9-2010

ENHANCING BUSINESS-INTELLIGENCE TOOLS WITH VALUE-DRIVEN RECOMMENDATIONS

Adir Even
Ben-Gurion University of the Negev, Israel, adireven@bgu.ac.il

Yoav Kolodner
Ben-Gurion University of the Negev, Israel, kolodner@bgu.ac.il

Roy Varshavsky
Microsoft Israel R&D Center, royv@microsoft.com

ENHANCING BUSINESS-INTELLIGENCE TOOLS WITH VALUE-DRIVEN RECOMMENDATIONS

Adir Even, Ben-Gurion University of the Negev, Israel, adireven@bgu.ac.il

Yoav Kolodner, Ben-Gurion University of the Negev, Israel, kolodner@bgu.ac.il

Roy Varshavsky, Microsoft Israel R&D Center, royv@microsoft.com

Abstract

Business-intelligence (BI) tools are broadly adopted in organizations today, supporting activities such as data analysis, decision making, and performance measurement. This study investigates the integration of feedback and recommendation mechanisms (FRM) into BI tool, defining FRM as visual cues that are embedded into the tools and provide the end-user with usage guidelines. The study focuses on FRM that are based on assessment of previous usage. It introduces the concept of value-driven usage metadata - a novel methodology for tracking and communicating the usage of data resources, linked to a quantitative assessment of the value gained. A laboratory experiment tested FRM-integration with 200 participants and confirmed our assumptions that FRM integration will improve the usability of BI tools and increase the benefits that can be gained from data resources. It also highlighted the potential benefits of collecting value-driven usage metadata and using it to generated usage recommendations.

Keywords: Business Intelligence, Recommender Systems, Data Warehouse, Metadata

1 INTRODUCTION

Data repositories, along with the information systems (IS) utilizing them, had long been recognized as critical organizational resources. Recent years have witnessed a major transition toward extended usage of data resources for business analysis, performance measurement, and managerial decision support. This transition is driven by the notion that decisions based on evidence of what works within the company and elsewhere are likely to be better and will help the organization thrive (Pfeffer and Sutton, 2006). Davenport (2006) provides examples of firms that have gained strong competitive advantage by investing in the development of data analysis capabilities and data-driven analytics. This transition toward data-driven management is well-supported by the rapid progress in the capacity and the performance of information and communication technologies (ICT) for utilizing large data resources. Most notable is the broad adoption of business intelligence (BI) platforms and tools, which permit rapid development and distribution of data analysis and decision support utilities – to a point that in a 2008 Gartner survey, ~1500 CIO's ranked BI as their #1 top priority for coming years.

High complexity is a major limitation of current BI environments. The common end-user, in search of an answer to a business question, often finds large data repositories too difficult to navigate for reaching the right data, and BI tools too difficult to use for answering the question. Furthermore, it is even not uncommon for end-user to know neither the right business question to ask, nor the full range of capabilities offered by data repositories and BI tools. This study investigates the integration of feedback and recommendation mechanisms (FRM) into BI tools. We define FRM as textual and/or graphical visual cues that guide the end-user to consider using certain data subsets and/or analysis forms. We suggest that the FRM integration can improve the usability of BI tools and increase the benefits that end-users and organizations can gain from data resources, this by facilitating effective and efficient navigation, and by helping to reveal undiscovered potential of unused data and analysis forms, and thus add business value. The experiment described later investigates value-driven FRM – a novel form of a recommender system, based on quantitative assessments of business-value gains and their attribution to the data resources being used.

Our study makes a few contributions. First, it presents the concept of integrating FRM into BI and highlights a few possible approaches for generating them. Second, it proposes a novel methodology for tracking the use of data, termed as value-driven usage metadata, which integrates in assessments of both the frequency of use and the associated value gains toward generating FRM. Finally, it explores the potential contribution of collecting value-driven metadata and generating value-based FRM through a comprehensive laboratory experiment, in which participants were asked to evaluate different variants of FRM, integrated into a BI tool. An in-depth analysis of the experiment's results supports the assumption that value-based FRM may significantly improve the effectiveness of using BI tools and provides some additional important insights. In the reminder of this paper, we first provide the background to our work, and introduce the two novel concepts that underlie it – the integration of FRM into BI tools, and the collection of value-driven usage metadata. We then describe in detail the laboratory experiment, present the results and discuss the findings. To conclude, we highlight the potential contributions of the new concepts that we present and discuss directions for future research.

2 BACKGROUND

Our study aims at improving the usability and effectiveness of BI systems. The software market offers a plethora of commercial BI platforms, which typically offer a variety of presentation capabilities (e.g., tables, charts, statistics, and advanced analytics), rapid-development utilities, and administrative tools. BI platforms permit different forms of data usage such as reports, spreadsheets, OLAP (On-Line Analytical Processing), dashboards, and data mining. This variety of presentation and analysis forms confers the flexibility to use the same data for supporting different analytic tasks and to adapt the presentation style to end-users' preferences. BI solutions often use a data warehouse (DW) as an infrastructure. A DW stores historical data about past business behavior, patterns and trends, which is imported from different sources and covers a broad range of business perspectives and activities.

The increasing popularity of DW/BI environments can be attributed to benefits such as gaining broad business coverage, leveraging data-collection investments, and shortening implementation cycles (Counihan et al., 2002; March and Hevner, 2005; Even et al., 2006). BI systems enable analytical data usage toward supporting important decisions such as evaluation of corporate strategies (Cooper et al., 2000), transforming business processes (Wixom et al., 2008), and managing customer relations (Even and Shankaranarayanan, 2008). Yet, exploiting DW/BI environments is challenging both technically, due to the many components and the complexity of configuration decisions involved (Shankaranarayanan and Even, 2006), and organizationally, due to the substantial managerial support and financial resources needed (Wixom and Watson, 2001). Moreover, DW/BI design and configuration decisions are often associated with substantial cost-benefit tradeoffs (Even et al., 2006). Despite the increasing popularity of DW/BI implementations in recent years, so far these concepts have attracted only limited academic research aimed at the challenge of increasing the effectiveness of DW/BI utilization from the end-user's perspective.

2.1 Feedback and Recommendation Mechanisms (FRM)

As a contribution to improving the effectiveness of DW/BI, we propose to integrate FRM capabilities in a manner that would help the end-user navigate through complex data resources and highlight new usage directions with a high benefit potential, while still maintaining simple and easy-to-learn functionality. FRM, in the form visual cues, would provide the end-user with feedback on the actions taken so far, and some guidelines for further actions to consider – e.g., approach certain data subsets and/or apply certain analysis forms. FRM can be seen as a form of a recommender system – an automated mechanism which provides end-users with some rating of information items not seen so far, which have been shown to have great influence on the end-user's decisions (Adomavicius and Tuzhilin, 2005). Recommender systems aim at improving usability and decision-making outcomes, enhancing the end-user's experience, and reducing information overload (Wei et al., 2005; Sun et al., 2008). Such systems are a common practice today in commercial and social websites (Adomavicius and Kwon, 2008), and in information retrieval systems such as digital libraries (Song et al., 2006). Adomavicius and Tuzhilin (2005) identify two categories of methods for generating recommendations – content-based methods, driven be the choices made by the user in the past, and collaborative-filtering methods, driven by the choices made by other users with similar preferences. They also suggest that some methods introduce a hybrid between these two approaches – the recommendation form that we test in our experiment, described further later, can be seen as such a hybrid.

The integration of recommender systems into BI tools is not a common practice today, and has not been significantly explored so far, although some commercial software vendors have introduced recommendation utilities to an extent (e.g., Bissantz, in their data mining tool, http://www.bissantz.com/deltamaster). We would argue that the motivations that drive the integration of recommender systems into websites and digital libraries - improving decision-making outcomes, enhancing end-user's experience, and reducing information overload – are all strong motivations in BI environments. We suggest that similar enhancement of BI systems may have important contribution to a better usage of the BI tools, and improve the decisions made.

Figure 1 offers a simplified illustration of integrating FRM into a BI tool that lets the end-user navigate through sales data, and slice it along certain customer characteristics (e.g., Income, Occupation, and Gender), towards identifying profitable segments. The FRM-enhanced version of the BI tool provides color-coded numeric rating of the different attributes, and of the data values in each attributes. These rates provide certain recommendations to the user on how to slice the data further at any given point of time (the rating method used in our experiment will be described in detail later). In this illustrative example – a high characteristic rate would recommend the user to slice the data along that characteristic (here, "Education"), indicating that the underlying data values differentiate well between customers with high profit potential, versus less-profitable customers. Similarly, a high data-value rate (here, the age group of "30-60") would identify a customers segment which is more likely to be profitable and worth further investigation. Notably, navigation decisions in this tool are left to end-users; however, they are now provided with visual cues on how to navigate more effectively.

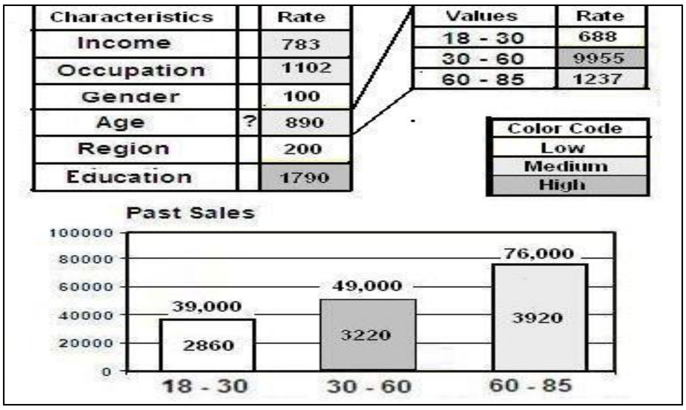

Figure 1. A Business-Intelligence Tool with FRM

Obviously, there are other possible forms for visualizing FRM besides color-coding (e.g., textual/graphical pop-up messages, “mouse-over” tool-tips, and side bars). Such FRM forms could indicate, in addition to the actual recommendations, the level of confidence and relevance of each recommendation based on the parameters that construct it. In this study we explore a novel approach for generating FRM, based on past usage of the underlying dataset and the associated value gains. We discuss other possible FRM forms in our concluding section.

2.2 Value-Driven Usage Metadata

Usage has long been identified as an important factor in explaining IS success (DeLone and McLean, 1992). Deveraj and Kohli (2003), suggest the actual usage of IS strongly explains its success and payoff – even more than the level of ICT-investments made. Burton-Jones and Straub (2006) review a few conceptualizations of IS usage in past research – e.g., based on the extent, the nature, and the frequency of use. In this study, we contrast frequency-based measurement of use, with outcome-based measurement of use, linked to the business-value gained.

We particularly observe the usage of data in complex environments, such as DW/BI implementations. Such environments are often described as a manufacturing process, consisting of interconnected acquisition, processing, storage, retrieval and usage stages (Ballou et al., 1998). This manufacturing process can be conceptualized as having two high-level stages – data administration versus data consumption (Figure 2) - each associated with different stakeholders, goals, motivations and tasks. Data administration addresses technical aspects – providing the ICT capacity needed to store and process data (e.g., hardware, databases, and back-end processes), and the tools for implementing information-product outcomes (e.g., datasets, reports, and analyses). Data consumption, on the other hand, would seek to transform data resources and information products into business value, through their effective usage, less concerning about the technical aspects associated with managing them.

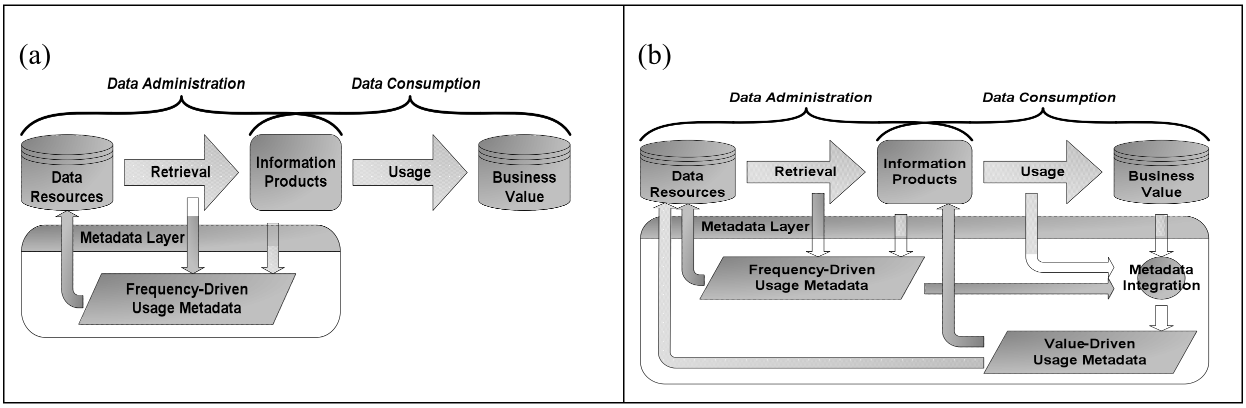

Figure 2. (a) Frequency-Driven versus (b) Value-Driven Usage Metadata

Tracking the use of data subsets (e.g., tables, attributes, and records) and applications in DW/BI has been identified as an important form of metadata (Shankaranarayanan and Even, 2006). Usage-tracking utilities are offered by some specialized commercial solutions and, to an extent, by database management and BI platforms. We term the common approach implemented by today's solutions as frequency-driven usage metadata (Figure 2a). This approach is based on tracking data-retrieval requests and identifying the data subsets being most-frequently used (e.g., by parsing the underlying SQL statements). Frequency-driven metadata collection may provide important inputs to the data administrator, toward improving system design and prioritizing data and system administration efforts. It is common in databases that some records and attributes are accessed more frequently than others. The assumption that drives this approach is that frequent usage reflects higher importance. Accordingly, the results of frequency-driven usage tracking may lead to important data management decisions such as giving frequently-used data subsets higher priority in terms of data-quality improvement – i.e., watch these data subsets more closely, detect and correct defects, and make sure to keep them up-to-date. A data administrator may also consider transferring less-frequently used data subsets to archives, to reduce load on active systems, and improve data-retrieval performance.

While seeing the merits of collecting frequency-driven usage metadata for data administration, we question - does it truly benefit data consumers? Frequent usage may reflects higher significance of certain data subsets to data consumers; hence, to an extent, higher value-contribution potential. Conversely, one could argue that frequent usage reflects certain stagnation and tendency to "dig into the same well" – i.e., re-using certain data subsets repetitively, while possibly ignoring unused subsets with high contribution potential. Therefore, a potential risk with basing data management decisions solely on frequency-driven metadata is a possible loss of opportunity to benefit from data subsets that consumers have neglected to use so far - e.g., due to removing of less-frequently used datasets, or neglecting to maintain their quality at a high level. There is possibly no "clear cut" answer to this question, as it largely depends on business context, the timing, and the nature of usage tasks. However, we suggest that important insights can be gained from tracking not only the frequency of data usage, but also the decisions made based on data retrieval and the associated value gains. We term this novel approach value-driven usage metadata (Figure 2b).

The notion underlying this approach is that the purpose of using data and information resources is generating value. Moreover, Shapiro and Varian (1999) argue that information resources and products don't possess "stand-alone" value, but can rather be attributed with value only through usage and experience. The benefits gained from the use of information products have been conceptualized as utility (Ahituv, 1980), which can be measured in terms of revealed value (objective, measurable), perceived value, and/or willingness to pay. Capturing value measurements (e.g., production increase, sales activity, revenues and costs) and storing them in dedicated data repositories is a common practice in organizations today. However, organizations rarely link these value measurements to the data resources and the decision-support tools that were used in the process of value generation. Aral et al. (2006) argue for the importance of creating such a link, and show that it can help explaining information-workers productivity.

Utility assessments have been used to optimize the configuration of data processes (Ballou et al., 1998), DW datasets (Even et al., 2007), and data-quality improvement policies (Even and Shankaranarayanan, 2008) – typical data administration tasks. We suggest that, beyond the benefits offered to data administration, collecting quantitative value assessment as a form of metadata can improve data consumption as well. The novel usage-tracking approach promoted in this study (Figure 2b), extends the current common approach (Figure 2a), by combining measurements of both the frequency of use and of the associated business-value assessments. In certain cases, value assessments can be based on data that exists within the same data resource (e.g., sale transactions, linked to marketing campaigns that were based on analysis of previous sales). In other cases – such assessments may reside in other information resources such as CRM and accounting systems. The integration is done by observing data at the decision-task level – observing the value gained by decision task outcomes and linking it to the queries that have supported each task.

Once the link between decision tasks and queries is established, different methods can be considered for attributing value to specific data subsets. For illustration, we describe here a relatively simple method, which assumes that value is attributed to the last in a sequence of queries that support a decision task. We assume that to support a certain decision, users query repetitively a certain tabular dataset with N records indexed by [n] and M attributes indexed by *[m]*. We consider Q queries indexed by *[q]*, each associated with a business value V^q. The binary indicator $R^q{}_n$ indicates whether record *[n]* was retrieved by query *[q]* $(R^q{}_n = 1)$, or not $R^q{}_n$. Similarly, $R^q{}_m$ indicates whether attribute *[m]* participated in query *[q]* or not. The value of a certain query (V^q) is attributed between the participating data items, using a certain value-attribution function $V^q{}_{n,m}=u(V^q,\ R^q{}_n,\ R^q{}_m)$, such that $V^q=\Sigma_n\Sigma_m V^q{}_{n,m}$. For simplification, we use here an equal attribution of value among all participating data items. Accordingly, the overall value attributed to a certain data items $V_{n.m}$ is given by:

$$V_{n,m} = \sum_{q=1..Q} V^q_{n,m} = \sum_{q=1..Q} u\left(V^q, R^q_n, R^q_m\right) = \sum_{q=1..Q} R^q_n R^q_m V^q \Big/ \left(\sum_{n=1..N} \sum_{m=1..M} R^q_n R^q_m\right), \text{ where} \quad (1)$$

Q, q -	The number of queries performed, and the corresponding index, respectively
M, N -	The number of attributes (indexed *[m]*) and records (indexed *[n]*), respectively
V^q, $Vq_{n,m}$, u -	Query *[q]* value, its attribution to data item *[n,m]*, and the attribution function used
$R^q{}_n$, $R^q{}_m$ -	Binary indicators of the participation (=1) of record *[n]* and attribute *[m]* in query *[q]*

To illustrate value-driven metadata, we use a simplified example of a table used by marketing associates to decide which customers will be approached when promoting a new product. An associate would use a BI tool to investigate previous sale transactions. The tool generates queries directed to the tables, such as those demonstrated, and we assume that each query has led to a certain promotion campaign in which a group of customers has been approached. Some customers may have responded to the campaign by making purchases, and the overall value attributed to a query is the total purchase amount. As illustrated, this value proxy may vary among queries. We now allocate it (Eq. 1) to assess the relative value of each data object (the cumulative value is indicated per record and per attribute). For comparison, we also calculate the frequency-driven metadata – i.e., the number of times that each record or attribute participated in queries (this number is indicated in brackets, near the value score).

Customers

#	Customer	Gender	Income	Children	Status	…	Value
1	Abraham	Male	High	0	Single		1000 (1)
2	Sarah	Female	Low	1	Married		510 (2)
3	Isaac	Male	Medium	2	Married		50 (1)
4	Rebecca	Female	Low	0	Single		10 (1)
5	Jacob	Male	Medium	3	Married		50 (3)
6	Lea	Female	High	2	Married		1510 (3)
7	Rachel	Female	Low	4	Single		0
…							
Value		515 (3)	2000 (1)	60 (2)	500 (1)		

Value	Color
0	
< 100	
< 1000	
≥1000	

Queries

SQL-WHERE condition	Attributes Used	Records Used	Value
Gender = 'Male' and Children > 0	Gender, Children	[3], [5]	100
Gender = 'Female' and Children < 3	Gender, Children	[2], [4], [6]	30
Gender = 'Female' and Status = 'Married'	Gender, Status	[2], [6]	1000
Income = 'High'	Income	[1], [6]	2000

Figure 3. Frequency-Driven and Value-Driven Usage Metadata Assessment

As illustrated by color-coding – some records and attributes possess higher value than others. Moreover, the "value-map" looks significantly different than the one based on the attribution of usage frequency. For example, the Income attribute, which was less frequently used, is associated with the highest value, while the Children attribute, which was more frequently used, is associated with lower value. Insights as such can be transformed into valuable recommendations for marketing associates the next time they plan to run a similar campaign.

The experiment described next tested FRM driven by usage-tracking (both frequency-driven and value-driven). As a preliminary step, we have successfully implemented a prototype of a metadata module, which permits a front-end BI tool to access the usage-tracking scores on demand through function calls. These scores can be integrated into front-end tools, enhance the visual presentation, and communicate important information to both data consumers and administrators. A key challenge with the value-driven metadata collection approach is the fact that most data environments today are not designed to establish an explicit link between decision outcomes and the underlying data and queries. An implicit link can be possibly created through inference mechanisms - e.g., by comparing user identifiers and time-stamps (e.g., in (Even and Shankaranarayanan, 2008)). We acknowledge this challenge as a major limitation of our study that would certainly requires more research.

3 LABORATORY EXPERIMENT

Our study can be seen as design-science research. Design science is a problem-solving paradigm that targets the creation of new artifacts toward improving IS implementation and use (March and Smith, 1995). The success of design-science research is judged by the quality, the contribution, and the impact of the developed artifacts (Hevner et al., 2004). To back our arguments about the contribution potential of the new concepts presented (FRM integration into BI tools, and value-driven usage tracking), we have performed a laboratory experiment for studying and assessing them further in a controlled environment. The experiment was guided by the following questions:

- *Does FRM integration improve decision outcomes?* We assume that FRM integration will significantly improve decision-making outcomes; however, repetitive use is other important factor that may explain such improvement. As end-users repetitively use a decision-support tool, and get more familiar with the underlying data and the decision task – it is reasonable to assume that some improvement to the decision outcome will be observed, regardless FRM integration. However, our assumption is that decision-outcome improvement with FRM integration will be above and beyond the improvement gained with repetitive use alone.
- *Does FRM integration affect usage style?* We assume that FRM integration will significantly change the way people use a BI tool, in terms of making data navigation more focused. Again, we also assume here that repetitive usage will play an important role in changing usage style, but that the changes detected with FRM integration will be significantly greater, above and beyond the changes caused by repetitive usage alone.
- *Do end-users recognize the contribution of FRM integration?* End-users' perception plays an important role in the success of IS artifacts. The TAM (technology acceptance model) and other theoretical models suggest that a sense of usefulness and ease of use will increase the likelihood of acceptance and adoption (Venkatesh et al., 2003). We hence see importance in assessing the end-users' perception of FRM integration, assuming that its contribution to better usability and performance will be well recognized (the design of the survey instrument we used to measure end-users' perception was indeed inspired by previous TAM studies)
- With respect to all the questions above*, would different forms of FRM lead to significantly different results?* When discussing frequency-driven versus value-driven collection of usage metadata earlier, we have suggested that the latter form is likely to be superior to the former. In our experiment, we explore whether FRM generated by value-driven usage tracking will indeed outperform frequency-driven. Further, we explore whether certain forms of value assessment and attribution will significantly outperform others.

3.1 Experiment Settings

Our experiment simulated a marketing task – given a list of customers, choose segments that will be targeted in a promotion campaign. To aid this task, users were provided with a BI tool, similar to the one shown in Figure 1, and a simulated dataset of past sale transactions, which could be analyzed by the tool (the task, the tool, and the underlying database are further described in Appendix A). The experiment was conducted with 200 participants, all undergraduate engineering students around the same age. Most participants were majoring in IS engineering, in their 3rd year of study or later. Some participants have indicated previous exposure to marketing tasks and/or BI applications; hence, we have controlled for these effects as well. All sessions were conducted in labs with similar room conditions, each with identical personal computers, and were all scheduled to similar hours. All participants received some course credit. Additional cash prizes were offered to 5 participants picked by a raffle, in which the chance of winning was correlated to the decision outcome.

Each participant was asked to attend two one-hour sessions. In the beginning of each session, the participants were given a scripted description of the task and instructions on using the tool. After this introduction, the participants were asked to perform the same decision task 6 times repetitively, given a maximum of 5 minutes per repetition, after which the performance measures per repetition (units sold, costs, and net-benefit) were recorded. We have recorded a few measures per repetition that reflect users' interaction - the time spent per decision task, the number of customers and segments chosen, and the number of mouse-clicks made. Upon completing 6 repetitions, the participants were asked to fill out a survey, containing questions on a Likert scale of 1-7 (Appendix A)

In the first session, all participants were provided with the identical BI-tool version with no FRM included. This session served a few purposes – first, to serve as a baseline for assessing the impact of FRM usage. Second, to familiarize the participants with the task and with the BI tool, and third, to collect usage metadata – as soon as each task repetition was completed, the decision value (the net-benefit), and the segments selection were passed to a usage-tracking metadata module. Using these inputs, the module calculated how frequently each data item in the customer dataset was used, and the value attribution among data item, using the attribution method described earlier.

For the 2nd session, participants were divided randomly into 5 groups, each provided with a different FRM variant:

1. *No FRM* – participants in this control group received the same BI tool in both sessions.
2. *Frequency-Driven FRM* - this version uses FRM based on usage frequency. The rate per characteristic value reflected the number of times this value was used for defining a segment.
3. *Value-Driven FRM* – here, the FRM reflects the relative value contribution, based on the net-benefit achieved in each decision task. The rate reflects the cumulative value attributed to each characteristic category.
4. *Expert FRM* – this version is similar to the value-driven FRM, but here the value attribution is based only on the best 20% scores achieved (the rate, 20% was chosen arbitrarily).
5. *Subjective FRM* – here, we used the users' perceived performance in the first session as a value proxy. Unlike the 3 other FRM form that rely on objective value, this kind of FRM rely on users' subjective assessment. We will note an important difference in the scoring of this group - while the objective value scores were collected one per repetition, the subjective score was collected only once per session; hence, we have attributed the same score to all the 6 tasks within a session – what biases the results of this group to an extent.

FRM Group	**Participants (M/F)**	**Age μ/σ**	**3rd Year**	**IS major**	**English Fluency**	**Native**	**Marketing Exp.**	**BI Exp.**
1. None	39 (20/19)	26.3/1.8	87%	95%	100%	95%	21%	15%
2. Frequency	42 (21/21)	26.2/1.2	86%	93%	98%	88%	17%	17%
3. Value	40 (15/25)	25.8/1.6	100%	95%	100%	88%	10%	20%
4. Expert	40 (18/22)	25.9/1.8	95%	90%	97%	93%	10%	10%
5. Subjective	39 (18/21)	26.1/1.7	93%	90%	94%	92%	10%	10%
Overall	**200 (92/108)**	**26.0/1.7**	**92%**	**93%**	**98%**	**91%**	**14%**	**86%**

Table 1. Experiment Groups (μ: Avg. , σ: STDEV).

The participant groups and the associated demographics are summarized in Table 1. All participants completed the assignments successfully and no data collection issues could be detected. The experimental task, tools, and procedures were all tested in a pilot study done with 6 research students. The pilot ensured that all experimental procedures work properly, and helped testing the survey toward identifying unclear and/or ambiguous questions.

From the collected data, we have measured the following variables (Table 2) - some objective, performance-based measurements (average along 6 decision tasks per session), and others perceptive, based or on survey questions:

Objective: Decision Value, Time, Segments, and Clicks – the average net-benefit, time (in seconds) spent on a decision, number of segments chosen, and number of mouse-clicks made, respectively

Perceptive: Performance, Usefulness, Ease-of-Use, Acceptance – based on the questionnaire given in both sessions (the FRM-related questions were given only in session B)

	Variable	**S1: μ**	**S1: σ**	**S2: μ**	**S2: σ**	**FRM: μ**	**FRM: σ**
Objective	**Decision Value**	30.90	498.00	574.52	593.11		
	Time	163.74	90.02	156.24	94.85		
	Segments	2.47	1.43	2.30	1.32		
	Clicks	82.37	39.76	46.75	24.89		
Perceptive	**Performance**	4.88	1.12	5.19	1.14	5.10	1.78
	Usefulness	4.82	1.20	4.81	1.18	5.24	1.50
	Ease of Use	4.79	1.28	4.80	1.24	5.23	1.40
	Acceptance	4.37	1.35	4.32	1.34	5.38	1.43

Table 2. Summary Statistics (S1/2: Session 1 or 2, μ: Avg. , σ: STDEV).

Testing these variables against the control variables (Gender, Age, Year of Study, Major, English Fluency, Country of Birth, Exposure to Marketing Task, Exposure to BI Tools), has shown no significant effect.

3.2 FRM Integration and Decision Outcomes

To assess this question, we have used the "Decision Value" variable. Our base assumption that repetitive use would improve decision outcomes was indeed supported: considering the control group alone, which received the same BI tool in both sessions (group 1, 29 participants), the average decision value increased from 93.7 in the first session to 381.2 in the second. An ANOVA test shows that the increase was indeed significant (F-value: 5.90, P-Value: 0.017).

When all 5 groups considered (Table 3) – the decision-value increase between sessions is even more significant (F-Value: 98.54, P-Value: ~0). As expected, with respect to first-session performance there is no significant difference between the groups (F-Value: 0.399, P-Value: 0.809). However, the difference among groups in the average second-session performance is significant (F-Value: 23.47, P-Value: ~0). With group 3 (Value) and 4 (Expert) – the increase was sharp and highly significant, where the latter increase is slightly greater than the former. With group 2 (Frequency), the increase was significant, but only marginally higher than the increase gained by the control group. With group 5 (Subjective), the increase was insignificant and even lower than the control-group's increase.

We next evaluate the survey-based perceived performance, noticing its high, positive and significant correlation to the decision-value (0.419 in session 1, 0.487 in session 2). For the control group, the perceived performance slightly increased from 4.93 in the 1st session to 5.10 in the 2nd; however, insignificantly (F-value: 0.42, P-Value: 0.521). When all 5 groups considered, the overall performance-score increase from 4.87 to 5.18 is significant (F-value: 7.41, P-Value: 0.007). As expected, the difference among the groups in session 1 is insignificant (F-Value: 0.68, P-Value: 0.608), but the differences are significant in the second-session scores (F-Value: 4.96, P-Value:

0.001), and in the level of score-increase (F-Value: 8.175, P-Value: ~0). When observing the increase per group, only group 3 had a significant increase in the score from 4.63 to 5.70 between the sessions (F-Value: 23.76, P-Value: ~0), group 1, 2, and 4 had only a minor and insignificant increase, and group 5 had some insignificant decrease – what means that the fact that the overall increase was significant, can be attributed mainly to the increase in group 3.

FRM Group	S1: μ	S1: σ	S2: μ	S2: σ	Δ: μ	Δ: σ	Δ: F	Δ: P.V.
1. None	93.72	504.19	381.20	540.02	287.48	511.02	5.90	0.017
2. Freq.	36.98	508.13	359.63	560.24	322.65	600.88	6.76	0.011
3. Value	28.19	408.95	968.13	372.66	939.94	544.22	107.97	~0
4. Expert	-46.35	526.68	987.69	370.71	1034.04	561.46	99.13	~0
5. Subj.	43.55	546.87	171.79	574.20	128.24	577.89	0.68	0.410
Overall	**30.90**	**497.99**	**574.52**	**593.1**	**543.62**	**666.89**	**98.54**	**~0.000**

Table 3. Decision Value Analysis (S1/2: Session 1 or 2, μ: Avg. , σ: STDEV, Δ: Difference, F: F-Value, P.V.: P-Value).

3.3 FRM integration and Usage Style

To assess the effect on usage style, we have used the “Time”, “Segments”, and “Clicks” variables. We assumed that, as users get more familiar with the decision tool through repetitive usage, they will tend to reach a decision faster, refine the decision by extending the number of segments chosen, and reduce navigation intensity – i.e., use the mouse less often to change data segmentation. Further, we assumed that FRM-inclusion will even increase these effects. The findings for “Time” and “Segments” did not support our assumption. The correlation between these variables was positive, high and significant (0.89 in the first session, 0.83 in the second). However, their scores did not show significant difference neither between the two sessions, nor among FRM groups within a session.

Conversely, the change in “Clicks” was apparent (Table 4). Considering the control group alone, the average clicks per task decreased significantly from 86.33 in the 1st session to 64.33 in the 2nd (F-Value: 5.80, P-Value: 0.018). When considering all 5 groups, the decrease in clicks between sessions is even more significant (F-Value: 115.34, P-Value: ~0). As expected, there was no significant difference between groups in 1st-session clicks (F-Value: 1.92, P-Value: 0.109); however, the difference between groups in 2nd-session clicks is significant (F-Value: 7.68, P-Value: ~0). Significant decrease in the number of clicks could be detected for each group that was provided with FRM, where all decreases are greater than the decrease in the control group. Interestingly, the number of clicks decreased significantly with no apparent correlation to the quality of the decision-outcome improvement offered by the FRM – for example, the decrease in group 2 was a lot greater than the decrease in group 4, although the decision-value improvement in group 4 was much higher (Table 3). This may imply that when FRM are provided, end-users tend to accept the recommendations and reduce their navigation intensity, regardless the quality of the recommendations.

FRM Group	S1: μ	S1: σ	S2: μ	S2: σ	Δ: μ	Δ: σ	Δ: F	Δ: P.V.
1. None	86.33	47.83	64.23	31.53	-22.10	5.94	5.80	0.018
2. Freq.	90.58	39.56	39.31	16.20	-51.27	5.35	60.41	~0
3. Value	88.10	39.82	45.18	20.01	-42.92	5.71	37.09	~0
4. Expert	70.87	39.10	46.77	27.28	-24.10	5.20	11.75	~0
5. Subj.	75.44	33.09	38.84	18.65	-36.60	4.74	36.23	~0
Overall	**82.36**	**39.75**	**46.74**	**24.88**	**-35.62**	**2.52**	**115.34**	**~0**

Table 4. Clicks Analysis (S1/2: Session 1 or 2, μ: Avg. , σ: STDEV, Δ: Difference, F: F-Value, P.V.: P-Value).

To assess the extent to which interaction style affects decision value, we ran a regression of decision-value scores in session 2 versus time, segments, and clicks (Table 5). When lumping together the 5 groups, all 3 regressions are significant with positive slopes – meaning that the decision-value increases with the more time spent, with a larger number of segments chosen, and with higher navigation intensity. However, the adjusted R-Square for these regressions is small, meaning that the decision performance is explained only to a small extent by these factors.

FRM Group	Time				Segments				Clicks			
	α	**T**	**P.V.**	**A-R^2**	**α**	**T**	**P.V.**	**A-R^2**	**α**	**T**	**P.V.**	**A-R^2**
1. None	2.01	2.73	0.01	0.15	127	3.02	0.01	0.17	3.38	1.22	0.23	0.01
2. Freq.	0.37	0.37	0.71	-0.02	84	1.33	0.19	0.02	9.72	1.84	0.07	0.06
3. Value	1.59	2.03	0.05	0.07	217	4.07	~0	0.29	7.13	2.55	0.01	0.12
4. Expert	1.16	1.82	0.07	0.06	140	3.48	~0	0.22	2.61	1.11	0.23	0.01
5. Subj.	1.41	1.73	0.09	0.05	220	2.08	0.04	0.08	0.93	0.18	0.85	-0.26
Overall	**1.49**	**3.44**	**~0**	**0.05**	**149**	**4.95**	**~0**	**0.11**	**3.79**	**2.26**	**0.02**	**0.02**

Table 5. Value Regression against Time, Segments, and Clicks (Session 2; α: Slope, T: T-Test, P.V.: P-Value, A-R2: Adjusted R-Square).

When treating each group individually, time seems to have a significant effect on the decision value only for the control group. The other time effects on decision outcome are less significant, with a smaller regression slope. The positive effect of clicks on the decision was significant only for group 3 (Value FRM). For most groups (with the exception of group 2), the number of segments significantly affected the decision value. This result not surprising, as maximizing the net-benefit in the given decision task would require selecting a large number of small segments. Interestingly, this optimal decision policy was not explained explicitly to the participants. However, participants who understood and adopted this policy, regardless the form of FRM provided – indeed performed better.

3.4 Do End-Users Recognize the Contribution of FRM Integration?

To evaluate end-user perception, participants were asked to fill in a survey at the end of each session. The survey measured four constructs (besides assessing the control variables) – Performance, Usefulness, Ease-of-use, and Acceptance – each operationalized with a few questions (Appendix A). Some questions were asked in both sessions while others, which were more specific to the FRM inclusion, were included only in the second session.

	Usefulness				Ease of Use				Acceptance			
FRM Group	**μ**	**σ**	**Val. C.**	**P-P C.**	**μ**	**σ**	**Val. C.**	**P-P C.**	**μ**	**σ**	**Val. C.**	**P-P C.**
1.	4.51	1.23	.20	**.50	4.68	1.34	*.39	**.79	3.92	1.53	.31	**.55
2.	4.47	1.28	**.45	**.51	4.52	1.32	.29	**.56	3.83	1.42	*.37	.28
3.	5.09	.88	.18	*.40	5.35	.95	.17	**.81	4.39	1.13	.19	*.39
4.	5.26	1.26	**.41	**.69	5.04	1.19	*.32	**.78	4.84	1.16	.28	**.62
5.	4.71	1.00	.19	**.44	4.41	1.16	.05	**.68	4.65	1.17	.01	.23
Overall	**4.81**	**1.17**	****.37**	****.53**	**4.80**	**1.23**	****.34**	****.73**	**4.34**	**1.34**	****.26**	****.38**
Variability	**F-Value:3.81 , P-Value:.05**				**F-Value:4.13 , P-Value:.03**				**F-Value:4.67 , P-Value:.01**			

*Table 6. Usefulness, Ease of Use, and Acceptance Analysis (Session 2, μ: Avg. , σ: STDEV, C.: Correlation, Val- Value, P-P: Perceived Performance; *Significant with P-Value < .05, ** - Significant with P-Value < .01).*

With respect to the three constructs that are typically examined in TAM studies – Usefulness, Ease of Use, and Acceptance – there was no significant difference between the sessions, neither when examining each FRM group separately, nor when lumping together all the groups. However, some significant variability could be detected between the groups in the scores of the 2nd session (Table 6). The scores for groups 3 and 4 were generally higher, what coincides with the fact the decision-value improvements for these two groups were higher than the rest. The correlations between usefulness, ease-of-use, and acceptance versus perceived performance were mostly positive, high, and significant – both when observing each group separately, and when lumping them together. However, this wasn't the case with respect to correlations with the objective decision value – the correlations were all positive, but mostly insignificant, and lower than the corresponding correlations with performance.

With respects to the scores based on FRM-related questions, a significant difference between the groups could be detected for usefulness, ease-of-use, and acceptance (Table 7). In this case too, the scores for groups 3 and 4 were higher than the others – what corresponds to the higher decision improvement in those groups. The correlations to the perceived performance are consistently high, positive and significant. Conversely, the correlations to the corresponding decision values are lower, insignificant, and sometimes even negative (particularly, in groups 2 and 5, who received the less-helpful FRM variants). This inconsistency in the correlations is surprising, considering the fact the all average scores agreed with the performance-improvement.

	Usefulness				**Ease of Use**				**Acceptance**			
FRM Group	**μ**	**σ**	**Val. C.**	**P-P C.**	**μ**	**σ**	**Val. C.**	**P-P C.**	**μ**	**σ**	**Val. C.**	**P-P C.**
2. Freq.	4.10	1.41	-.17	**.86	4.82	1.35	.00	**.42	4.51	1.42	-.15	**.84
3. Value	6.19	1.10	.18	**.91	5.92	0.98	.07	**.60	6.05	1.07	.13	**.73
4. Expert	5.95	0.98	.20	**.68	5.60	1.30	.27	**.57	5.98	1.12	.07	**.61
5. Subj.	4.75	1.39	-.26	**.71	4.59	1.52	-.10	**.54	4.98	1.44	-.1	**.71
Overall	**5.24**	**1.50**	****.26**	****.83**	**5.23**	**1.40**	****.25**	****.61**	**5.37**	**1.42**	****.21**	****.78**
Variability	**F-Value:25.76 , P-Value:~0**				**F-Value:9.34 , P-Value:~0**				**F-Value:14.38, P-Value:~0**			

*Table 7. Usefulness, Ease of Use, and Acceptance Analysis (Session 2, FRM-related questions; μ: Avg. , σ: STDEV, C.: Correlation, Val- Value, P-P: Perceived Performance; * Significant with P-Value < .05, ** - Significant with P-Value < .01).*

3.5 Discussion

Overall, we see this experiment as successful. The number of participants was relatively large, and the data collection procedures all worked well, what permitted getting a clean and complete dataset and results with high significance. The results provide some interesting and thought-provoking results. Some support our preliminary assumptions to a great extent, and some to a lesser extent, but in either case – they provide some important and useful insights that may guide future investigation.

A key assumption of our study was confirmed by the finding – certain forms of FRM-integration into a BI tool indeed improved decision outcomes significantly, above and beyond the improvement gained with repetitive use alone. The improvement was recognized by end-users – the survey-based scores given by groups who worked with FRM versions that permitted decision improvement were consistently and significantly higher than the scores given by the control group, or by groups who received lesser FRM variants. Moreover, FRM-integration has not also affected the decision outcome, but also navigation intensity – based on the clicks-count, it appears to be that participants who received recommendations tended to use them, and navigate less. On one had – this may point out a certain advantage of FRM-integration, in terms of promoting a more focused and efficient navigation. On the other hand, the results may reflect some risk – it appears that users tended to follow the recommendations even thought they have recognize that the recommendations do not lead to any major improvement. This underscores the need for further investigation, that would detect which forms of FRM work better, and which should be avoided.

Another key assumption of our study - that value-driven FRM will outperform frequency-driven FRM – was also supported to a great extent. With members of groups 3 and 4, who received recommendations based on value-mapping, the increase in performance was high and significant. This was true for both variants – value attribution that considered the results of all 1st-session tasks, as well as value attribution that considered only the 20% best scores. Interestingly, the improvement gained with "expert-based" recommendations was only marginally greater than the improvement gained by considering all scores, as we expected that the gain would be significantly greater. The small difference could be incidental, however it raises a question – would it be sufficient to create value-driven usage metadata for generating FRM generate based on a sample of tasks, rather than on the whole population? If yes, would be the optimal sampling policy? Answering this would require some more analytical and empirical investigation. The low magnitude of improvement gained by using frequency-driven FRM was somewhat surprising. Although we assumed value-driven FRM to outperform frequency-driven FRM, we still expected to see some improvement with the latter, beyond the improvement gained by repetitive use, and some more investigation will be required to detect whether these results are incidental. Notably, implementing frequency-based usage tracking is less demanding technically (and likely, less expensive) than implementing value-driven usage tracking – indeed, usage-tracking in data environment today, if done at all, is frequency-driven in nature.

A result that has particularly raised our attention is the low improvement gained with FRM that were based on users' subjective assessment - the group that received this form of FRM had the lowest gain in the second session (even lower than the gain achieved by the control group), even though the perceived performance scores were highly correlated with the actual decision value. The reason for this lower performance is not clear, and the use of subjective assessments for generating FRM indeed requires some more investigation. The different nature of this FRM can be possibly attributed in part to the averaging effect, caused by the different value-attribution method used, as underscored earlier. We would hasten to say that this result is particularly interesting, due to the fact that many common collaborative filtering recommender systems are based on the user's subjective assessments. This result may indicate that, in the BI context, be cautious about using subjective assessments to generate recommendations.

4 CONCLUSIONS

Our research investigated the integration of feedback and recommendation mechanisms (FRM) into BI tools. The experiment that we conducted had confirmed our assumption that such integration can improve decision-making outcomes, when the right forms of FRM are provided. Recommendation mechanisms have been investigated in the context of E-commerce and digital libraries, and we would argue that similar motivations to those that have directed recommender-system research so far ought would promote integration of similar systems in BI environment - improving decision-making outcomes, enhancing end-user's experience, and reducing information overload. Another key contribution of our study, which links to the previous, is the introduction of a novel approach for usage tracking in data environments. This approach suggests that integrating quantitative assessments of usage-frequency together with the associated value gained would offer substantial benefits to both data administration and consumption. Joint frequency and value assessments can help identifying unused data subsets with high value-contribution potential and, consequently, motivate new usage forms. Complementing frequency assessments with value assessments may "close a broader loop", in terms of providing feedback based on usage performance, and reducing the potential risks. First, value allocation gives higher weight to past usages with high contributions. Second, it can reflect variability in the importance of different subsets depending on the usage context. Lastly, it can help detecting data subsets with high contribution potential that have not been frequently used. The experiment indeed confirmed our assumption that recommendations based on value-driven assessment of previous data usage would outperform frequency-driven recommendations.

Obviously, some research will be needed to address some key limitations with our study. First is the issue of quantifying data value. As stated often in past research, quantifying the value of information resources is a challenging task, particularly with respect to data resource. Organizations maintain performance measurements (e.g., productivity, income, and profitability) that can be possibly linked to decision tasks. However, decision performance may depend on resources other that data – e.g., human knowledge and financial assets. Further, the value depends on the data-usage context, and value assessment for a certain type of use does not necessarily apply to others. Further, value varies with time, as data that can be used effectively at a certain point of time, might become obsolete later. We hasten to say that the value-allocation methodology, which we apply in this study, appears to be a better fit to operational environments in which decision tasks have a high degree of repetition, and in which causal relations between data usage and business performance are easier to establish. Promotion-campaign management, such as in our illustrative example would be a good representative for this type of decision-making. Financial-investment decisions would be another example for data-driven decisions, in which outcomes are measurable (e.g., the change in the value of the financial asset) and linkable to the data resources being used. Conversely, quantifying the value of decision outcomes might turn out to be challenging in strategic decision scenarios, which are not repetitive in nature and often relay on information resources other than organizational data repositories.

Another issue that needs addressing is a method for linking value to specific queries. Performance assessments are rarely linked explicitly to the data resources and tools used. Our preliminary prototype includes inference mechanisms for creating implicit links – e.g., based on the user name, and/or time proximity. Obviously, implicit links cannot be absolutely precise and might bias the value allocation significantly. Establishing explicit links will require stronger metadata integration between systems and, likely, redesign of data environments (e.g., joint codes that link each decision task and queries). One could question whether or not making such a high investment in redesigning data environments and BI tools would justify the benefits gained. There is also a need to explore further how to attribute value to specific data objects, as the attribution method may have a critical impact on the results. In our experiment, we have attributed value only to the last query in the sequence that led to the decision, and distributes the value equally among all the data items involved. A different allocation method may consider, for example, allocating the usage value among all queries and/or consider unequal allocation.

Further extensions may be also by testing the integration of FRM based on other recommender mechanisms. One possible approach would be to consider task and user characteristics, as the same DW/BI environment can be used to support a plethora of business processes and tasks, each with very different data usage needs. FRM capabilities can take into account such needs, by creating either task profiles that capture specific task characteristics or by asking expert users to identify certain data elements or analysis results that are more useful and relevant for a given task. FRM capabilities can also be driven by analysis of the data using algorithmic data-mining techniques. Data mining algorithms typically explore the information contained in data sources automatically and provide end-users with quantitative assessments and to enable them to extract and evaluate knowledge from this information. Data mining can suggest alternatives to decisions and actions that are about to be taken and allow users to re-consider these decisions and actions. A possible drawback of using statistical analysis and data mining to derive FRM is the risk that recommendations will be based solely on the data resource, without considering the context in which it is used.

Finally, future research should also address organizational and economic perspective of the proposed concepts. As organizations today do not possess the appropriate infrastructure for value-driven collection of usage metadata, adopting this approach may mean a substantial investment. One could ask – would the benefits of adopting value-driven usage tracking and using it for FRM integration justify the cost? Our experiment indicates that usage-driven FRM can indeed improve decision outcome; however, the results were obtained under a simulated and well-monitored environments; hence, cannot be generalized to real-world scenarios. Evaluating and testing value-driven metadata and FRM integration in real-world environments would therefore be an important follow-up step in furthering this line of research, toward gain more knowledge and insights about the challenges that firms may face when implementing these new concepts that we propose.

APPENDIX A – EXPERIMENT TASK AND TOOLS

The experiment task was defined as – given a list of customers, choose those who will be targeted in a promotion campaign that offers a certain product. Due to certain costs (e.g., printing charges, mailing fees, and call time), promoting the product to the entire list would be sub-optimal, as only some customers are likely to make any purchase. An optimal decision would therefore be to target only those customers with a high likelihood to purchase enough units to justify the promotion cost. This decision can be formulated as maximizing an objective function:

$$V = \left(\sum_{m=1..M} I_m \left(PQ_m - C^V\right)\right) - C^F \text{ , where} \tag{2}$$

V -	Net-benefit, the decision value
$M,\{I_m\}$ -	The total number of customers (indexed *[m]*), and the set of binary variables, each indicating whether to include customer *[m] (=1)* or not *(=0),* respectively
P, Q_m -	Unit price, and the expected units that customer *[m]* purchase $(Q_m \geq 0)$, respectively
C^F, C^V-	Fixed campaign cost, and variable promotion cost per customer, respectively

The optimal decision would be to include only customers for which expected revenue is greater than promotion cost (i.e., $I_m=1$, when $P^Q{}_m > C^v$). However, while costs and unit price are typically known, the expected quantity of units is subject to uncertainty. Marketing professionals often estimate purchase intents by analyzing past sales, and identifying segments of customer who are more likely to accept a promotion offer. Segments can be defined along certain customer characteristics and the associated values (e.g., "target male customers with no children who are between 30-45 years old"). Accordingly, the task was defined as – given (a) a list of customers, each with a known set of characteristics values (e.g., age, gender, occupation, etc.), (b) data on past purchases, and (c) unit price and cost parameters, choose the customer segments that would maximize the campaign's net-benefit.

To aid this task, participants were provided with a BI tool similar to the one shown in Figure 1, offered in two base versions – with and without FRM. The version with no FRM permitted the following BI functionality: (a) "Slicing": upon selecting a customer characteristic, the bar graph at the bottom summarizes the number of customers under each category, and the total sales associated with those customers. The tool presented 12 categorical customer characteristics, each with 3 possible values (e.g., High/Medium/Low Income). To prevent biased pre-assumptions on the predictive power of each characteristic, the characteristic names were coded with capital letters (A, B, C, …, L) and the associated value with enumerated lower-case letters - e.g., instead of High/Medium/Low Income, users would see a characteristic named A with associated values {a1, a2, a3}, (b)"Filtering": when slicing along a certain characteristic, a user would have the choice to limit the presentation to show only certain values (e.g., under characteristic F, show only values f1 and f3), and (c) "Drilling": after slicing along a certain characteristic, a user would have the choice to slice the data further along others. The numbers would then be summarized and presenters along the value-combinations of all the characteristic included – for example, if H(h1, h2) and G(g1, g3) where chosen, the tool will present summaries for the combinations {h1, g1}, {h1, g3}, {h2, g1}, and {h2, g3}.

The FRM version included all the above BI functionality. However, a certain rate was added per characteristic and value, indicating a certain recommendation, which reflects potential contribution to a better decision (the rate-calculation method will be explained later). A high characteristic-value rate (r) would indicate a favorable customer segment - e.g., if, under D, r(d1) >> r(d2), the recommendation would be to prefer customers who belong to category d1. A high characteristic rate (R) would indicate a high variance among the rates among characteristic values; hence, a higher likelihood to single out better customer segments. If, for example, R(J) >> R(K), would mean a high variability between the rates of {j1, j2, j3} versus low variability between the rates of {k1, k2, k3}; hence, a recommendation to prefer slicing the data along J rather than along K. Importantly, the FRM-enhanced tool recalculates the rates dynamically, depending on the characteristic-value combination observed, meaning that at each point of time the user would get a recommendation how to proceed which depends on the "Slicing", "Filtering" and "Drilling" choices made so far.

The simulated database included a list of 1000 customers, each with 12 associated characteristic values (e.g., {a1, b3, …, k2, l1}, which were randomly drawn using given value distributions (e.g., a1:0.3, a2:0.5, a3:0.2). Using a set of rules, which associated certain value combinations with certain levels of purchase intent, each customer [m] was assigned with a set of likelihood numbers $\{P_{m},z\}$ of purchasing z (between 0 and 5) units in a given campaign, such that $\Sigma zPm,z,=1$, and $Q_{m}=\Sigma_{z}ZP_{m},z$, is the expected quantity of units purchased. We then generated a list of sale transactions per customer, by simulating a sequence of campaigns and randomly drawing the quantity of items purchased per campaign (including only transactions with quantity greater than 0). This random draw used the set of purchase-likelihood numbers per customer, with some level of randomness added.

Users were also provided with a utility for selecting one or more customer segments to be targeted. Upon completing the selection – the campaign performance would be evaluated using Eq. 2. The participant would then get a performance report, including the number of customers targeted, units sold, costs, and the net benefit. Table 8 lists the survey-instrument questions and Cronbach's-Alpha scores per session (all relatively high). The questions were presented in a mixed order, and those related to FRM were given only in the second session. The questions listed in table 8 are translated, as the survey was given in the native language of the country where the experiment was conducted.

	Both sessions (S1, S2)	**α (S1/2)**	**FRM-related (S2 Only)**	**α (S2)**
Performance	• I have been successful • The assignment was well understood, and I knew how to perform it well • I wasn't clear about what I was doing • I found the assignment too difficult	0.683, 0.768	• The FRM provided helped improving my performance • The FRM provided were meaningless	0.830
Usefulness	• Using the BI tool I was provided with have helps performing the task • If I was a marketing person, BI tool as such would improve my work • If I was a marketing person, a BI tool would have made my work easier • If I was a marketing person, BI tool as such would have not helped me	0.803, 0.820	• The FRM provided can help using the tool • I didn't use the FRM provided by the tool	0.760
Ease-of-Use	• I easily learned how to operate • I was clear about the information presented by the BI tool • The BI tool used in the experiment was too complex to use • The BI tool I user could not support the actions I needed to perform	0.737, 0.808	• The FRM were presented clearly and in an understandable manner • I couldn't get how to use the FRM	0.777
Acceptance	• I would have recommend a company to invest in a BI tool • If I was a marketing person in a firm, I would have requested to use • If I was a manager, I would have recommended to invest in such a tool	0.904, 0.937	• I would propose to invest in FRM • A company should not invest in FRM such as those provided	0.770

Table 8. Survey.

ACKNOWLEDGMENTS

This project was supported by a grant from Microsoft R&D, Israel.

REFERENCES

Adomavicius, G., and Tuzhilin A. “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions,” IEEE Tr. on Knowledge and Data Engineering (17:6), June 2005, pp. 734-749

Adomavicius, G., and Kwon, Y. “New Recommendation Techniques for Multicriteria Rating Systems,” IEEE Intelligent Systems (22:3), May-June 2007, pp. 48-55

Ahituv N. “A Systematic Approach towards Assessing the Value of Information System,” MIS Quarterly (4:4), Dec. 1980, pp. 61-75

Aral, S., Brynjolfsson, E., and Van Alstyne, M. “Information, Technology and Information Worker Productivity,“ in Proceedings of the International Conference on Information Systems (ICIS), Milwaukee, WI, USA, Dec. 2006

Ballou D. P., Wang R., Pazer H., and Tayi G. K. “Modeling Information Manufacturing Systems to Determine Information Product Quality,” Management Science (44:4), April 1998, pp. 462-484

Cooper B. L., Watson H. J., Wixom B. H., and Goodhue D. L. “Data Warehousing Supports Corporate Strategy at First American Corporation,” MIS Quarterly (24:4), Dec. 2000, pp. 547-567

Counihan, A., Finnegan, P., and Sammon, D. “Towards a Framework for Evaluating Investments in Data Warehousing,” Information Systems Journal (12), 2008, pp. 321-338

Davenport, T.H. “Competing on Analytics,” Harvard Business Rev. (84:11), January 2006, pp. 99-107

DeLone, W.H., and McLean, E.R. "Information Systems Success: The Quest for the Dependent Variable," Information Systems Research (3:1), March 1992, pp 60-95.

Devaraj, S., and Kohli, R. “Performance Impacts of Information Technology: Is Actual Usage the Missing Link?,” Management Science (49:3) 2003, pp. 273-289.

Even, A., Shankaranarayanan, G. and Berger, P. D. “The Data-Warehouse as a Dynamic Capability: Utility-Cost Foundations and Their Implications for Economics-Driven Design,” in Proceedings of the International Conference on Information Systems (ICIS), Milwaukee, WI, USA, Dec. 2006

Even, A., Shankaranarayanan, G., and Berger, P. D. “Economics-Driven Design for Data Management: An Application to the Design of Tabular Datasets,” IEEE Tr. on Knowledge and Data Engineering (19:6), June 2007, pp. 818-831

Even, A., and Shankaranarayanan, G. “Comparative Analysis of Data Quality and Utility Inequality Assessments,” in Proceedings of the European Conference on Information Systems (ECIS), Galway, Ireland, June 2008

Hevner, A.R., March, S.T., Park, J., and Ram, S. “Design Science in Information Systems Research,” MIS Quarterly (28:1), 2004, pp. 75-105

March S.T., and Smith G.F. “Design and Natural Science Research on Information Technology,” Decision Support Systems (15), 1995, pp. 251-266

March, S. T. and Hevner A. R. “Integrated Decision Support Systems: A Data Warehousing Perspective,” Decision Support Systems (43:3), July 2005, pp. 1031-1043

Pfeffer, J. and Sutton, R.I. "Evidence-Based Management," Harvard Business Review (84:1), Jan. 2006, pp. 62-74.

Shankaranarayanan, G., Even, A. (2006), “The Metadata Enigma”, Comm. of the ACM (49:2), Feb. 2006, pp. 88-94

Shapiro C., and Varian H.R., Information Rules, Harvard Business School, Cambridge, MA, 1999

Song, X., Tseng, B.L., Lin, C.Y., Sun, M.T., “Personalized Recommendation Driven by Information Flow,” In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, USA, 2006

Sun, Y., Li, H, Councill, I.G., Huang, J., Lee, W.C., Giles, L. “Personalized Ranking for Digital Libraries Based on Log Analysis, “ in Proceeding of the 10th ACM workshop on Web Information and Data Management (WIDM), Napa Valley, CA, USA, 2008

Venkatesh, V., Morris, M.G., Davis, G.B., and Davis, F.D., “User Acceptance of Information Technology: Toward a Unified View,’ MIS Quarterly (27:3), Sept. 2003, pp. 426-478
Wei, Y.Z., Moreau, L., and Jennings, N.R., “Learning Users’ Interests by Quality Classification in Market-Based Recommender Systems,“ IEEE Trans. on Knowledge and Data Engineering (17:12), Dec. 2005, pp. 1678-1688
Wixom, B.H., and Watson, H.J. “An Empirical Investigation of the Factors Affecting Data Warehousing Success,” MIS Quarterly (25:1), March 2001, pp. 17-41
Wixom, B.H., Watson, H.J., and Hoffer, J.A. “Continental Airlines Continues to Soar with Business Intelligence,” Information Systems Management (25:2), 2008, pp. 102-112