KPMG

Forensic Focus

kpmg.ca/forensicfocus

Finding the evidentiary needle demonstrating fraud in the big data haystack

By Samir Syed, Senior Manager, KPMG Forensic Services, KPMG in Canada

"Data is useless without context." – Nate Silver, *The Signal and the Noise, 2012.*

Nate Silver may have been talking about politics and baseball, however forensic accountants using data analytics to look for signs of fraud would do well to pay attention to his credo.

In the not-so-distant past, organizations faced a major hurdle in leveraging analytics to fight fraud: technical skills. People who could manipulate large data sets were scarce. Recently, riding the latest "Big Data" wave, analytical tools have exploded in availability, and improved dramatically in terms of affordability and usability. The ability to analyze data is now widespread.

Yet wide-scale deployments of tools such as Alteryx, Tableau and Qlikview eventually revealed another issue: Generic, one-size fits all analytics were very good at finding false positives, and often terrible at fighting fraud. The missing ingredient in most analyses: context.

While a pro baseball general manager can use analytics to evaluate a player, he should also look at the context in which that data was produced. This will inform the decision of what metrics (science) mean the most, and allow him to make adjustments based on his judgment (art). If two players have comparable home runs in college, but one of them played in higher altitude where balls travel further, should this be factored into that player's evaluation?

This article provides examples of how analytics are used to fight fraud, and how context is a crucial determinant in the success of the analytics.

Case Study 1: The Facilities Manager

KPMG was asked by a public organization to investigate a facilities manager who appeared to be circumventing procurement procedures to give maintenance work to his contractor friends.

One of the main concerns management had in this situation was that this practice led to his friends billing the company for fictitious work.

For certain known vendors, the company analyzed several years of billing data and devised simple metrics that may be indicative of potential irregularities and fraud: total daily billings divided by hourly rates.

However, being indicative of fraud (the haystack) is weaker and less actionable than being demonstrative of fraud (the needle).

KPMG's review of the client work revealed two major issues:

- The client falsely assumed that all invoiced costs related to labor. The costs actually included materials re-billed to the company.
- The client assumed all labor occurred on the invoice date. Contractors would often detail days worked in the invoice line item descriptions or in supporting timesheets. Some invoices covered several days of work. Therefore, flagging invoices with over 24 hours billed produced a lot of noise in the form of false positives and distracting from the true issue that existed.

As a result, when analytics were performed, both issues resulted in erroneous estimates of hours.

Despite having industrial analytics tools at its disposal, the client's corporate compliance function lacked a practical understanding of how contractors worked on the ground level, i.e. the "context". Their analysis had provided little to no value in detecting fraud.

KPMG used Optical Character Recognition ("OCR") and unstructured text analysis to index every contractor invoice and distinguish material from labor costs. KPMG also created an overall labor database from the data in the invoices, to identify what days contractors were billing hours on.

Finally, from the building physical access system, KPMG obtained the log of all magnetic card swipes ("Ins" and

"Outs") for the period under review. This allowed KPMG to improve the client's analysis in order to identify:
days where the contractors actually billed an impossible number of hours (i.e. over 24 hours in a day);
instances where work containing the same date and description was charged on 2 invoices;
days where contractors did not "swipe in" but where they billed labour.
By taking the time to understand the context, and by using advanced tools to analyze years of invoice data, time sheets and access logs, KPMG's work provided sufficient evidence to rely upon when identifying and taking further investigative steps and remedial action. Ultimately, in this case, it resulted in termination of the facilities manager.

Case Study 2: Benford's Law

Fraud professionals have been applying Benford's Law to large data sets for decades. For the un-initiated, Benford's Law is a mathematically-proven observation about the frequency of leading digits in naturally occurring sets of numbers. Because fraudulent transactions are often not normally occurring, judicious application of Benford can identify patterns of transactions that may indicate fraud. The method is so prevalent that many analytical tools come with native Benford functionalities. They allow analysts to apply the law to data populations to identify items meriting additional scrutiny. However, it is typically a glaring mistake to import a massive data set and blindly apply Benford.

Applying Benford generically, without adjusting for context, typically leads to either of two outcomes:

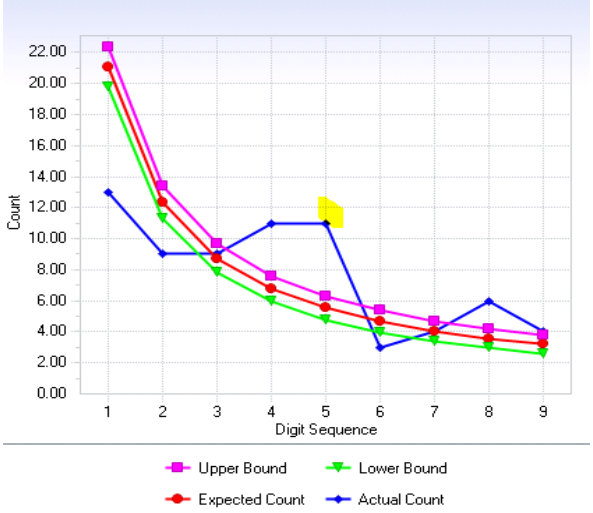
1. No anomalies are found, because even if fraudulent transactions are within the data, the size of the population will "drown out" the result. That is to say, if a fraudster has been creating a fraudulent recurring transaction of $5,000 per month, thus skewing his frequency for leading 5's, it will barely register in a population of millions of transactions posted by hundreds of users.
2. Large, recurring transactions are identified and explained away. Classic examples of this are: recurring rent, recurring phone bills, and recurring car payments. Clearly, in identifying these types of transactions, we are no closer to identifying fraud.

To this end, it serves the auditor well to understand the population and to perform several Benford analyses, on various subsets of the population.
In the example above, if the Benford analysis had been applied to transactions processed by the fraudster in isolation, and he was found to have made 300 transactions, an extra 12 of which begin with a 5, this result would have brought the investigator closer to uncovering fraud (see the graphical depiction of this exception).

Common ways in which to run Benford on sub-populations include:

- By User or Individual
- By Date (Year, month, day, etc.)
- By Location
- By Type of transaction
- By division
- By Stratified amounts
- Combinations of the above

Benford Analysis showing an over-representation of leading 5's, occurring 12 times.



Unfortunately, there is no hard-and-fast rule as to how to subdivide a population for Benford analysis. The context of each investigation is unique, and must be considered. The more the analyst can weave context into his or her analysis, the better and more relevant his or her results will be.

Putting it all together

The best analyses, those that truly zero in on the proverbial needle, are those that use the context to feed the analysis. They consider many factors, both qualitative and quantitative, and adjust their models accordingly.
While new tools and software suites have given investigators an unprecedented ability to analyze large data sets, care must be taken to always customize and adjust the analysis for the purpose or issue at hand. Running analytics in a generic, "one size fits all" fashion more often leads to noise than to a true signal of fraud.

Samir Syed
Senior Manager
Forensic Services
KPMG in Canada
T: 514-840-2696
E: samirsyed@kpmg.ca

For more information, visit kpmg.ca/forensic or

Contact us

Montréal
Stéphan Drolet
T: 514-840-2202
E: sdrolet@kpmg.ca

Myriam Duguay
T: 514-840-2161
E: myriamduguay@kpmg.ca

Dominic Jaar
T: 514-840--2262
E: djaar@kpmg.ca

Ottawa
Kas Rehman
T: 613-212-3689
E: kasrehman@kpmg.ca

Greater Toronto Area
Peter Armstrong
T: 416-777-8011
E: pearmstrong@kpmg.ca

Colleen Basden
T: 416-777-8403
E: cbasden@kpmg.ca

Enzo Carlucci
T: 416-777-3383
E: ecarlucci@kpmg.ca

Joe Coltson
T: 416-777-8786
E: jcoltson@kpmg.ca

Corey Fotheringham
T: 416-218 7974
E: coreyfotheringham@kpmg.ca

Southwestern Ontario
Karen Grogan
T: 519-747-8223
E: kgrogan@kpmg.ca

Calgary
Paul Ross
T: 403-691-8281
E: pross1@kpmg.ca

Vancouver
Suzanne Schulz
T: 604-691-3475
E: saschulz@kpmg.ca