EXPLORING COMMUNICATION SUCCESS FACTORS IN DATA SCIENCE AND ANALYTICS PROJECTS

Author: Deanne Larson

Abstract

This research explored commonly held communication principles about the management of successful data science and analytics projects. These principles include that the project builds upon an alignment with strategic outcomes, projects are fully evaluated before initiation, business objectives for the project are clearly stated, and project goals are clearly established in advance. A questionnaire was distributed to a non-stratified convenience sample to collect viewpoints on these key issues from the perspective of 60 experienced data science and analytics project managers and analysts. A qualitative Likert-scale evaluation of responses showed that operating managers rarely understand how projects align with strategic goals, projects are not fully evaluated in advance, and neither business objectives nor project goals are clear. These findings are significant because they show serious communication gaps between formal project management theory and the application of the theory to data science and analytics projects.

Keywords: analytics, data science, data mining, project management, analytics methodology, communications

Introduction

Data science and analytics is a field that extracts insights from data in various forms (Dhar, 2013). Data science and analytics is an evolution of other data analysis fields including data mining, statistics, and machine learning (Leek, 2013). Organizations use data science and analytics based on the value it brings as this practice has been documented to decrease fraud, reduce customer and product churn, improve target marketing, identify new markets, improve operations, and increase profitability (Sim, 2014). Extracting insights using data science and analytics involves many different steps and skill sets to achieve a result, thus most of these initiatives are completed as projects (Sim, 2014).

Understanding what makes data science and analytics projects a success is a challenge. Sim (2014) provided an initial study on data mining success factors in 2003 and republished the results in 2014 to a wider audience based on the original study. According to Sim (2014), the 2003 study represents a basis for data mining Critical Success Factors (CSFs) that decision makers and researchers can use to increase data mining success. The study was based on 2003 survey data, and Sim outlined that with the changes in the industry, new research should be pursued to better understand success factors. Knowing more about the success factors can also assist in preventing project failure.

Gartner Research reported that the market for data science and analytics will be a top priority for organizations through 2017 (Gartner, 2013). Gartner's research reported that data science and analytics projects had a failure rate of greater than 50% where projects failed to deliver the benefits agreed on at the start of the project. As part of Gartner's report, multiple failure points were listed including not having a clear strategy which defines the expected outcomes.

The general problem is that with the digitized world, more data exists than ever before and organizations are attempting to mine this data for a competitive advantage; however, many organizations struggle as data science and analytics projects deal with transforming data into information to be used in decision-making (Bole, Popovič, Žabkar, Papa, & Jaklič, 2015; Davenport, 2014, 2015; Davenport & Harris, 2013; Sim, 2014). Highly competitive organizations view data science and analytics as a capability to enhance decision-making by discovering new knowledge to gain a competitive advantage (Bole et al., 2015; Davenport, 2014, 2015; Sim, 2014). Understanding what makes a data science and analytics projects successful can assist organizations in developing this as an organizational capability.

Data Mining to Data Science

Data mining is widely considered a subfield of computer science (KDD, 2014). Data mining is defined as an interdisciplinary field involving artificial intelligence, statistics, databases, and machine learning (Sim, 2014). Data science and analytics emerged as the next generation of data mining due to the increased need to use analytical abilities to find, extract, and interpret large amounts of data. Data science and analytics is different than data mining due to the need to: manage large disparate sets of data across hardware and software constraints; ensure consistency of data; wrangle data sets; merge data sets; use visualization to understand data; incorporate statistics and algorithms using data; and present findings. Data scientists are expected to produce results in days versus months by completing exploration and iterative analysis to produce results in business context (Davenport, 2014).

The emergence of data science can be traced back to the growth of the field of statistics (Cleveland, 2001). Cleveland introduced the field of "data science" and outlined it as a new discipline using computer science and data mining. According to Cleveland, data science promotes innovation and statisticians partner with computer scientists to promote advances in computing with data. To reinforce the emergence of data science as a field, the *Data Science Journal* (https://datascience.codata.org/) was launched in April 2002 to publish research on the use of databases and data systems in science and technology. *The Journal of Data Science* (www.jds-online.com) focuses on the application of statistical methods and addresses all aspects of data use include collecting, analyzing, modeling, and applying.

The addition of "analytics" to the field of data science emerged in 2005 in a report from the Babson College Working Knowledge Research Center (Davenport, Cohen, & Jacobson, 2005). In the report, analytics was described as a new form of competition focusing on the use of data and fact-based decision-making. Organizations are employing statistics, quantitative analysis, and prediction in place of competing in traditional areas.

Yau (2009) addressed the rise of data science and analytics which combined different areas of expertise including math, statistics, data mining, computer science, and information visualization. Yau highlighted that the role of the data scientist publically emerged in 2009. Loukides (2010) further expanded on data science and analytics to include an aspect of entrepreneurship and the ability to iterate to create a data product. Data science and analytics includes defining a problem, collecting data, and conditioning to draw a conclusion. Mason and Wiggins (2010) defined a taxonomy of data science and analytics as obtain, scrub, explore, model, and interpret data which includes a blend of statistics and machine learning.

The use of data science and analytics is no longer limited to scientific fields (Davenport, 2014. The use of data has evolved into the "data economy" where organizations are focused on being data-driven (Davenport, 2014; Davenport & Harris, 2013). This has led to the development of data services and products which are created for customers based on data discovery. Examples of data products include customer path analysis, social network analysis, and service offerings based on the Internet of Things (IoT) (Davenport, 2014; Davenport & Harris, 2013). Data products such as recommendation engines (i.e., Amazon and Netflix) or quoting engines (Geico Insurance) are examples of data products that provide a competitive advantage.

Data Science and Business Benefits

The value of data science and analytics projects are defined by their business impact (Dhanrajani, 2015). Data science and analytics at a high-level focuses on fostering new thinking, exploring unknown patterns in data, challenging the status-quo, improving continuously, and identifying business drivers. Dhanrajani outlined multiple examples of business benefits made possible through the use of data science and analytics.

Amazon launched its recommendation engine which is a data science product that recommends products to a buyer. Dhanrajani (2015) approximated that 15% to 20% of Amazon's business comes from recommendations; customers rely on the recommendation engine to explore other related products or packaged deals. United Parcel Service (UPS) created the On-Road Integrated Optimization and Navigation (ORION) system which optimized routes for UPS drivers using many different data sources. UPS improved its routing schedules resulting in better customer service and saving millions of dollars (Dhanrajani, 2015).

Other examples of business benefits include improving product categorization. Online retailers may not classify products the way that customers think about them. Data science and analytics projects seek to improve product categorization by all product features such as shape, purpose, look, and product text descriptions. Airlines have predicted more accurately the percentage of passengers who purchase a ticket but fail to take the flight, which enables airlines to sell more tickets. This prediction minimizes lost revenue and reduces the risk of overselling a flight (Dhanrajani, 2015). Many industries are seeing benefits from data science and analytics. Retailers, insurance companies, financial institutions, and telecom companies are refining customer segmentation to increase customer profitability, drive customer behaviors, and increase engagement (Dhanrajani, 2015). Optimization of prices, reduction of risk, and propensity to buy are other capabilities enabled by data science and analytics. Deriving business benefits from data science projects is at risk due to the lack of insight on how to improve the success rate of these projects.

Data science and analytics projects are on the rise due to perceived business benefits; however, the success and failure factors of data science projects are not well known or researched (Chang, Kaufman, & Kwon, 2014; IBM, 2010; Kambatla, Kollias, Kumar, & Grama, 2014; Li, Thomas, & Osei-Bryson, 2016). Data science and analytics projects are more complex than data mining projects due to: larger volumes, variety, and speed of data; the need to have scalable analytic solutions; the increase in data science and analytical projects; the lack of deep data science and analytical skills sets; the need to shorten the data acquisition to decision cycle; and an organization's analytical maturity (Chang et al., 2014; IBM, 2010; Kambatla et al., 2014; Li et al., 2016).

> **"Communication is a success factor in project management and in data science and analytics projects, particularly contributing to attaining business benefits."**

Communication is a success factor in project management and in data science and analytics projects, particularly contributing to attaining business benefits (Demirkan & Dal, 2014; IBM, 2010; Marr, 2017; PMI, 2013; Roberts, 2015; Sharma & Osei-Bryson, 2009). Leveraging data mining success factors can be a basis for data science planning, but the research here is limited. Success factors are very broad, so the focus of this research is to explore the contribution of communication practices to attaining business benefits. Thus, the scope of project communication in this research will focus on how business benefits are defined and communicated in data science and analytics projects.

Project Management, Inc. (PMI) (2013) outlined that, to improve communication, organizations should close the gap around communicating business benefits. Consistent communication protocols and knowledge sharing enables project teams to meet business goals efficiently. Some of the best practices to improve communication include clearly defining project business objectives, ensuring projects are aligned with organizational strategy, defining clear project goals, and evaluating project expectations and success factors up front. These best practices are the focus of this research.

Project Failures and Communication

Demirkan and Dal (2014) outlined several reasons why data science and analytics projects fail, and one of the top reasons listed is the lack of identifying clear business need and value. Before the investment in a data science and analytics project occurs, organizational leaders should have a clear idea of the business outcomes or problems to be addressed. Some suggestions made by Demirkan and Dal to address this failure point include ensuring project management ownership, clear project alignment with the organizational strategy, and business stakeholder involvement.

Demirkan and Dal (2014) addressed other failure points including having a departmental focus versus a strategic focus on data science and analytics, having islands of analytics, not addressing data quality, having no clear communication plan on analytics, not planning for data quality, and not seeing data science and analytics as a core capability requiring an ecosystem of technology, people, and process. Most of the failure points align with lack of strategy and communication of strategy – which includes communication of the benefits of data science and analytics.

Roberts (2015) posited that data science and analytic projects often start with the wrong questions to be answered. According to Roberts, data science and analytics projects start with the expectation that something valuable will be discovered, and this becomes the business case for the project. Data science and analytics projects that start as an exploratory project fail because the scope is too broad to drive useful analysis. Roberts outlined that the better approach is to start a project with established goals that map to creating business benefits. Projects with established goals follow a hypothesis-testing approach that begin with a set of defined questions. By approaching data science and analytics projects with a clear goal, business justification is paired with business action.

Marr (2017) outlined that data science and analytic projects start with high expectations, but a high number of projects fail. One failure point outlined by Roberts is that projects start out without having clear business objectives. Roberts argued that project teams start with the "how" without understanding the "why." Before embarking on the "how," project teams need to understand what problem is to be solved. Other project failures are linked to not having a clear business case (which impacts having clear business objectives) and poor communication. Poor communication is linked to not clearly outlining the "why" of the project and the ability of project leaders to guide the project based on business objectives.

Historically, communicating the business benefits of data mining has not been a focus in data mining projects. Sharma and Osei-Bryson (2009) highlighted the lack of focus and formality on the business understanding phase of data mining (CRISP-DM), which traditionally has been implemented in an ad hoc fashion. Sharma and Osei-Bryson's research exposed that little research existed that provided a detailed description of how this phase was implemented. The business understanding phase is where business objectives, hypothesis formation, and project goals and planning occur. Quite the opposite is true of other phases of the data mining process such as modeling, where much research exists. Sharma and Osei-Bryson outlined that the business understanding phase is the most important as this phase influences the decisions made in subsequent phases such as data preparation, data understanding, modeling, evaluation, and deployment (Davenport, 2014; Davenport & Harris, 2013).

Role of Communications in Projects

According to *Forbes* (2011), nine out of 10 organizational leaders outlined that communication is a critical success factor in strategic initiatives, and half of the leaders surveyed identified communication as a key component for strategic planning and execution success (PMI, 2013). Project managers identified that stakeholder communication is the most critical success factor in project management. According to PMI (2013), two of five projects do not meet the intended business benefit, and 50% of these failed projects are caused by communication failures. Ineffective communications lead to fewer successful projects and fewer projects meeting original goals.

Project communication includes defining the project business benefit and contributing to organizational strategy (PMI, 2013). When organizations align the strategy with execution, projects are more successful. Organizations that focus project communications on business benefit have more successful projects versus organizations that do not communicate this information or do it less frequently. Project leaders are able to focus project teams on the right outcomes when business benefits are relayed to the teams frequently and clearly, thus providing the context for the project.

Projects are delivered via a methodology belonging to communication practices (PMI, 2013). Data science and analytical projects are iterative and exploratory in nature and follow data mining methodologies. While multiple methodologies exist for data science and analytical projects, the generally accepted methodology used is the Cross Industry Standard Process for Data Mining (CRISP-DM) approach (Gartner, 2013; Sharma & Osei-Bryson, 2009).

CRISP-DM is a data mining process model that conceptually describes the stages that are used to tackle data mining problems. CRISP-DM was originally created to align with data mining, but it has organically evolved into the primary approach used by data scientists. CRISP-DM is broken into six stages which appear to be in sequence; however, the stages are not strictly sequential, and iterating through the stages is expected (Marbán, Mariscal, & Segovia, 2009). The six stages are business understanding, data understanding, data preparation, modeling, evaluation, and deployment; communication of business benefits is covered in the business understanding stage. The business understanding stage contains the steps to define business benefits and objectives. A review of literature has uncovered seven Critical Success Factors (CSF) in data mining, two of which are business mission and communication (Sim & Cutshall, 2003). Both these areas are typically addressed as part of the "business understanding" phase of CRISP-DM, and little research exists on how this phase is accomplished (Sharma & Osei-Bryson, 2009). The review of literature established the four key areas to focus on to explore the communication of project business benefits: ensuring projects are aligned with organizational strategy, defining clear project goals, and evaluating project expectations and success factors up front.

Study Description

This study examined communication practices in data science and analytical projects, specifically the practices used to communicate the purpose of the project and expected business benefits. These areas were identified as CSFs and also as a failure point at the start of projects (Sharma & Osei-Bryson, 2009; Sim, 2014; Sim & Cutshell, 2003). Communicating the purpose of the project and outlining business benefits were further defined based on existing research and divided into the subcategories of project selection and alignment, initial project evaluation, setting clear business objectives, and defining clear project goals. These four areas were used to develop hypotheses and design the survey instrument. By focusing on these areas, data science and project leaders can understand how to communicate business benefits to increase project success. Based on the lack of research and the literature review, the following hypotheses were formed to examine communication practices used to communicate business benefits in data science and analytics projects:

H1: Project Alignment: Hypothesis: Data science and analytical projects are selected based on alignment with organizational strategic outcomes.

H2: Project Evaluation: Hypothesis: Data science and analytical projects are fully evaluated before initiated.

H3: Business Objectives: Hypothesis: Business objectives for data mining projects are clear.

H4: Project goals: Hypothesis: Goals to be explored in analytical projects are clearly established ahead of time.

Based on these four areas, the following research questions were formulated:

R1: Are data science and analytical projects selected based on alignment with organizational strategy and outcomes?

R2: Are data science and analytical projects fully evaluated before project initiation?

R3: Are business objectives for data science and analytical projects clear?

R4: Are data science and analytical project goals clear prior to project start?

The research questions explored areas that define communicating business benefits for data science and analytical projects. Gaining an understanding of an organization's existing practices in this area established a baseline and context for data science and analytical project communications, thus contributing to the identification of communication CSFs. The research questions were used to frame the questions used in the creation of a survey.

Research Results

Four hypotheses were investigated and drove the survey design, the collection of data, and the data analysis. These hypotheses were used to create the research questions in the area of understanding data science and analytics project communication in the areas of project alignment, project evaluation, business objectives, and project goals – all of which contribute to understanding the business benefits of the project. The survey instrument was posted in LinkedIn professional groups with membership that had experience completing surveys. The survey instructions included a description of the experience required of three or more years of participating or leading data science and analytics projects before taking the survey. The total membership of the professional groups where the survey was placed was approximately 500,000 members.

According to Beamish (2010), there are two primary groups of individuals that join online professional groups – contributors and lurkers. Contributors are those that post and respond to posts on online groups, and lurkers are those that browse and read messages but rarely or never contribute. Based on participation studies, on average, lurkers represent over 90% of the membership in online groups and contributors on average represent only 1-10% of the group (Beamish, 2010). Using the research from Beamish, the population size could vary from 5,000 to 50,000 individuals.

The survey consisted of 4 sections and 15 questions. The sections aligned with the research questions, and under each section 3 to 4 survey questions were provided to participants using a 5-point Likert scale of strongly disagree (1), disagree (2), neutral (3), agree (4), and strongly agree (5). Neutral is treated as neither disagreeing nor agreeing to the questions in the survey. The best measure of central tendency for a Likert scale is the mode versus the mean or the median due to skewness (Fowler, 2013). However, a one sample t-test was used to evaluate whether the results of each question was greater than the neutral value of 3. Hypotheses testing was completed using the Kolmogorov-Smirnov test and the chi-square test.

The survey was posted in the following LinkedIn professional groups on three different days over a four-month period: Data Science Central; Machine Learning and Data Science; and Big Data, Analytics, Business Intelligence and Visualization Experts Community. These groups were chosen based on the high number of members and the high number of daily contributions which were observed over a six month period (June to December, 2016) of ten or more contributions a day. After four months of posting the survey, 60 respondents had completed the survey.

Data Analysis

Survey data using a Likert scale is categorized as ordinal data, where data can only be used to determine that one score is ranked higher or lower than the other. Based on the data type, in this case ordinal, only certain analyses was appropriate (Sullivan & Artino, 2013). Fowler (2013) outlined that ordinal scales are encountered often in research and the typical parametric tests for hypotheses testing do not typically apply. The first reason is that a parametric test relies on data adhering to an interval or ratio scale, and the second reason is that samples are drawn from a population with a known distribution. Because the survey uses an ordinal scale and the population distribution is unknown, nonparametric methods are used in the data analysis (Sullivan & Artino, 2013).

Each research question had three to four survey questions, and each survey question was analyzed focusing on median and frequency analysis as appropriate for ordinal data. The question analysis was then grouped based on the categories of project alignment, project evaluation, business objectives, and project goals.

The first nonparametric approach used was a Kolmogorov-Smirnov test (KS test), which tests for goodness of fit when the measurement scale is ordinal and examines whether frequencies of observations are aligned with the frequencies expected under a null hypothesis. The Kolmogorov-Smirnov test is used to accept or reject the null hypotheses (Sullivan & Artino, 2013).

The second test completed was a chi-square test which was used with nominal data. The five response categories were combined into two nominal categories – agree and disagree where neutral responses were ignored. The chi-square test was used to determine if a relationship exists between two nominal variables in a sample and to accept or reject the null hypothesis (Sullivan & Artino, 2013).

Thus, three different statistical tests were performed on the survey results. A one sample t-test was used to evaluate if the result of each question was greater than the neutral value of 3, and two hypotheses tests were completed – the KS-test and the chi-square test – to accept or reject the null hypothesis.

Findings

The research questions pertained to four areas that comprised the category of data science and analytics project communication (a) project alignment, (b) project evaluation, (c) business objectives, (d) and project goals. Each question was analyzed to determine if the null hypothesis should be accepted or rejected. The results of each question and support for each research question and hypothesis follows.

Hypothesis 1

H1: Project Alignment: Hypothesis: Data science and analytical projects are selected based on alignment organizational strategic outcomes.

H1 Null: Project Alignment: Hypothesis: Data science and analytical projects are not selected based on alignment with organizational strategic outcomes.

Out of the four questions exploring project alignment, the conclusions support that organizations design data science and analytics projects with a strong focus on strategic outcomes; however, operating managers are not clear on how projects align with the goals of the organization, project teams do not frequently meet to discuss project alignment, and data science and analytical projects are not typically part of strategic project portfolios. The overall results support that data science and analytical projects are not selected based on alignment with organizational strategy and outcomes, thus accepting the null hypothesis.

R1: Are data science and analytical projects selected based on alignment with organizational strategy and outcomes? Not supported based on the results of the hypothesis tests and the t-test.

Area of Analysis	Survey Question	T-Test (greater than the neutral value of 3)	KS Test	Chi Square	Final Conclusion
Project Alignment	Q1: Organizations design data science projects with a strong focus on strategic outcomes.	Supported	Reject Null Hypothesis	Reject Null Hypothesis	Reject Null Hypothesis
	Q2: Operating managers have a good understanding of the linkage between their analytical projects and the goals of the overall organization.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis
	Q3: Project teams frequently meet after starting a project to discuss whether they are maintaining alignment with strategic outcomes.	Supported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis
	Q4: Data science and analytical projects are typically part of a portfolio of projects created as part of strategic planning.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis

Hypothesis 2

H2: Project Evaluation: Hypothesis: Data science and analytical projects are fully evaluated before initiated.

H2 Null: Project Evaluation: Hypothesis: Data science and analytical projects are not fully evaluated before initiated.

Out of the three questions exploring project evaluation, the conclusions support that data science and analytical projects do not frequently start with a clear business case, data science and analytical projects do not have measurable success factors established prior to project initiation, and these projects do not have a project sponsor that partners with the team on planning and initiation. The overall results support that data science and analytical projects are not fully evaluated before project initiation, thus accepting the null hypothesis.

R2: Are data science and analytical projects fully evaluated before project initiation? Not supported based on the results of the hypothesis tests and the t-test.

Area of Analysis	Survey Question	T-Test (greater than the neutral value of 3)	KS Test	Chi Square	Final Conclusion
Project Evaluation	Q5: Data science and analytical projects frequently start with a clear approved business case.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis
	Q6: Data science and analytical projects have measureable success factors established before project initiation.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis
	Q7: Data science and analytical projects typically have a project sponsor who partners with project teams on project planning and initiation.	Supported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis

Hypothesis 3

H3: Business Objectives: Hypothesis: Business objectives for data science and analytical projects are clear.

H3 Null: Business Objectives: Hypothesis: Business objectives for data mining projects are vague.

All four questions supporting business objectives were not supported which corroborates that the business objectives for data science projects tend to be vague. Project charters are not created, project teams are not clear on business objectives, and business objectives are not clear and measurable, thus accepting the null hypothesis.

R3: Are business objectives for data science and analytical projects clear? Not supported based on the results of the hypothesis tests and the t-test.

Area of Analysis	Survey Question	T-Test (greater than the neutral value of 3)	KS Test	Chi Square	Final Conclusion
Business Objectives	Q8: Project teams typically create project charters for data science and analytical projects.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis
	Q9: Project sponsors and project teams define business outcomes for data science and analytical projects.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis
	Q10: Data science and analytical project teams have a good understanding of business objectives linked to analytical projects.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis
	Q11: Business objectives for data science and analytical projects are specific and measurable.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis

Hypothesis 4

H4: Project goals: Hypothesis: Goals to be explored in analytical projects are clearly established ahead of time.

H4 Null: Project goals: Hypothesis: Goals to be explored in analytical projects are not clearly established ahead of time.

The four questions that explored project goals resulted in one question being supported and the remaining three not. Data science and analytical projects frequently start with a problem statement or question of interest; however, project outcomes are not clearly established that support business objectives, project sponsors and teams do not define project outcomes, and the scope of data science and analytical projects are not clear as part of project planning. Data science and analytical project goals are not clear prior to project start, thus accepting the null hypothesis.

R4: Are data science and analytical project goals clear prior to project start? Not supported based on the results of the hypothesis tests and the t-test.

Area of Analysis	Survey Question	T-Test (greater than the neutral value of 3)	KS Test	Chi Square	Final Conclusion
Project Goals	Q12: Project teams typically establish project outcomes that support the required business objectives for data science and analytical projects.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis
	Q13: Data science and analytical projects frequently start with a problem statement or question of interest.	Supported	Reject Null Hypothesis	Accept Null Hypothesis Accept Null	Reject Null Hypothesis
	Q14: Project sponsors and project teams frequently define project outcomes for data science and analytical projects.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis
	Q15: The scope of data science and analytical projects is documented and clear as part of project planning.	Unsupported	Accept Null Hypothesis	Accept Null Hypothesis	Accept Null Hypothesis

Significance of Findings

This study examined communication practices in data science and analytical projects. The scope of communication practices focused on the areas of communicating the purpose of the project and outlining business benefits as this area was identified as a CSF and also a failure point in the literature review (Sharma & Osei-Bryson, 2009; Sim, 2003, 2014). The purpose of the project and presentation of business benefits were further defined based on existing research and divided into the subcategories of project selection and alignment, initial project evaluation, clear business objectives, and clear project goals. All four research null hypotheses were accepted, concluding that communication is not a primary focus in data science and analytical projects:

H1 Null: Data science and analytical projects are not selected based on alignment with organizational strategic outcomes.

H2 Null: Data science and analytical projects are not fully evaluated before initiated.

H3 Null: Business objectives for data mining projects tend to be vague.

H4 Null: Goals to be explored in analytical projects are not clearly established ahead of time.

As outlined, data science and analytic project investment is increasing, and communication is a large contributor to project failure.

The failure of communication in data science and analytical projects has strategic and financial implications. Strategic implications include developing data science and analytics as a capability. Developing this capability is complicated by: larger volumes, variety, and speed of data; the need to have scalable analytic solutions; the increase in data science and analytical projects; the lack of deep data science and analytical skills sets; the need to shorten the data acquisition to decision cycle; and an organization's analytical maturity (Chang et al., 2014; Kambatla et al., 2014; Li et al., 2016). The lack of clear alignment with organizational strategic goals complicates the development of data science and analytics as a capability and delays the attainment of strategic goals.

As a result of not developing data science and analytics as a capability, another strategic implication is an organization's ability to maintain a competitive advantage. As previously stated, the use of data science and analytics is no longer limited to scientific fields (Davenport, 2014). The "data economy" is the norm, where organizations are using data science and analytics to be data-driven (Davenport, 2014; Davenport & Harris, 2013). The "digital" lifestyle has led to every action and interaction being captured in data, resulting in new opportunities and insights to study through the use of data science and analytics (Larson & Chang, 2016). By 2016, data science and analytics was considered a core competency organizations needed to cultivate to remain competitive (Larson & Chang, 2016).

The result of not having clear business objectives and project goals is the lack of business value. This has both strategic and financial implications. As noted earlier, the value of data science and analytics projects are defined by their business impact (Dhanrajani, 2015). The lack of business impact will directly relate to the success of an organization's strategic goals and attaining desired financial outcomes.

Communication is a key success factor in project success (PMI, 2013). Project failures are costly. Once a project fails, the investment made is lost, and work to determine failure and next steps incur additional cost. Project failure also results in the loss of time to market, which will add to the financial loss. Several failure points for data science and analytics projects studied by Demirkan and Dal (2014) have been outlined such as not addressing data quality, not clearly communicating a plan on analytics, and not seeing data science and analytics as a core capability requiring an ecosystem of technology, people, and process, to name a few. Most of the failure points align with lack of strategy and communication of strategy – which includes communication of the benefits of data science and analytics.

The significance of this research to leadership concludes that to increase the probability of business benefits from data science and analytics projects, additional attention needs to be focused on the project initiation and planning stages. Focusing on project selection and alignment, initial project evaluation, clear business objectives, and clear project goals can increase the likelihood of data science and analytics projects delivering business benefits. This is a clear gap, and closing this can improve data science and analytics project success.

This research identified the lack of focus on communication in data science and analytics projects; however, the root causes that contribute to this lack of focus was not in the scope of the research. Future research to be considered is to collect and analyze the contributing factors to this lack of focus on communication in data science and analytics projects. Other future research areas to consider are the use of agile methodology to improve collaboration and communication in data science and analytics projects as current research on agile methodology suggests potential improvements in project delivery and business benefits realization.

The findings of this research highlight the importance of communication to the success of data science and analytics projects. Failure to focus on communication in these projects has strategic impacts to the organization's competitive advantage and financial outcomes. The practices of aligning data science and analytical projects with organizational strategy, fully evaluating projects, defining clear business objectives and project goals are not being practiced based on the findings of this research. Data science and analytical projects may be new to some organizations; however, the communication practices used in project management can have benefits and should be adapted.

References

Beamish, A. (2010). Contributors and lurkers: Obstacles to content creation in a professional online community. In *Collaborative information behavior: User engagement and communication* (pp. 36-54). Hershey, PA: IGI Global. doi:10.4018/978-1-61520-797-8

Bole, U., Popovič, A., Žabkar, J., Papa, G., & Jaklič, J. (2015, April). A case analysis of embryonic data mining success. *International Journal of Information Management*, 35(2), 253-259.

Cleveland, W. (2001) Data science: An action plan for expanding the technical areas of the field of statistics. Retrieved from http://cm.bell-labs.com/cm/ms/departments/sia/doc/datascience.pdf

Chang, R., Kauffman, R., & Kwon, Y. (2014, July). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems, 63*, 67-80.

Davenport, T. (2014, October). *The rise of data discovery* [Teradata white Paper]. Retrieved from http://www.nimbusninetyignite.com/downloads/teradata/Davenport%20Rise_of_Data_Discovery.pdf

Davenport, T. (2015, October 21). 5 essential principles for understanding analytics. *Harvard Business Review.*

Davenport, T., & Harris, J. (2013). Competing on analytics: The new science of winning. *Smart Business Pittsburgh, 20*(6), 15.

Davenport, T. H., Cohen, D., & Jacobson, A. (2005, May). *Competing on analytics.* Babson Park, MA: Babson Executive Education. Retrieved from http://www.babsonknowledge.org/analytics.pdf

Demirkan, H., & Dal, B. (2014, July-August). Why do so many analytics projects still fail?: Key considerations for deep analytics on big data, learning and insights. *INFORMS Analytics*, 44-52.

Dhanrajani, S. (2015). Demystifying data analytics, decision science and digital topical trends, themes and perspectives. Retrieved from https://sameerdhanrajani.wordpress.com/2015/12/07/sameer-dhanrajani-how-data-science-projects-deliver-business-impact/

Dhar, V. (2013). Data science and prediction. *Communications of the ACM, 56*(12), 64. doi:10.1145/2500499

Forbes Insights Strategic Initiatives. (2011). Adapting corporate strategy to the changing economy. Retrieved from https://www.forbes.com/forbesinsights/FDStrategy_2011/index.html

Fowler, F. J. (2013). *Survey research methods* (5th ed.). Thousand Oaks, CA: Sage.

Gartner. (2013, December 16). Gartner predicts business intelligence and analytics will remain top focus for CIOs through 2017. Retrieved from http://www.gartner.com/newsroom/id/2637615.

IBM. (2010). Planning successful data mining projects. Retrieved from http://www.besmart.company/wp-content/uploads/2014/12/Planning.pdf

Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014) Trends in big data analytics. *Journal of Parallel and Distributed Computing, 74*, 2561-2573.

KDD. (2014). Data mining curriculum. ACM SIGKDD.

Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management, 36*, 700-710. doi:10.1016/j.ijinfomgt.2016.04.013

Leek, J. (2013, December 12). The key word in "data science" is not data, it is science. Retrieved from https://simplystatistics.org/2013/12/12/the-key-word-in-data-science-is-not-data-it-is-science/

Li, Y., Thomas, M. A., & Osei-Bryson, K. (2016, November). A snail shell process model for knowledge discovery via data analytics. *Decision Support Systems, 91*, 1-12. doi:10.1016/j.dss.2016.07.003

Loukides, M. (2010). What is data science?: The future belongs to the companies and people that turn data into products. Retrieved from https://www.oreilly.com/ideas/what-is-data-science

Mason, H., & Wiggins, C. (2010). A taxonomy of data science. Retrieved from http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

Marbán, O., Mariscal, G., & Segovia, J. (2009). A data mining and knowledge discovery process model. In J. Ponce (Ed.), *Data mining and knowledge discovery in real life applications* (pp. 438-453). Vienna, Austria: I-Tech.

Marr, B. (2017). Why big data projects fail. *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2015/03/17/where-big-data-projects-fail/#5b7efe08239f

Roberts, G. (2015, July 22). Stop hiring data scientists until you're ready for data science. *VentureBeat*. Retrieved from https://venturebeat.com/2018/05/31/qure-ai-can-identify-15-different-forms-of-tuberculosis-with-90-accuracy/

Sharma, S., & Osei-Bryson, K. (2009). Framework for formal implementation of the business understanding phase of data mining projects. *Expert Systems with Applications, 36*, 4114–4124.

Sim, J. (2014). Consolidation of success factors in data mining projects. *GSTF Journal on Computing, 4*(1), 66-73. Retrieved from http://ezproxy.library.nyu.edu:2048/login?url=http://search.proquest.com/docview/1642381546?accountid=12768

Sim, J., & Cutshall, R. (2003, January 1). *Critical success factors in data mining projects: Proceedings of the Annual Meeting of the Decision Sciences Institute* (pp. 907-912).

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education, 5*(4), 541-542. Retrieved from http://doi.org/10.4300/JGME-5-4-18

Yau, N. (2009). Rise of the data scientist. Retrieved from http://flowingdata.com/2009/06/04/rise-of-the-data-scientist/

About the author

Deanne Larson is an active data science practitioner and academic. Her areas of research include data strategy, data science, and project management. She holds Project Management Professional (PMP), Certified Business Intelligence Profession (CBIP), and Six Sigma certifications. She attended AT&T Executive Training at the Harvard Business School focusing on IT leadership and at Stanford University focusing on data mining and holds a master's in analytics from NYU. She presents several times a year at conferences for TDWI. Dr. Larson is Principal Faculty at City University of Seattle and has consulted for several Fortune 500 companies. She holds a PhD from the International School of Management.