BakerSCI™ BAKER STREET SCIENTIFIC Intelligence through relevance.™

400 Broad St, Suite 221B Rome, GA | a
info@bakersci.com | e
706.331.5455 | p

Three "I" s of Data Bias:
Overcoming Traditional Analysis Shortfalls via Visual Associative Analysis

Background: Financial Services environments can be incredibly dynamic. While quantitative influencers may change the underlying facts, qualitative influencers steer the story. What is needed for true financial analysis is a way for the information to be analyzed without fear of infrastructural, instructional, and/or institutional bias. One bias is enough to alter the true data story and each additional one only serves to compound the issue: Bias-riddled analyses today dictate error-prone decisions tomorrow.

Problem Statement: In any data environment, even with the technological advances in Big Data techniques, on average data analysts spend 80% of their time *searching* data for relevant events and 20% of their time actually *analyzing* those events. To an untrained eye, this inefficiency is chalked up to being just part of the process as they do not recognize that this issue is manifested silently through three levels of bias; none of which help internal productivity goals nor emphasize external client expectations/services:

- Infrastructural Bias – Bias hard-coded into organizational software (whether custom built or COTS) whereby decisions as to what is and is not relevant information (aka "noise") to be kept in the system for analysis was made by personnel outside the entity analyzing the data.
- Instructional Bias – Bias that is inherent in all analysts that uniquely influences their decision making based on a multitude of different factors: experience level, type of experience, education level, position of authority, risk aversion thresholds, and, yes, even human emotions.
- Institutional Bias – Bias that is indwelt in an organization perpetuated by company thought leaders whereby decisions are made based on external factors unbeknownst to the analyst, outside of the data analysis itself, that further curtail objectivity to accomplish competing priorities.

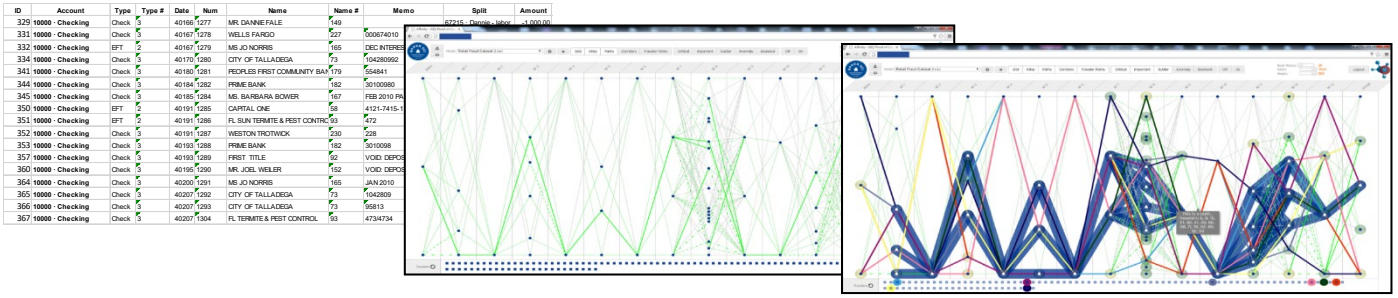
Deliverable A: Financial Service analysts must be able to quickly review fast-paced, volume-heavy datasets for trends, outliers and anomalies for timely event identification and analysis.

BakerSCI Solution – BakerSCI's RITR solution employs a unique methodology using associative analysis whereby the various data elements within the dataset are monitored by *how they behave with each other.* Identification of optimal trend patterns, while helpful, are only a part of the solution. Understanding how those patterns can (and will) change as new information is added or subtracted from the dataset is crucial to revealing why trends are occurring and IF those trends will continue or not. More important to a data analysis is how those trends can be affected by relevant outliers and/or anomalies that RITR identifies in the "noise" of the dataset; the same noise that is deemed irrelevant by conventional systems and therefore removed from the analysis. When an analyst can see *into the noise* and understand *how the data behaves as a whole*, then the depth of the current (and future) analyses take on new meaning.

From Volume-Heavy Transactions to Visual Behavior Revelation and Confirmation:

Use Case #1: Accounting Transaction Analysis

BakerSCI was asked by a local CPA Firm to review its internal books for evidence of unusual activities. BakerSCI began with a download of the raw transactions from the accounting information system (below left). After a brief ingestion period, BakerSCI ran the data through its associative analysis system for a visual representation of the data prepped from analysis and manipulation. Wireframing of the data can be seen below showing how the transactions "moved" through the system (below center). Heavy saturation of results as well as interesting outliers were identified for further investigation with the client (below right). *Client was able to visually interpret spreadsheet data, that had become voluminous, cumbersome, and time-consuming, allowing for quicker identification of events of interest that often times get lost during in-depth spreadsheet analysis.*



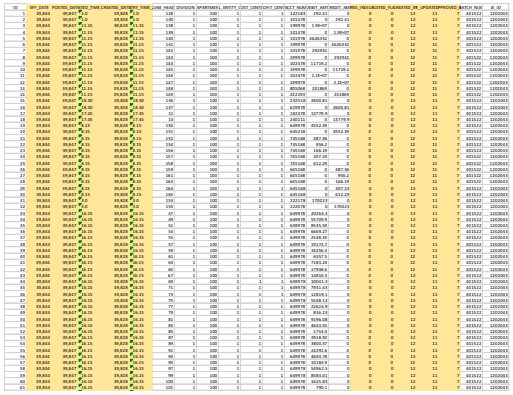
ID	Account	Type	Type #	Date	Num	Name	Name #	Memo	Split	Amount
329	10000 · Checking	Check	3	40166	1277	MR. DANNIE FALE	149		67215 · Depcie - labor	-1,000.00
331	10000 · Checking	Check	3	40167	1278	WELLS FARGO	227	000674010		
332	10000 · Checking	EFT	2	40167	1279	MS JO NORRIS	165	DEC INTERES		
334	10000 · Checking	Check	3	40170	1280	CITY OF TALLADEGA	73	104280992		
341	10000 · Checking	Check	3	40180	1281	PEOPLES FIRST COMMUNITY BAN	179	554841		
344	10000 · Checking	Check	3	40184	1282	PRIME BANK	182	3010098		
345	10000 · Checking	Check	3	40185	1284	MS. BARBARA BOWER	167	FEB 2010 PA		
350	10000 · Checking	EFT	2	40191	1285	CAPITAL ONE	58	4121-7415-1		
351	10000 · Checking	EFT	2	40191	1286	FL SUN TERMITE & PEST CONTRO	93	472		
352	10000 · Checking	Check	3	40191	1287	WESTON TROTWICK	230	228		
353	10000 · Checking	Check	3	40193	1288	PRIME BANK	182	3010098		
357	10000 · Checking	Check	3	40193	1289	FIRST TITLE	92	VOID: DEPOS		
360	10000 · Checking	Check	3	40195	1290	MR. JOEL WEILER	152	VOID: DEPOS		
364	10000 · Checking	Check	3	40200	1291	MS JO NORRIS	165	JAN 2010		
365	10000 · Checking	Check	3	40207	1292	CITY OF TALLADEGA	73	1042809		
366	10000 · Checking	Check	3	40207	1293	CITY OF TALLADEGA	73	95813		
367	10000 · Checking	Check	3	40207	1304	FL TERMITE & PEST CONTROL	93	4734734		

BakerSCI © 2018

400 Broad St, Suite 221B Rome, GA | a
info@bakersci.com | e
706.331.5455 | p

Use Case #2: Journal Entry Testing – Fraud Examination

BakerSCI was given an extremely large journal entry dataset from a Big Four Accounting Firm to review for unusual activities in an alternative way than what is currently used by their internal data group. BakerSCI began with a download of the raw file from the accounting information system (below left). After a brief ingestion period, BakerSCI ran the data through its associative analysis system for a visual representation of the data prepped from analysis and manipulation. Results (one segment of many is below right) were created for several different events of interest, such as: last five days and first five days of a reporting period; journal entries booked and/or approved during the hours between 11pm and 6am; journal entries booked and/or approved on weekend days; and journal entries with material thresholds hitting the same accounts at different times of the day, month, etc. *Client is interested in using this type of visual approach to identify events of interest that they believe will make their professionals more efficient in finding the* starting points *for further investigation, thereby reducing overall analysis time on search and allowing them to spend that valuable (and expensive) time on figuring out why the event is happening.*

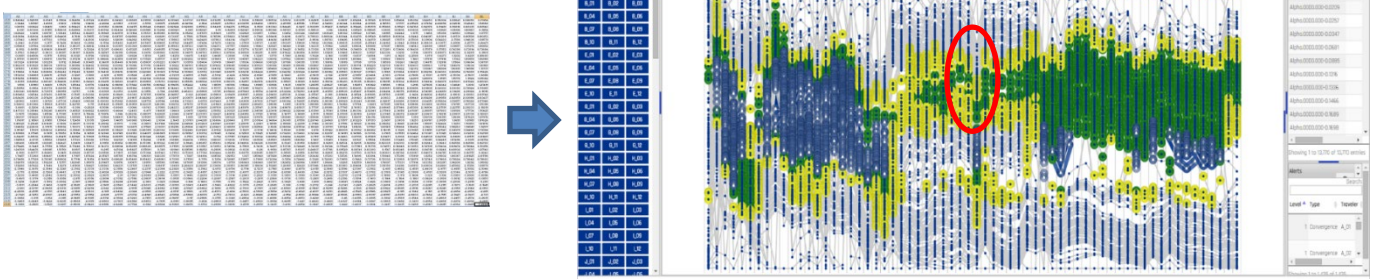


Deliverable B: Financial Service analysts need a system that shifts the burden of search to a machine learning environment so analyst efficiency is focused on analysis and speed of analysis is increased without sacrificing accuracy.

BakerSCI Solution – Through associative analysis, BakerSCI's RITR solution plods through datasets that, at first glance, appear to have too much going on whereby a human analyst would struggle with any type of initial analysis, let alone deeper knowledge and perception. Because BakerSCI looks into the behavior of how data elements interact with each other (and reveals those interactive events to the subject matter expert analyst), new insights become available for review and discussion that have the potential to change how decisions are made. For example, finding important anomalies in very large data sets, and/or trends in financial performance that could indicate larger events are about to take place.

Use Case #1: Finding "It" – Revealing Anomalies in Data

BakerSCI was approached by a government agency interested in new technologies that could help find "it". Agency was testing to see if new data analysis methods could see an anomaly that they already knew was in the data, but had yet to find new methods that could find "it" (without already knowing *where "it" was).* BakerSCI found "it". *Agency opened doors to other agencies and departments so BakerSCI's tech could be shown to their counterparts for both research and investigation purposes. Discussions are ongoing.*

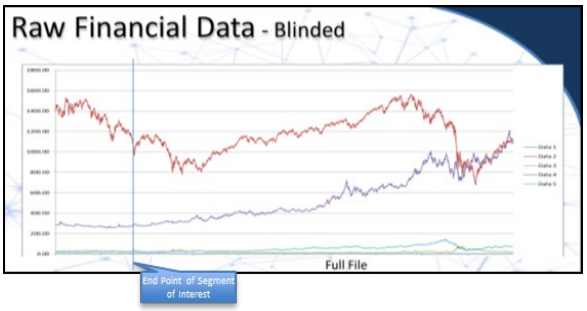


BakerSCI © 2018

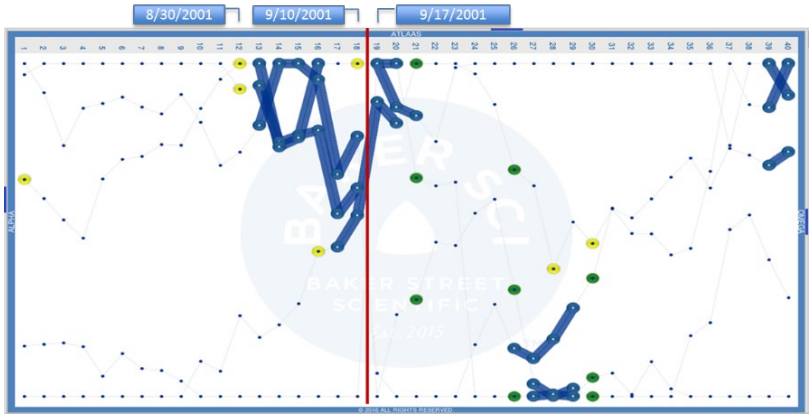
400 Broad St, Suite 221B Rome, GA | a
info@bakersci.com | e
706.331.5455 | p

Use Case #2: Financial Anticipatory Intelligence – Predictive Analytics based on Financial Transactions

BakerSCI was given a blinded dataset made up of ten (10) years of financial data. BakerSCI was not told anything about the dataset other than it was financial transactions. The three-letter agency wanted to see if BakerSCI could help them predict global events by "following the money" *before* an event occurred. BakerSCI ran the data, returned the results to the Agency and was afterwards notified of what we had found: the warm up of events in the financial markets leading up to the 9/11 attacks. *Agency discussed the possibilities of using BakerSCI to help them create an Early Alert System that would learn and monitor the* behavior *of certain financial transactions (amongst other items of interest) to help predict when nefarious bad actors may be transitioning from planning to engagement.*



Results graphed on typical MS Excel:
Excel line graph of 10 years of raw, BLINDED financial data; Large Y-axis hides information; No behavior interaction can be seen.



BakerSCI Results Outcome:
REVEALED: The system identified a saturation increase in associativeness of data **10 days prior** to the event. **After** the submission of these results, the entity provided the dates of the transactions. **Note: 9/11/2001 is the red line.**

BakerSCI Science & Technologies (S&T) Capabilities: BakerSCI's S&T address and mitigate/circumvent various hindrances of conventional algorithmic, statistical, and data analytic methods and systems through proprietary methods generally titled, Integrated Multi-Paradigm Processing™ (IMPP™). IMPP augments conventional systems' abilities by enabling them to discover substantially deeper data context about the micro and macro data environments. IMPP enables significant amplification of existing capabilities, Data Environment Association and Interaction™; and implementation of new capabilities such as auto-optimizing processing components and self-improving its metacriteria and metaconnections resulting in higher CL decision making.

These capabilities applied in the volume-heavy environment of financial transactions allow for a quick, visual interpretative tool that can assist data analysts in identifying critical, relevant events of interest for additional investigation.

Generally, BakerSCI's applicable capabilities to this problem can:

- test and determine data, algorithm and signature fitness in both conventional system and BakerSCI system implementations;
- employ *n*--point feature analysis™, where *n is a very large range of variables for the number of data points and their organization that can be analyzed by BakerSCI's S&T and subsequently taught to financial data analysts,*
- operate linear-based processing methods which can be run at high speeds in a wide variety of environments;
- increase CL by allowing SMEs to "tune" BakerSCI's unbiased algorithms with viable interest and modeling bias;
- reveal events of interest events that are based on the associatedness and behavior of the processing results rather than trying to attribute meaning to results that have already drawn conclusions (aka. analyze without bias, then add it later if necessary)
- run with proprietary algorithms and data evaluating processes (herein collectively defined as "algorithms") which are: 1) foundational, small and nimble; 2) atypical to conventional statistical and geometric algorithms 3) data / orientation agnostic; 4) unbiased; 5) associatedness and emergence-based/driven;
- run in new and novel software engine processes and designs,

BakerSCI Applicable Use Case Examples (Upon Request)
IMINT; SIGINT; MASINT; HUMINT; Healthcare Imaging and Genomic Research, Diagnostics and Compliance; Financial Intelligence, Transactions and Compliance; Energy Management & Distribution; Ideological Impact & Modeling; Human Performance; Logistics & Transportation.

BakerSCI © 2018