Blazing a Trail for the Benford's Law of Words

December 15, 2016

RICHARD B. LANZA, CFE, CPA, CGMA

WWW.RICHLANZA.COM

WWW.AUDITSOFTWAREVIDEOS.COM

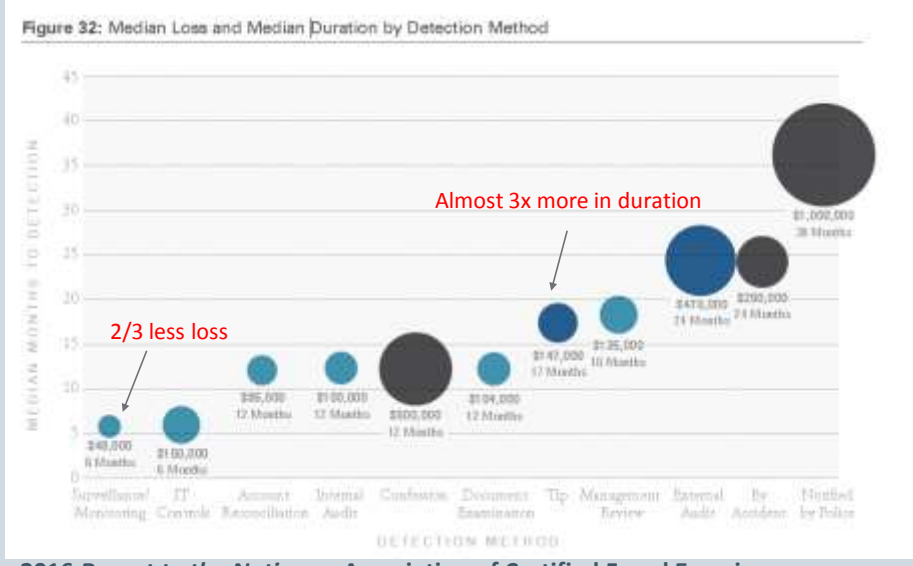
Richard B. Lanza, CFE, CGMA

- Assists clients daily in the use of analytic software
- Nearly 25 years of ACL and excel software usage
- Has written and spoken on the use of audit data analytics for over two decades
- Received the outstanding achievement in business award by the Association of Certified Fraud Examiners for developing the publication *Proactively Detecting Fraud Using Computer Audit Reports* as a research project for the IIA

- Recently was a contributing author of:
  - Global Technology Audit Guide (GTAG #13) Fraud In An Automated World – Institute Of Internal Auditors.
  - Data Analytics – A Practical Approach - research whitepaper for the information system accountability control association.
  - Cost Recovery – Turning Your Accounts Payable Department Into A Profit Center – Wiley And Sons.
- In 2015, discovered a new textual analytic technique using letters called the Lanza Approach to Letter Analytics (LALA)™

Please see full bio at **www.richlanza.com**

Surveillance is the quickest; lowers fraud impacts



2016 *Report to the Nations* – Association of Certified Fraud Examiners

3

PredPol http://www.predpol.com/

Predictive Modeling To Improve Police Detection

Santa Cruz experienced:

- 27% decrease in burglary
- 11% decrease in robbery
- 56% increase in arrests

http://bit.ly/1VyQPQY

"PredPol **does not replace the experience and intuition of our great officers**, but is rather an invaluable added tool that allows our police force to use their patrol time more efficiently and helps stop crime before it happens." ***Chief Mark Yokoyama***

4

http://bit.ly/1gP3meq

EY Global Forensic Data Analytics Survey 2014

450 executives surveyed

- 72% of respondents believe that emerging big data technologies can play a key role in fraud
- Only 7% of respondents are aware of any specific big data / Only 2% are using them
- 12% utilize visualization / 26% apply key word searches
- 62% of respondents indicate that they need to improve management's awareness of the benefits of analytics

The Population of Data Type

Structured Data	Unstructured Data	External Data
▪ Accounting records	▪ Documents (Excel, PDF, Word)	▪ Geomap Service
▪ Sub ledger details	▪ Emails	▪ OFAC, SAM.Gov Watch Lists
▪ Monthly performance measures	▪ Network Logs	▪ IRS Tax ID Match

90% of Data is Text Based When Did You Last Investigate Text?

It works fast to quickly gain a perspective of the business process data:

- Can work in real-time with the data while talking to the client – no prep needed…meaningful questions in seconds
- Look for deviations over a 3-year moving average to the current period

If digital analysis/Benford's Law is latitude, letter analytics is longitude

- Text is far richer in business value and providing a picture than simple digit theory
- The unique word chart provides a more normalized view of activity while total word occurrences by letter provides a more dynamic chart
- The trends can be seen quickly to ask relevant questions and to highlight fraud

Why not use another approach, such as Benford's Law, to look at ALL data?

7

Red Flag (Key) Word Phrases

Red Flag Word Phrases/Words

- One could build a database of suspicious words and then search the entire data file for these words, looping back to the table to get the next word:
  - bribe
  - fraud
  - plug
  - etc.
- Summaries can be done by person and collectively for additional collusion reviews

Lessons from WorldCom/ MCI

The fraud was accomplished primarily in two ways:
1.Booking **"line costs" (**interconnection expenses with other telecommunication companies) as capital expenditures on the balance sheet instead of expenses.
2.Inflating revenues with bogus accounting entries from **"corporate unallocated revenue accounts".**

In 2002, a small team of internal auditors at WorldCom worked together, often at night and secretly, to investigate and reveal $3.8 billion worth of fraud….

Per Wikipedia – MCI Inc.

THE WALL STREET JOURNAL.

PAGE ONE
FROM THE ARCHIVES: October 30, 2002

How Three Unlikely Sleuths Exposed Fraud at WorldCom

By SUSAN PULLIAM and DEBORAH SOLOMON
Staff Reporters of THE WALL STREET JOURNAL

CLINTON, Miss. -- Sitting in his cubicle at WorldCom Inc. headquarters one afternoon in May, Gene Morse stared at an accounting entry for $500 million in computer expenses. He couldn't find any invoices or documentation to back up the stunning number.

"Oh my God," he muttered to himself. The auditor immediately took his discovery to his boss, Cynthia Cooper, the company's vice president of internal audit. "Keep going," Mr. Morse says she told him.

A series of obscure tips last spring had led Ms. Cooper and Mr. Morse to suspect that their employer was cooking its books. Armed with accounting skills and determination, Ms. Cooper and her team set off on their

Key Words/Phrase Survey Summary Results – Upd. 2015

Unique Responses
- 4,320 response phrases / 2,153 unique phrases
- Average of 17 phrases per response

Phrase Occurrences

Unique phrases	1,424 (66%)
2 to 4 occurrences	574 (27%)
5 to 19 occurrences	144 (7%)
20 and Over	11 (1%)
	2,153

Phrase Letter Length
- Average of 10, Max of 75 and Min of 2 letters

Per AuditNet® Key Words Survey

http://bit.ly/1XyMwch

11

Survey Phrase Summary Results Top Words - 2014

GIFT(S)	52
BRIBE	34
FRAUD	28
CORRECT	26
WRITE OFF	25
ERROR	24
FACILITATION	21
COMMISSION	20
ENTERTAINMENT	20

GREY AREA	20
ILLEGAL	19
MISCELLANEOUS	35
PLUG	18
WRITEOFF	18
CASH	17
REVERSE	17
ADJUST	16
COVER UP	16

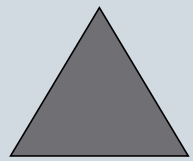
KICKBACK	16
RESERVE	16
ADJUSTMENT	15
DISCOUNT	15
OFF THE BOOKS	15
PROBLEM	15
OTHER	14
OVERRIDE	14
SPREAD	14

Page 12

Fraud Triangle Phrase Examples

Pressure
FROM PETER TO PAY PAUL
MANAGE EARNING
TICKING TIME BOMB
WALL STREET EXPECTATIONS

Opportunity
CHANNEL STUFFING
COOKIE JAR RESERVES
FRAUDULENT SHAM
QUID PRO QUO
SECRET ACCOUNT



Rationalization
DON'T KNOW WHAT ELSE TO DO
PIECE OF THE ACTION
TREAT ME THIS WAY
WORRY ABOUT IT LATER
THEYLL BE SORRY

Per AuditNet® Survey
http://bit.ly/1XyMwch

2015 Updates to Key Words

AuditNet® LLC (Jim Kaplan) added:

- More Key Words
- Spam Words
- SEC and Terrorist Words
- Social Media Terms

Stephen Valance enhanced:

- The classification of the 2014 survey key words

Coney B.V., Amsterdam (Joris Joppe & Pieter de Kok)

- Translated the 2014 survey to Dutch

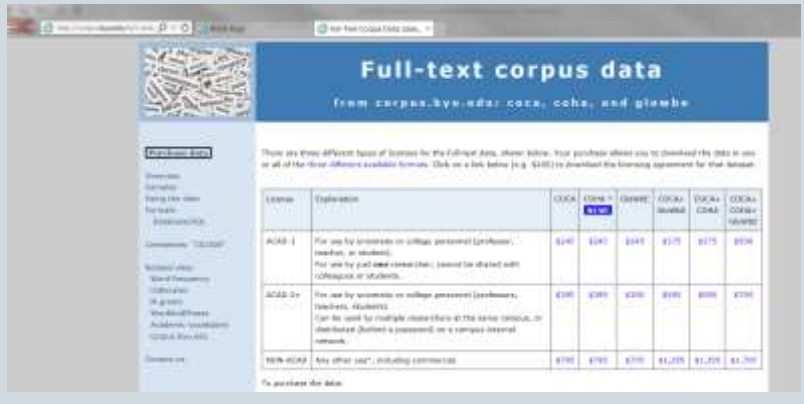
Rich Lanza

- Finalized the list and organized all data for use

Moving Beyond "Bad" Words to Other Word Lists

THE CORPUS OF CONTEMPORARY AMERICAN ENGLISH (COCA)

450 MILLION WORDS, 1990-**2012**



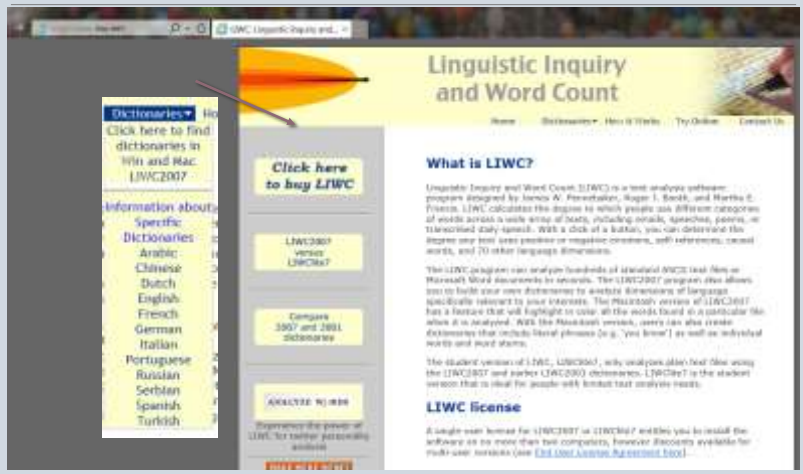
Brown Corpus – Top 5,000 Makes up 73% of Used English

http://www.wordfrequency.info

Word Analysis Just Copy and Paste

http://www.wordandphrase.info/analyzeText.asp

Secret Life of Pronouns
LIWC



Secret Life of Pronouns
LIWC – Dictionary Page Example

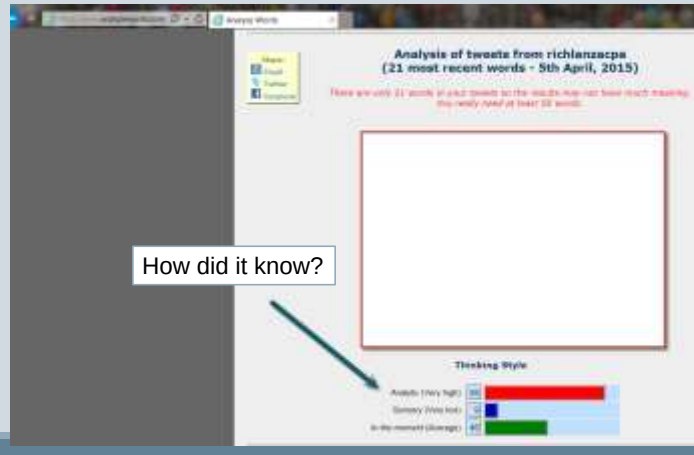
(www.liwc.net)

31 Posemo			32 Negemo			
hugs	promising	wealthy	):	dismay*	ignorant	p
humor*	proud	welcom*	:(	disreput*	ignore	p
humour*	prouder	well	abandon*	diss	ignored	p
hurra*	proudest	wellbeing	abuse*	dissatisf*	ignores	p
ideal*	proudly	wellness	abusi*	distraught	ignoring	p
ily*	radian*	win	ache*	distress*	immoral*	p
importance	readiness	winn*	aching*	distrust*	impatien*	p
important	ready	wins	advers*	disturb*	impersonal	p
importantly	reassur*	wisdom	afraid	domina*	impolite*	p

Secret Life of Pronouns

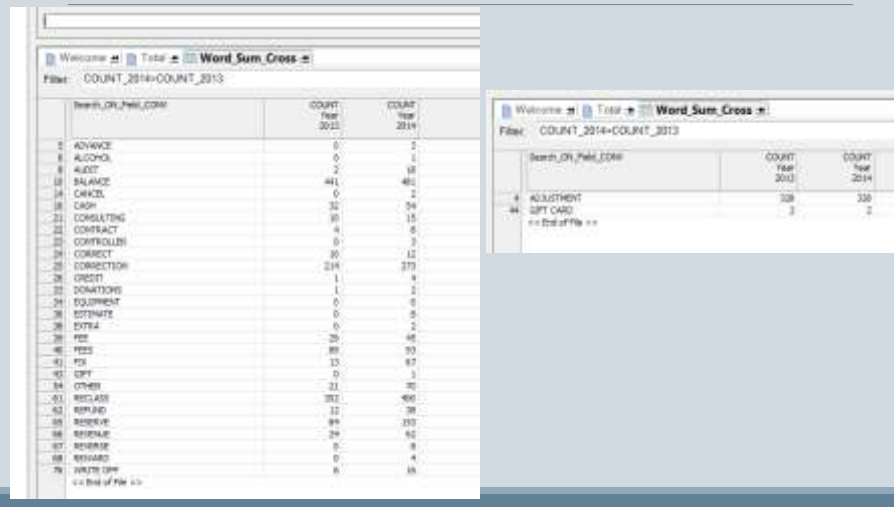
Twitter Analysis

AnalyzeWords.com – Twitter Analysis



Word Summarization and Review

Words On The Rise / Words Equal



Summarize Words Analysis

What You Need

- Table With Description Fields

How You Do It

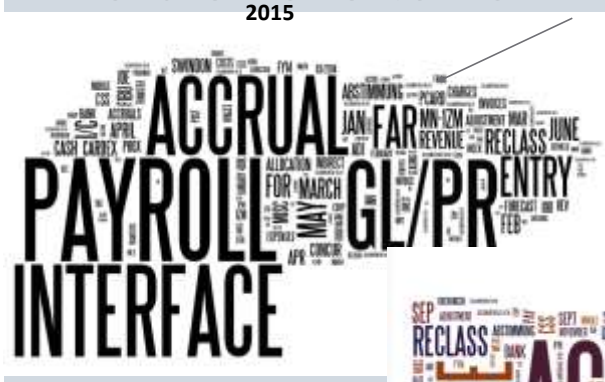
- SPLIT the field to be searched into words
  - 10 to 20 words should work well
- Extract every split field to a new table
- SUMMARIZE on the field to understand usage
  - Sum by month/year as well to trend
  - Sum by enterer

Creating the Wordle www.wordle.net

Wordle www.wordle.net

Wordle 2015 to 2014

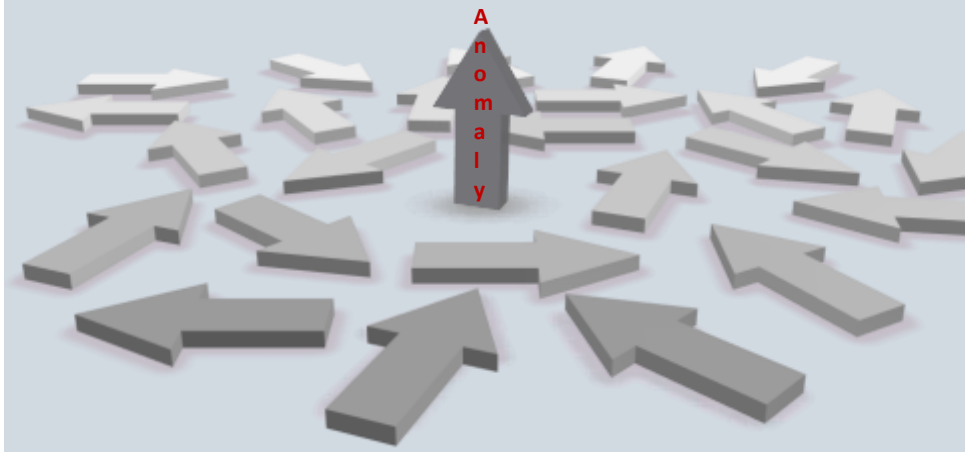
2015



2014

Getting to the Specific Word Deviations Fast!!!!

But, Isn't It About Finding the Deviations?



29

Can You Read This?

It deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae.

30

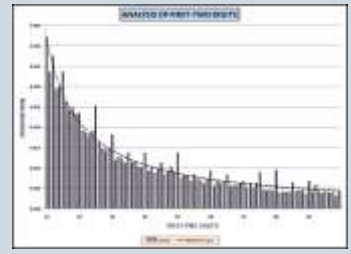
Benford's Law – The Basis of Digital Analysis

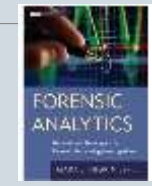
TABLE 5.1 The Expected Digit Frequencies of Benford's Law

	Position in Number			
Digit	1st	2nd	3rd	4th
0		.11968	.10178	.10018
1	.30103	.11389	.10138	.10014
2	.17609	.10882	.10097	.10010
3	.12494	.10433	.10057	.10006
4	.09691	.10031	.10018	.10002
5	.07918	.09668	.09979	.09998
6	.06695	.09337	.09940	.09994
7	.05799	.09035	.09902	.09990
8	.05115	.08757	.09864	.09986
9	.04576	.08500	.09827	.09982

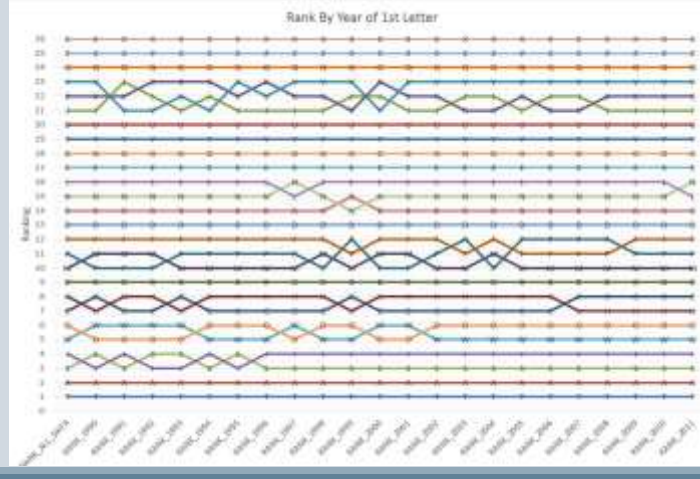
Source: Nigrini, M. J., 1996, "A taxpayer compliance application of Benford's Law," The Journal of the American Taxation Association 18 (Spring): 72-91.

The table shows that the expected proportion of numbers with a first digit 2 is 0.17609 and the expected proportion of numbers with a fourth digit 4 is 0.10002.

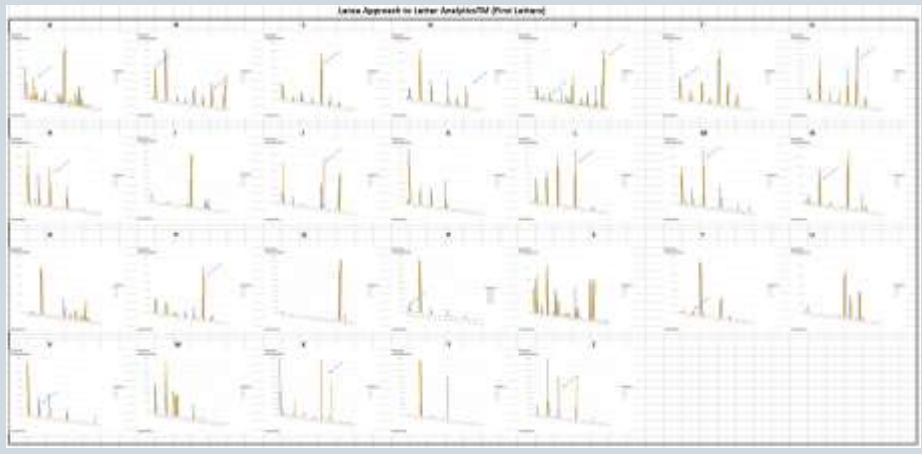




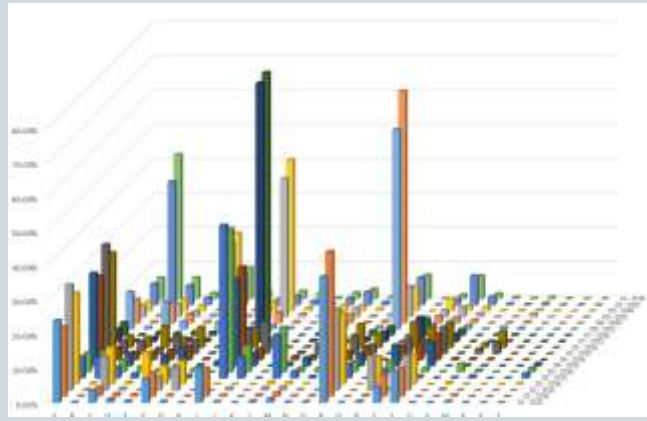
COCA Ranking – In First Letters The Benford's Law of Words?



A Benford's Law For Letters and Words?



General Ledger Fingerprint



50% From 1st and Last Letters

	A	B	(A) x (B)
# of Letters in a Word	Word Occurrences per COCA	% of Letters Analyzed in First and Last Letters	Final %
1	3.51%	100%	3.51%
2	16.02%	100%	16.02%
3	20.71%	67%	13.88%
4	17.26%	50%	8.63%
5	11.29%	40%	4.52%
6	8.53%	33%	2.81%
7	7.74%	22%	1.70%
8	5.40%	25%	1.35%
	90.46%		52.42%

35

A Benford's Law For Words The Dashboard

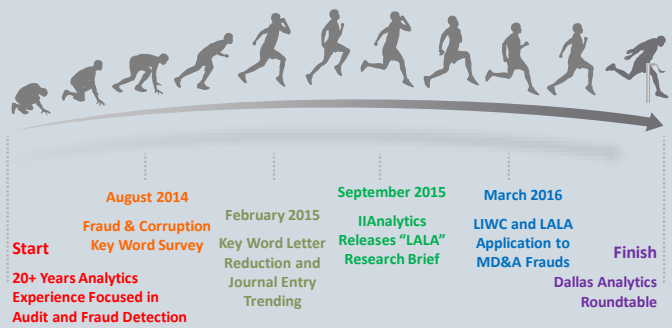
Lanza Approach to Letter Anaytics™ ("LALA")

Identifies word deviations swiftly by relating letter frequency patterns to benchmarks of the English language and prior period letter occurrences. Focus is placed on:

- ✓ First letter (26 letters)
- ✓ Last letter (26 letters)
- ✓ First two letters (702 letters)
- ✓ Last two letters (702 letters)

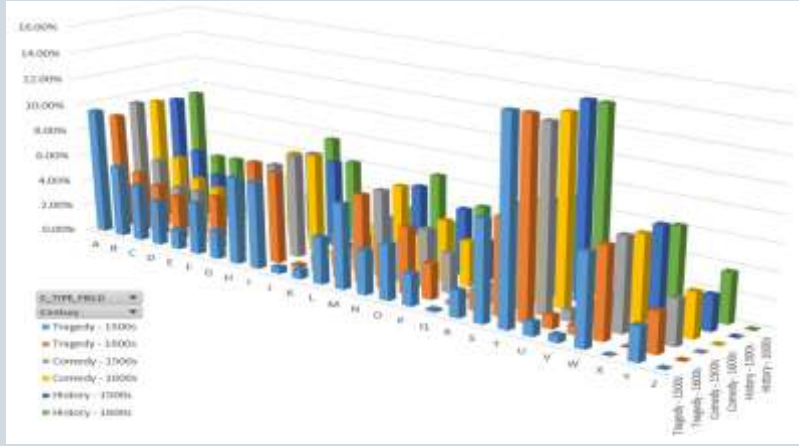
37

Letter Analytics Lifecycle 2014 to Present

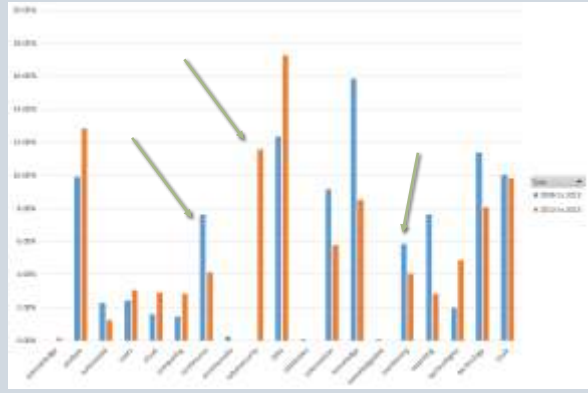


British Pop Songs - Britburn

Shakespeare's Plays By Century



Protiviti Reports on Internal Auditing 2006 to 2012 vs. 2013 to 2015 – Key Words



http://www.protiviti.com/IAsurvey

LALA - Where Can It Be Used

Use Casev	Specific Analytics
• Gain insight to a business process and its deviations	• Learn new facts about the process through organizing and trending description fields from purchase and sales orders. • Trend business process documentation over time for specific items (i.e. travel policy) and holistically across all documents • Identify new areas of risk from customer feedback blogs, questionnaires, Emails and social media postings • Analyze test result documentation in GRC description fields • Pinpoint the common threads between safety and manufacturing shutdown reports
• Assess journal entry risks and financial accounting trends	• Trend word usage in journal entry names, and line descriptions to better visualize the monthly activity • Map ledger-focused unusual key words to identify entries worthy of discussion • Determine the letter fingerprint of the monthly journal entry titles, and their rate of change throughout the year

42

LALA - Where Can It Be Used

Use Case	Specific Analytics
• Profile employees for corruption and collusion	• Determine network links between employees by trending the words in employee Emails • Assess hourly payroll time descriptions to gain a new perspective of what everyone is working on through their words • Perform key word searching of travel expense business descriptions
• Pinpoint computer application issues and concerns	• Trend employee web page access pages, searches terms and documented posts to social media through company networks • Summarize file directory and file names by department • Review error log tables over time to identify new error patterns or areas of increased exposure

43

Useful Links on LALA

- **http://bit.ly/1jFD87b** - Blog announcing the discovery of letter analytics.
- **http://bit.ly/1RZpolz** - Research Paper #1 – Focused on explaining the letter analytic concept with reference to a benchmark for the English Language and an analysis of British song titles from 1960 to 1999.
- **http://bit.ly/1QebYkL** - Research Paper #2 – Provides a more in-depth analysis of the population of text data and how letters can explain text variations over time more quickly than word summaries. Three examples are provided including Shakespeare's plays, Berkshire Hathaway shareholder reports and my personal Emails.
- **http://bit.ly/1W0CAZO** - Predictive Analytics Times article on how Word clouds analysis could improved with letter analytic visualizations
- ✓ **http://bit.ly/1TGwvPS** and **http://bit.ly/21mEbsU** - ACFE Fraud Magazine articles on "The Benford's Law of Words – Parts 1 and 2"
- ✓ **http://bit.ly/28LVoLd** - A Better Way To Win At Audit Wheel of Fortune Using Letter Analytics

44

Thank You!

Richard B. Lanza, CFE, CGMA
Cash Recovery Partners, LLC
Phone: 973-729-3944
Email: rich@richlanza.com
www.AuditSoftwarePros.com