

## Big Data, Big Headaches: An Agile Modeling Solution Designed for the Information Age

Mariam Seirafi, Cornerstone Group of Companies

### ABSTRACT

The surge of data and data sources in marketing has created an analytical bottleneck in most organizations. Analytics departments have been pushed into a difficult decision: either purchase “black-box” analytical tools to generate efficiencies or hire more analysts, modelers, and data scientists. Knowledge gaps stemming from restrictions in black-box tools or from backlogs in the work of analytical teams have resulted in lost business opportunities. Existing big data analytics tools respond well when dealing with large record counts and small variable counts, but they fall short in bringing efficiencies when dealing with wide data. This paper discusses the importance of an agile modeling engine designed to deliver productivity, irrespective of the size of the data or the complexity of the modeling approach.

### INTRODUCTION

We live in an era where the exponential growth in big data is common knowledge. Words such as petabyte, exabyte, zettabyte, ... are starting to enter common business lexicon. In an article published by Analytics Magazine, IDC reported an increase in big data technology and services from \$3.2 billion in 2010, to \$16.9 billion dollars in 2015 (IDC). Pressure on analytics teams to evaluate this ever-growing flow of data in search of insights is intensifying. IDC further identifies a shortage of analytics resources as “an inhibitor of adoption” (IDC).

In an article recently published by Forbes Magazine, McKinsey Consulting addresses the challenge of achieving better marketing analytics. Organizations have the problem of “trying to analyze ever greater data sets in an effort to uncover a killer insight. Or they look for a tool that can solve every problem.” However, the core issue at hand, is a disconnect between analysts and decision makers. According to McKinsey, effective application of analytics can free up 15%-30% of total marketing budget. Forbes concludes that the key to a successful marketing analytics is to use tools that are **comprehensive, easy to use and fast**... (Thomas Bauer, Lars Fiedler, Jeff Jacobs, and Dennis Spillecke, 2016)

While the Forbes paper carries on discussing the need for an overarching tool to evaluate marketing efforts to concentrate on higher valued projects, this paper approaches the challenge from a different angle. It focuses on the importance of rebalancing analytical resources, so they can shift energy towards more strategic and data-driven insights with their business counterparts.

As a step towards this objective, this paper also presents a SAS®-based tool to safely free-up resources in more tactical analytics such as a campaign targeting. It outlines the problem with many existing tools when dealing with big data with very high “variety” of information, and proposes a solution to consider.

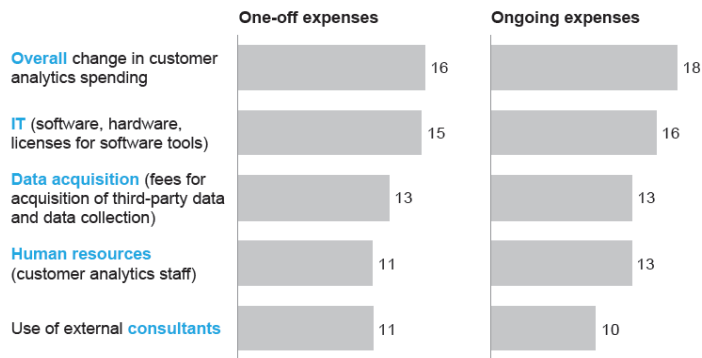
### RESOURCE ALLOCATION PARADOX

Explosion in data sources and organizations' growing investment in data has lead to a continuous growth in analytics demands. Typically, request for analytics grows exponentially as data sources grow. Higher variety in data raises more challenges from business.

### CHALLENGE OF THE GROWING RESOURCE MODEL

According to a DataMatics survey by McKinsey (international survey of large firms), organizations are increasingly more focused on customer-centric and data-driven insights for customer-centric decisioning. Surveyed firms pointed out to increased investments in all aspects of data and technology, including human resources. Unfortunately, most organizations identified that their investments in big data have fallen short of delivering noticeable improvement in performance (DataMatics, 2014).

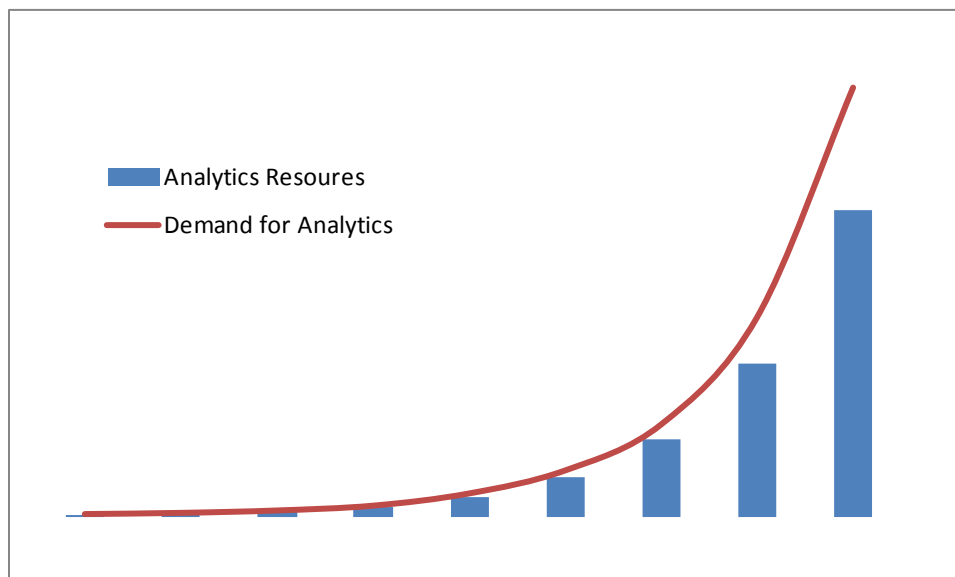
**Expected increase in spending over the next 12 months**  
Percent<sup>1</sup>



<sup>1</sup> Based on "Compared with the past 12 months, how will your company's/business unit's customer analytics spending change (as a percentage) over the coming 12 months?".  
SOURCE: McKinsey, DataMatics 2013

**Figure 1 Expected annual increase in investments in data and technology among large organizations**

Facing an ever growing demand for analytics, many organizations associate the lack of performance to tight resources, and respond to this growing need by investing in incremental resources, including an increase in analytical staff. However, unless the demands on data-driven insights curbs, this business model will eventually become unsustainable as demand will outpace resource supply (Figure 2). In addition to higher training costs, increases in resources have to also carry higher costs of management and monitoring of quality control, knowledge consistency, data stewardship, data governance, ...

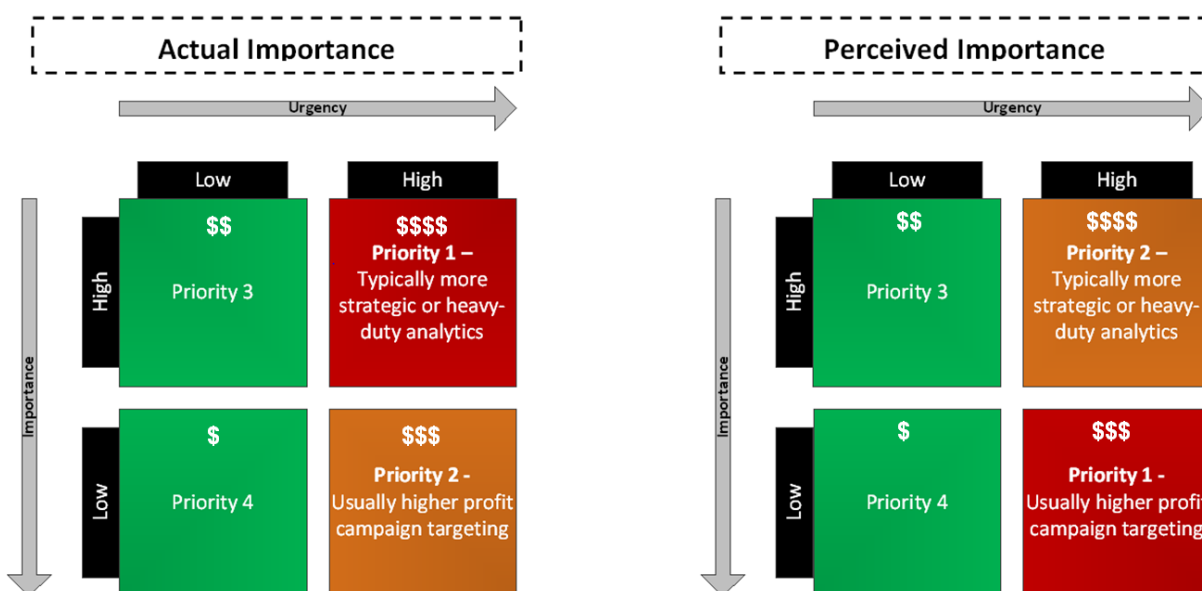


**Figure 2. Growing Analytical Resource Demand and Supply**

According to McKinsey Consulting, the lackluster delivery of insights from big data is due to organizations failing to embark the issue optimally. (DataMatics, 2014)

## THE DIMENSION TENSION

Prioritization becomes a key management requirement when dealing with high workload. Analytics tasks are usually of 2 natures: tactical (campaign support) or strategic (holistic customer-base evaluation). Strategic analytics are often deemed to have a higher expected financial benefit, and should receive higher importance over tactical objectives. Nonetheless, in practice, most organizations tend to give higher priority to tactical jobs. This is partly due to the urgent nature of the campaigns. This urgency overshadows the over-arching importance of strategic tasks; maybe because businesses prefer to bank on small short-term revenue gains over bigger, yet more distant revenue streams (Figure 3, right chart).



**Figure 3. Resource Allocation Paradox**

As a result of this perpetual quandary, often times, powerful strategic analytics takes lower priority over campaign modeling and targeting. These tasks are analytically less complex, but in the long-term are also bound to bring less revenue.

This resourcing paradox has led many organizations to invest in purchasing or developing new automated analytical tools. According to the same study, McKinsey has identified that organizations who were able to successfully deploy analytics (measured as improved business performance), had managed to use tools to fully automate data processing and analytical processes. They had also provided their business users with an easy-to-use software interface for more mundane analytical tasks, freeing more experienced analytical resources to focus on more complex and groundbreaking work. (DataMatix, 2014)

Organizations have the option of either creating their own internal automated tools, purchasing pre-built or outsourced solutions. The key is to select the right one! Not all tools are created equal.

## A PEAK INTO AUTOMATED ANALYTICAL SOLUTIONS

### BIG DATA RESULTED IN GENERATION OF A BIG VARIETY OF TOOLS

Often, exposure to “big data” in large organizations is around social media, web data, or even existing transactional data. In these cases, the data can often be characterized as having massive record counts, but relatively small variable counts (number of columns). The content of this data also often changes, leading to the need for frequent re-evaluations.

As such, many automated tools are designed around this data structure. Products like Hadoop and SAS's in-memory calculations enable analyzing data with higher volume and velocity; but what about variety? Imagine dealing with a dataset with over 10,000 columns of input into the model... Are the tools designed to handle this level of variety?

Privacy laws in Canada are one of the tightest in the entire world. As a result, analysts have major limitations accessing valuable individual-level information for targeted acquisition marketing campaigns. In most cases, the best alternative is to access various external data sources with geographic overlay (such as Census). Just by looking at 4-5 overlay datasets, it is easy to amass thousands of columns of information, to be evaluated for a model to predict future acquisitions.

To speed up the modeling process, we evaluated a number of automated tools. Almost all these tools offered a nice GUI interface, visual display, sophisticated graphics, built on reliable analytical engine, sound modeling steps, and clean output delivery.

In contrast, here are some common challenges we faced:

1. Lack of ability to properly reduce data.
2. Black box solution, leaving the modeler with many unanswered questions, or even highly over-fitted models.
3. Lengthy manual setup process. The time savings from running the model in an automated way, was overshadowed by the duration of setup. Product was designed for much fewer variables in mind.
4. Carpal tunnel syndrome!! Many tools only worked with a GUI-interface. Proper decision-marketing for treatment of data at a variable-level required an inordinate number of mouse clicks, given the number of variable inputs.
5. Inability to incorporate custom analytical processes within the code.
6. Difficulty in utilizing final data and outputs for alternative analytical solutions beyond the solutions offered by tool.
7. Not an end-to-end tool.
8. Designed for a single audience: Technical modeler, or non-technical business user.

Faced with these limitations, we developed our proprietary predictive modeling solution.

## **AUTOMATED PREDICTIVE MODELING (APM) SOLUTION**

Prior to development of this tool, we developed a list of must-have requirements:

- Easy and fast setup
- Ability to minimize need for modeler's time without compromising modeling result
- Ability to handle very wide data
- Cater to different audiences (business or technical user)
- Ability to allow for analytical continuity
- Ability to setup on any client environment with pre-existing SAS license

APM was positioned as a marketing campaign targeting tool and uses various regression approaches to predict likelihood of an event, or financial outcome of a campaign (such as, incremental revenue from an event).

## **APM Performance Evaluation**

### ***Quality of performance***

Initial goals for APM were to deliver a solution that meets at least 80% of the performance met by an experienced modeler (lift, fit, accuracy, robustness, actionability, ...).

In every experiment, the APM model easily surpassed the 80% threshold (often, meeting 90th percentile of models developed by experienced modelers). This threshold was particularly hard to meet with alternative automated tools that we evaluated, when facing big data with very high variety of sources and variables.

### ***Speed of performance***

The APM lowered modeler's preparation time to under 30 minutes, even when facing *data with very high variety of variables*... Evaluation of the distribution of thousands of input variables, or decisions to properly impute or treat a high count of input columns can be a daunting task with other tools.

This was a very welcoming change over other tools studied. In some instances, the build of model in a traditional manner, or through the use of some automated tools took anywhere from 60 to 200 times longer (measured as human-time invested in model build). Most tools evaluated:

- Either made wrong assumptions about data distribution and the desired data treatment, leading to **lower model performance**
- Or they consumed an inordinate amount of modeler's time on data preparation, leading to **higher modeling cost**

The key to time-savings from APM was to:

1. Focus on modeler's time and not just run-time -- Run-time could be reduced by deploying a slew of technologies such as Hadoop and in-memory computing
2. Capture the modeler's wish list in the fastest and most accurate manner -- Wish list would be assessed and tweaked automatically, during the automated data evaluation
3. Incorporate a comprehensive set of evaluation check-points to automate modeler's decision-making during the intermediate steps of model build

By meeting these two gains from APM while ensuring the delivery of a high quality model, modelers were able to increase efficiency. The time savings were re-allocated on higher analytics, such as potential further fine-tuning of high ROI model to ensure utmost actionability, offer and channel optimization, evaluation of health of product, ...

### ***Business Users Access to APM***

The APM was pushed forward one more step to ensure the model can be managed and run by non-technical users. Additional automation captured the most common modeling requirements to provide a basic and non-technical setup GUI front for business users (demonstrated below).

This was achieved with maintaining the need for a wish list per variable. Guidelines and basic education on creation of the wish list was created. This ensured that the business knowledge about the data overrides a tool making false assumptions about the content of the data at hand. The modeler became

advisory and was brought in mainly after the model was created to discuss learnings and implications from the model.

The following section evaluates APM's delivery vis-à-vis the initial requirements set prior to the build of the model.

## Evaluation of APM's requirements

### *Easy and fast setup*

Characteristic	Description
Default setup uses Excel as easy setup	Excel DECISION CENTRE serves as an intake form to create a wish-list
Can be changed to other input format if desired	Ability to easily filter, edit, copy and paste requirements across makes it an easier setup
	Model is triggered using a simple GUI interface

**Table 1. Automated Predictive Modeling Solutions - Easy and Fast Setup Characteristics**

Similar to all automated tools, APM Solution requires the data to be in a flattened format prior to starting the modeling job. At that stage, the modeling job starts. APM limits the modeler's initial involvement to 2 steps:

- An Excel-based Decision Centre to capture the modeler's wish list
- A SAS-based GUI interface to trigger the model and deliver the results

Excel is still recognized as the most versatile and commonly used BI tool. The wish list is a list of input variables in the model, and a place to tell SAS how to treat the variables. When dealing with a high variety of data, Excel's capability in capturing modeler's requirements is a welcome change over a GUI interface that requires too many mouse clicks. By combining 2 very powerful tools in the market (SAS and Excel), APM balances capturing modeler's requirements, while introducing tremendous automation and speed.

Variable Layout						Data Source / Category	Initial Var Prioritization	Cat Vars	Imputation	Zero Flogs	Cap Outliers	Non-L				
Name	Type	Length	Format	Inform	Label	Source	Init Setup: Use In Model	Init Setup: Priority for Model Use	Init Setup: Type = Class?	Has Missing	Impute to	Requires M. Flag	Create Flag for Value = 0	Cap Outlier	Requires O. Flag	Requires non-Linear Trans
74 w	Numeric	8				Std_Bus_Var	N	10								
75 w	Numeric	8				Std_Bus_Var	N	10								
76 C	Numeric	4				StatsCan	Y	1		Y	MEAN					
77 C	Numeric	6				StatsCan	Y	1		Y	MEAN					Y
78 C	Character	1				StatsCan	Y	5	Y	Y	—					
79 C	Character	1				StatsCan	Y	5	Y	Y	—					
80 C	Character	1				StatsCan	Y	5	Y	Y	—					
81 TL	Numeric	8	BEST12		Li	StatsCan	Y	1		Y	MEAN		Y			Y
82 TL	Numeric	8	BEST12		M	StatsCan	Y	1		Y	MEAN		Y			Y
83 TL	Numeric	8	BEST12		%	StatsCan	Y	1		Y	MEAN		Y			Y
84 TL	E	Numeric	8	BEST12	Av	StatsCan	Y	1		Y	MEAN		Y			Y
85 TL	M	Numeric	8	BEST12	Av	StatsCan	Y	1		Y	MEAN		Y			Y
86 TL	M	Numeric	8	BEST12	Ni	StatsCan	Y	1		Y	MEAN		Y			Y
87 TL	M	Numeric	8	BEST12	Av	StatsCan	Y	1		Y	MEAN		Y			Y
88 TL	R	Numeric	8	BEST12	BEST32	Av	StatsCan	Y	1	Y	MEAN		Y			Y
89 TL	R	Numeric	8	BEST12	BEST32	M	StatsCan	Y	1	Y	MEAN		Y			Y
90 TL	O	Numeric	8	BEST12	BEST32	Av	StatsCan	Y	1	Y	MEAN		Y			Y
91 TL	E	Numeric	8	BEST12	BEST32	Ni	StatsCan	Y	1	Y	MIN		Y			Y
92 TL	OE	Numeric	8	BEST12	BEST32	Ni	StatsCan	Y	1	Y	MAJ		Y			Y
93 TL	DE	Numeric	8	BEST12	BEST32	Ni	StatsCan	Y	1	Y	MEAN		Y			Y
94 TL	E	Numeric	8	BEST12	BEST32	Ni	StatsCan	Y	1	Y	MEAN		Y			Y
95 TL	N	Numeric	8	BEST12	BEST32	Ni	StatsCan	Y	1	Y	MEAN		Y			Y
96 TL	N	Numeric	8	BEST12	BEST32	F <sub>1</sub>	StatsCan	Y	1	Y	MEAN		Y			Y
97 TL	N	Numeric	8	BEST12	BEST32	F <sub>1</sub>	StatsCan	Y	1	Y	MEAN		Y			Y
98 TL	N	Numeric	8	BEST12	BEST32	F <sub>1</sub>	StatsCan	Y	1	Y	MEAN		Y			Y
99 TL	N	Numeric	8	BEST12	BEST32	F <sub>1</sub>	StatsCan	Y	1	Y	MEAN		Y			Y
100 TL	N	Numeric	8		F <sub>1</sub>	StatsCan	Y	1		Y	MEAN		Y			Y
101 TL	N	Numeric	8		F <sub>1</sub>	StatsCan	Y	1		Y	MEAN		Y			Y
102 TL	N	Numeric	8		F <sub>1</sub>	StatsCan	Y	1		Y	MEAN		Y			Y
103 TL	N	Numeric	8		F <sub>1</sub>	StatsCan	Y	1		Y	MEAN		Y			Y
104 TL	N	Numeric	8		F <sub>1</sub>	StatsCan	Y	1		Y	MEAN		Y			Y
105 TL	N	Numeric	8		F <sub>1</sub>	StatsCan	Y	1		Y	MEAN		Y			Y

Figure 4. View of automated

Figure 4 above shows the list of all variables to be used in the model, and a series of preferred decisions per variable. Decisions determine variable role, importance, type (categorical, continuous, ...), imputation preference, ... *Note that these are just a wish list. The preferred action will only take place if the data supports such action (explained below).*

Rapid Predictive Model Launcher

Cornerstone

GROUP OF COMPANIES

Welcome to Cornerstone Automated Predictive Modeling Solution!

Basic Setup | Overlay Data | Advanced Setup

Project Name

Project Folder

...

Time required for campaign execution (between selection and campaign drop):

month(s)

Input Data Name

...

Save Changes

Verify Input

Build Automated Model

Quit

Figure 5. Initial view of APM

Beyond the Decision Centre the above GUI interface is all that is required to kick-off a model. It requires

simple and basic information such as location of project folder (which would contain the Decision Centre), and data set name. In a few clicks, the model is setup and ready to delivery results.

### ***Minimize modeler's time without compromising modeling result***

Characteristics	Description
Designed around a fundamental mindset:	
<ul style="list-style-type: none"> <li>Upfront use of DECISION CENTRE ensures use of business insight in model setup</li> </ul>	Users can change roles and treatment of variables using their business knowledge
<ul style="list-style-type: none"> <li>Speed in delivery of final result is through elimination of modeler's need to check intermediate results</li> </ul>	A series of modeling steps are automated using the most industry-standard evaluation techniques
<ul style="list-style-type: none"> <li>SMART reduction of deadweight time of data preparation</li> </ul>	APM uses a SMART approach. Decision Centre requests are evaluated and potentially dismissed or tweaked according to data content <i>with proper explanation</i>
<ul style="list-style-type: none"> <li>Incorporates all common checkpoints a modeler does during model build (SMART data prep)</li> </ul>	Product is built in SAS product. Not a black box. Access to all intermediate data prep reports
<ul style="list-style-type: none"> <li>Incorporation of custom tools</li> </ul>	Allows for modification or incorporation of additional components, such as additional reporting, alternative variable reduction process, ...
<ul style="list-style-type: none"> <li>Verification of final model result against actual data</li> </ul>	Only variables with signs in similar direction to raw data are kept in final model

**Table 2. Automated Predictive Modeling Solutions - Minimize Modeler's time**

The APM tool offers a wide array of data evaluation options, such as evaluation of categorical data, SMART dummy variable creation, evaluation of numeric variables, imputation, outlier capping, ...

The Decision centre wish list is just a start. Data treatment is based on content and distribution of variables. For instance:

- When dealing with categorical variables, most tools automatically evaluate all valid values in an alphabetical order. In many instances, the remainder against which the data is being compared (the last valid value) is a very small subset of, say, 0.02% of the data. This jeopardizes the quality of the model. APM's SMART evaluation ensures that only valid values with sufficient distribution are evaluated in the model. In addition, they are also compared to a remainder that has a sufficient distribution in data.
- Another simple example of APM's SMART data evaluation is its consideration of variable inclusion in the model. At the onset, the modeler can comfortably set all desired variables to be considered in the model. APM evaluates this decision and over-turns it if required. Reasons for exclusion can be items such as all constant, high percentage of missing, ... Figure 6 below, shows an example of a variable being dismissed early in the modeling process because of a very higher percentage of missing content, and the remainder of the variable appearing as constant.



Variable	Max	Mean	Min	N	NMiss	StdDev	fail_allmiss	fail_constant	Pct_Missing	Fail_HighMissing	Fail_AllMeans
AI:		0	0	0 3492895	291105		0		1 0.076930497		
DA		1 0.026843552		0 3784000	0	0.161626057				0	

**Figure 6. Sample report from a SMART evaluation of data in APM**

As demonstrated in Figure 6 above, intermediate reports capture all these decisions for modeler's knowledge and information. APM was not designed to be a black box. These reports are kept in form a SAS Output, SAS Dataset, and an Excel sheet for ease of future access.

### ***Ability to handle very wide data***

Characteristic	Description
Use of Excel as an input form	Decision centre allows easier information setup for modeler, while avoiding too many mouse clicks  Also, provides easier navigation and search capabilities
Incorporates multiple powerful variable reduction approaches	With thousands of variables, a powerful process of variable reduction is key
Access to intermediate reports	Access to all intermediate reports provides a quick ability to evaluate data content and characteristics prior to final approval of model

**Table 3. Automated Predictive Modeling Solutions - Ability to handle wide data**

As described in detail above, the key to the APM is its upfront setup of a wish list in the decision centre. The selection of Excel allows for very fast navigation of the data. This is key when dealing with a very wide dataset with thousands of variables.

While APM is not limited to the use of Excel as its Decision Centre, it was the reality of a wide data that drove us to using this tool. No other BI tool still offers the same flexibility for an interactive content evaluation. Easy filtering, copy and paste, powerful search, wide view (not limited to a narrow window for viewing field name), data entry, etc. made this tool the preferred setup tool for our modelers.

Wide data requires intensive data reduction. All data reductions are beneficial, but they use a different technique and have their influence or assumption on the data. To remedy the situation, when facing such wide big data, APM incorporates and overlaps multiple data reduction techniques:

- Standard unsupervised principle components
- Proprietary supervised variable reduction
- LARS-influenced variable reduction

Variable	label	mean_resp_0	mean_resp_1	mean_overall	index	absIndex	toppick_re
LENGTH_F	LENGTH_F	0.095367561	0.111702128	0.095575317	0.168734052	0.168734052	0
ClusterN	Variable				RSquareRatio	MinRSquare	lal
86 D	no				0.2637	0.263747465	DA
79 E	G4				0	0	ET
43 F					0.1268	0.126835753	FL

**Figure 7. Sample report from APM Data Reduction**

As demonstrated in Figure 7 above, intermediate reports capture all outputs of all steps including various data reduction techniques. APM is not designed to be a black box. These reports are kept in form a SAS Output, SAS Dataset, and an Excel sheet for ease of future access.

***Cater to different audiences (Business or technical user)***

Characteristic	Description
Option to kick-off model using a simple interface	While all the APM code is in SAS, a simple, user-friendly interface calls the modeling tool
Interface provides a "Basic" and an "Advanced" option	Basic view of setup leverages recommending default setting  Advanced tab in the setup provides power users to change key statistics used in modeling as desired

**Table 4. Automated Predictive Modeling Solutions - Cater to different audiences**

Figure 5 shown earlier (also seen as the left and side of Figure 8 below) is the simple, minimalistic interface for a business users. It is limited to basic information such as the file and project name, and name of dataset.

All analytical decisions about setup and build of the model are created and incorporated as a default within the code. Even decisions about % of sampling and over-sampling of data are automated within the tool based on output requirements and data distribution.

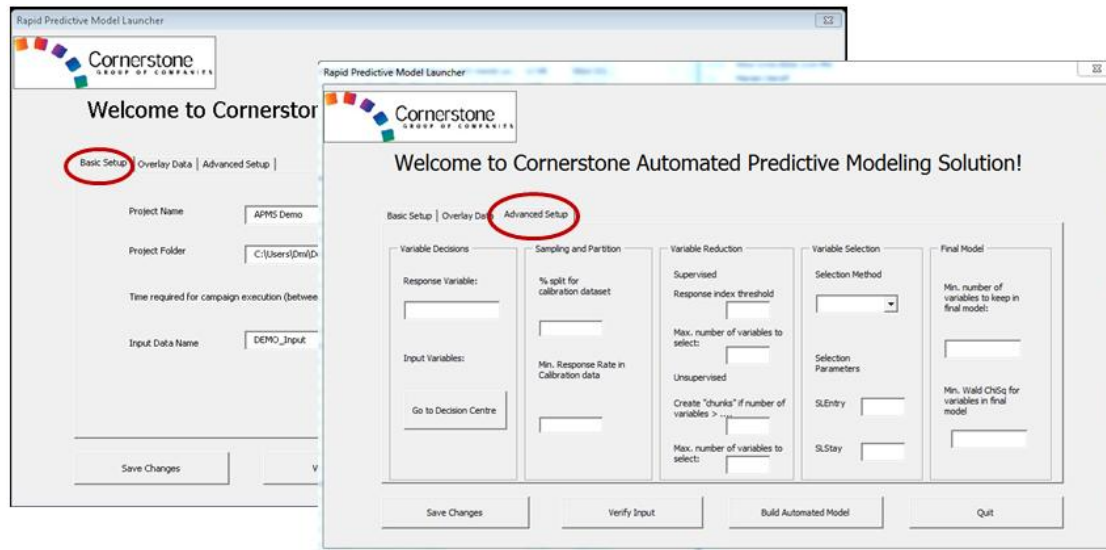


Figure 8. View of Basic and Advance APM model setup

The picture in the right hand side of Figure 8 above, is the Advanced Setup. It allows for the modeler to further influence the model build process, from sampling and variable reduction, to variable selection within the model, and final modeling output.

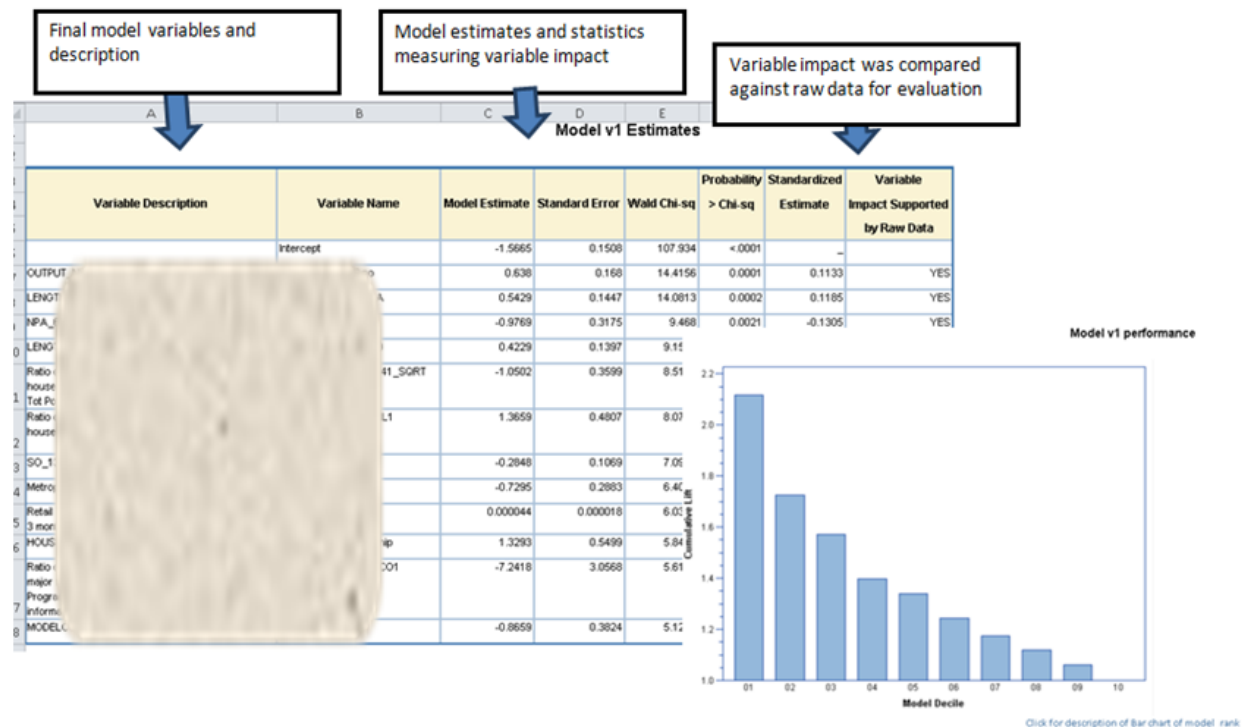


Figure 9. Sample output of final model for APM

Figure 9 above shows sample outputs from final model delivery. Again, with 2 users in mind, the detailed variable description on the left hand side is for the technical user. In addition to the typical output of the

regression, it offers confirmations such as, *ensuring that the direction of the variable in the model matches the direction from the raw data*.

The output on right hand side of Figure 9 is for both technical and business users. In addition, the output provides various additional financial information such as ROI or CPA per model rank for business decisioning.

#### ***Allow for analytical continuity***

Characteristic	Description
Access to all outputs	SAS datasets (calibration / validation / ...) all available in an Output folder
Ability to further build on model	Modeler can also take data and combine with additional analytical tools

In addition to saving all intermediate reports about the data content, the APM saves all datasets and variables used in the model build process. Datasets contain all input data, calculated variables, ... including fields generated to explain variable distribution.

For sophisticated models, or models with a high revenue expectation, modelers typically use the tool to build a base model. If required (additional time is warranted by business) output is available for model fine-tuning.

Also, because the dataset is saved as a SAS dataset, the output is ready to launch into the next phases of analytics; say, from response model (APM) to a price-sensitivity model (more sophisticated model, beyond the scope of APM).

In both instances, APM was used as a way to expedite delivery of the model.

#### ***Ability to setup on any client environment with pre-existing SAS license***

Characteristic	Description
Minimal setup required	Basic setup utilizes Base SAS, SAS Stat, SAS Access
Tool highly portable	Current input tool is Excel, but can be changed to accommodate client requirements

Combination of 2 very highly powerful tools in the industry (SAS and Excel) ensured higher portability of the tool. In addition to using APM in-house to support our clients' analytical needs, APM was very quickly implemented within any organization that had SAS license to provide them with an internal automated analytical tool.

## **CONCLUSION**

An exponential growth in data within a very short period of time, is putting growing pressure on analytical teams to generate more and better insights. The demand for higher analytics has pushed some organizations to continuously invest in higher resources; however, this increase in investment must be managed in a sustainable manner. Facing this increased demand, many analytical teams fall in a dimension tension, allocating resources to more urgent, yet less financially beneficial projects (e.g., more

tactical tasks of campaign targeting). This jeopardizes success of analytical teams by ignoring more strategic, financially lucrative data and insight discovery tasks that require the skills of strong, experienced analytical teams.

To alleviate this pressure, a various number of automated tools and services have been built to help analytics departments. While many tools meet their objectives by producing run-time speed to deal with high volumes and velocity in data, they often fail dealing with data with high variety. Cornerstone's Automated Predictive Modeling Solution provides a successful alternative to noticeably decrease modeling deadweight time when dealing with very wide data, while ensuring quality of model is not compromised. The tool is also available for business users (internal or at client site) to provide direct access to data and insights driven from the model.

## REFERENCES

Analytics Magazine. "IDC forecasts big growth for Big Data." Analytics Magazine Special Edition. March/April 2016. Available at <http://www.analytics-magazine.org/special-articles/546-idc-forecasts-big-growth-for-big-data->

Thomas Bauer, Lars Fiedler, Jeff Jacobs, and Dennis Spillecke. "The Secret To Great Marketing Analytics? Connecting With The Decision Makers." Forbes. January 6, 2016. Available at <http://www.forbes.com/sites/mckinsey/2016/01/06/the-secret-to-great-marketing-analytics-connecting-with-decision-makers/#1c92217752ab>

DataMatics. "Using customer analytics to boost corporate performance." McKinsey & Company. January 2014. Available at <http://www.mckinseyonmarketingandsales.com/sites/default/files/pdf/Datamatics.pdf>

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Mariam Seirafi  
Cornerstone Group of Companies  
416-932-9555 x 2147  
[mseirafi@cstonecompanies.com](mailto:mseirafi@cstonecompanies.com)  
[www.cstonecompanies.com](http://www.cstonecompanies.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.