

ASSESSING THE FACTORS INFLUENCING ANALYTICS INVESTMENT PRIORITIES ACROSS ORGANIZATIONS

Bruce Pollack-Johnson, Department of Mathematics and Statistics, Villanova University, 610-519-6926, bruce.pollack-johnson@villanova.edu

Suzanne Heller Clain, Department of Economics, Villanova University, 610-519-6556, suzanne.clain@villanova.edu

Matthew J. Liberatore, Department of Management and Operations, Villanova University, 610-519-4390, matthew.liberatore@villanova.edu

ABSTRACT

The purpose of this study is to understand and assess the factors influencing priorities for investment in analytics software and services in organizations. It is based on data obtained from a series of intensive workshops held by IBM with potential clients over 2009 - 2011. The factors considered include assessments of current analytics maturity and aspirations within defined analytics areas, as well as industry, geography, year, and workshop type. Univariate analysis and predictive models were employed to better understand the importance of the various factors on analytics investment priorities. The results show significant differences across industries, geographies, and time, as well as the effect of current analytics assessments on future priorities.

Keywords: analytics, business intelligence, IT investment

INTRODUCTION

IBM's Smarter Analytics approach offers a broad, integrated portfolio of information and analytics capabilities, spanning software, hardware and services. The benefit is better and faster decisions, and process automation. IBM has offered workshops for organizations to analyze, strategize, and plan for investments in analytics. Beginning in 2009, these workshops were called Information Agenda Roadmap Workshops. Starting in 2011, additional workshops with the same format were also offered, targeted at functional divisions of an organization, and were called Functional Assessment workshops. Through each engagement, IBM gathered extensive details on their clients' current maturity capabilities and future goals, and provided their recommendations for clients' information agenda. From summary workshop documents, a data base was developed that contains a capability analysis and assessment for each workshop. The purpose of this research is to analyze the prioritized capability assessment (PCA) data to determine if there are differences across organizations in terms of geography, industry, and time, as well as to determine the effect of maturity and aspirations within defined analytics areas. The results provide insights into the determinants of organizations' analytics investment priorities.

DATA COLLECTION AND METHODOLOGY

The data include the organization's Account name, the Workshop type, an Industry code (16 categories), a Geography code (7 categories), the Year of the workshop, 12 numerical values on a 0 to 10 scale evaluating the current (As Is) maturity level, aspirations (To Be, or 2B), and the difference between these (Gap) in four areas: Define and Govern (DG), Analytics & Optimization (AO), Trusted Information (TI), and Information Foundation (IF). The

subcategories defining each of the four areas are given as Table 1. We define an organization as having a Prioritized Capability Area (PCA) in DG if at least one of the subcategories of DG as listed in Table 1 was indicated as a priority. The same definition is used for the other three capability areas (AO, TI, and IF). For example, if Information Integration was reported as a priority, then TI is a PCA. The data were collected from organizations all over the world over a three-year time frame (2009 – 2011). The data formats were standardized and researchers gathered as much as possible in the English language.

A total of 771 records were collected, with 393 having usable data for the analysis. Extensive procedures were followed to clean and validate the data prior to analysis. It is important to note that the data set is a *representative sample* of organizations working with IBM and is *not a random sample* of organizations in the marketplace. However, given the sample size and the diversity of the organization size, industries, and geographies represented, the data present a good picture of the current state of analytics.

After some preliminary analysis, the 16 industry codes were reduced to nine industry sectors: Banking & Financial Markets, Insurance, Communications, Government, Healthcare, Industrial & Chemical, Retail, Consumer Goods, and Travel & Transportation. Similarly, the seven geographic codes were reduced to six geographic regions: North America, Western Europe, Central & Eastern Europe & Middle East & Africa (CEE/MEA), Latin America, Japan, and Asia Pacific.

Univariate analysis was employed to understand the basic relationships across the data, followed by regression analysis. All analyses were performed using IBM SPSS software.

Table 1: Subcategories Defining Each of the Four Analytics Capability Areas

Information Foundation

- **Data Management** – access, scale, flexibility and optimal query performance
- **Metadata Management** – establish, publish, maintain central repository of data
- **Content Management** – capture, store, manage, deliver unstructured data

Trusted Information

- **Information Integration** – extract, capture changed data, transform information
- **Information Quality** – monitor, standardize, match, accurate data
- **Master Data Management** – consistent view of entity, authoritative & trusted source of information
- **Business Process Management** – automate response to business events
- **Records Management** – retain, archive or delete data for compliance

Define & Govern

- **Strategy** - capture, integrate, process information
- **Architecture** – document standards, policies & principles
- **Stewardship** – data governance and definitions
- **Processes** – data flow and quality control

Analytics & Optimization

- **Business Intelligence & Performance Management** – access to analyze, tailor & deliver timely, reliable forecasts, plans, reports, queries

- **Advanced Analytics** – statistical modeling and data mining to find patterns & probabilities to make predictions on future events

FINDINGS

Based on the results of regression analysis, the principal findings are as follows:

- Industry plays some role in the determination of PCA.
 - A client in the communication, insurance and travel & transportation sectors is more likely to name IF a PCA, in comparison to otherwise similar clients.
 - A client in the industrial & chemical, insurance, communication, government and consumer goods sectors is less likely to name DG a PCA, in comparison to otherwise similar clients.
- Geography plays some minor role in the determination of a PCA.
 - A client located in Japan is less likely to name DG a PCA, in comparison to otherwise similar clients located in North America.
 - A client located in Western Europe is more likely to name DG a PCA, in comparison to otherwise similar clients located in North America.
- Year plays some role in the determination of a PCA.
 - A client in 2009 was more likely to name IF a PCA, in comparison to otherwise similar clients in later years.
- Nature of workshop plays some role in the determination of PCA.
 - A participant in an Information Agenda Roadmap workshop is more likely to name IF and/or DG a PCA, in comparison to otherwise similar clients participating in Functional Assessments.
- Capability measures (in some form) play a role in the determination of PCAs. There is mixed support for a maturity path that identifies AO as a PCA of mature clients
 - The size of the gap in an area is directly linked to the tendency to name that area a PCA. (A client with a higher IFGap is more likely to name IF a PCA, in comparison to otherwise similar clients. A client with a higher DGGap is more likely to name DG a PCA, in comparison to otherwise similar clients. A client with a higher TIGap is more likely to name TI a PCA, in comparison to otherwise similar clients. A client with a higher AOGap is more likely to name AO a PCA, in comparison to otherwise similar clients.)
 - A client with a higher level of DGAsIs is more likely to name DG a PCA, in comparison to otherwise similar clients.
 - There are also effects across areas.
 - A client with a higher AOAsIs is less likely to name DG a PCA, in comparison to otherwise similar clients.
 - A client with a higher AOGap is less likely to name TI a PCA, in comparison to otherwise similar clients.
 - A client with a higher DGGap or higher TIGap is less likely to name AO a PCA, in comparison to otherwise similar clients. Unexpectedly, a client with a higher IFGap is more likely to name AO a PCA, in comparison to otherwise similar clients.

SUMMARY AND CONCLUSIONS

The results show significant differences in priorities for analytics capabilities across time, industries, and geographies. They also show which areas of analytics maturity and aspiration are significant in explaining analytics capability priorities. The findings from this paper provide insight into analytics investment priorities within organizations. They highlight the interrelatedness of four different areas that might require analytics investment if organizations wish to improve their capabilities to remain competitive.

ACKNOWLEDGEMENT

The authors would like to acknowledge and thank IBM for supporting this research, and would especially like to recognize Carolyn Martin, Will Reilly, Michele Shaw and Rebecca Shockley of IBM for their assistance and contributions.