



DRIVING BETTER DATA ANALYSIS

Moving Beyond Spreadsheets



AN IDEA® DATA ANALYSIS E-BOOK

JANUARY 2015



Table of Contents

Data Analysis - Then & Now	1
Changing of the Guard	2
New Generation	4
Core Data Analysis Tasks	6

Data Analysis- Then & Now

Due to their convenience, spreadsheets remain one of the most popular applications for auditing and fraud detection. Originally designed for financial modeling and budgeting, they are still widely used for database manipulation and internal audit activities. However, as best practices and mandates of the internal audit and fraud detection professions have matured, a reliance on spreadsheets has diminished in favour of more sophisticated tools.

With new directives from executive management to provide accurate, timely and fact-based assessments, the savings typically associated with spreadsheets have become more of an illusion than reality. Even though the investment for spreadsheet technology is low and the tool is familiar, the issues of data errors, record input limitations, and slow speed of analysis represent much greater hidden costs to the organization.

With the increasing volume of transactions flowing through organizations today, the accurate examination of individual transactions is of paramount importance. Often times, this cannot be accomplished using only spreadsheet analysis. Without the complete picture of raw data, organizations are more susceptible to fraud and abuse, possibly resulting in negative consequences.

"As a result of all of these adjustments, the company's third-quarter loss increased by \$9 million... there were too many zeros added to the employee's accrued severance. But it was an accrual. There was never a payment."

Company Spokesperson, Multi-national equipment and services company, commenting on a severance error traced to a faulty spreadsheet that resulted in a \$9 million loss and restructuring of the company earnings.

For more insights on companies who have endured costly spreadsheet errors, visit: <http://www.eusprig.org/stories.htm>

Changing of the Guard

Attributes such as ease of data input and manipulation that have made spreadsheets such a gift for data entry professionals have in turn become a curse for those needing to perform thorough data analysis. Some of the primary reasons to look beyond spreadsheets for data analysis are:

1. Lack of Data Integrity

Data values can be easily altered, in a number of ways, often times by accident and sometimes deliberately. Formula errors across spreadsheets also make the analysis logic prone to mistakes. Even the simple process of pulling data from different systems and formats into a spreadsheet can create errors.

2. Slow Data Analysis

Analyzing millions or even billions of columns of data with spreadsheets is often sluggish or non-existent, due to its record limitations. Today's data comes in varying sizes and types; merging data across spreadsheets to get the "true" view of the data or for simply conducting analysis remains a complex task and requires many steps. The process also lends to manual errors.

Example: Aging Formula

Use the Age Band column as the column field in the Pivot Table.

The oldest (i.e., first) date should be older than the oldest record. For simplicity, use "1" (1/1/1900) as the date. This will represent the "X days +" band.

`D2 =VLOOKUP(C2,Sheet2!:B:6,2,TRUE)`
With the following table in Sheet2:

	A	B
1	Start Date	Age Band
2	01/01/1900	90 days +
3	03/10/2010	61-90 days
4	02/11/2010	31-60 days
5	02/12/2010	0-30 days
6	01/01/2011	Future

Changing of the Guard

3. No Security, Non-Compliance

Due to the inherently risky and unsecure nature of spreadsheets, they pose a threat to proper retention of records. Most audit and compliance programs have strict standards and guidelines regarding retention and reliability of information. As a spreadsheet gets emailed along, users tend to have their own duplicates and possibly conflicting copies.

There is no proper way to track the copying and pasting of data and formulas across spreadsheets or data input from multiple sources. The inability to track edits hampers spreadsheets in meeting industry standards for achieving high levels of assurance or compliance.

"In order to be suitable, information must be relevant, reliable and timely."

COSO Guidance on Monitoring Internal Control Systems, Volume II: Application

New Generation

Internal auditors and fraud detection analysts have started using audit purpose built solutions to aid in analysis. These solutions help manage the new responsibilities of capturing and providing more frequent business insights and can provide:

1. Reliable and Accurate Data

Data Access:

Get data from any source, such as text files (flat files generated from ERP stems), MS Access®, MS Excel®, SAP®, Oracle®, SQL®, JD Edwards, or native PDF formats.

Data Integrity:

Source data is protected and data access is read-only. This ensures reliable data.

2. Quick Analysis

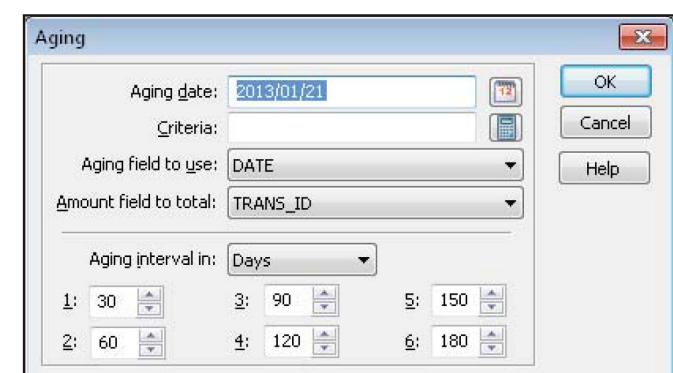
Audit Specific Tasks:

Without needing to program intricate macros or multiple pivot tables, algorithms exist to readily perform tests, simply by selecting a task. Examples include tasks to do duplicate and gap detection, join and relate databases, stratify, or Benford's Law analysis.

Processing Power:

An infinite number of records can be read and processed in seconds.

Example: Aging

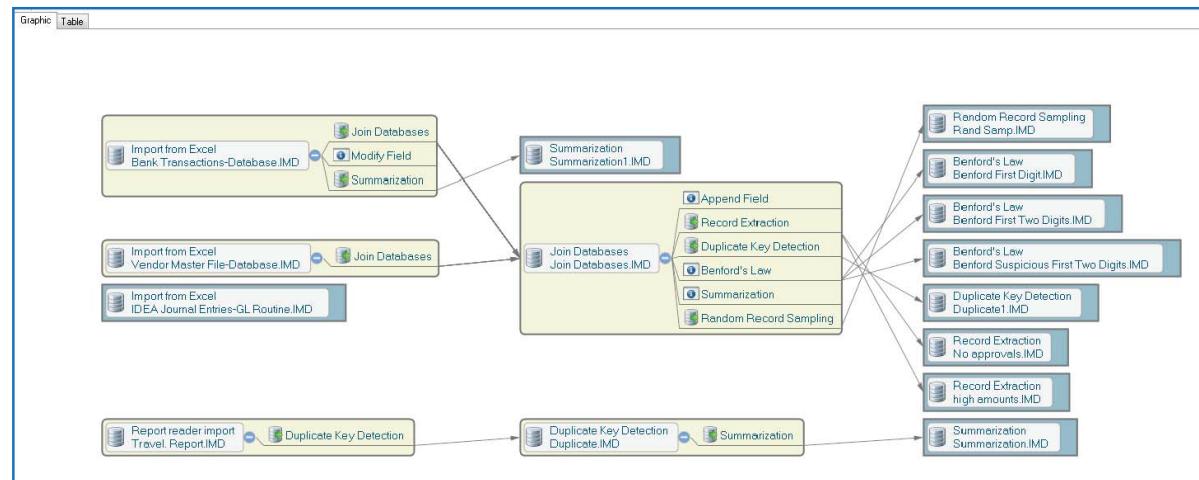


New Generation

3. Security & Compliance

Audit Trails:

Keep a record of all changes made and maintain an audit trail or log of all operations carried out on a database, including file and format imported, the type of analysis performed, and result created.



"Once imported, analyzing and extraction can be performed easily on the data to target the audit risks. You can manipulate the data without affecting the integrity of the original database and allows you to search for anomalies in an entire population, instead of just a sample of the population."

Director, Audit Firm, on the capabilities of Audit Purpose Built solutions

Core Data Analysis Tasks

Tasks	Description	Spreadsheets	IDEA®
Append/Merge	Combines two files with identical fields into a single file. For example, merge two years' worth of accounts payable history into one file.	✓	✓
Calculated Field/ Functions	Creates expressions or computed fields in order to calculate or recalculate key values. For example, the net pay to an employee could be recalculated using the gross pay field and deducting any withholdings/taxes.	✓	✓
Cross Tabulate	Allows you to analyze Character fields by setting them in rows and columns. By cross-tabulating Character fields, you can produce various summaries, explore areas of interest, and accumulate Numeric fields.	✓	✓
Export	Creates a new file that can be used in another software format. For example, export information to MS Word.	✓	✓
Extract/Filter	Extracts specified items from one file and copies them to another file, normally using an IF statement. For example, extracting all account balances over a predefined limit.	✓	✓
Index/Sort	Sorts a file in ascending or descending order. For example, sorting a file by social security number to see if any blank or "999999999" numbers exist.	✓	✓
Summarize	Accumulates numerical values based on a specified key field. For example, summarizing travel and entertainment expense amounts by employee to identify unusually high payment amounts.	✓	✓
Aging	Produces aged summaries of data based on established cutoff dates.	✗	✓
Benford's Law	Finds abnormal duplications of specific digits and round numbers in corporate data, based on a deviation from the expected frequencies as inferred from Benford's Law.	✗	✓
Duplicates	Identifies duplicate items within a specified field in a file. For example, identify duplicate billings of invoices within the sales file.	✗	✓
Gaps	Identifies gaps within a specified field in a file. For example, identify any gaps in check number sequence.	✗	✓
Join/Relate	Creates a new data file using a common field to combine two separate data files. This task is used to create relational databases on key fields and identify differences between data files.	✗	✓
Sample	Creates random or monetary unit samples from a specified population.	✗	✓
Stratify	Categorizes the data into various strata, or ranges, for a given Numeric field.	✗	✓



CaseWare Analytics
469 King Street West, Suite 200
Toronto, Canada
M5V 1K4

IDEA is a registered trademark of CaseWare International Inc.

1-800-265-4332 Ext: 2803
Fax: 613-842-8129
casewareanalytics.com