Getting Active with an Excel Add-In to Boost Your Investigative Proficiency

David C. Hayes, PhD
James H. Irving, PhD, CPA
*Matthew R. Reidenbach, PhD, CPA (inactive)**

I. Introduction

A recent article by Lee et al. (2018) shows that surveyed accounting practitioners distinctly identify Microsoft Excel as their most frequently used software application, irrespective of their years of accounting experience or accounting specialty area. The results of this survey are reproduced below.

Survey item: How FREQUENTLY do you use this software / tool as part of your current job (or most recent accounting-related position)?

Scale: 1: "Not at all" 2: "Infrequently" 3: "Somewhat Infrequently" 4: "Neither Frequently Nor Infrequently" 5: "Somewhat Frequently" 6: "Frequently" 7: "Very Frequently"

		Years of Accounting Experience[b]			Sole Primary Accounting Area[c]			
Software	All Participants[a]	Less Than 4 Years	4 – 7.5 Years	Greater Than 7.5 Years	Audit	Tax	Advisory	Corporate
	(n=190)	(n=64)	(n=60)	(n=64)	(n=49)	(n=27)	(n=13)	(n=45)
Excel	**6.92[d]**	**7.00[d]**	**6.85[d]**	**6.89[d]**	**7.00[d]**	**6.93[d]**	**7.00[d]**	**6.98[d]**
Adobe Acrobat	**6.34**	**6.20**	**6.38**	**6.44**	**6.41**	**6.85**	**6.39**	**6.00**
PowerPoint	**4.27**	**3.63**	**4.63**	**4.63**	**4.04**	2.93	**5.62**	**4.89**
Acctg. / ERP Software[e]	**3.68**	3.38	**3.33**	**4.25**	3.27	**3.93**	**4.46**	**3.24**
FASB Codification	3.38	**3.52**	3.07	3.47	**5.18**	2.15	2.77	3.11
Tax Research	2.33	2.53	1.95	2.52	1.74	**4.85**	1.23	1.84
Audit Software[f]	2.25	2.63	2.03	2.06	2.98	1.89	2.23	1.93
Access	1.91	1.77	1.88	2.09	1.61	1.44	2.23	1.84
Microsoft Project	1.60	1.45	1.73	1.56	1.61	1.04	2.62	1.62
VBA	1.43	1.53	1.30	1.45	1.39	1.07	2.08	1.47
Tableau	1.36	1.20	1.38	1.50	1.29	1.04	2.15	1.20

Source: Table 3 of Lee et al., 2018

While accounting practitioners often utilize a variety of data analysis tools, Excel is far and away the most prevalent of these tools (Felski, 2023). Based on Felski's survey, accounting practitioners use Excel on a daily basis and expect job candidates to be well-versed in Excel prior to graduation. Yet, in the authors' anecdotal observations, few students have been exposed to the wide-ranging capabilities of Excel prior to enrolling in college. Accordingly, many accounting programs at the college and university level have designed a curriculum that focuses heavily on Excel in order to fulfill the expectations of accounting employers. This article contributes to the accounting literature by showcasing how data analysis and investigative skills (1) commonly taught in an accounting curriculum and (2) commonly utilized to complete routine accounting tasks can be learned and applied using the Active Data for Excel® add-in (hereafter ActiveData).

The remainder of this article is organized as follows. Section II summarizes the extant literature related to data analytics and forensic accounting. Section III describes the ActiveData tool as well as its availability for instructors and students. Section IV demonstrates five key tools offered by ActiveData, using a step-by-step approach. Section V provides additional guidance for instructors and introduces teaching materials valuable for forensic analysis, such as identifying data anomalies and evaluating potentially fraudulent activity. Section VI concludes the article.

II. Background Literature

*James Madison University: Professor of Accoutning; Associate Professor of Accounting; and Assistant Professor of Accounting; respectively.

"Data analytics" is revolutionizing the way that accounting tasks are completed and how accounting practitioners solve business problems. Cost-effective access to data has resulted in the emergence of specialized software tools to help accounting practitioners better identify trends, patterns, and anomalies.

This data analytics revolution has stimulated a large number of accounting journal articles. Two streams of literature have emerged in accounting scholarship: (1) articles examining curriculum innovation and integration in the U.S. (e.g., Dzuranin et al., 2018; Richardson and Shan, 2019; Andiola et al., 2020) as well as abroad (e.g., Rezaee and Wang, 2022), and (2) articles providing educational resources for accounting instructors using data analytics tools including Alteryx (e.g., O'Brien and Stone, 2021), Excel (e.g., Borthick and Schneider, 2023), and PowerBI/Tableau (e.g., Cainas et al., 2021; Samuels and Snow, 2024).

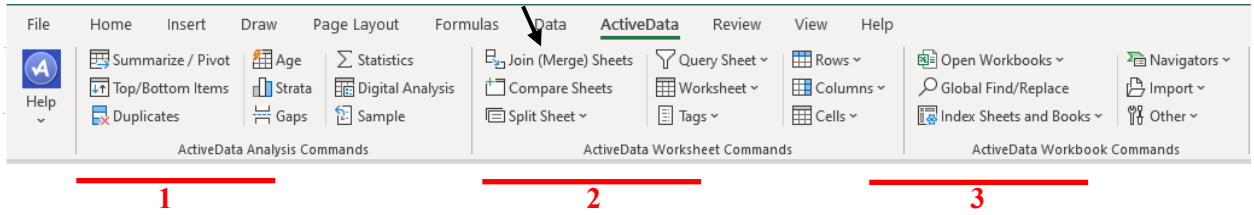
Specifically in accounting education, there has been a shift to include data analytics alongside discussions of relational databases using quantitative statistical analyses as well as simple analyses of textual properties of accounting disclosures (e.g., Appelbaum et al., 2021). Within data analytics, data scientists spend the majority of their time performing the Extract/Transform/Load (ETL) process, yet accounting pedagogy emphasizes both data visualization and interpretation (e.g., Raschke and Charron, 2021).

Forensic accounting has a rich history of analyzing financial information for misrepresentations of past performance. While this is often correlated with fraud examination, it is also much broader and includes topics such as valuation and financial legal disputes (Huber and DiGabriele, 2014). Broadly, this involves evaluating "evidentiary matter" to reach a conclusion on the most likely cause of, or value for, a particular set of information (e.g., Rechtman, 2020). Forensic accountants are particularly valuable when uncovering latent misconduct such as related party schemes (e.g., Ameer and Othman, 2023) and when investigating financial statement irregularities (e.g., Grandstaff and Solsma, 2021; Comunale et al., 2022).

III. What is ActiveData?

Since accounting students and accounting practitioners are very familiar with Microsoft Excel, it is logical to obtain additional proficiency with investigative tools using Excel as opposed to learning an entirely new computer language or purchasing expensive supplementary software. An application that accomplishes this objective is ActiveData for Excel by InformationActive Inc.[1,2] Relative to more sophisticated accounting data analysis software (e.g., ACL, IDEA), ActiveData offers comparable functionality, is reasonably priced, and leverages the powerful Excel environment (Aldhizer, 2017).[3]

Once purchased and installed, the ActiveData menu option appears in the Excel ribbon.



ActiveData has three main menu options (underlined and labeled above for emphasis):

1. Analysis Commands
2. Worksheet Commands
3. Workbook Commands

[1] The InformationActive website is https://www.informationactive.com/ia.cgi?f=home-en. As noted on the website, the founder of InformationActive Inc. was the original developer of the IDEA audit platform. A comparison between the features offered by ActiveData, ACL, and IDEA is available at https://www.informationactive.com/ia.cgi?f=adxl-acl-idea-en.

[2] The authors of this article are not affiliated in any way with InformationActive, Inc. or ActiveData. A former student introduced us to this software.

[3] Instructors who certify that they are employed by a higher education institution will be provided a complimentary license for the Professional Edition of ActiveData to use for educational purposes. Instructors should email support@informationactive.com to obtain this license. Students can receive a free trial of ActiveData for up to 90 days. The free trial is available at https://www.informationactive.com/ia.cgi?f=adxl-download-form-en.

Within these main menu options, there are many investigative tools available. Section IV of this article demonstrates five of these key investigative tools. For each tool, the specific main menu option and ribbon command follow the tool name in parentheses.

- Tool 1: Benford's Law (Analysis Commands – Digital Analysis)
- Tool 2: Query Sheet (Worksheet Commands – Query Sheet)
- Tool 3: Pivot Table (Analysis Commands – Summarize/Pivot)
- Tool 4: Fuzzy Matching (Worksheet Commands – Join [Merge] Sheets)
- Tool 5: Keyword Searching (Worksheet Commands – Query Sheet)

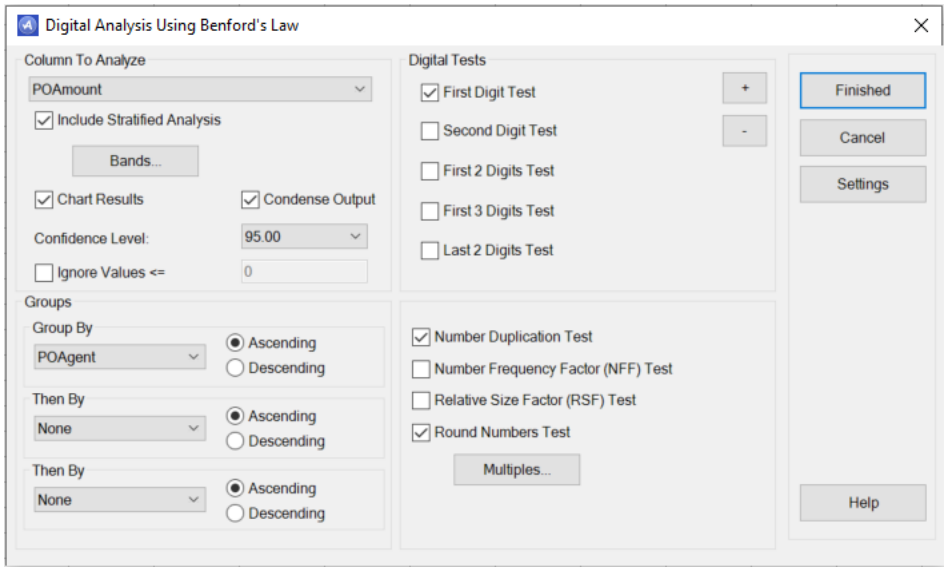
IV. Tool Demonstrations

Tool 1: Benford's Law

Benford's Law is used to search for unusually high levels of digits in a list of numbers. There are many options available with Benford's Law (first digit, first two digits, last digit, etc.), but this article will keep it simple and just demonstrate the first digit search.[4] Assume that we are a wine distributor, and we have five purchasing agents that purchase our inventory. For the current year we have 9,979 purchases and would like to analyze the purchases for any suspicious transactions. Assume that purchasing agents do not collude and would be working alone to purchase items from "preferred vendors." Here is a small sample of the purchases journal.[5]

	A	B	C	D	E	F
1	PO Number	PO Agent #	PO Agent	Region	PO $ Amount	Vendor / Supplier
2	597	24	Johnny Blaze	Northeast	$ 182.02	Bad Brad's Beer
3	598	24	Johnny Blaze	Northeast	$ 122.10	Roof's Premium Wine
4	599	24	Johnny Blaze	Northeast	$ 557.14	Bad Brad's Beer
5	600	18	Bill Wright	Northwest	$ 453.31	Haunted Spirits
6	601	24	Johnny Blaze	Northeast	$ 826.29	Bad Brad's Beer
7	602	24	Johnny Blaze	Northeast	$ 850.53	Quit Your Wining
8	603	19	Bob Jones	Northwest	$ 154.39	Fly By Night Liquor
9	604	22	Mary Smith	Northeast	$ 637.83	Premium Wines
10	605	18	Bill Wright	Northwest	$ 176.49	ABC Liquor
11	606	22	Mary Smith	Northeast	$ 546.40	Discount Liquor

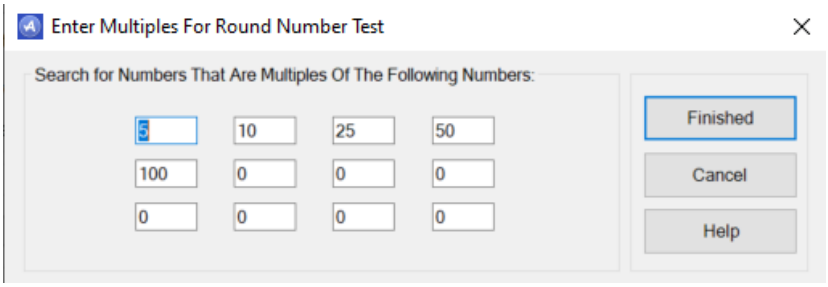
To analyze the data, select the "ActiveData" menu, then select "Digital Analysis" under "ActiveData Analysis Commands." At the following menu, we selected "POAmount" as the "Column to Analyze."



[4] Link to video on Benford's Law: https://www.informationactive.com/iacrm.cgi?v=Benford&x=play&p=ad%2Fvideos&la=&i=adVideos.csv.

[5] The data file is available upon request from the authors.

We then selected "POAgent" to "Group By" and selected "First Digit Test" as the "Digital Tests." We also selected "Number Duplication Test" to see if there might be any duplicate purchases placed. We also selected "Round Numbers Test" and clicked on "Multiples" to select the following round numbers to test for:

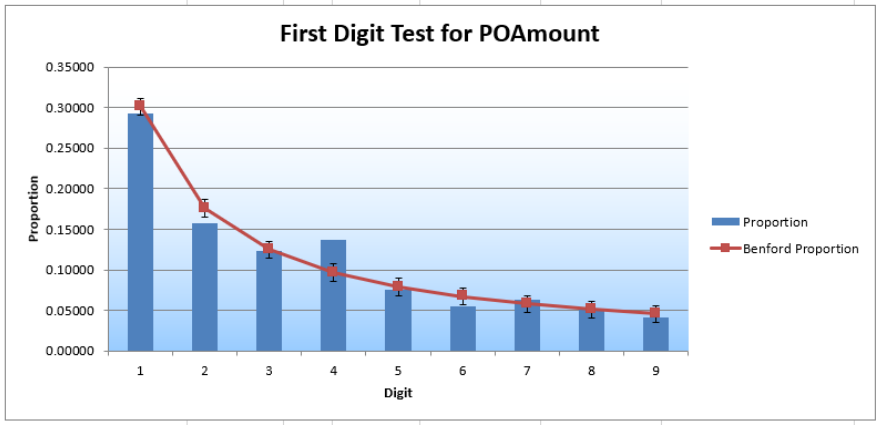


After selecting "Finished" and running the analysis, ActiveData creates a new worksheet with the results. In cell A1 you can see an activity log of information (tests run and data used) by hovering over the cell. This feature can be useful to determine how the worksheet was created and who created it.[6]

7/26/2024 11:50:30 AM
ActiveData Benford's Analysis
From Purchases_Students.xlsx!Purchases
Column POAmount
Group By POAgent /Ascending
Digital Tests
First Digit Test
Number Duplication Test
Number Frequency Factor (NFF) Test
Round Numbers Test
 # Rows: 281
//

Bill Wright, Bob Jones, Mary Smith, and Norm Adams did not show any suspicious Benford's Law activity based on their Z Statistics. Johnny Blaze had the following results:

First Digit Test	POAgent	Count	First Digit	Count	Proportion	Benford Proportion	Difference	Upper Bound	Lower Bound	Z Statistic
	Johnny Blaze	2839	1	832	0.29306	0.30103	0.0079691	0.31808	0.28398	0.90521
			2	446	0.15710	0.17609	0.0189937	0.19028	0.16190	**2.63231**
			3	350	0.12328	0.12494	0.0016559	0.13728	0.11260	0.23846
			4	389	0.13702	0.09691	0.0401101	0.10797	0.08585	**7.19242**
			5	216	0.07608	0.07918	0.0030981	0.08929	0.06907	0.57659
			6	158	0.05565	0.06695	0.0112934	0.07632	0.05758	**2.37008**
			7	180	0.06340	0.05799	0.0054107	0.06677	0.04922	1.19330
			8	150	0.05284	0.05115	0.0016830	0.05943	0.04287	0.36444
			9	118	0.04156	0.04576	0.0041936	0.05362	0.03789	1.02441
			Total	2,839		**Mean Absolute Difference**	**0.0104897**			



[6] The first row of the audit trail data shows who created the worksheet. We have cropped this row from the screen shot to preserve author/reviewer anonymity.

Of the 9,979 purchases, Johnny accounted for 2,839 of them. The Benford's First Digit test shows that the first digits of two, four, and six are statistically suspicious based on a Z statistic of 1.96 or higher (p-value of .05). However, the two and six digits are below the expected number of occurrences ("Benford Proportion" line in graph above), and they are dropped from investigation because you cannot look for something that is not there in the primary source data (i.e., understatement error). So, the suspicious digit in this instance involves purchase amounts where the first digit is a four. By clicking on the four, ActiveData does a drill-down listing of all 389 items. Unfortunately, the drill-down view is not easy to manipulate but does give a quick peek at the suspicious results. Since we now have an idea of what to look for, we will use ActiveData's "Query Sheet" function to further analyze the data ("Query Sheet" is featured below as Tool 2). The following two items are the results from the additional items requested (round numbers and number duplication tests) when running the Benford's Law.

DrillDown

ActiveData Drill Down: DigitalAnalysis
From *'Purchases'
POAmount FirstDigit=4 Group By POAgent=Johnny Blaze

	PO Number	PO Agent #	PO Agent	Region	PO $ Amount	Vendor / Supplier
1	668	24	Johnny Blaze	Northeast	476.42	Quit Your Wining
2	673	24	Johnny Blaze	Northeast	471.28	Quit Your Wining
3	682	24	Johnny Blaze	Northeast	464.01	Bad Brad's Beer
4	699	24	Johnny Blaze	Northeast	486.80	Quit Your Wining
5	817	24	Johnny Blaze	Northeast	464.87	Quit Your Wining
6	818	24	Johnny Blaze	Northeast	458.07	Bad Brad's Beer
7	822	24	Johnny Blaze	Northeast	451.08	Roof's Premium Wine
8	823	24	Johnny Blaze	Northeast	472.76	Quit Your Wining
9	837	24	Johnny Blaze	Northeast	443.11	Quit Your Wining
10	846	24	Johnny Blaze	Northeast	431.38	Roof's Premium Wine
11	869	24	Johnny Blaze	Northeast	445.47	Quit Your Wining
12	891	24	Johnny Blaze	Northeast	445.73	Quit Your Wining
13	909	24	Johnny Blaze	Northeast	485.89	Quit Your Wining
14	920	24	Johnny Blaze	Northeast	480.37	Bad Brad's Beer
15	1004	24	Johnny Blaze	Northeast	473.65	Bad Brad's Beer
16	1021	24	Johnny Blaze	Northeast	443.98	Quit Your Wining
17	1044	24	Johnny Blaze	Northeast	491.10	Roof's Premium Wine
18	1073	24	Johnny Blaze	Northeast	499.70	Quit Your Wining
19	1097	24	Johnny Blaze	Northeast	426.27	Roof's Premium Wine
20	1110	24	Johnny Blaze	Northeast	434.06	Quit Your Wining
21	1119	24	Johnny Blaze	Northeast	475.49	Quit Your Wining
22	1158	24	Johnny Blaze	Northeast	464.15	Quit Your Wining
23	1189	24	Johnny Blaze	Northeast	477.37	Quit Your Wining

R:1 C:1 389 rows

The "Round Numbers Test" results reflected Bob Jones had 10 out of his 2,655 purchases, Johnny Blaze had six out of his 2,839 purchases, Norm Adams had three out of his 1,201 purchases, Bill Wright had three out of his 1,473 purchases, and Mary Smith had two out of her 1,811 purchases. These results suggest that as a percentage of purchases that Bob Jones 10 round numbers should be investigated. By clicking on the "10" we see the following round numbers for Bob Jones; of which nine of them were to "Fly By Night Liquor". Note that each of the purchasing agents only have a few suppliers they buy from, so this may not be too surprising, but should be further investigated.

DrillDown

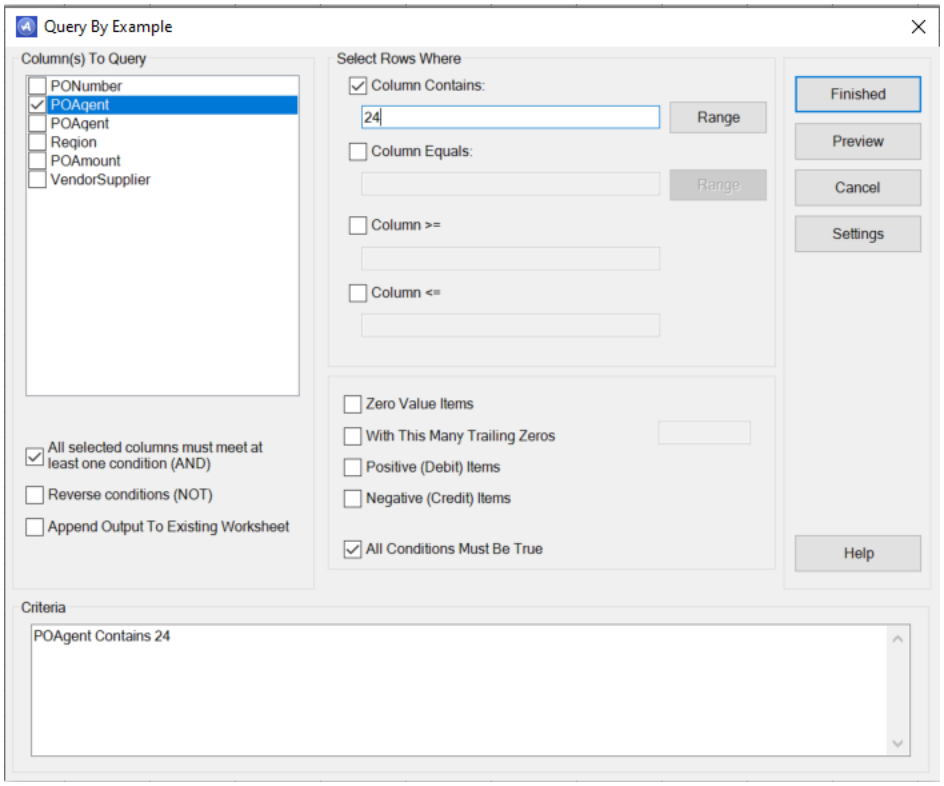
ActiveData Drill Down: DigitalAnalysisRounded
From *'Purchases'
POAmount Rounded To 5, 10, 25, 50, 100 Group By POAgent=Bob Jones

	PO Number	PO Agent #	PO Agent	Region	PO $ Amount	Vendor / Supplier
1	1698	19	Bob Jones	Northwest	180	Cheers to You
2	1842	19	Bob Jones	Northwest	195	Fly By Night Liquor
3	1844	19	Bob Jones	Northwest	315	Fly By Night Liquor
4	2908	19	Bob Jones	Northwest	125	Fly By Night Liquor
5	5902	19	Bob Jones	Northwest	685	Fly By Night Liquor
6	6711	19	Bob Jones	Northwest	295	Fly By Night Liquor
7	6964	19	Bob Jones	Northwest	195	Fly By Night Liquor
8	7594	19	Bob Jones	Northwest	100	Fly By Night Liquor
9	7952	19	Bob Jones	Northwest	725	Fly By Night Liquor
10	8803	19	Bob Jones	Northwest	575	Fly By Night Liquor

The "Number Duplication Test" only showed two or three items duplicated for each of the purchasing agents and did not suggest any further investigation.

Tool 2: Query Sheet

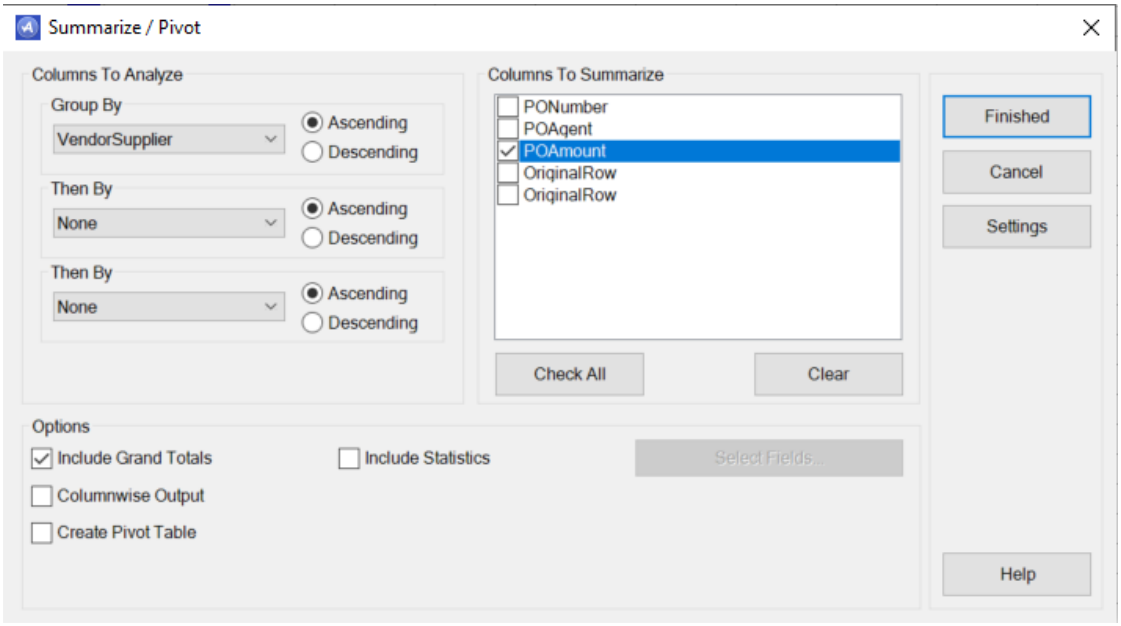
Query Sheet can be used to search and filter the data for suspicious activities including atypical journal entry amounts (e.g., large dollar transactions; credits to expense accounts). The results from the Benford's test above suggest that Johnny Blaze had excess purchases that began with the digit four. The first thing to do is to filter the data down to purchases for only Johnny Blaze by selecting "Query Sheet" and "Query By Example" and selecting the "POAgent" in the "Column(s) to Query" and to select the "Column Contains:" and enter Johnny's PO Agent number of "24" and then "Finished".



The newly created worksheet contains only the purchases for Johnny Blaze. We will filter that down by running another "Query By Example" on the results from the last query and this time selecting the "Column >=" and entering "400" and selecting the "Column <=" and entering "499.99" to get all the purchases in the $400 to $499.99 range. After selecting "Finished," the results created in the new worksheet reflect the 389 purchases identified by the Benford's Law analysis from Tool 1. Next, we will use the Pivot Table tool to compare these results to Johnny's total purchases to see if there is an abnormally large number of purchases from a specific supplier.

Tool 3: Pivot Table

Utilizing the results from the "Query Sheet," select "Summarize/Pivot" under "ActiveData Analysis Commands" and then select "VendorSupplier" in the "Group By" option and finally check the "POAmount" in the "Columns to Summarize" option as shown below.



The newly created worksheet shows that Johnny has three main suppliers that he purchases from and that approximately 56 percent of his purchases that begin with the first digit of four were from "Quit Your Wining".[7]

	A	B	C	D	E
1	**Vendor / Supplier**	**Count**	**PO $ Amount.Total**	**% of Count**	**% of PO $ Amount**
2	Bad Brad's Beer	103	$ 46,733.04	26.5%	26.4%
3	Quit Your Wining	219	$ 99,707.69	56.3%	56.4%
4	Roof's Premium Wine	67	$ 30,436.84	17.2%	17.2%
5	Totals	**389**	$ 176,877.57	100.0%	100.0%

If this is aligned with the company's normal business practice, the proportion of vendor orders with a four starting dollar value should be consistent with the full population of Johnny's purchases. To compare that, we are going to use the worksheet tab that filtered the purchases data to just Johnny's 2,839 purchases (with all the digits) and run another "Summarize/Pivot".

	A	B	C	D	E
1	**Vendor / Supplier**	**Count**	**PO $ Amount.Total**	**% of Count**	**% of PO $ Amount**
2	Bad Brad's Beer	1137	$ 454,631.02	40.0%	39.8%
3	Quit Your Wining	1002	$ 413,640.28	35.3%	36.2%
4	Roof's Premium Wine	700	$ 274,500.26	24.7%	24.0%
5	Totals	**2839**	$ 1,142,771.56	100.0%	100.0%

The results from all of Johnny's purchases appear to be very different from his purchases where the PO amount start with the digit four. The supplier "Quit Your Wining" is 56 percent of the purchases with the first digit four compared to only 35 percent of all the purchases. It might be wise to inquire with the owner if they have a PO approval policy of $500; this would suggest Johnny does not want to get approval for purchases above $500 and is just below the limit with Quit Your Wining.

Tool 4: Fuzzy Matching

Fuzzy Matching can be used to search for similar characters between two lists of data. For example, the following data reflects possible street address names/suffixes and abbreviations:

[7] Column D and E "% of Count" and "% of PO $ Amount" were manually added to the results to calculate the % of transactions for each Supplier.

	A	B	C	D
1		List 1	List 2	Difference
2		Avenue	Ave	3
3		Avenue	Ave.	3
4		Drive	Dr	3
5		Road	Rd	2
6		Street	St	4
7		Ave	Drive	3

The Difference column refers to how closely the item in List 1 is to List 2 based on the difference in the count of characters. For example, in rows two and three, Avenue is within three characters of Ave and Ave. (i.e., it is missing "nue"). In row seven, Ave is also within three characters of Drive. When fuzzy matching is utilized, it takes the first item in List 1 and compares it to all the items in List 2 and then takes the second item in List 1 and compares it to all the items in List 2 until all the items are compared for both lists.[8] The items are then sorted based on how close of a match (difference) they are. The lists are color-coded so that you can easily see where these items appear in the results on page 11. The data in the screen captured above was split into two worksheets as follows to demonstrate fuzzy matching:

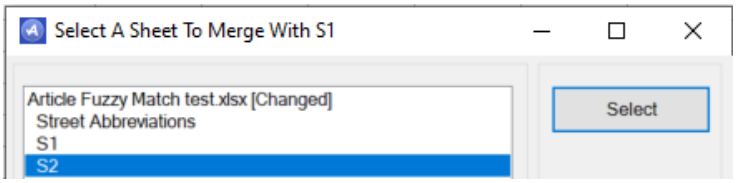
Worksheet S1

	A	B
1	List 1	Difference
2	Avenue	3
3	Avenue	3
4	Drive	3
5	Road	2
6	Street	4
7	Ave	3

Worksheet S2

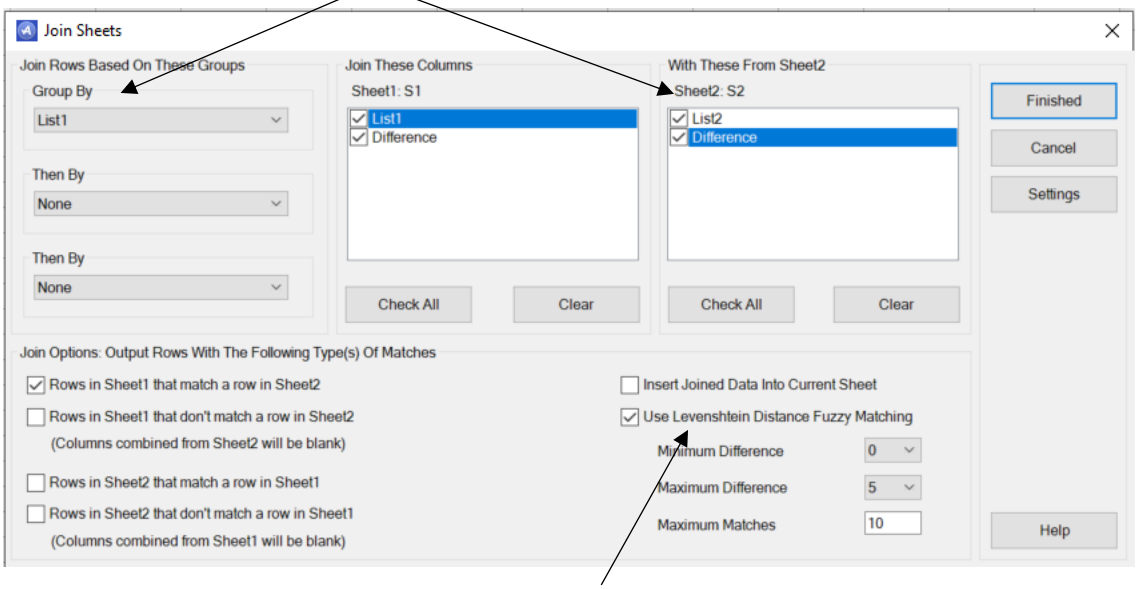
	A	B
1	List 2	Difference
2	Ave	3
3	Ave.	3
4	Dr	3
5	Rd	2
6	St	4
7	Drive	3

While in worksheet S1, select "Join (merge) Sheets" from the worksheet commands. Next, select the worksheet S2 as the sheet to merge with S1.[9]

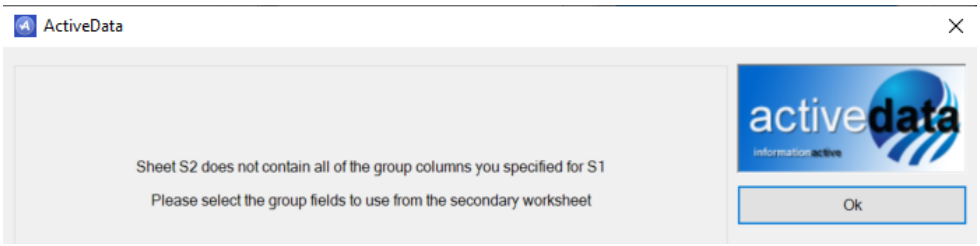


In the "Join Sheets" menu option, select "List1" as the "Group By" and then select both the items in the "With These From Sheet 2" field.

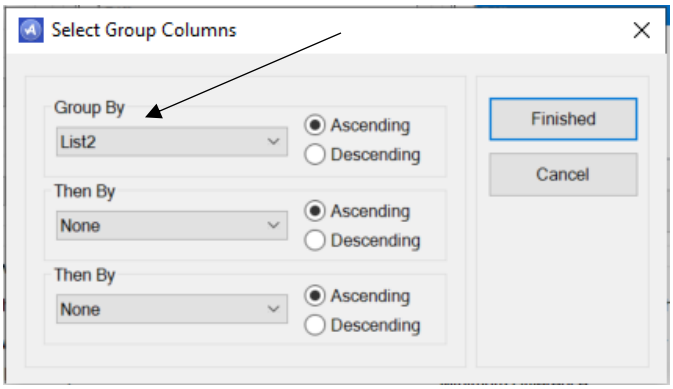
[8] The lists cannot be in the same worksheet. They must be in separate worksheets (or even in separate files). Also, you need at least two columns of data; one that you are matching on and a second column of additional data.
[9] The data file is available upon request from the authors.



To run our fuzzy matching test, we check the "Use Levenshtein Distance Fuzzy Matching," select "0" (exact match) as the "Minimum Distance", enter "5" as the "Maximum Difference", "10" as the "Maximum Matches" and "Finished". Since the two lists do not have an identically named row header, at the following popup, click on "OK" so we can tell it what to match on.



At the "Select Group Columns" menu, we need to select "List2" as the "Group By" option and click on "Finished".



The newly created worksheet results are automatically sorted by column C (Difference) with the closest matches listed first. For the color-coded items (rows 5, 6, 7, 12, 16, and 23) the difference in column C matches our predicted difference in column D. So, if we want to search for similar addresses, we should consider using a distance of at least four or alternatively perform additional data cleaning procedures to standardize these address suffixes.[10]

[10] Row 23 reflects that "Street" and "St" have a Difference of four. One method to perform this alternative data cleaning is through the USPS abbreviations: https://pe.usps.com/text/pub28/28apc_002.htm

	A	B	C	D	E	F
1	List1	S2.List2	Difference	Difference	S2.Difference	Tags
2	Drive	Drive	0	3	3	4
3	Ave	Ave	0	3	3	7
4	Ave	Ave.	0	3	3	7
5	Road	Rd	2	2	2	5
6	Avenue	Ave	3	3	3	2
7	Avenue	Ave.	3	3	3	2
8	Avenue	Ave	3	3	3	3
9	Avenue	Ave.	3	3	3	3
10	Drive	Ave	3	3	3	4
11	Drive	Ave.	3	3	3	4
12	Drive	Dr	3	3	3	4
13	Ave	Dr	3	3	3	7
14	Ave	Rd	3	3	2	7
15	Ave	St	3	3	4	7
16	Ave	Drive	3	3	3	7
17	Drive	Rd	4	3	2	4
18	Road	Ave	4	2	3	5
19	Road	Ave.	4	2	3	5
20	Road	Dr	4	2	3	5
21	Road	St	4	2	4	5
22	Road	Drive	4	2	3	5
23	Street	St	4	4	4	6
24	Avenue	Drive	5	3	3	2
25	Avenue	Drive	5	3	3	3
26	Drive	St	5	3	4	4
27	Street	Ave	5	4	3	6
28	Street	Ave.	5	4	3	6
29	Street	Dr	5	4	3	6
30	Street	Rd	5	4	2	6
31	Street	Drive	5	4	3	6

Next, we will use fuzzy matching to see if any of our employees have set up an account as a supplier/vendor and may be selling to our company at inflated prices. Our employee data has 40 rows of employee data and we use the letter "E" at the beginning of the headers to reflect that it is employee information.[11]

	A	B	C	D	E	F	G
1	E-ID	E-First	E-Last	E-Street	E-City	E-State	E-Zip
2	1	Empl First 1	Empl Last 1	111 Canyon Street	Monroe	NC	28110

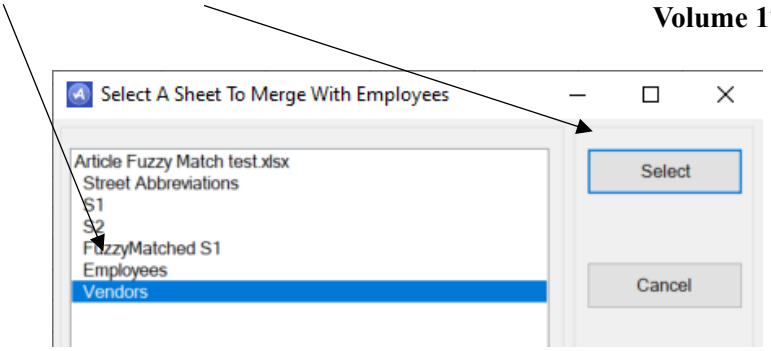
Our vendor/supplier data has 60 rows of data and we use the letter "V" at the beginning of the headers to reflect that it is vendor/supplier information.[12]

	A	B	C	D	E	F	G
1	V-ID	V-First	V-Last	V-Street	V-City	V-State	V-Zip
2	1	Vend First 1	Vend Last 1	150 Gardenia Street	Danville	VA	24540

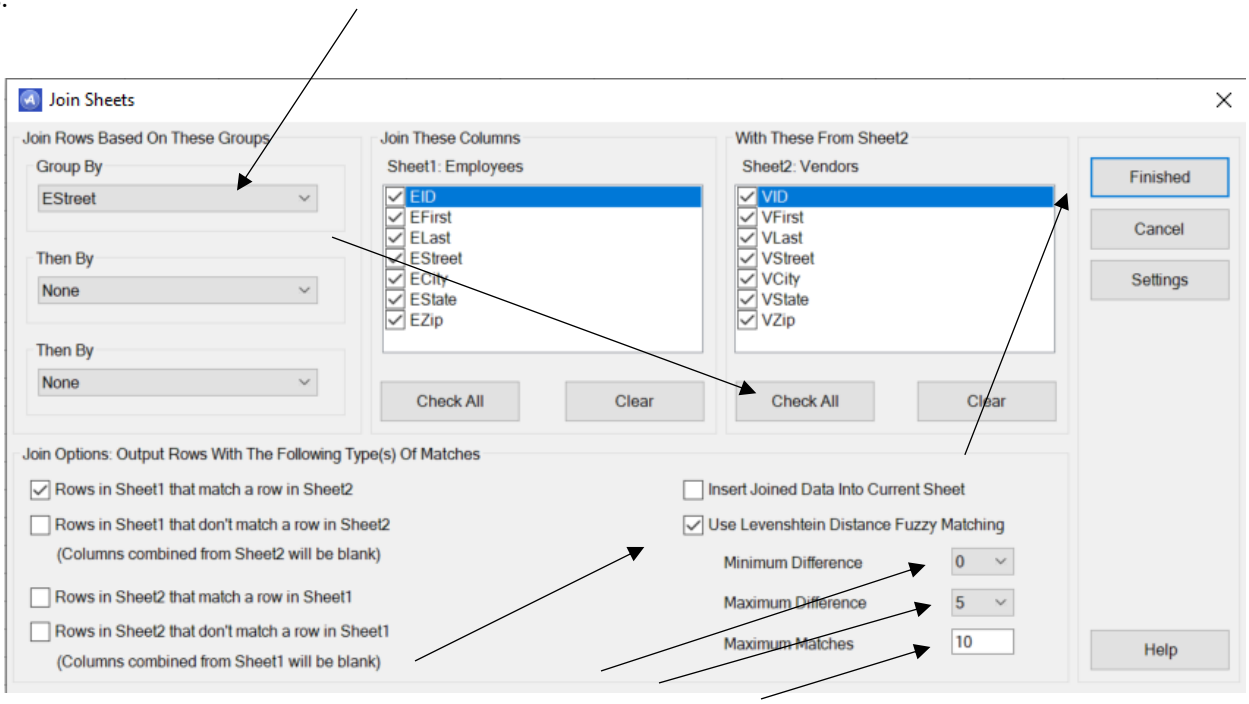
Select the "Employees" tab and in the "Worksheet Commands" select "Join (Merge) Sheets," then select "Vendors" and "Select."

[11] The data file is available upon request from the authors.

[12] If the data that you want to compare is not similar in format, in the "Worksheet Command" you can use the "Columns" and either "Split Columns" or "Combine Columns" to get the data similar to analyze. For example, if the employee address was split into separate columns as displayed, but if the vendor data reflected the street address, city, state, zip in one column, then you may want to split the vendor data into separate columns (street, city, state, zip).



On the "Join Sheets" screen, select "EStreet" as the "Group By," then click on "Check All" to include all the fields for the Vendors.



Select the "Use Levenshtein …" and "0", "5", and "10" options and "Finished". Click "OK" at the popup that informs us that the Vendors does not have the same header as the employees (EStreet vs. VStreet). In the "Select Group Columns" menu select "VStreet" in the "Group By" option and then "Finished."

The newly created worksheet shows that 14 of the employees' street addresses were similar to the 60 vendor/supplier street addresses.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	EStreet	Vendors.VStreet	Difference	EID	EFirst	ELast	ECity	EState	EZip	Vendors.VID	Vendors.VFirst	Vendors.VLast	Vendors.VCity	Vendors.VState	Vendors.VZip	Tags
2	383 Sandstone Street	383 Sandstone Street	0	17	Empl First 17	Empl Last 17	Rockingham	NC	27601	18	Vend First 18	Vend Last 18	Rockingham	NC	27601	18
3	388 Main St	788 Main St	1	18	Empl First 18	Empl Last 18	Huntersville	NC	28078	46	Vend First 46	Vend Last 46	Rockingham	NC	27601	19
4	344 Basswood Street	544 Westwood Street	4	14	Empl First 14	Empl Last 14	Danville	VA	24540	29	Vend First 29	Vend Last 29	Winston-Salem	NC	27101	15
5	348 Bay Street	248 Ebony Street	4	15	Empl First 15	Empl Last 15	Charleston	WV	25301	7	Vend First 7	Vend Last 7	Alexandria	VA	22302	16
6	612 Ashwood Street	619 Yewwood Street	4	26	Empl First 26	Empl Last 26	Newport News	VA	23601	35	Vend First 35	Vend Last 35	Portsmouth	VA	23701	27
7	788 Main Street	788 Main St	4	33	Empl First 33	Empl Last 33	Rockingham	NC	27601	46	Vend First 46	Vend Last 46	Rockingham	NC	27601	34
8	461 Boxwood Street	619 Yewwood Street	5	21	Empl First 21	Empl Last 21	Oxford	NC	27565	35	Vend First 35	Vend Last 35	Portsmouth	VA	23701	22
9	461 Boxwood Street	788 Waxwood Street	5	21	Empl First 21	Empl Last 21	Oxford	NC	27565	47	Vend First 47	Vend Last 47	Clarksburg	WV	26301	22
10	612 Ashwood Street	204 Wood Street	5	26	Empl First 26	Empl Last 26	Newport News	VA	23601	4	Vend First 4	Vend Last 4	Chesapeake	VA	23320	27
11	628 Brook Street	204 Wood Street	5	27	Empl First 27	Empl Last 27	Concord	NC	28025	4	Vend First 4	Vend Last 4	Chesapeake	VA	23320	28
12	738 Bell Street	7 Yew Street	5	32	Empl First 32	Empl Last 32	Winston-Salem	NC	27101	41	Vend First 41	Vend Last 41	Newport News	VA	23601	33
13	788 Main Street	833 Fig Street	5	33	Empl First 33	Empl Last 33	Rockingham	NC	27601	50	Vend First 50	Vend Last 50	Portsmouth	VA	23701	34
14	921 Cedarwood Lane	939 Redwood Lane	5	38	Empl First 38	Empl Last 38	Lumberton	NC	28358	56	Vend First 56	Vend Last 56	Lumberton	NC	28358	39
15	980 Acorn Street	918 Hawthorn Street	5	40	Empl First 40	Empl Last 40	Alexandria	VA	22301	54	Vend First 54	Vend Last 54	Charleston	WV	25301	41

In the above results on row two, there is an exact street address match (0 difference) for Employee #17 and Vendor #18. Many cities have similar street names (Main Street for example) and so we should also compare Column G (Employee City) with Column M (Vendor City) and they match for row two. Similarly, row seven is suspicious in that the only difference is

"Street" vs. "St" – the difference of four. Note that row three is within one for the street address, but the cities are different.[13] The final results suggest that two of the employees should be investigated as to why their addresses are also in the Suppliers list.

Tool 5: Keyword Searching

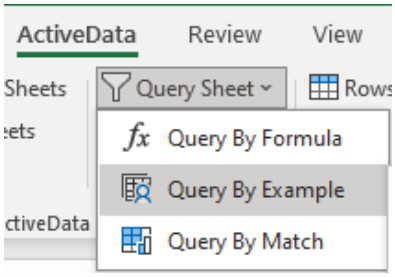
Keyword Searching is valuable when searching for potentially suspicious journal entries. In Excel, "Ctrl" and "F" can be used to search for individual items. However, ActiveData contains a more powerful keyword search that allows you to search for numerous keywords all at once. As part of the investigation, we will search journal entry descriptions based on the perceived level of fraud risk we assigned to the following list of words:

Level of Risk			
Very Low	**Low**	**Moderate**	**High**
Cash	Adj	Entertainment	Bribe
Discount	Adjust	Gift	Cover Up
Commission	Adjustment	Gifts	Fraud
	Misc	Override	Illegal
	Miscellaneous	Problem	Kickback
	Error	Write Off	Theft
	Correct	Writeoff	Off the books
	Fix	Spread	Gray Area
		Reserve	
		Plug	
		Terminated	
		Fired	

Our Journal Entry data has 211 journal entries (530 rows) in the following format:[14]

	A	B	C	D	E	F	G	H	I	J	K
1	**Trans#**	**Account #**	**Account**	**Trans-Date**	**Journal**	**Journal**	**Amount**	**Dr/Cr**	**Emp #**	**Recorded by**	**Entry Description**
2	13,398	104	Inventories	1/2/2024	1	Purchases Journal	$ 7,881.13	Dr	1	Purchasing Agent	Buy inventory on Acct
3	13,398	200	Accounts Payable	1/2/2024	1	Purchases Journal	$ 7,881.13	Cr	1	Purchasing Agent	Buy inventory on Acct

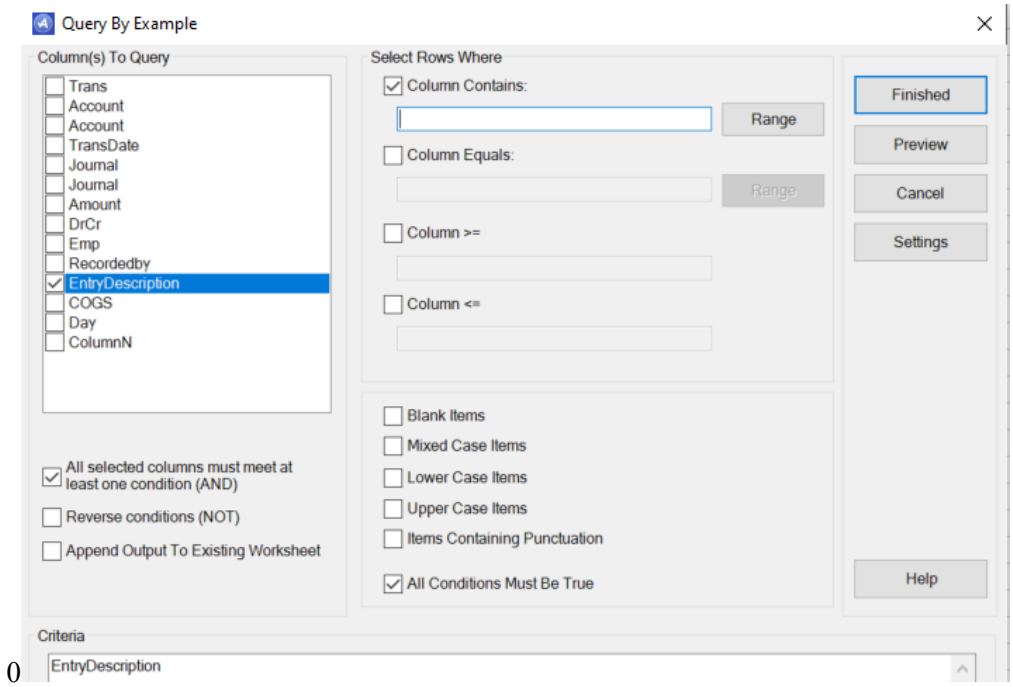
To search the "Entry Description" for the list of key words, select ActiveData menu, then "Query Sheet" and "Query By Example."



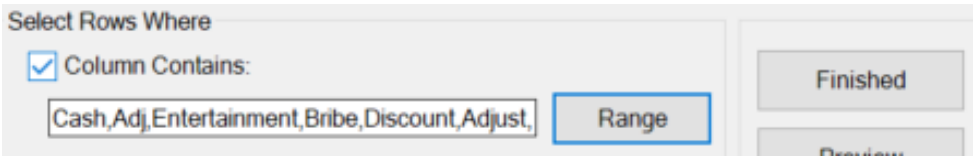
At the "Query By Example" menu select "EntryDescription" in the "Column(s) To Query" and the "Column Contains" in the "Select Rows Where" and then select "Range."

[13] These results could have been narrowed down by selecting "EStreet" in the "Group By" and then combining that with "ECity" in the "Then By". Also, we would have to select "VStreet" in the "Group By" and then combining that with "VCity" in the "Then By" for the selection of what to match on after the popup menu warning. The results of this fuzzy matching would only result in rows 2, 7, and 14 from the prior fuzzy matching results reflecting the two suspicious addresses and one false positive (row 14).

[14] The data file is available upon request from the authors.



When the "Select Range" menu is displayed, highlight all the words in the list and "OK."[15] The "Column Contains:" now shows some of the words selected from the list.



The "Criteria" at the bottom of the "Query By Example" menu shows all of the items selected.

The results from this keyword search resulted in about 10 percent of the journal entries being flagged and most due to the word "Cash" in entries where payment was received on account.[16] The keyword search was performed again with the revised list:

The results show that there were 10 journal entries (20 rows of data) that should be investigated.

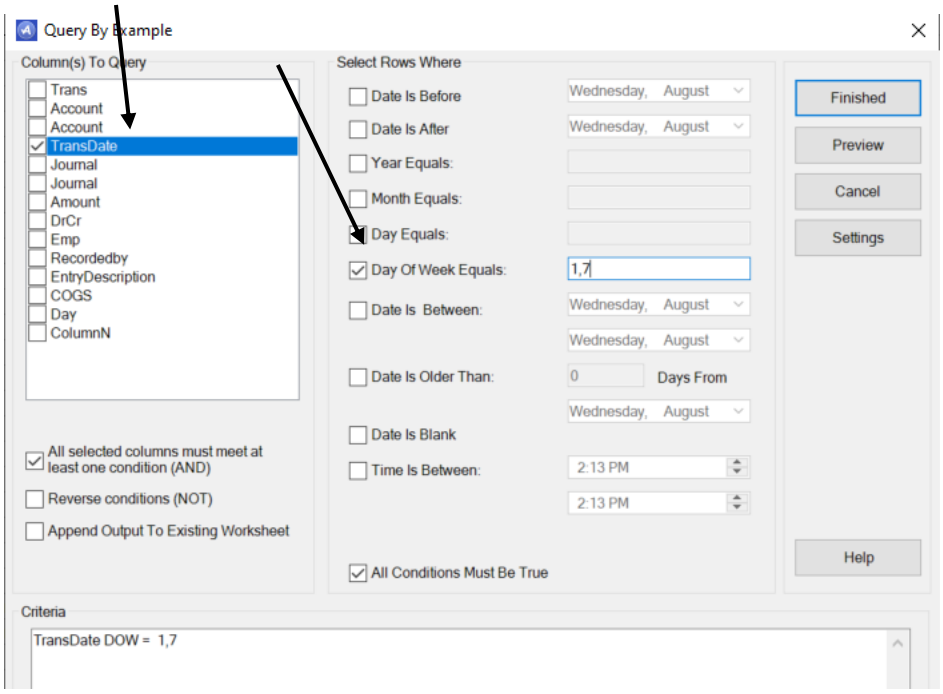
[15] The list of key words can be in a different tab or even a different file.
[16] The list was revised to exclude "Cash". Additionally, several similar words were eliminated. For example, "Adj" was kept, but "Adjust" and "Adjustment" were excluded because both of the words contain "Adj".

Trans#	Account #	Account	Trans-Date	Journal	Journal	Amount	Dr/Cr	Emp #	Recorded by	Entry Description
13413	506	Operating Expense	1/5/2024	5	General Journal	577.62	Dr	5	Controller	Adjust cash to bank rec
13413	100	Cash & Equivalents	1/5/2024	5	General Journal	577.62	Cr	5	Controller	Adjust cash to bank rec
13423	506	Operating Expense	1/11/2024	2	Cash Disbursements	257.06	Dr	4	Cash Disbursements	Purch Misc Supplies
13423	100	Cash & Equivalents	1/11/2024	2	Cash Disbursements	257.06	Cr	4	Cash Disbursements	Purch Misc Supplies
13428	200	Accounts Payable	1/12/2024	2	Cash Disbursements	1425.78	Dr	4	Cash Disbursements	Paid for EOY Christmas Gifts on acct
13428	100	Cash & Equivalents	1/12/2024	2	Cash Disbursements	1425.78	Cr	4	Cash Disbursements	Paid for EOY Christmas Gifts on acct
13433	204	Accrued Expenses	1/15/2024	1	Purchases Journal	1678.99	Dr	7	Payroll Clerk	Paid Commission Bonus
13433	100	Cash & Equivalents	1/15/2024	1	Purchases Journal	1678.99	Cr	7	Payroll Clerk	Paid Commission Bonus
13472	506	Operating Expense	1/31/2024	5	General Journal	1387.07	Dr	5	Controller	Adj PPDs for Jan
13472	106	Prepaid Expenses	1/31/2024	5	General Journal	1387.07	Cr	5	Controller	Adj PPDs for Jan
13496	506	Operating Expense	2/8/2024	5	General Journal	667.96	Dr	5	Controller	Adjust cash to bank rec
13496	100	Cash & Equivalents	2/8/2024	5	General Journal	667.96	Cr	5	Controller	Adjust cash to bank rec
13536	506	Operating Expense	2/27/2024	5	General Journal	1387.07	Dr	5	Controller	Adj PPDs for Feb
13536	106	Prepaid Expenses	2/27/2024	5	General Journal	1387.07	Cr	5	Controller	Adj PPDs for Feb
13547	510	Non-Operating Income/Expense	3/1/2024	3	Sales Journal	33884.62	Dr	5	Controller	Reclass COGS due to Kickbacks to Purch Agent fired.
13547	500	Cost of Goods Sold	3/1/2024	3	Sales Journal	33884.62	Cr	5	Controller	Reclass COGS due to Kickbacks to Purch Agent fired.
13560	506	Operating Expense	3/7/2024	5	General Journal	1765.76	Dr	5	Controller	Adjust cash to bank rec
13560	100	Cash & Equivalents	3/7/2024	5	General Journal	1765.76	Cr	5	Controller	Adjust cash to bank rec
13607	506	Operating Expense	3/29/2024	5	General Journal	1387.07	Dr	5	Controller	Adj for March PPDs
13607	106	Prepaid Expenses	3/29/2024	5	General Journal	1387.07	Cr	5	Controller	Adj for March PPDs

Another useful investigative feature is searching for transactions that were created on a specific day of the week. ActiveData utilizes the following numbers when searching for days of the week:

Day of Week #	Day of Week
1	Sunday
2	Monday
3	Tuesday
4	Wednesday
5	Thursday
6	Friday
7	Saturday

For example, to search for transactions that were recorded on a weekend (Saturday = 7 or Sunday = 1), select "Query Sheet" and "Query By Example". In the "Query By Example" menu, select "TransDate" in the "Column(s) to Query" and "Day of Week Equals:."



Type the numbers "1,7" in the input box to the right of "Day of Week Equals:" and select "Finished". The results reflect that one transaction by the Purchasing Agent was made on a weekend day (2/10 was a Saturday) and should be investigated.

Trans#	Account #	Account	Trans-Date	Journal	Journal	Amount	Dr/Cr	Emp #	Recorded by	Entry Description
13500	104	Inventories	2/10/2024	1	Purchases Journal	21839.38	Dr	1	Purchasing Agent	Bought Inv on Account
13500	200	Accounts Payable	2/10/2024	1	Purchases Journal	21839.38	Cr	1	Purchasing Agent	Bought Inv on Account

ActiveData can also be used for numerous other elements of journal entry testing such as searching for journal entries by unapproved users, round dollar value transactions, etc.

V. Additional Guidance and Teaching Materials for Instructors

Several accounting courses, at both the undergraduate and graduate levels, are suitable for introducing students to ActiveData. These courses include accounting information systems, auditing, data analytics, and forensic accounting. The authors introduced their undergraduate accounting information systems students to ActiveData during the Fall 2023 semester. The knowledge, skills, and abilities a student acquires by working with ActiveData are practical given the increased attention and emphasis on data analytics by the AACSB (in Accounting Standard A5) as well as the AICPA (in its Model Curriculum and CPA Exam blueprints).

The authors allocated one class session to provide students with initial training with respect to ActiveData functionality.[17] Students were also taught how to match datasets via key variables using both exact matching and fuzzy matching. Students then completed an in-class assignment using ActiveData to assess their initial proficiency in identifying data anomalies and performing basic forensic accounting tasks. The five tasks they were asked to complete included: (1) eliminating blank spaces and "ghost characters", (2) identifying numbers stored as text fields, (3) determining if employees left or joined a company, (4) detecting duplicate employees, and (5) using fuzzy matching to uncover if employees are also customers. This in-class assignment is included as Appendix 1.

Following the initial training and in-class assignment, students were given a homework assignment using ActiveData. Students were asked to complete more advanced tasks in the homework assignment than in the in-class assignment. Data was initially created using the GADGET data generator (Holt and Lang, 2021) for a full set of revenue data: customer list, employee list, inventory cost list, sales order list, sales invoice list, and a cash receipts list. Errors were seeded into the data including likely related-party relationships (e.g., customer-supplier), non-standard pricing, and duplicated employees. Forensic accounting tasks embedded into the homework assignment included: (1) identifying potential concerns with the data input process ("ETL", or Extract, Transform, and Load), and (2) listing at least two potential unusual activities that may be taking place in the business. Focusing specifically on an advanced forensic accounting task in the homework assignment, most students were able to identify at least one of the embedded potential related-party transactions via fuzzy matching. Students had to match individuals based on last name, street address, and zip code. This suggests that even novice users are capable of learning how to perform basic forensic accounting tasks using ActiveData with relative ease. This homework assignment is included herein as Appendix 2.[18]

The authors noted several observations and best practices as they experimented with ActiveData and introduced it to their students. First, ActiveData requires significant computing power when analyzing datasets larger than 20,000 records. While ActiveData still functions all the way up to the native Excel maximum worksheet size, users may experience slower processing times with larger datasets. Second, ActiveData is best described as a "jack of all trades" tool that performs numerous tasks efficiently and effectively. While some functions can be replicated using free Excel add-ins (e.g., Microsoft Fuzzy Lookup) or more expensive data analysis software (e.g., PowerBI), we found that there is great value in an inexpensive tool with numerous capabilities for forensic analysis. Third, ActiveData has well-developed internal logging capabilities for forensic analyses. Both the "audit trail" feature and the time-stamped notes provides a verifiable record of tasks performed within a dataset. Finally, while ActiveData provides online video tutorials, they are short, simplistic demonstrations. These tutorials are satisfactory for getting started and introduced to a tool; however, developing mastery of the tool may require some additional practice.

VI. Conclusion

Accounting students are universally trained to be proficient with Microsoft Excel. ActiveData is conveniently embedded within Excel, and it provides an opportunity for students to increase their aptitude with identifying data anomalies and evaluating potentially fraudulent activity. The tools provided in ActiveData are a significant enhancement to the native Excel application, offering comparable functionality to competing accounting data analysis software (e.g., ACL, IDEA).

[17] Students were also directed to the short tutorial videos for many of the ActiveData tools, available at https://www.informationactive.com/iacrm.cgi?x=play&v=&p=ad/videos&i=adVideos.csv

[18] For the homework assignment (Appendix 2), as well as the in-class assignment from the previous paragraph (Appendix 1), the data files, suggested solutions, and grading rubrics are available upon request from the authors.

In this article, we showcase five of the key tools available in this Excel add-in (Benford's Law, Query Sheet, Pivot Table, Fuzzy Matching, and Keyword Searching). We then provide additional guidance and teaching materials for instructors who would like their students to learn and master this tool. ActiveData is available at no charge for instructors and students who use it for educational purposes.

References

Aldhizer III, G. 2017. Visual and Text Analytics: The Next Step in Forensic Accounting and Auditing. *The CPA Journal* 87(6): 30–33.

Ameer, R., and Othman, R. 2023. Workaround Opportunities in Concealing Revenue and Expense Misstatements. *Journal of Forensic and Investigative Accounting* 15(3): 437–462.

Andiola, L. M., Masters, E., and Norman, C. 2020. Integrating Technology and Data Analytic Skills into the Accounting Curriculum: Accounting Department Leaders' Experiences and Insights. *Journal of Accounting Education* 50: 1–18.

Appelbaum, D., Showalter, D. S., Sun. T., and Vasarhelyi, M. 2021. A Framework for Auditor Data Literacy: A Normative Position. *Accounting Horizons* 35(2): 5–25.

Borthick, A. F., and Schneider, G. P. 2023. Getting Students Ready for Accounting Spreadsheets: Training for Basic Spreadsheet Skills with Pre/Post Assessments. *Issues in Accounting Education* 38(2): 63–84.

Cainas, J. M, Tietz, W. M., and Miller-Nobles, T. 2021. KAT Insurance: Data Analytics Cases for Introductory Accounting Using Excel, Power BI, and/or Tableau. *Journal of Emerging Technologies in Accounting* 18(1): 77–85.

Comunale, C. L., Hayes, D. C., and Irving, J. H. 2022. Keeping an Investigative Eye on the Financial Pulse of a Company Using Data Analytics. *Journal of Forensic and Investigative Accounting* 14(3): 514–528.

Dzuranin, A. C., Jones, J. R., and Olvera, R. M. 2018. Infusing Data Analytics into the Accounting Curriculum: A Framework and Insights from Faculty. *Journal of Accounting Education* 43: 24–39.

Felski, E. 2023. Audit Technologies Used in Practice and Ways to Implement These Technologies into Audit Courses. *Journal of Accounting Education* 62: 1–15.

Grandstaff, J. L., and Solsma, L. L. 2021. Financial Statement Fraud: A Review from the Era Surrounding the Financial Crisis. *Journal of Forensic and Investigative Accounting* 13(3): 421–437.

Holt, M., and Lang, B. 2021. GADGET: An Accounting Data Generator. *Journal of Emerging Technologies in Accounting* 18(1): 113–129.

Huber, W. D., and DiGabriele, J. A. 2014. Research in Forensic Accounting – What Matters? *Journal of Theoretical Accounting Research* 10(1): 41–71.

Lee, L., Kerler, W., and Ivancevich, D. 2018. Beyond Excel: Software Tools and the Accounting Curriculum. *AIS Educator Journal* 13(1): 44–61.

O'Brien, A., and Stone, D. N. 2021. A Case Study in Managing the Analytics "Iceberg": Data Cleaning and Management using Alteryx. *Journal of Emerging Technologies in Accounting* 18(2): 221–245.

Raschke, R. and Charron, K. 2021. Review of Data Analytic Teaching Cases, Have We Covered Enough? *Journal of Emerging Technologies in Accounting* 18(2): 247–255.

Rechtman, Y. 2020. The Past, Present, and Future of Forensic Accounting. *The CPA Journal* 90(1): 10–12.

Rezaee, Z., and Wang, J. 2022. Integration of Big Data into Forensic Accounting Education and Practice: A Survey of Academics in China and the United States. *Journal of Forensic and Investigative Accounting* 14(1): 133–150.

Richardson, V. J., and Shan, Y. 2019. Data Analytics in the Accounting Curriculum. Calderon, T.G. (Ed.) *Advances in Accounting Education: Teaching and Curriculum Innovations* Emerald Publishing Limited, Leeds, 23: 67–79.

Samuels, J. A., and Snow, M. C. 2024. Analyzing Depreciation: A Financial Accounting Data Analytics and Visualization Case. *Issues in Accounting Education* 39(3): 165–174.

Appendix 1: Example In-Class Assignment Using ActiveData

Use the "HIR-Data.xlsx" Excel file to answer Questions 1–5 below. Note that there are three worksheet tabs in this Excel file (Customers 2023, Employees 2023, and Employees 2022).

	Question	ActiveData menu option and tool used to answer the question
1.	Do you observe any blank spaces in the Customers 2023 worksheet tab? If so, list all cell references where you discover these blank spaces.	**Worksheet Commands menu option** **Cells – Convert Selected Cells – Replace Multiple Spaces With Single Space** **(can also remove most blank spaces with Cells – Convert Selected Cells – Remove Leading/Trailing Spaces)**
2.	Do you observe any numeric fields stored as text fields in the Customers 2023 worksheet tab? If so, list all cell references where you discover a number stored as text.	**Worksheet Commands menu option** **Cells – Convert Selected Cells – Cells to Numbers**
3.	For this question, use the Employees 2023 and Employees 2022 worksheet tabs. Did any employees join the company or leave the company during 2023? If so, list separately the employees that joined and left the company.	**Worksheet Commands menu option** **Compare Sheets**
4.	Do you observe any possible duplicate employees in the Employees 2023 worksheet tab? If so, list all employees who appear more than once.	**Analysis Commands menu option** **Duplicates**
5.	For this question, use the Employees 2023 and Customer 2023 worksheet tabs. Do you observe any employees who may also be a customer? If so, list all employees who may also be a customer.	**Worksheet Commands menu option** **Join (Merge) Sheets – Use Levenshtein Distance Fuzzy Matching**

Appendix 2: Example Homework Assignment Using ActiveData

Instructions: This assignment covers basic database concepts and reporting processes using case materials modified from Holt and Lang (2021) and data generated by GADGET.

You must complete this assignment using only Microsoft Excel and the ActiveData for Excel add-in. Once you complete the assignment, upload the following files to your Learning Management System:

1. Data file including all your work (i.e., Excel files showing ActiveData log activity notes).
2. Word file with your written responses.

Background: Your role as the new staff accountant is to analyze the company's profitability by store and provide recommendations to the CFO on improving the business. Based in New York City, this fictitious beverage distributor aims to become the largest wine wholesaler in the U.S. This company has loyal employees, uses QuickBooks for its accounting recordkeeping, and maintains three corporate bank accounts for optimizing its cash management. To maximize profitability, salespeople are salaried but must follow standard pricing arrangements when selling merchandise to customers.

Part 1: Extract, Transform, and Load Data (15 points total)

1. Ensure that the "Audit Log" feature in ActiveData is turned on. Load each data file into ActiveData. For each file, identify the following: (1) any difficulties loading the file, (2) primary key, (3) any foreign key(s), and (4) identify any potential database anomalies and/or data coding issues that could exist in the files (e.g., sequence gaps in primary key fields; duplicates in key fields; keys and field types).
2. By hand or electronically, create a visual representation of the accounting information system for the above files showing the interconnection between data files using primary and foreign keys.

Part 2: 2023 Financial Performance Analyses (15 points each – 45 points total)

For each question, list the data files used and what steps you performed to answer the question.

3. Use the sales order detail file to calculate the total sales and total profits by brand in 2023. *Identify the brands with the highest and lowest amounts for each metric (sales and profits).*
4. Use the invoice summary file (and any other files you deem necessary) to calculate the amounts collected from each customer for 2023 sales. *What amount of discounts were provided to each customer?*
5. Review the invoice and cash receipts records for the company's three primary bank accounts and create an aging schedule by customer of unpaid invoices as of February 29, 2024. *Are there any particular customers that have a problematic payment history?*

Part 3: Investigation of Unusual Activities (20 points)

6. The CFO has tasked you with identifying whether you believe any fraudulent activity is occurring in the business. List at least two potential unusual activities that may be taking place in the business based upon evaluating employees, customers, and their relationship to your analyses in Part 2.
    - *Fuzzy matching may help you complete this task.*

Part 4: Evaluation of Business Activities (20 points)

7. Write a formal one-page business memorandum summarizing your findings from Parts 1–3 and covering the following topics:
    - *Are there any changes that should be made to the database structure in the AIS?*
    - *Are there any customers whom you believe we should discontinue business activity?*
    - *Are there any employees who you believe are taking advantage of the company? If so, how should the company respond?*