when the data is not valid										
71.	When there is invalid data, you try to improve data to										
#	Statement										
	the process of decision making										
72.	Having valid data is a group effort that affect all staff in the organization										
73.	A data that is valid for making one decision is not necessarily valid for another										
74.	The organization works on achieving the highest level of data validity for all its employees in there different										
	Data relevance										
75.	Data relevancy affects the decision making process in the organization										
76.	It is difficult to reach a proper decision when the data is not relevant										
77.	When there is irrelevant data, you try to improve data to facilitate the process of decision making										
78.	Having relevant data is a group effort that affect all staff in the organization										
79.	A data that is relevant for making one decision is not necessarily valid for another										
80.	The organization works on achieving the highest level of data relevancy for all its employees in there										

81.	Obtaining high data quality is not important to the organization’s success										

Annex (3): Arabic Questionnaire

الجامعة الاسلامية غزة

عمادة الدراسات العليا

كلية التجارة – قسم ادارة الاعمال

السادة الموظفين الكرام

تحية طيبة وبعد...

الاستبانة التي بين ايديكم هي اداة دراسة لاستكمال متطلبات درجة الماجستير في ادارة الاعمال, وعنوانها " تأثيرجودة البيانات على العمليات الادارية واتخاذالقرار – دراسة حالة شركات التأمين في قطاع غزة" والتي تهدف لقياس مدى تأثيرجودة البيانات المستخدمة على عمليات المؤسسة ونجاحها والوصول لأفضل درجات الدقة والاحترافية في اتخاذ القرارات.

لذا نرجو من سيادتكم التكرم بالإجابة على الاسئلة المرفقة بوضع اشارة (X) في المربع الذي يمثل خياركم مع العلم ان 1 تمثل غير موافق بشدة و10 تمثل موافق بشدة, وأن جميع المعلومات التي سيتم الحصول عليها ستستخدم لأغراض البحث العلمي فقط.

شاكرة لكم حسن تعاونكم

الباحثة

حنان حسن الطلاع

الجزء الاول:

أولا: البيانات الشخصية

الجنس	أنثى		ذكر				
العمر	أقل من 25 سنة		من 25 الى اقل من 35 سنة		من 35 الى اقل من45 سنة		أكبر من 45 سنة
المؤهل العلمي	دبلوم		بكالوريوس		ماجستير		دكتوراة
المستوى الوظيفي	رئيس قسم		نائب مدير فرع		مدير فرع		موظف
عدد سنوات الخبرة	أقل من 5 سنوات		من 5 الى 9 سنوات		من 10 الى 14 سنة		15 سنة فأكثر
هل لديك علم مسبق بمصطلح جودة البيانات	نعم		لا				
هل تعمل في مستوى تنفيذي في الشكرة وتقوم باتخاذ القرار	نعم		لا				

ثانيا: تعريف المصطلحات المستخدمة

العمليات الادارية business processes: هي العمليات الادارية اليومية التي تقوم بها المؤسسة لتحقيق الهدف من وجودها, كتقديم الخدمات للجمهور والعمليات المالية, غيرها.

اتخاذ القرار decision making: عملية اتخاذ قرار لتسيير عمل المؤسسة والوصول الى اهدافها.

اكتمال البيانات completeness: أي ان تكون البيانات كاملة لتنفيذ العملية او اتخاذ القرار المرتبط بها.

دقة البيانات accuracy: أي ان تكون البيانات المستخدمة خالية من الاخطاء الاملائية والكتابية والاختصارات العشوائية.

التناغم consistency: أي ان تكون البيانات موحدة في جميع الأغراض المستخدمة فيها.

صحة البيانات validity of data: أي أن البيانات المستخدمة صحيحة للغرض الذي ستستخدم له.

ذو علاقة relevance: أي أن البيانات المستخدمة لها علاقة مباشرة بالغرض الذي ستستخدم له.

الجزء الثاني: فقرات الاستبانة

#	العبارة	1	2	3	4	5	6	7	8	9	10
	أولا: العمليات الادارية										
	اكتمال البيانات Data completeness										
.1	البيانات المستخدمة في عمليات المؤسسة اليومية تكون كاملة للغرض التي ستستخدم له										
.2	عند وجود مشكلة في اكتمال البيانات تقوم بالبحث عن مصدرها للتأكد من اكتمالها										
.3	عند وجود مشكلة في اكتمال البيانات تقوم بالتبليغ عن وجود مشكلة										
.4	عند وجود مشكلة في اكتمال البيانات تتابع العمل بدون أخذ النقص في البيانات في عين الاعتبار										
.5	تعمل المؤسسة عل بتحسين جودةالبيانات المستخدمة فيها لكافة الاغراض										
.6	عدم اكتمال البيانات يؤدي لحدوث مشاكل في العمليات الادارية اليومية										
.7	عند وجود مشكلة في اكتمال البيانات تتوقف عن العمل لحين حل المشكلة										
.8	عدم اكتمال البيانات يؤثر على سير العمليات الادارية اليومية										
.9	اكتمال البيانات ضروري لسير العمليات الادارية في المؤسسة										
10	كلما كانت البيانات المستخدمة كاملة كلما زادت جودة المخرجات من العمليات الادارية										
	دقة البيانات Data accuracy										
11	البيانات المستخدمة في عمليات المؤسسة اليومية تكون دقيقة للغرض التي ستستخدم له										
12	عند وجود مشكلة في دقة البيانات تقوم بالبحث عن مصدرها للتحقق من دقتها										
13	عند وجود مشكلة في دقة البيانات تقوم بالتبليغ عن وجود مشكلة										
14	عند وجود مشكلة في دقة البيانات تتابع العمل بدون أخذ الخلل في الدقة بعين الاعتبار										
15	عدم دقة البيانات يؤدي لحدوث مشاكل في العمليات الادارية اليومية										
16	عند وجود مشكلة في دقة البيانات تتوقف عن العمل لحين حل المشكلة										
17	عدم دقة البيانات يؤثر على سير العمليات الادارية اليومية										
18	دقة البيانات ضرورية لسير العمليات الادارية في المؤسسة										
19	كلما زادت دقة البيانات المستخدمة زادت جودة العمليات الادارية										
	تناغم البيانات Data consistency										
20	البيانات المستخدمة في عمليات المؤسسة اليومية تكون متناغمة مع الغرض الذي ستستخدم له										
21	عند وجود مشكلة في تناغم البيانات تقوم بالبحث عن مصدرها للتحقق من تناغمها										
22	عند وجود مشكلة في تناغم البيانات تقوم بالتبليغ عن وجود مشكلة										
23	عند وجود مشكلة في تناغم البيانات تتابع العمل بدون أخذ الخلل في التناغم بعين الاعتبار										

24	عدم تناغم البيانات يؤدي لحدوث مشاكل في العمليات الادارية اليومية									
25	عند وجود مشكلة في تناغم البيانات تتوقف عن العمل لحين حل المشكلة									
26	عدم التناغم بين البيانات يؤثر على سير العمليات الادارية اليومية									
27	تناغم البيانات ضروري لسير العمليات الادارية في المؤسسة									
28	كلما زاد تناغم البيانات المستخدمة زادت جودة العمليات الادارية									
	صحة البيانات Validity of data									
29	البيانات المستخدمة في عمليات المؤسسة اليومية تكون صحيحة للغرض الذي ستستخدم له									
30	عند وجود مشكلة في صحة البيانات تقوم بالبحث عن مصدرها للتحقق من صحتها									
31	عند وجود مشكلة في صحة البيانات تقوم بالتبليغ عن وجود مشكلة									
32	تعتقد ان انخفاض جودة البيانات يؤثرعلى عمليات المؤسسة واتخاذ القرار فيها									
33	عند وجود مشكلة في صحة البيانات تتابع العمل بدون أخذ الخلل في صحتها بعين الاعتبار									
34	عدم صحة البيانات يؤدي لحدوث مشاكل في العمليات الادارية اليومية									
35	عند وجود مشكلة في صحة البيانات تتوقف عن العمل لحين حل المشكلة									
36	تعتقد ان العمل على رفع جودة البيانات عملية ثانوية وغير مهمة للمؤسسة									
37	عدم صحة البيانات يؤثر على سير العمليات الادارية اليومية									
38	صحة البيانات ضرورية لسير العمليات الادارية في المؤسسة									
39	كلما كانت البيانات المستخدمة صحيحة زادت جودة العمليات الادارية									
	علاقة البيانات بالغرض المستخدمة له Data relevance									
40	البيانات المستخدمة في عمليات المؤسسة اليومية تكون ذات علاقة بالمحتوى التي ستستخدم له									
41	عند وجود مشكلة في علاقة البيانات بالمحتوى تقوم بالبحث عن مصدرها للتحقق من صحتها									
42	عند وجود مشكلة في علاقة البيانات بالمحتوى تقوم بالتبليغ عن وجود مشكلة									
43	عند وجود مشكلة في علاقة البيانات بالمحتوى تتابع العمل بدون أخذ الخلل بعين الاعتبار									
44	عدم وجود علاقة بين البيانات والمحتوى يؤدي لحدوث مشاكل في العمليات الادارية اليومية									
45	عند وجود مشكلة في علاقة البيانات بالمحتوى تتوقف عن العمل لحين حل المشكلة									
46	عدم وجود علاقة للبيانات بالمحتوى يؤثر على سير العمليات الادارية اليومية									
47	وجود علاقة للبيانات بالمحتوى ضرورية لسير العمليات الادارية في المؤسسة									
48	كلما كانت البيانات ذات علاقة بالمحتوى زادت جودة العمليات الادارية									
	ثانيا: عملية اتخاذ القرار Decision making									
	اكتمال البيانات Data completeness									
49	يؤثر اكتمال البيانات على عملية اتخاذ القرار في المؤسسة									
50	يكون من الصعب الوصول لقرار صحيح عند وجود نقص في البيانات									
51	عند وجود نقص في البيانات تعمل على تجميع البيانات الناقصة لتسهيل عملية اتخاذ القرار									

52	تعمل المؤسسة عل بتحسين جودة البيانات المستخدمة فيها لكافة الاغراض									
53	تعتبر ان الوصول لبيانات كاملة مهمة جماعية تعني كل الافراد العاملين في المؤسسة									
54	تعتبر ان البيانات مكتملة لاتخاذ قرار معين ولكن قد لا تكون مكتملة لاتخاذ قرار اخر									
55	تعمل المؤسسة على الوصول الى اعلى مستوى من اكتمال البيانات لجميع العاملين فيها باختلاف وظائفهم									
	دقة البيانات Data accuracy									
56	تؤثر دقة البيانات على عملية اتخاذ القرار في المؤسسة									
57	يكون من الصعب الوصول لقرار صحيح عند انخفاض دقة البيانات									
58	عند وجود مشكلة في دقة البيانات تعمل على تعديل الاخطاء لتسهيل عملية اتخاذ القرار									
59	تعتبر الوصول لدقة بيانات عالية مهمة جماعية تعني كل الافراد العاملين في المؤسسة									
60	تعتقد ان تحسين جودة البيانات في المؤسسة يرفع من قيمة المؤسسة ككل									
61	ان تكون البيانات دقيقة لاتخاذ قرار معين لا تعني انها دقيقة لاتخاذ قرار اخر									
62	تعمل المؤسسة على الوصول إلى أعلى مستوى دقة البيانات لجميع العاملين فيها باختلاف وظائفهم									
	تناغم البيانات Data consistency									
63	يؤثر تناغم البيانات على عملية اتخاذ القرار في المؤسسة									
64	يكون من الصعب الوصول لقرار صحيح عند وجود مشكلة في تناغم البيانات									
65	عند وجود مشكلة في تناغم البيانات تعمل على تعديل الاخطاء لتسهيل عملية اتخاذ القرار									
66	يعتبر الوصول لدرجة عالية من تناغم البيانات مهمة جماعية تعني كل الافراد العاملين في المؤسسة									
67	أن تكون البيانات متناغمة لاتخاذ قرار معين لا تعني انها متناغمة لاتخاذ قرار اخر									
68	تعمل المؤسسة على الوصول إلى أعلى مستوى في تناغم البيانات لجميع العاملين فيها باختلاف وظائفهم									
	صحة البيانات Validity of data									
69	تؤثر صحة البيانات على عملية اتخاذ القرار في المؤسسة									
70	يكون من الصعب الوصول لقرار صحيح عند وجود مشكلة في صحة البيانات									
71	عند وجود مشكلة في صحة البيانات تعمل على تعديل الاخطاء لتسهيل عملية اتخاذ القرار									
72	يعتبر الوصول لدرجة عالية من صحة البيانات مهمة جماعية تعني كل الافراد العاملين في المؤسسة									
73	أن تكون البيانات صحيحة لاتخاذ قرار معين لا يعني بالضرورة انها صحيحة لاتخاذ قرار اخر									
74	تعمل المؤسسة على الوصول إلى أعلى مستوى من صحة البيانات لجميع العاملين فيها باختلاف وظائفهم									
	علاقة البيانات بالغرض المستخدمة له Data relevance									
75	أن تكون البيانات ذات علاقة بالمحتوى يؤثر على عملية اتخاذ القرار									

76	يكون من الصعب الوصول لقرار صحيح عند وجود مشكلة في علاقة البيانات بالمحتوى									
77	تعمل على تعديل الاخطاء لتسهيل عملية اتخاذ القرار عند وجود مشكلة في علاقة البيانات بالمحتوى									
78	ان تكون البيانات ذات علاقة بالمحتوى مهمة جماعية تعني كل الافراد العاملين في المؤسسة									
79	ان تكون البيانات ذات علاقة باتخاذ قرار معين لا يعني بالضرورة انها ذات علاقة لاتخاذ قرار اخر									
80	تعمل المؤسسة على الوصول إلى أعلى مستوى من ترابط البيانات وعلاقتها بالمحتوى المستخدمة له لجميع العاملين فيها باختلاف وظائفهم									
81	تعتقد ان العمل على رفع جودة البيانات عملية ثانوية وغير مهمة للمؤسسة									