A Beginner's Guide to Fraud Detection With Data Analytics

5 Essential Data Analytics Tests for Fraud Investigations

IDEA

The Persistent Problem of Fraud

We are living in an age of explosive new technology and abundant data. This has helped businesses around the globe grow and prosper in new and exciting ways. But it has also provided fertile ground for something that has been around for nearly as long as currency itself: financial fraud.

Fraud remains a persistent problem across all industries, a fact recently confirmed in PwC's 2020 Global Economic Crime and Fraud Survey, where over 5,000 respondents reported fraud totalling $42 billion in losses. Nearly half indicated they had experienced fraud in the 24 months preceding the report. Similarly, The Association of Certified Fraud Examiner's (ACFE) 2020 Report to the Nations on occupational fraud and abuse found over $3.6 billion in annual losses and a $125,000 median loss per case across 2,504 individual cases.

*5,000 respondents reported fraud totalling **$42 billion** in losses last year*

– PwC's 2020 Global Economic Crime and Fraud Survey 2020

Fraud is everywhere, fraud is expensive, and fraud is not going anywhere, but it can be fought effectively now more than ever before.

While much fraud activity is found initially through tips, internal audit plays a large role in detecting and investigating fraud. Internal audit once had a limited capacity to analyze the massive sets of data used to find fraud in enterprise organizations. But in addition to transforming the world of business, this era of explosive technology and abundant data has also handed auditors a whole new set of tools to find and fight fraud in the form of data analytics.

In this beginner's guide to fraud detection with data analytics, we look at the basics of how auditors and fraud examiners can detect, analyze, and even prevent fraud by using data analytics tools.

Fraud and Data Analytics

Common Types of Fraud

Fraud can take many different forms, but some of the most common financial fraud takes the shape of.

- **Fraudulent financial statements:** Concealed liabilities, fictitious revenues, or improper valuation.
- **Corruption schemes:** Bribes, kickbacks, hidden interests, improper gratuities and other related activities.
- **Asset misappropriation:** Theft of inventory and other assets. This tends to represent 85 to 90% of fraud cases.

When fraud investigators look at possible fraud of any type in large organizations, they are faced with massive sets of data to make sense of. In the past, auditors would look at only a small random sample of transactions and base their conclusions on sample analysis results. But samples cannot sufficiently capture irregularities across large organizations with vast amounts of data. Also, traditional electronic spreadsheet tools like Microsoft Excel are limited in their capacity to compute and analyze large datasets. As a result, many fraud cases go undetected.

Fortunately, the widespread emergence of data analytics within auditing has changed the game. In a recent paper on internal audit and data analytics, we reported that most respondents around the globe had either already incorporated data analytics into their internal audit approach or were planning to adopt data analytics in 2020. This suggests that 92% of respondents identify the value of data analytics in their audit approaches and have either adopted it or plan to adopt it this year.

This is a good indication for the profession at large and fraud investigations specifically. As the science of analyzing large datasets, data analytics allows for 100% coverage over a testing area and can help auditors detect and investigate fraud incredibly efficiently, going beyond meagre sample sizes for comprehensive, big-picture insight.

In areas like fraud investigations, being able to use data analytics is critical to ensuring auditors have the best chance to capture all of the fraud that might have occurred, and to look at new areas where a client might not have even suspected fraud.

Fraud examiners do not have to be data scientists to perform data analysis within fraud investigations. Data analytics software tools have made it easy for auditors to analyze massive sets of data quickly and efficiently.

Learn how this CPA used data analytics to help fight fraud with the US Government Accountability Office (GAO).

5 Essential Data Analytics Tests for Fraud Investigations

If you are just beginning your foray into data analytics for fraud investigations, there are a number of critical tools and tests to familiarize yourself with.

Here we present some of the data analytics tests, tools, and approaches auditors and fraud examiners can use to support their fraud investigations.

1. Benford's Law Analysis

Benford's Law *states that the leading digit of a large set of naturally occurring numbers is likely to be small*

Benford's Law analysis is one approach that can be used effectively within a fraud investigation. The law upon which it is based states that the leading digit of a large set of naturally occurring numbers is likely to be small.

Frank Benford, the law's namesake, was a physicist who observed that the first few pages of his logarithm tables were more worn than the last few pages. This led him to believe that he looked for logarithms beginning with low digits much more frequently than those beginning with high digits.

He then looked at 20 lists of numbers with 20,000 records as diverse as the surface areas of rivers, US populations, physical constants, molecular weights, numbers in a Reader's Digest issue, street addresses, and death rates. The results, published in a 1938 paper showed 30.6 of the numbers had a leading digit of 1 and 18.5 percent of the numbers started with a 2. In other words, 49% of the numbers started with a 1 or a 2, with the remaining digits appearing as the first digit in decreasing frequency from 3 to 9.

It is important to reiterate that Benford's Law applies to number sets that reflect the magnitude of some phenomenon, not where there are built-in maximum and minimum values (e.g. some parts of tax returns, such as actual income and expenses) or where numbers are used as labels or set at thresholds (e.g. highway numbers, social security numbers, phone numbers, postal codes).

However, since Benford's Law applies in frequency distributions across an array of real-life data, it is unsurprising that it can be found in financial data, including:

- Accounts payable transactions
- Credit card transactions
- Customer balances and refunds
- Disbursements
- Inventory prices
- Journal entries
- Loan data
- Purchase orders
- Stock prices, T&E expenses, etc.

Since Benford's Law can be used within sets of financial data, it can be especially useful to help detect invented numbers when one individual has fabricated all of the numbers, or when many different people have incentive to manipulate the numbers in the same way (e.g. tax returns).

The example below is an example of a Benford's Law analysis on a database of over 10,000 vendor invoices. In a Benford's Set, the first digit would be 3 only about 12% of the time, but here it appears as the leading digit 26% of the time, a considerable deviation. This anomaly would be an obvious target for further investigation.

First Digit	Benford's Set	Data Set X	Deviations
1	30.10%	24.00%	0.06
2	17.61%	18.00%	0.00
3	**12.49%**	**26.00%**	**-0.14**
4	9.69%	11.00%	-0.01
5	7.92%	5.00%	0.03
6	6.70%	7.00%	0.00
7	5.80%	5.00%	0.01
8	5.12%	2.00%	0.03
9	4.58%	2.00%	0.03

Benford's Law tests can be applied in some of the following situations:

Forensic audits: Cheque fraud, bypassing permission limits, improper payments, etc.

Corporate finance/company evaluation: Examining cash-flow forecasts for profit centres.

Financial statement audits: Manipulation of cheques, cash on hand, etc.

Anti-fraud programs: From risk assessment to monitoring.

For auditors, Benford's Law is a very useful high-level test of reasonableness. It can indicate abnormal duplications or find irregularities between one period's financial data against that of a previous period.

Some key Benford's Law tests used within data analytics tools include:

- **First-digit test:** A high-level comparison of a given data set against a Benford's Set. This only provides a glimpse of the obvious and should not be used to select audit samples.
- **Second-digit test:** A high-level test that helps identify conformity.

- **First-two-digits test:** A more focused test that examines the frequency of the numerical combinations 10 through 99 on the first two digits of a series of numbers. It can be used to select audit targets for preliminary review.
- **Summation test:** Detects large numbers compared to the rest of the numbers of a population.
- **Second-order test:** Looks at the differences between numbers and indicates whether something is problematic with the data. This test will sort a numeric field from smallest to largest where the value differences between each pair of consecutive records should follow Benford's Law digit frequencies.
- **First-three-digits test:** A highly-focused test that can be used to select audit samples and tends to identify number duplication.
- **Last-two-digits test:** Helps detect invented, overused, or rounded numbers (can be used in coupon counts, inventory counts, odometer readings, etc.)

Learn More About Benford's Law for Fraud Detection

Benford's Law is no panacea when it comes to fraud investigations, but as with all fraud tools, it can provide a useful start to get a better understanding of the data before you. Here are a couple of resources to better understand the role of Benford's Law in fraud investigations:

- The Application of Benford's Law Within IDEA
- Practical Applications of Benford's Law

2. Joins and Correlation Analysis

Joining Databases

One critical feature of data analytics is its ability to compare and analyze two different databases that are not normally compared with one another. Data analytics tools for auditing should include a "join" feature, an incredibly simple yet powerful function for joining databases. It is used to:

- Combine fields from two databases into a single database for analysis.
- Test data for matches (or the lack of a match) across databases.

The join function can be used to reconcile data between two databases, such as bank statements or sales reports and invoices, to identify errors or fraudulent and suspicious activity. Once two or more databases have been joined, fraud investigators can take their investigation a step further with correlation analysis.

Correlation Analysis

Correlation analysis is another data analytics tool that can be used to determine whether there is a connection between different pieces of information in two or more sources, such as income statements and accounts payable transactions. If a strong correlation between the two is found over a period of time, there may be an indication of fraud. Correlation analysis can also be used in areas that most people would not consider in fraud testing, such as with the correlation between the outside temperature and heating expenses. In general there should be an opposite correlation between the two, in that when the temperature decreases the heating cost should increase. If this is not the case, these entries should be flagged for further investigation.

With correlation, a regression analysis can be applied to datasets to flag any abnormalities in the relationship between the two compared to history, industry norms, and other benchmarks. This can help identify any critical changes that can be further investigated for signs of fraud.

Correlation analysis can be performed manually on smaller datasets with spreadsheets, but larger datasets require the power of audit-specific data analytics tools to analyze numbers swiftly, efficiently, and comprehensively.

3. Same-Same-Same and Same-Same-Different Tests

The purpose of the same-same-same (SSS) and the same-same-different (SSD) tests is to identify abnormal duplications as potential indicators of errors or fraud.

The SSS test identifies records that contain fields of information which are exact duplicates of other records. This test helps detect duplicate expenses claimed, occurrences of the same payment to vendors made in error, multiple warranty claims, or duplicated service fees paid by private or government health plans, for example.

Alternatively, the SSD test is used to identify records with near duplicates for fields selected by the users. The SSD test is valuable for detecting errors and fraud, and especially useful for detecting errors in accounts payable data.

For example, when a payment is made to an incorrect vendor initially and later the correct vendor is properly paid, we can find records with the same invoice number and amount but different vendors. If a business system does not process orders in real time, customers may be attempting to split their orders to avoid exceeding their credit limit (e.g. same invoice date, same customer number, same product code, and different sales representatives).

4. Gap Detection

*Gap detection **is an easy way to find fraud** by identifying missing items in a sequence.*

Often, a telltale sign of fraud is the information auditors and fraud examiners cannot find. Series such as invoice numbers, cheques, and purchase orders are normally sequential and without gaps. For example, invoice numbers should not repeat or be skipped entirely within a numerical sequence.

Gap detection, commonly used as a test for completeness, can be used to identify missing items in a numerical sequence or a range of dates in numeric, character, or date fields in a database. A gap indicates missing items and this test is commonly used to test for completeness.

Series such as purchase orders, invoice numbers, and cheque numbers are typically sequential and any gaps should be accounted for. Gap detection can be run against character strings, number sequences and dates to search for missing information, such as missing invoices.

Gap detection can also be used to look for missing transactions. An example of this type of use can be found in the restaurant business, where auditors might look for unexpected gaps in the transactions that could indicate where transactions were not entered or removed, indicating unrecorded sales. This is an especially useful test for tax authorities.

5. Fuzzy Matches and Duplicates

Fuzzy duplicate tests are an easy way to cast a spotlight on duplicate information that is otherwise notoriously difficult to find.

Fuzzy matches and fuzzy duplicates are key detection approaches used to find matching and similar records in character fields. The operative word here is "similar"; there are already many tools for finding exact matches.

Once records are identified as having similarities, they can be gathered into groups called fuzzy groups and ranked according to their similarity degree. The more similar the data, the higher the similarity degree. For example, if an exact match like the names Smith and Smith is considered to have a similarity degree of 1, two very similar names like Schmidt and Schmid will have a similarity degree of 0.85.

Fuzzy duplicate tests are useful for:

- Fields that contain single words (e.g. stock market IDs, foreign exchange codes, etc.) as well as character strings where the sequence is important, such as a phone number with a national code and an area or regional code as well as the local number.

- Phrases or short sentences where word order is important (e.g. a business name or an address.)

In practical terms, the tool can help find entries with slight differences, such as a spelling error, or variations introduced during data entry. An example of a data variation entry can be seen in the words "road" or "street" when used in an address. It's common to see these words abbreviated, but period use is often inconsistent. Even though we recognize these entries as variations that represent the same entity, this minor difference is enough for St and St. to be identified as unique keys in a field. These kinds of duplicates are notoriously difficult to find, but a fuzzy duplicate search makes finding them much easier.

Additional Considerations

The sobering reality is that fraud is virtually inescapable today, and traditional spreadsheets and sample-based approaches are ill-equipped to handle it effectively. However, there should be a great sense of optimism surrounding the anti-fraud capabilities presented in the powerful world of data analytics. This science and discipline represents a quantum leap for fraud investigations and allows auditors and fraud examiners to make sense of massive amounts of data from different sources quickly and efficiently.

In this paper, we have provided an introduction to some of the data analytics tools used to support fraud investigations. We caution that simply having a tool is not enough, and offer some of the following takeaways as you explore the applicability of data analytics in fraud detection, analysis, and prevention.

Think people, process & technology

Data analytics approaches are invariably supported by data analytics technology. When you either implement a new data analytics software solution or optimize the use of an existing solution, it is important to remember people, process, and technology all play central roles in utilization. Software is only effective when it is adopted by users. You do not have to be a data scientist or programmer to use the best data analytics solutions for auditing and fraud detection. Find a user-friendly tool that incorporates best practices and can be easily used by entry level auditors and experienced fraud examiners alike.

There's no one-click fraud investigation

Sometimes auditors expect data analytics technology to analyze data and detect fraud in a few keystrokes. While data analytics software contains powerful fraud detection capabilities, fraud investigations will always be an art and science. This is promising for auditors, for while technology will always empower them to take fraud investigation further, it will never take their jobs. The critical mind of a cunning fraud examiner will always be essential to an effective fraud investigation. In this way, the powerful tests outlined above (Benford's Law, correlation analysis, gap detection, fuzzy duplicates, etc.) are but tools auditors can master to improve their ability to detect fraud.

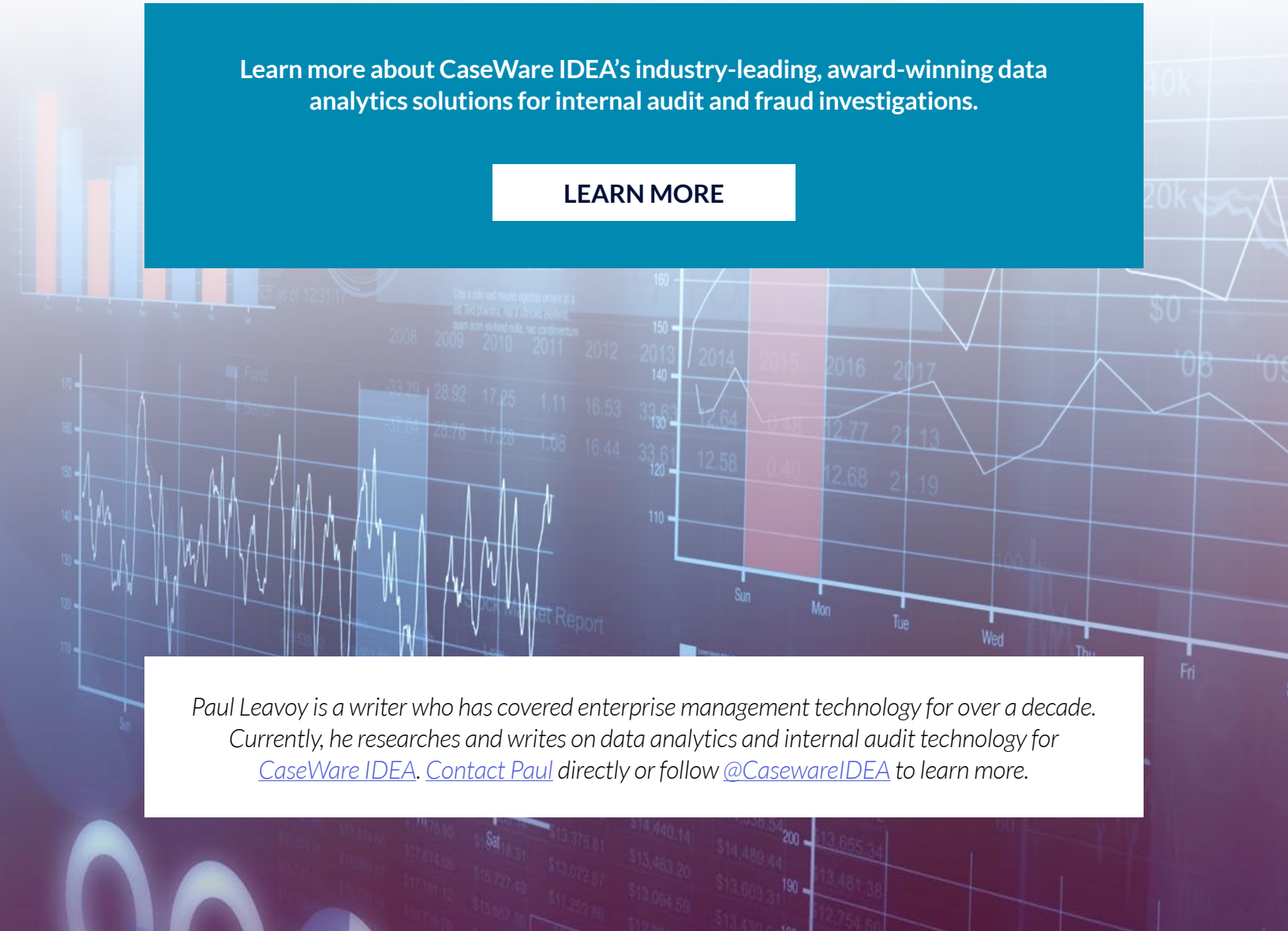
Stay current with data analytics technology

Because both the audit landscape and technology are rapidly changing, auditors must periodically take stock and re-evaluate how they do their work. The sheer volume of data and transactions today is massive, so auditors must find the most effective ways of navigating through the data to find anomalies. This can be challenging in a field that is sometimes criticized for its perceived aversion to new technologies, even though traditional approaches and spreadsheet-based tools are insufficient for comprehensive fraud investigations.

Data analysis technology for auditing is not new. For example, CaseWare IDEA has been helping clients detect, analyze, and prevent fraud for three decades. However, data analytics continues to completely change the way we investigate fraud. Once upon a time sample analysis was viewed as sufficient, but between the sheer power of data analytics tools and the massive amounts of data available to us today, sampling is a thing of the past and data analytics needs to be the new normal in the ongoing fight against fraud.

Learn more about CaseWare IDEA's industry-leading, award-winning data analytics solutions for internal audit and fraud investigations.

LEARN MORE



Paul Leavoy is a writer who has covered enterprise management technology for over a decade. Currently, he researches and writes on data analytics and internal audit technology for CaseWare IDEA. Contact Paul directly or follow @CasewareIDEA to learn more.

CaseWare International Inc.

1 Toronto St, Suite 1400, Toronto, ON M5C 2V6 Canada

416-867-9504 | sales@caseware.com | www.caseware.com

About Us

CaseWare IDEA is an internationally recognized data analytics software company that provides cutting-edge solutions for accounting firms, corporations, and governments. A leader in the audit and accounting industries for over 25 years, IDEA® Data Analysis Software equips auditors, accountants and other finance professionals with a toolkit that allows them to perform data analysis quickly for various audit-related tasks. IDEA uses artificial intelligence and machine learning to change the way we think about and work with data. The result: measurable returns on efficiency, quality, and value. CaseWare IDEA is a division of CaseWare International, which has led the industry for over 30 years, with solutions supported in 16 languages and used by more than 500,000 people across 130 countries.

To learn more visit idea.caseware.com.