

Managing Procurement Spend using Advanced Compliance Analytics

*Pawan Chowdhary, Markus Ettl, Amit
Dhurandhar, Soumyadip Ghosh, Gopikrishna
Maniachari*

Business Analytics and Math Sciences

IBM T J Watson Research Center

Yorktown Heights, NY 10598

{chowdhar, msettl, adhuran, ghoshs, gopi@us.ibm.com}

Bruce Graves, Bill Schaefer, Yu Tang

Business Analytics and Math Sciences

IBM Global Process Services

{bruceg, wss, tangyu@us.ibm.com}

Often the processes for purchasing commodities and services within a business enterprise are centralized into a procurement organization. These purchases are often sourced from one or more suppliers, or vendors, based on contract terms and conditions (such as price, payment terms etc.), availability, and quality or legacy habit of purchasing service with known vendors. We have found that many organizations lack appropriate processes and disciplines to drive demand to preferred suppliers. Thus these enterprises are unable to leverage the value of the pre-negotiated contracts due to lack of process education, approval process steps or appropriate purchasing tools that could result in significant amounts of spending that would be considered not compliant (not being sourced through preferred suppliers). Depending upon the size of the organization, such transactions range from several million dollars to billions of dollars. Manually sifting or employing typical query tools to review large amounts of spend transaction data with multiple attributes to identify the level of non compliant spend and identify areas to take action is a daunting task.

In this paper, we discuss a software solution for spend compliance analytics that includes measurements of cost savings due to increased compliance and identification of areas where spend tends to be non compliant. We have developed a web enabled advanced analytical solution called Compliance Analytics Tool (CAT) that embeds a two phase methodology for compliance management. In the first phase, we use advanced data mining techniques to segment a large amount of historical spend transactions to quickly identify promising areas of improvement, exploiting a multitude of purchasing attributes such as business unit, procurement category, suppliers, etc. The second phase employs portfolio optimization techniques to further focus on specific segments that provide maximum benefit based on desired compliance targets or available budget. We also discuss the solution architecture that integrates business analytics along with business intelligence tools, dashboards, and data warehousing.

Compliance Analytics, Procurement, Spend Analysis, Data Mining, Optimization, SPSS Modeler, CPLEX.

I. INTRODUCTION

The fundamental monetary units of analysis in an enterprises can be characterized into two major parts; spend and revenue. Companies generate revenue by selling

goods or services. The cost to the company to generate revenue can be broadly classified as spend. This can come in the form of buying raw-materials as input to the goods or services being sold for manufacturing oriented companies to the technology services, software, hardware and salaries to personnel for non-manufacturing entities.

The types of spending generally can be of classified into direct spending (that impacts the product being manufactured such as raw materials, computer memory chip; i.e., cost of goods sold) and indirect spending (such as office supplies, maintenance services, travel, marketing, communication; i.e., sales, general & administrative expenses). This spending is often sourced from multiple suppliers/vendors and many organizations have negotiated contracts with suppliers that determine service levels, quality, price, etc. Some contracts provide back-end rebates where the purchasing organization qualifies for a percentage of spend as a rebate when it achieves a predetermined spend target, or special pricing for volume spend. Procurement spend can run into billions of dollars for Fortune 500 companies to hundred of millions for mid size organizations. Only very few companies have implemented proper processes to audit and monitor critical procurement purchases. Some organizations might have a level of visibility and control over a few procurement products that are core to their business, but beyond this the majority of procurement activities are ad-hoc in nature with minimal supervision. This can lead to severe spend leakage that often goes unnoticed due to lack of visibility or appropriate processes and controls. The ability to manage and control the procurement spending can significantly improve the bottom line of companies, especially during difficult and challenging economic conditions.

The types of procurement data that are typically available are purchase orders (PO), invoices, preferred suppliers lists, supplier contracts, and general ledger entries. There are tools available in the market that process PO and invoice transactions and provide spend visibility along a few standard dimensions such as product, vendor, organization, etc. [1]. Such tools often have a built-in rules engine to assess the quality of the data, cleanse and classify the data to standardize the vendor and product

categories and also identify the PO data as compliant spend (established process) or non-compliant spend (ad-hoc spend) [2, 3]. These are mainly reporting tools that provide drill-down visibility into spending across few dimensions, but they do not have capabilities to deep dive into data and determine the most problematic spend areas across a large number of dimensions [4, 5]. For example, today's spend analysis tools can help procurement managers to identify top non-compliant spend by vendor and product, but they do not further differentiate spend within one dimension by other attributes or provide further analysis into different areas of non-compliant spending across several dimensions. Thus, it often remains unclear where an organization should focus its attention to maximize the return on investment (ROI) when attacking non-compliance.

The Compliance Analytics Tool (CAT), the topic of discussion in this paper, is a web based solution that uses SPSS, iLOG, and Cognos business intelligence (BI) capabilities to deliver state of art data mining analytics to segment the spend transactions across several dimensions. It also identifies focus areas for compliance improvement using portfolio optimization techniques to determine the highest ROI for a given compliance target or investment. Furthermore the solution provides scenario management capabilities to create appropriate scenarios to evaluate spending using a mix of dimensions to segment the data space appropriately. In addition, it also creates optimization scenarios on top of each segmentation scenario to determine the right investment strategy to increase compliance and obtain maximized ROI. With companies spending millions of dollars to improve procurement compliance, having a right tool to analyze and identify focus areas for improvements can enable them to achieve increased savings [6].

The paper is organized as follows. Section 2 gives an overview of the compliance analytics engines and architecture. Section 3 presents a use case representing the proposed solution. Section 4 provides brief summary and proposed extensions to the current analytics, along with collaboration with other business processes.

II. TWO PHASE COMPLIANCE ANALYTICS PROCESS AND ARCHITECTURE

In this section we will discuss the analytical engines in the CAT tool and briefly overview the architecture. The analytical modules in CAT are designed to quickly identify areas of non compliance that are most relevant, and subsequently determine the highest savings potential given an overall compliance or investment target. This approach is a two-phase compliance analytics process which is described in detail below.

Two-phase compliance analytics process

The Compliance Analytics Tool (CAT) tool employs a two phase methodology to manage the spend compliance as shown in Figure 1.

- The first phase comprises of an advanced compliance visibility method that dives deep into spend transaction data using statistical segmentation and clustering techniques.
- The second step comprises of a compliance optimization method that employs portfolio optimization techniques to determine optimal savings given a compliance or budget objective, along with cost/benefit data and segment data from the first phase.

Advanced Compliance Visibility

The purpose of the Advanced Compliance Visibility module is to point out areas that are highly non-compliant across several spend dimensions. The module not only points out areas of non-compliance but also informs the user of factors/features that most likely lead to the observed non-compliance.

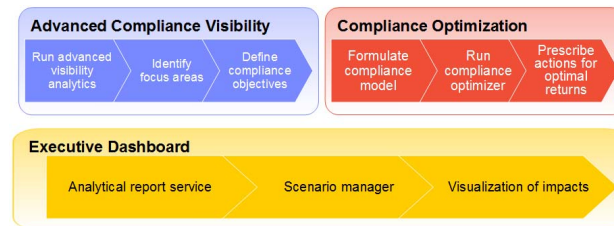


Figure 5. Two phase Compliance Management Methodology

The module is implemented in SPSS and uses a two-step clustering approach [7] to highlight areas that need further attention. The distance measure used in clustering is log-likelihood. We do not use standard distance measures such as Euclidean distance or other l-p norms since we have non-numeric string attributes in addition to real valued attributes. Log-likelihood is effective in such situations. We use the Schwarz Bayesian Criteria (BIC) [8] to decide when to consider different sets of points to signify different clusters. After the clusters have been identified we calculate the importance of each feature in every cluster by measuring its homogeneity in that cluster. The homogeneity is negatively correlated to the variance of that feature in the particular cluster and higher the homogeneity, greater the coherence and hence more significant the feature.

Using such an approach we effectively highlight areas that need attention which are then sent as inputs to the optimization module.

Figure 2 shows the Advanced Visibility module modeled as stream using the SPSS Modeler Client. The parameters for the node such as SQL and Type (dimensions that participate in clustering) can be supplied using the dashboard (discussed later in use case section). First the learning node (Exp 4) -- that is customized for our application -- is run that generates the model node (in yellow) and then this model is applied which categorizes each transaction into a specific cluster or segment. The

stream execution is programmatically invoked using the SPSS PS-API which is wrapped into the web services for any client to perform such analysis

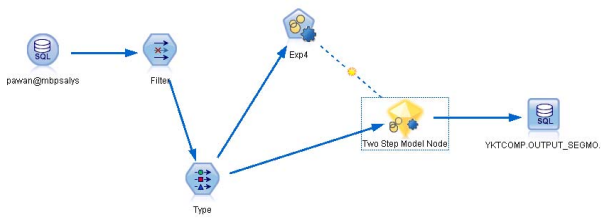


Figure 2: Advanced Visibility SPSS Stream Model

Compliance Optimization

The compliance optimization module provides the capability of optimizing the compliance improvement investment strategy. Multiple optimization models are supported: the user can minimize the investment required to achieve a specified compliance target for the entire (or parts of the) organization, or maximize the achievable compliance for a given maximum investment budget, or maximize the return-on-investment given a maximum investment budget and/or minimum compliance target. The cost of investing on compliance monitoring can be specified for any arbitrary part of the organization in terms of fixed investments (e.g. procuring a new compliance tool) or in terms of volume-based costs (e.g. 15-mins to review per transaction). Similarly, the benefit of increased compliance can be fixed or volume-based (e.g. volume discounts offered by preferred suppliers).

Figure 6 displays an example of how to capture the cost benefit data. The figure shows three tables (fixed cost, variable cost and savings from improvement). Each table rows define compliance level (v1 = 10%, v2 = 20%, v10 = 100%) and column defines the product categories. For each such combination one can provide the fixed and variable cost to bring the compliance to a particular level and savings that can be obtained to move to that particular compliance level. For example, for product category "Marketing and Sales" the fixed cost to move from level v6 to v7 (60% to 70%) is \$37,500 and variable cost is 0.85% of spend. The benefit to move to this level is about 10% of the spend. These costs could represent multiple actions such as introducing new software tool or establishing/enhancing current compliance process or introducing physical monitoring of purchase orders at a certain clip level, etc.

Category	Marketing & Sales	Logistics	Food Service Equipment	Human Resources	IT Hardware	IT Software	Fleet	MRO/ Operating Supplies	Benefits & Insurance	Facilities Management
Leverage Spend - V1	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000
Leverage Spend - V2	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000
Leverage Spend - V3	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000
Leverage Spend - V4	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000
Leverage Spend - V5	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000
Leverage Spend - V6	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000	\$30,000
Leverage Spend - V7	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500
Leverage Spend - V8	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500	\$37,500
Leverage Spend - V9	\$47,500	\$47,500	\$47,500	\$47,500	\$47,500	\$47,500	\$47,500	\$47,500	\$47,500	\$47,500
Leverage Spend - V10	\$80,000	\$80,000	\$80,000	\$80,000	\$80,000	\$80,000	\$80,000	\$80,000	\$80,000	\$80,000

Category	Marketing & Sales	Logistics	Food Service Equipment	Human Resources	IT Hardware	IT Software	Fleet	MRO/ Operating Supplies	Benefits & Insurance	Facilities Management
Leverage Spend - V1	0.75%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%
Leverage Spend - V2	0.75%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%
Leverage Spend - V3	0.75%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%
Leverage Spend - V4	0.75%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%
Leverage Spend - V5	0.75%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%
Leverage Spend - V6	0.75%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%	1.00%
Leverage Spend - V7	0.85%	1.10%	1.10%	1.10%	1.10%	1.10%	1.10%	1.10%	1.10%	1.10%
Leverage Spend - V8	0.85%	1.10%	1.10%	1.10%	1.10%	1.10%	1.10%	1.10%	1.10%	1.10%
Leverage Spend - V9	1.05%	1.30%	1.30%	1.30%	1.30%	1.30%	1.30%	1.30%	1.30%	1.30%
Leverage Spend - V10	1.15%	1.45%	1.45%	1.45%	1.45%	1.45%	1.45%	1.45%	1.45%	1.45%

Category	Marketing & Sales	Logistics	Food Service Equipment	Human Resources	IT Hardware	IT Software	Fleet	MRO/ Operating Supplies	Benefits & Insurance	Facilities Management
Leverage Spend - V1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Leverage Spend - V2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Leverage Spend - V3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Leverage Spend - V4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Leverage Spend - V5	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Leverage Spend - V6	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Leverage Spend - V7	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%
Leverage Spend - V8	11.00%	11.00%	11.00%	11.00%	11.00%	11.00%	11.00%	11.00%	11.00%	11.00%
Leverage Spend - V9	11.50%	11.50%	11.50%	11.50%	11.50%	11.50%	11.50%	11.50%	11.50%	11.50%
Leverage Spend - V10	11.75%	11.75%	11.75%	11.75%	11.75%	11.75%	11.75%	11.75%	11.75%	11.75%

Figure 3: Cost Benefit Input to Compliance Optimization

Such cost input can be further linked to other dimension such as business units such that each combination has its own cost and savings data that goes as input to the optimization engine.

As an example of an optimization scenario, consider the optimization model that maximizes the return-on-investment (ROI) for a given investment budget. The ROI of a business entity is defined as the increase in benefit derived from higher compliance less the additional investment required achieving the higher compliance. Then, the optimization module identifies a prioritized list of investments to make that are ordered in decreasing order of ROI. The list of all possible such orderings can grow exponentially with the number of business entities considered, as well as with the granularity of cost and benefit changes. The optimization module formulates each optimization model as a mathematical program called a mixed-integer linear program and uses IBM's ILog CPLEX optimization solvers to obtain the optimal solution. This is then converted into a prioritized list of investment options to implement. This module scales well with the number of business entities defined, and can identify candidate optimal solutions for problems with 100's of distinct business entities in a few seconds.

Architecture Overview

In this section, we describe how various components interact with each other to provide a seamless solution experience. Figure 4 shows the CAT web based solution architecture. It is a multi-tiered web enabled service-oriented architecture (SOA) based architecture that comprises of a presentation layer (BI Reports and scenario manager dashboard), a middleware layer (solution management, security component, data access beans, business optimization, etc) and a data storage layer (data cleansing component and physical data warehouse).

- **Information Integration Services:** This component receives the invoice and purchase order data, cleanses the data and finally stores the data in the Spend and Compliance Data Warehouse. This layer also applies a set of compliance rules to flag transactions as being compliant:

a) **Process Compliance:**

PO Date must precede the Invoice Date.
 PO transaction does not pass through the normal purchasing channel (Bypass).

b) **System Compliance:**

Purchase Order number exists unless exempt product category

c) **Vendor Compliance:**

Vendor Contract number exists for each transaction.
 If any of the rules a– c are violated, the transaction is flagged as non-compliant.

- **Spend and Compliance Data Warehouse:** A data warehouse that stores the raw PO and Invoice data. The database tables related to analytics and scenario management capabilities are also stored in this database.

Dashboard: This component is a combination of Cognos BI and Adobe Flex to deliver BI spend transaction reports as well as reports from the two analytical engines. The scenario manager, another component listed in the architecture, allows users to create new analytical scenarios and run them via a user interface.

- **Business Optimization Services:** This component consists of the two analytical engines discussed in the previous section. The engines are enabled using a web services interface that can be invoked with appropriate parameters.
- **Management Layer:** This component links all other components in the architecture and facilitates communication between them as needed. For example, a request to execute an analytics engine via the scenario manager dashboard is received by this component which in turns executes the analytics and stores the results in the data warehouse.

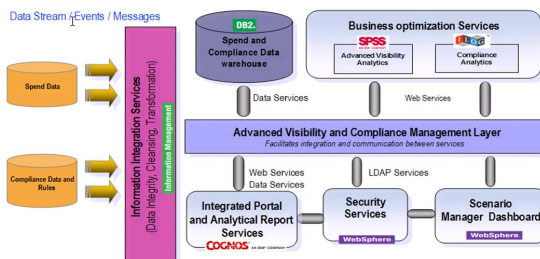


Figure 4: CAT Architecture

The use case section will provide more detail on these components and walk through the various steps of performing analytics, viewing the analytics results, and identifying areas of improvement that provide the highest return on investment.

III. USE CASE DISCUSSION

To better explain the concept we next walk through a use case to illustrate the benefits of the overall approach, and show how users interact with the solution.

ABC Auto is a fictitious company consisting of several business units (Auto, Truck, Financial Services, etc) that purchase various products and services as part of procurement spend across several suppliers to support the company’s business operations. The first step in the CAT tool is to obtain PO and Invoice data with a number of standard attributes such as PO Number, Business Units, Geography, Requester, Product Categories, Supplier, PO Date, Invoice Date, Supplier Contract, Profit Center and the dollar amount of spend. The next step is to store this information in the CAT data warehouse and apply the compliance rules. The data is summarized in Figure 5.

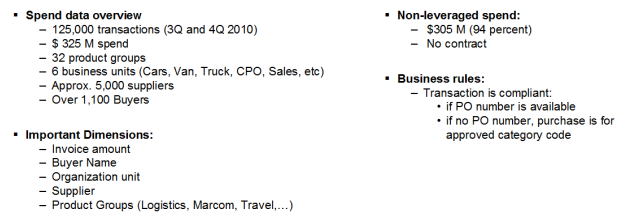


Figure 5. ABC Auto Spend data summary

Traditional Analysis

Figure 5 depicts high levels of non-compliance at this company. A user performing traditional spend analysis would try to identify the business units or product categories for which the non-compliance spend is largest. Figure 6 depicts the traditional approach a user might take to arrive at such a decision. As the two divisions “Auto SUV” and “Van” account for the largest spend, a user might choose to raise their compliance level. Similarly, the product categories “Marketing & Sales” and “Logistics” might be chosen to improve its compliance level. Although we discussed just two dimensions (business units and product category), a similar analysis could be performed on other dimensions such as Vendor, Requester, Profit and Cost Center, etc.



Figure 6. BI Report - Spend by Business units and product categories

Analysis using analytical models in CAT Tool

In this section we will walk through the CAT dashboard interface to interactively perform the analysis. A user would go to a web URL, select the Advanced Visibility Analytics view and then click on “Create Advanced Visibility Scenario Button” to bring up the screen as shown in Figure 7.

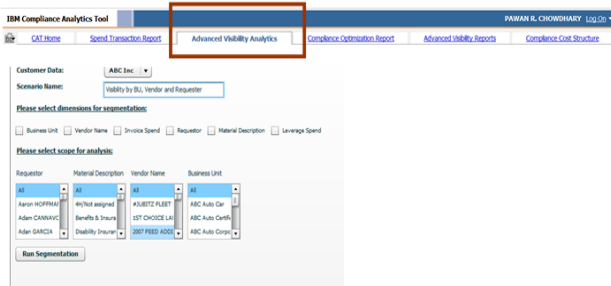


Figure 7. Advanced Visibility Analytics scope selection

The user is asked to enter a scenario name, select a set of dimensions for the segmentation, and define the scope of the data. For illustration, we have selected all dimensions and all data for the analysis. Clicking on the “Run Segmentation” button triggers a message to the CAT management layer which in turns invokes the web services with the segmentation and scope parameters to run the SPSS Model. The results of the advanced visibility analytics are shown below



Figure 8. SPSS Modeler Segmentation Result for Advanced Visibility Analytics

The analytics divide the data space into five segments based on the dimensions content and scope of the data. For the current analysis the dimensions considered for segmentation are business units, requestor, product categories, vendor and compliance indicator. The largest segment with non compliance spend (blue square) is Segment 1 that comprises of almost 60% of the total spend (red square). In this segment 97 percent of the transactions are non compliant. Another important characteristic to note is the average dollar amount per transaction of approximately \$6,700. This amount is much higher when

compared to other clusters (refer to the summary section for more information). The segment also shows that the ABC Auto Sales business unit accounts for the largest non compliant spend, with product category “Marketing and Sales” leading the non compliant spend. It also identifies other attributes such as Requestor name (Brian) and Vendor (Starcom) being responsible for the largest non compliant spend in this segment. Note that the spend attributes within each cluster are independent of the other attributes, and the figure displays the top contributors of a particular dimension. For example, requestor Brian may not be directly responsible for the largest spend within the Auto Sales business unit.

Finally the advanced visibility analysis results are stored in the CAT data warehouse for further analysis using the BI reporting capabilities of the tool (Cognos) where ABC Auto sales can be linked to other dimensions such as product categories or Requestor or Vendors for non-compliance spend.

Once an appropriate segment is identified to improve compliance (Segment 1 in this example) we can now perform the second phase of the analytics (compliance optimization) to further investigate what business units or product categories yield the highest ROI when raising the compliance target from its current level.

To execute the compliance optimization, a user can create several compliance optimization scenarios on top of a single advanced visibility scenario by varying the compliance target or by providing investment limits and finally compare them on the dashboard to make informed decisions.

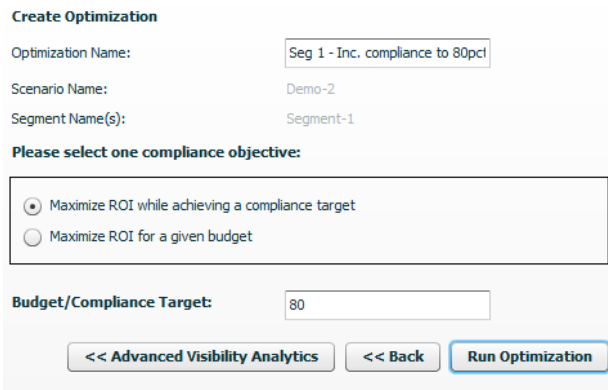
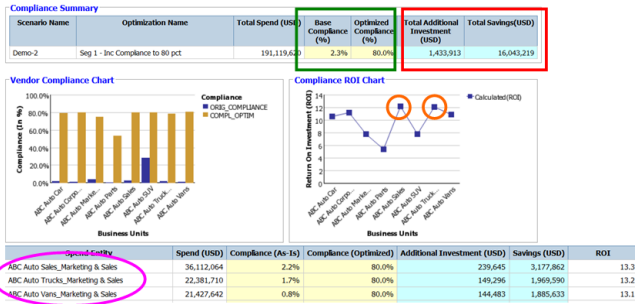


Figure 9. Compliance Optimization Scenario Creation

The result of the optimization run is shown in Figure 10. This report is generated using the Cognos BI tool. The report is divided into three sections (summary, compliance and ROI by business units and detailed ROI table). The summary view displays the total spend in Segment 1 (about \$191M) along with compliance (green box) and the investment of \$1.4M required to attain the 80 percent compliance target for this segment. The total potential savings that can be achieved is approx \$16M which results in the total aggregated ROI of eleven.



The bar chart displays the original compliance and optimized compliance by business units within Segment 1. The optimization chose to different compliance targets for the various business units so as to achieve an overall company-wide compliance level of 80 percent (for example Auto Vans compliance is 81%). This allocation depends upon the cost of raising compliance for each business unit from their original level. The optimization engine obtains appropriate inputs for cost and benefit attainable to move up the compliance levels by business units and product category. The line chart on the right displays the ROI by business unit. Business unit “Auto Sales and Auto Truck” provides the highest ROI of 12.2 for this segment. The report displays the ROI for each combination of business unit and product category. For example, in Figure 10 the ROI for business unit “Auto Sales” and product category “Marketing & Sales” is 13.2. So if it is desired to start with a limited investment, a user can start with the detailed report and select areas to raise compliance within a nominal budget limit (refer to the content circled in pink in Figure 10).

Use Case Summary

Using traditional analysis techniques, the two top non compliant business units “Auto SUV” and “Van” could have been selected as targets for compliance transformations, although these are not necessary the best selection from an ROI perspective. The CAT tool on the other hand identified the “Auto Sales” and “Truck” divisions as better targets for compliance improvement due to their potential of generating higher ROI. Although not displayed in Figure 10, the ROI for Auto SUV/Marketing & Sales is 5.5 which is much lower than the ROI of 13.2 for Auto Sales/ Marketing & Sales.

We mentioned in the earlier section that the average spend per transaction in Segment 1 is about \$6,700. This is an important observation coming out of the advanced visibility analytics engine as the compliance is typically enforced at the transaction level. Hence one can treat this average spend for given business units and product category identified in Segment 1 as an initial clip level for automatic approvals when POs are submitted for review. The absence of analytics-driven clip level information could result in setting inappropriate clip levels based on user experience which could lead to either high cost to

achieve compliance or spend leakage due to appropriate transactions not being audited.

IV. CONCLUSION AND NEXT STEP

Businesses that aim to improve their bottom line can take advantage of CAT tool to increase compliance within the procurement activities of their organization. To enable such analytics, it is imperative that the raw spend data is complete, i.e., all important fields are available in the spend transactions, and standard data dimensions such as Vendors, Products, etc are classified correctly [3]. This may require sophisticated techniques for data cleansing, data classification, completeness, and data enrichment. The need for data classification is important when expanding the CAT tool capability to analyze spend across several clients and identify cross client savings opportunities such as volume discounting. We are also investigating digitizing vendor contract data which can provide additional savings opportunities such as back-end rebates, better service levels, etc by routing spend to vendors that meet contracted spend levels.

REFERENCES

- [1] “Best Practices in Spending Analysis – Cure for Corporate Epidemic” Aberdeen Group 2004.
- [2] Singh, Moninder, et al. “Automated cleansing for spend analytics” Proceedings CIKM '05 Proceedings of the 14th ACM International Conference on Information and Knowledge Management
- [3] UNSPSC, The United Nations Standard Products and Services Code, <http://www.unspsc.org>
- [4] Greenfield, Rip, “Spend Analysis: The First Step in Strategic Sourcing”, Proceedings 90th Annual International Supply Chain Conference, May 2005
- [5] Bartels, Andrew, “Spend Analysis: Key App, Few Vendors” Forrester Research, July 2004
- [6] Kirit Pandit and H. Marmanis “Spend Analysis -The Window into Strategic Sourcing” 2008, pp. 219-235
- [7] Marina Meila and David Heckerman, “An Experimental Comparison of Several Clustering and Initialization Methods” In Machine Learning, 1998, pp. 386-395
- [8] Schwarz, Gideon E. (1978). "Estimating the Dimensions of a Model". Annals of Statistics 6 (2): 461-464