

Big data can tell unexpected stories: The chief financial officer who had a conflict of interest with a supplier to whom he had awarded a multimillion-dollar contract. The two employees who provided their company-supplied fuel cards to family members to refuel their personal vehicles. The executive who had an affair with a union official during wage negotiations.

Internal auditors never could have discovered such wrongdoing through traditional audit sampling, walk-throughs, or reliance on the representations of management. They were only found by using business intelligence tools to mine data sources that are now routinely available.

BUSINESS INTELLIGENCE FOR AUDITORS

Audits typically entail inquiries of management, walk-throughs, and transaction sampling as a basis for statistically inferring the effectiveness of each internal control attribute under review. To be generalizable within a given confidence interval, transaction samples need to be both large and randomized to represent the entire population. In doing so, internal auditors usually presume that the population conforms to a normal bell curve. This brings with it the risk

Auditing with self-service business intelligence tools can mine the organization's data sources to provide greater assurance.

Beneath the Data

**Christopher Kelly
James Hao**

Illustration by Edmon de Haro





that if the sample is too small, the tests are performed with insufficient care, or the population is skewed differently from a normal bell curve, the auditor may form the wrong conclusions about the control's true characteristics. If the population contains any erroneous or fraudulent transactions, it is unlikely they will turn up in a walk-through or random sample.

Today's self-service business intelligence tools expand internal audit's toolkit from mere questionnaires and sampling to mining entire data populations. These tools make it easier for auditors to mine data for errors such as anomalous transactions and fraudulent data correlations (see "Mining for

movement of thousands of nonhomogeneous goods. Inventory is vulnerable to receipting errors, barcode misreads, obsolescence, rot, and shrinkage.

Things often go wrong in inventory, and audits often have revealed downside errors of 10 percent of inventory value. Therefore, internal audit could focus on ensuring quantity and description data matches physical reality through accurate goods receipting into the accounting system, precise sales capture, and reliable stock-taking. Once inventory data reflects the physical goods on hand, data mining can assist with identifying:

- » Slow-moving and excessive inventory build-up.

organizations process thousands of supplier payments daily, so errors are likely. Data mining can include:

- » Matching supplier master data such as bank account numbers, addresses, and telephone numbers to employee and next-of-kin master data for unexpected relationships.
- » Isolation of purchase orders or payments just below authorization thresholds.
- » Erroneous duplicate invoice payments because of optical character recognition or human error when entering invoice references such as mis-entry of "I" instead of "1," or "S" instead of "5," or "/" instead of "\".
- » Historic credit notes that have never been offset against subsequent payments and remain recoverable from suppliers.

Audits using these tests have experientially revealed an average of 0.1 percent in errors, which enabled organizations to recover cash refunds from suppliers. Auditing over several prior years can result in material financial recoveries.

Beyond financial transactions, auditors can use business intelligence tools to access newly available data sources.

Errors" on page 45). In this way, auditors can pinpoint actual error, fraud, and cost savings that demand action.

Beyond financial transactions, auditors can use business intelligence tools to access newly available data sources such as telecommunications, email, internet usage, road tolls, time sheets, maintenance schedules, security incident logs, clocking on/off, and electronic point-of-sale transactions. Previously, many of these sources either were not auditable or were stored as manual records. Business intelligence tools open the door to a variety of audits.

Inventory For many organizations, inventory is a complex and poorly understood process. Organizations record movements in cash, debtors, and creditors within their financial systems. Yet, inventory data easily can get out of step with the physical daily

- » Book-to-physical adjustments pointing to shrinkage or theft by location.
- » Refundable stock that can be returned to suppliers.
- » Stock-outs where the organization lost sales because of insufficient demand analysis.
- » Negative quantities revealing goods receipting or similar process errors.

This kind of audit analysis demonstrates the informational value of having accurate inventory data. Such information can lead the organization to prioritize which inventory processes most need fixing.

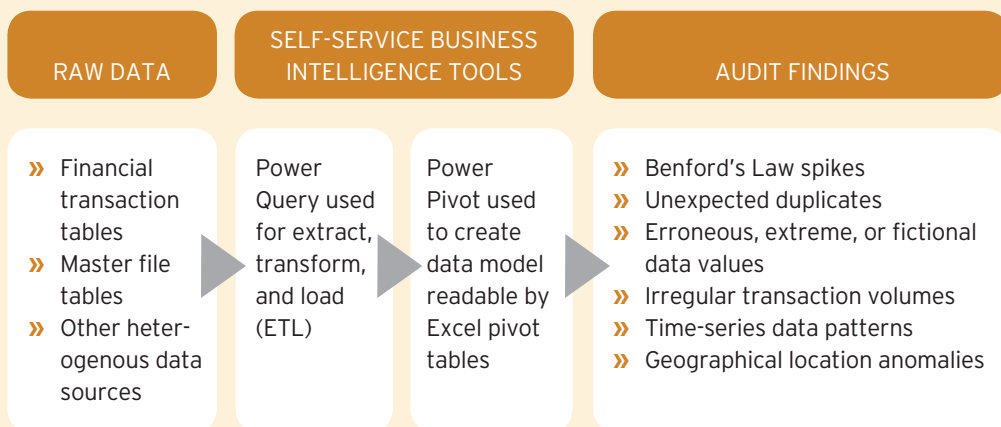
Supply Chain Organizations need to know supplier agreements do not conceal undeclared conflicts of interest and suppliers are paid no more than their contractual entitlements. Even small

Payroll For most organizations, payroll is the largest single cost. The board and audit committee need to know overpaying or underpaying employees is minimized. Payroll data mining can include comparing hours paid to hours actually worked by matching sick leave and holiday to other time- and location-stamped data such as building entry/exit data, cell phone metadata, and email data. In doing this, internal auditors can present management with compelling evidence that supports corrective action. Moreover, previous audits have uncovered savings of about 1 percent of total payroll cost from:

- » Claiming fictional hours on time sheets.
- » Falsely claiming to be working at home or on paid sick leave.

MINING FOR ERRORS

The diagram below summarizes the steps from raw data to audit findings when internal auditors use Excel's Power Query and Power Pivot features.



- » Missing scheduled training.
- » Finding repetitive patterns of fictitious sick leave taken on Mondays, Fridays, and the day before or after public holidays.

Company Motor Vehicles Auditors can mine data gathered from vehicles, including road tolls, refueling, traffic penalties, and insurance claims. This jigsaw puzzle of data can show auditors how vehicles are being used for business purposes, possible abuse of vehicles, and drivers with poor driving histories that result in unnecessary cost. This data can be obtained from external motor fleet providers and insurers. Such audits can recover around 5 percent of fleet costs.

Metadata While the content of company-issued cell phone calls and text messages is confidential, the accompanying nonconfidential metadata includes called numbers, durations, date and time stamps, and base station geographical locations. Auditors can discern employee activity, interconnections, and external relationships during work hours or while on paid sick leave by

matching this metadata to other sources such as the organization's telephone list and employee and supplier master files. Internet usage metadata provides similar insights. These data sources can help when investigating white collar conflicts of interest and fraud.

These are just a few areas where business intelligence opens new portals. Partnering with the chief information officer can help internal audit access the organization's databases. Once access is granted, auditors can use business intelligence tools with minimal assistance.

GETTING STARTED

With business intelligence, auditors are no longer constrained by Microsoft Excel's 1,048,576 row limit. Excel 2016 includes built-in business intelligence tools, Power Query and Power Pivot. Power Query is an extract, transform, and load (ETL) tool that reads source data and makes it available for Power Pivot for data modeling. This source data typically comes from comma- or tab-separated outputs from other systems. Auditors can access Power Query

under Excel 2016's Data ribbon, where it is also known as Get Data and, once opened, Query Editor.

Power Query and Power Pivot have formula languages that allow users to create new data columns specific to their own unique needs. Power Query uses M formula language and Power Pivot uses Data Access Expressions (DAX). Both languages differ from Excel formulas. Whereas Excel formulas are not case sensitive and usually do not distinguish among string, date, and numeric data types, M and DAX are sensitive to both text case and data type. This distinction is important when manipulating data and performing calculations.

Once internal auditors have loaded and edited the raw data down to only the needed columns in Power Query, they can add each table to the Power Pivot data model under the "Add to data model" option. Auditors can then access Power Pivot from Excel under "Manage data model." From there, they can use the "Diagram view" to link tables such as transaction files keyed to their corresponding master files. The data model can handle multiple external data

APPLYING BUSINESS INTELLIGENCE USING BENFORD'S LAW

The steps below illustrate how business intelligence tools can enable internal auditors to use Benford's Law to annotate original source data with leading digits. Leveraging Power Query, Power Pivot, M, DAX, and standard pivot tables together can produce audit insights.

STEP 1. Using a data-cleansed table created in Power Query, create additional field columns using "Add column/custom column" to capture the leftmost 1, 2, 3, etc. digits for Benford's Law analysis.

Data mining learning point: Being able to create custom columns in the data model is key to internal audit's ability to generate original insights. Auditors should not create too many new columns or the data model may become unmanageable within their computer's memory limits.

STEP 2. To avoid picking up the dollar-cent decimal points, multiply the amount field by 100 to convert it into whole cents and then use an M formula to convert the absolute amount to text and pick up the two (or three or more) leftmost digits. For example:

```
= Text.Start(Text.From(Number.
Abs([Amount]) * 100), 2)
```

Data mining learning point: In this M formula, auditors are isolating the two leftmost digits and "[Amount]" is the literal field heading from the source text file. Note that the formula syntax differs from Excel.

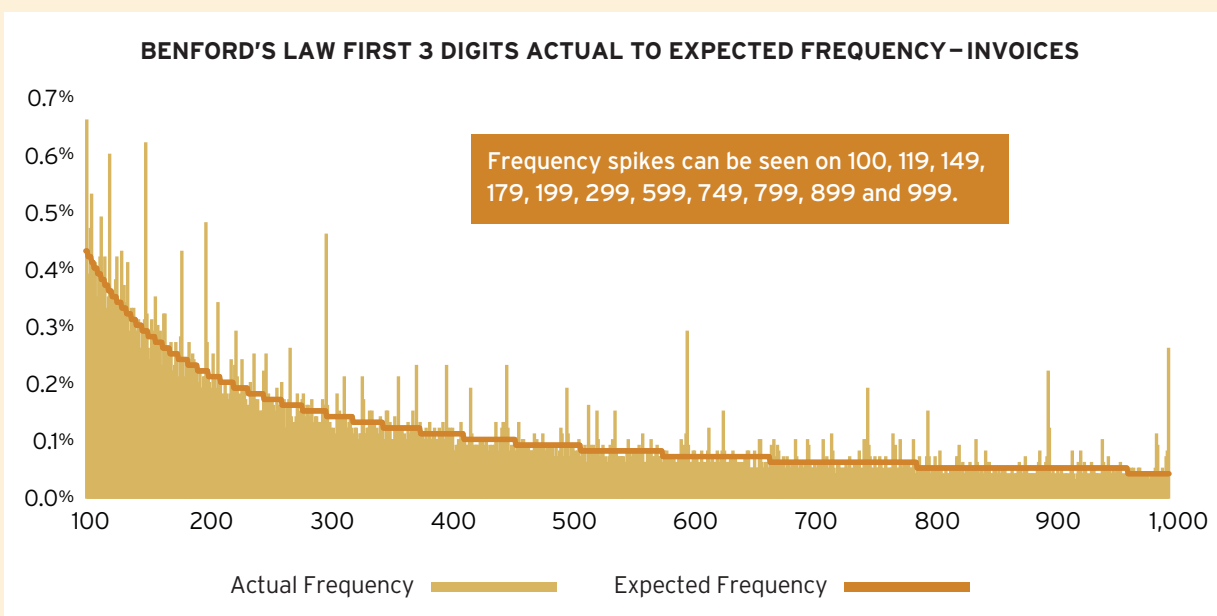
STEP 3. Once the desired Benford analysis columns have been created in Power Query, refresh the data model in Power Pivot.

Data mining learning point: Now that the raw data is in the Power Pivot data model, auditors can access the entire table, including the new column added in Step 1 within Excel's standard pivot tables.

STEP 4. Show the leftmost digits using the pivot table's "Show Values As/% of Grand Total" and compare this to the expected logarithmic frequency under Benford's Law. Then, visualize the resulting columns with a chart to highlight spikes between actual and expected frequency.

Deviations are most likely to have occurred where a systematic weakness has been exploited repeatedly.

Data mining learning point: Double clicking each deviant spike in the pivot table will display all the individual transactions that caused the spike, which auditors can then scrutinize for irregularities. In this way, Excel can instantly find the deviant transactions from a huge data population.



87% of organizations have low business intelligence and analytics maturity, yet Gartner estimates self-service analytics and business intelligence will produce more analysis than data scientists in 2019.

sources as well as normal Excel tables. This capability allows auditors to create multidimensional relational databases rather than two-dimensional flat files.

Power Pivot enables auditors to annotate the relational databases retrieved in Power Query with unique columns and measures specific to audit needs, which can be analyzed using Excel's pivot tables. "Applying Business Intelligence Using Benford's Law" on page 46 illustrates how Power Query, M, Power Pivot, and Excel can work together to search for irregularities.

DATA CLEANSING

Data files usually need to be cleansed before analysis. That is because over

objective of testing the entire population. If time allows, the auditor may cleanse the text files field-by-field in a spreadsheet or word processor by rejoining broken records, recalibrating misaligned fields, trimming stray characters or spaces, replacing known error values with blanks or zeros, and converting dates stored as text to real dates.


Further cleansing may be required if source files are fragmented across different years or subsidiaries and need to be joined into a single table, or if source files are tabulated differently from how internal audit wants to use them. In the first case, Power Query can append files into a single data source provided the field headings are identical. In the

Even with software, four trillion lookups could take several hours. Auditors can increase query efficiency by indexing, compartmentalizing a large query with efficient calculated fields, and filtering out unwanted columns or transactions that are blank or below a given materiality threshold.

SECURING DATA

To avoid internal audit being the source of a leak, or to limit the damage if the unthinkable occurs, auditors should take care with data. Auditors can exclude fields that identify living individuals, home addresses, or bank account numbers from downloads or replace them with codes such as an employee number instead of a name. They should be cautious when transmitting data to ensure USB drives are secure and electronic data is not emailed to unintended recipients. Auditors should check recipient email addresses before hitting "send." Password protection and encryption should be used when practical. As auditors only need to work on copy data—rather than live data—they usually can destroy their version and wipe USB drives after the audit is completed.

ORIGINAL INSIGHTS

Business intelligence tools unlock new ways to audit. With only a little new learning, business intelligence tools can expand internal audit's adventures into new pools of financial and operational data that may reveal risk and control insights. Moreover, because even the most innocuous transactions leave data trails, imaginative analysis can uncover errors, fraud, and cost savings that transform audit reports into compelling reading for executives and the board. 

CHRISTOPHER KELLY, DPROF, FCA, MIIA, is partner at Kelly & Yang in Melbourne, Australia.

JAMES HAO, CPA, is an associate at Kelly & Yang in Melbourne.

Internal auditors should keep a record of data cleansing actions in case future rework is required.

time, original source data is input by a variety of users whose training and attention to accuracy may be inconsistent. Some fields may hold invalid data as a result of being migrated from different systems or different versions of the same system. Moreover, stack overflow and other error types may lurk in historic data, the text files may have misaligned some fields, and records may be broken across two or more rows.

Comma-separated text files can present extra cleansing problems if users have input commas into individual fields. For example, "Kelly & Yang, Inc" would translate into two separate fields because of the comma, whereas "Kelly & Yang Inc" would translate into one field.

ETL tools will attempt to read all transactions from the raw data files. But if the tool encounters errors, it may exclude them from the upload, resulting in loss of data that dilutes the

second case, auditors can untabulate inappropriately tabulated source files back into a single column of data using Power Query's Unpivot command.

Internal auditors should keep a record of data cleansing actions in case future rework is required. Any updates to source data made in Power Query will need to be refreshed in the Power Pivot data model as well as in dependent pivot tables.

EFFICIENT QUERIES

Business intelligence tools are faster than previous versions of Excel, but internal auditors still need to be mindful of formula efficiency. If the auditor tries to add a new calculated field to a data model that requires a row-by-row lookup of each element in a two-million-row database, that could easily result in two million x two million = four trillion separate lookups.