

NY/NJ Intergovernmental Audit Forum

Trends in Data Analytics – Fraud Detection

May 4, 2016



Building a better
working world

Discussion Topics

- ▶ Introduction 5 min
- ▶ 2016 EY Fraud Survey 10 min
- ▶ Data Analytics Methodology 30 min
 - ▶ Analytics Design
 - ▶ Data Collection
 - ▶ Data Analytics
 - ▶ Explore & Risk Rank
 - ▶ Findings & Observations
- ▶ Case Study 45 min
- ▶ Questions & Answers 10 min

Data and analytics core businesses will be disrupted!

Information is growing exponentially

- 1.7 MB** New info. created every minute for every human being
- 80%** Data is unstructured; invisible to traditional computers
- 50 billion** “Things” will be connected to the internet by 2020

Computing power is increasing while cost is decreasing

- 18 months** For computing power (analytics) to double
- 14 months** For cost per gigabyte of storage to go down by half

The nature of work is changing ...

- 30%** Of consulting will be replaced by cognitive/AI technology
- 47%** Of today’s job could be automated in the next two decades 94% probability for accounting and Tax
- 50%** Of professional services will be procured as ‘managed service’

As a result of Open Data Initiative , government agency data are becoming more accessible

Human beings tend to overestimate in the near term but under-estimate in the longer term

- 1943** “I think there is a world market for maybe five computers.”
Thomas Watson
Chairman of IBM
- 1977** “There is no reason for any individual to have a computer in their home.”
Ken Olsen
Co-founder and former President of Digital Equipment
- 1998** “Touchscreen e-readers will never catch on.”
Bill Gates
Co-founder of Microsoft
- 1998** “The growth of the Internet will slow drastically ... most people have nothing to say to each other! By 2005 or so, it will become clear that the Internet’s impact on the economy has been no greater than the fax machine’s.”
Paul Krugman
Economist

2016 EY Fraud Survey



Current and emerging risks driving demand

- ▶ The fastest-growing threat in the fraud and investigative risk universe is from cyber breaches and insider threats.
- ▶ FDA demand is also being driven by increasing government and public scrutiny of fraud risk.

“At the SEC, we have made great strides in leveraging data and technology to detect and pursue misconduct. In the enforcement arena, the Commission is using data analytics to help identify wrongdoers and conduct streamlined investigations to optimize our resources.”

– Chair Mary Jo White, opening remarks at the 21st Annual International Institute for Securities Enforcement and Market Oversight, 2 November 2015

- ▶ C-suite respondents have shown a stronger tendency toward FDA adoption than other executives.

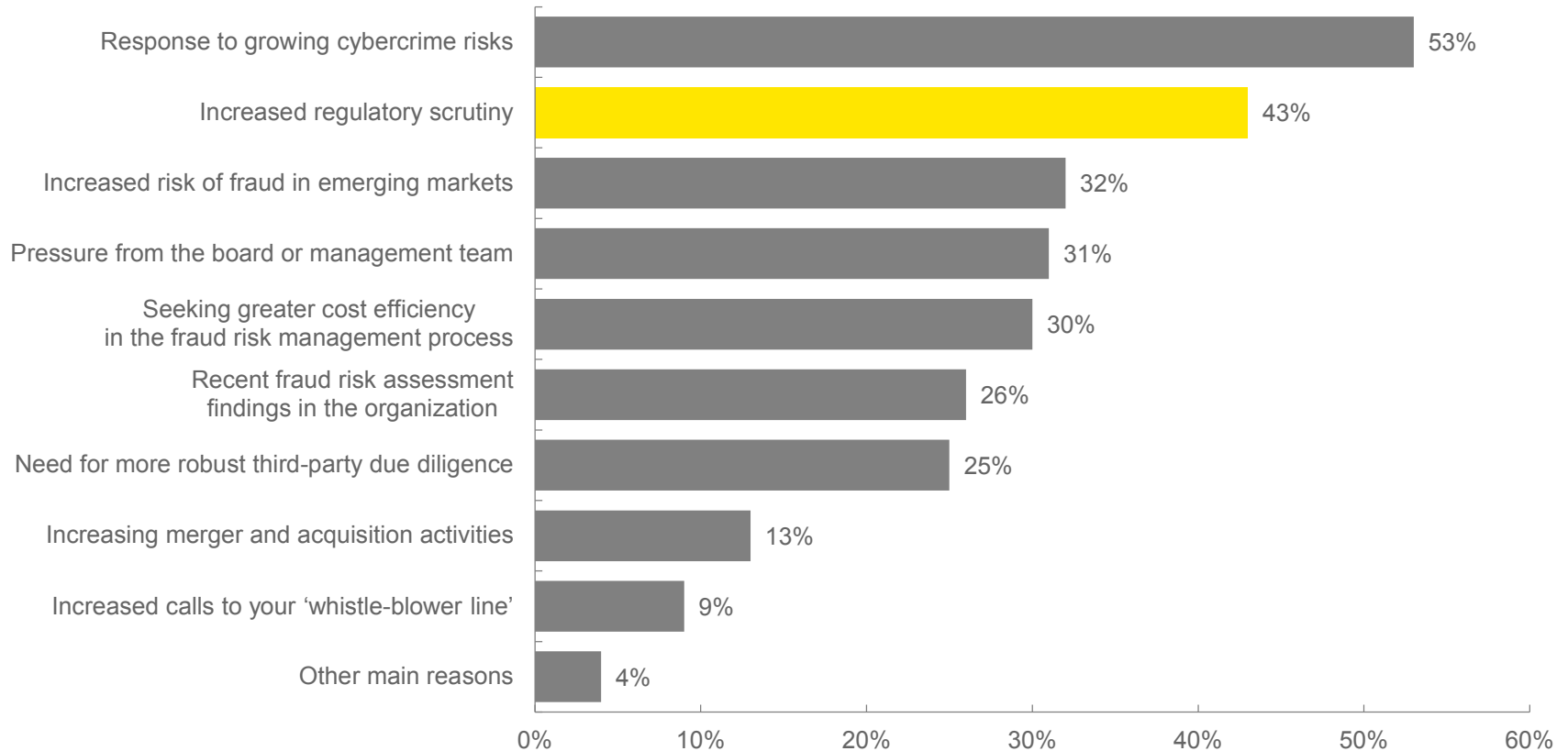
- ▶ While 69% overall agree, “we need to do more to improve our current anti-fraud procedures, including the use of FDA tools,” this number jumps to 74% for the C-suite cohort.



74%
of C-suite
respondents

“We need to do more to improve our current anti-fraud procedures, including the use of FDA tools.”

FDA demand is also driven by increasing government and public scrutiny of fraud risk



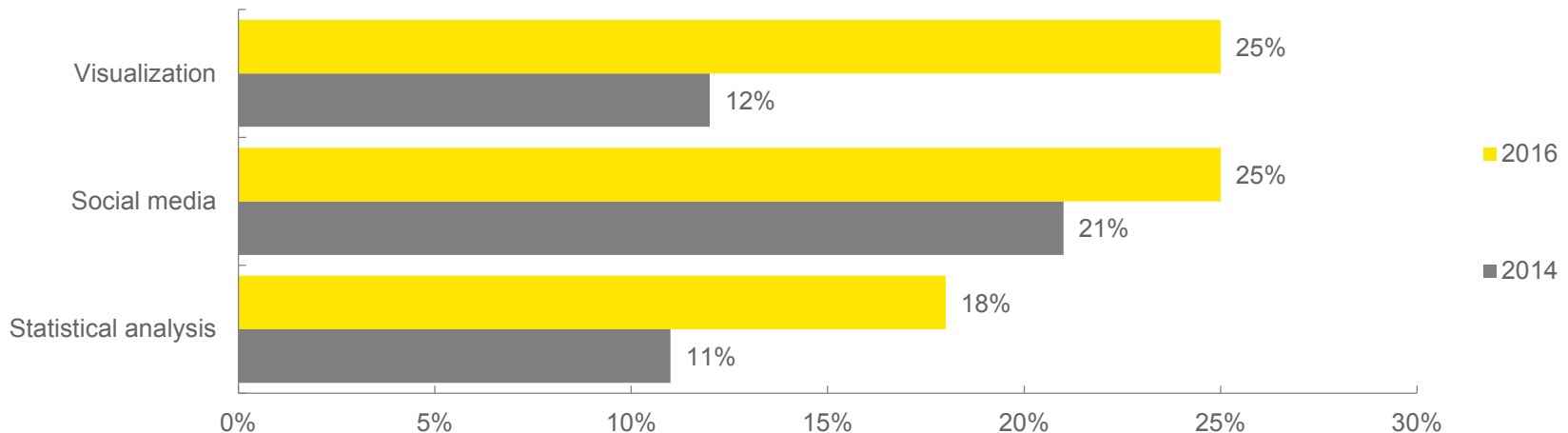
Q. What are the main reasons that you are planning to increase your investment in FDA capabilities?

Base: Respondents who plan to increase investment in FDA (405)

Multiple answers allowed, may exceed 100%.

Growing sophistication in technology and the use of data

- ▶ The use of visualization tools has doubled since our 2014 survey. There has also been increasing use of social and web monitoring tools and statistical analysis and data mining packages.
- ▶ 75% of respondents routinely analyze a wide range of structured and unstructured data.



Q. Which FDA tools do you utilize in managing fraud risk?

Base: 2016 all respondents (665); 2014 all respondents (466)

Multiple answers allowed, may exceed 100%.

Key challenges: building teams with the right skills

Successful deployment requires three distinct skill sets:

- ▶ **Technical skills** – to understand the organization’s systems and advise on acquiring additional technology
- ▶ **Domain knowledge** – familiarity with the relevant risk areas in the business and the ability to interpret analytics results in the context of the organization
- ▶ **Data analytics (e.g., data science) expertise** – mathematical, computer science and business intelligence techniques, such as pattern recognition, statistical analysis, query design and data visualization



Don't forget the human element of your FDA program

“In 39% of leading analytics organizations – versus 12% of the rest – analytics skills are recognized, effective, efficient, monitored and clearly used to support decisions. More than one-third of the top 10% also have well-defined competencies for each role and level, along with robust training programs that address potential skills shortages.”

EY and Forbes Insights, Data and Analytics Impact Index: don't forget the human element of analytics, 2015.

According to Gartner, the need for data scientists is growing at about three times that for statisticians and business intelligence analysts, and there is an anticipated talent shortage of 100,000 or more analytics personnel through 2020.

Key challenges: deploying the right technology

- ▶ Respondents continue to rate “challenges in combining data sources” as the top issue to overcome when implementing FDA.
 - ▶ Building data sets that talk to one another is the first step to successful analytics.
- ▶ One key obstacle to moving to more advanced tools appears to be the lack of funding.
- ▶ However, analytics technology has improved, making deployment and accessibility easier.
 - ▶ More “self-service” applications are available via cloud that require less customization to implement.
 - ▶ Significant improvements in computing power and scalability (i.e., Hadoop) combined with ever-decreasing storage costs make the use of FDA more cost effective.

Many have reported positive results or recoveries from the FDA tools



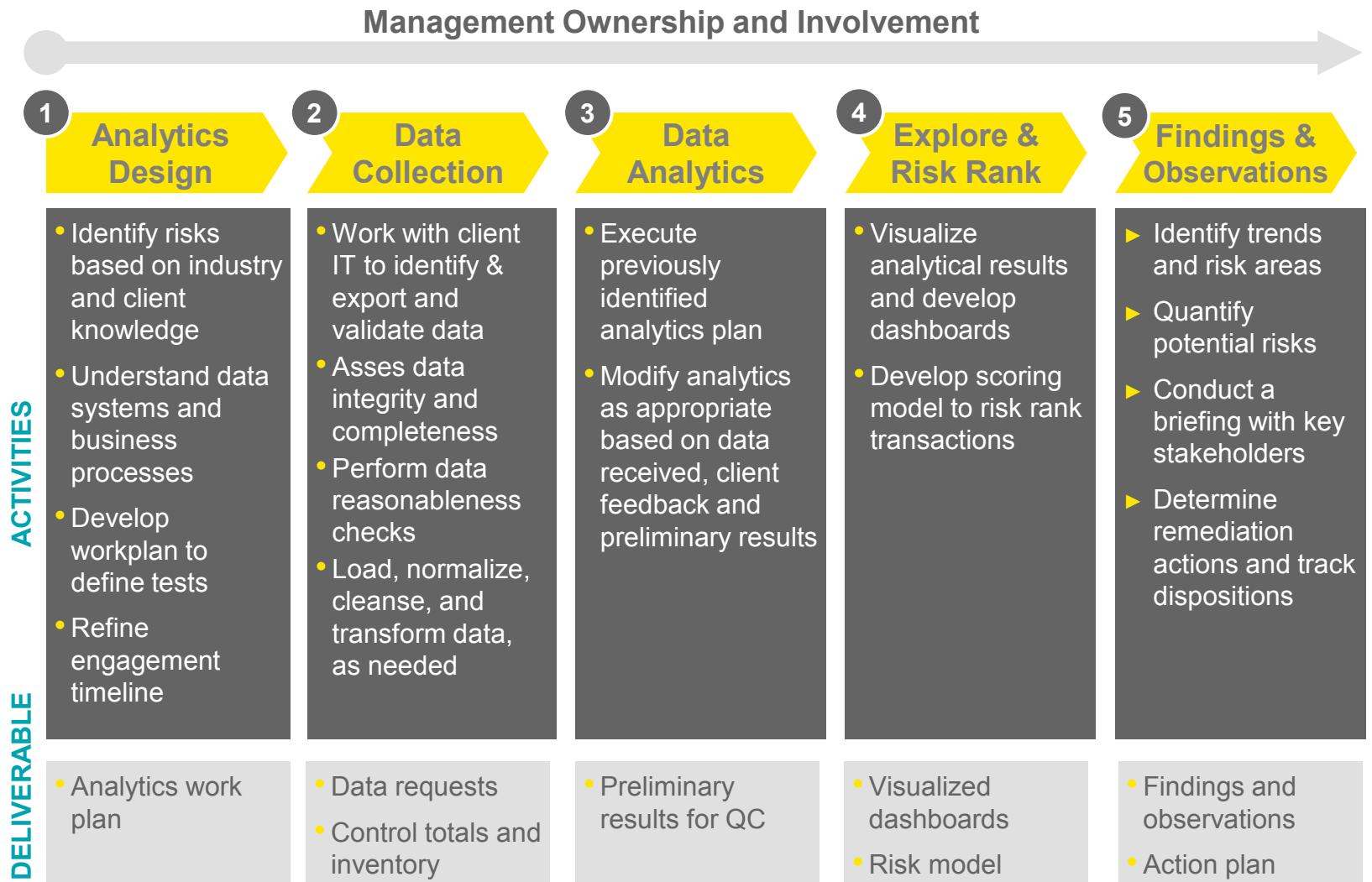
56%
agree

“We currently get positive results or recoveries from the FDA tools that we use.”

Data Analytics – Methodology



Analytics Implementation Methodology



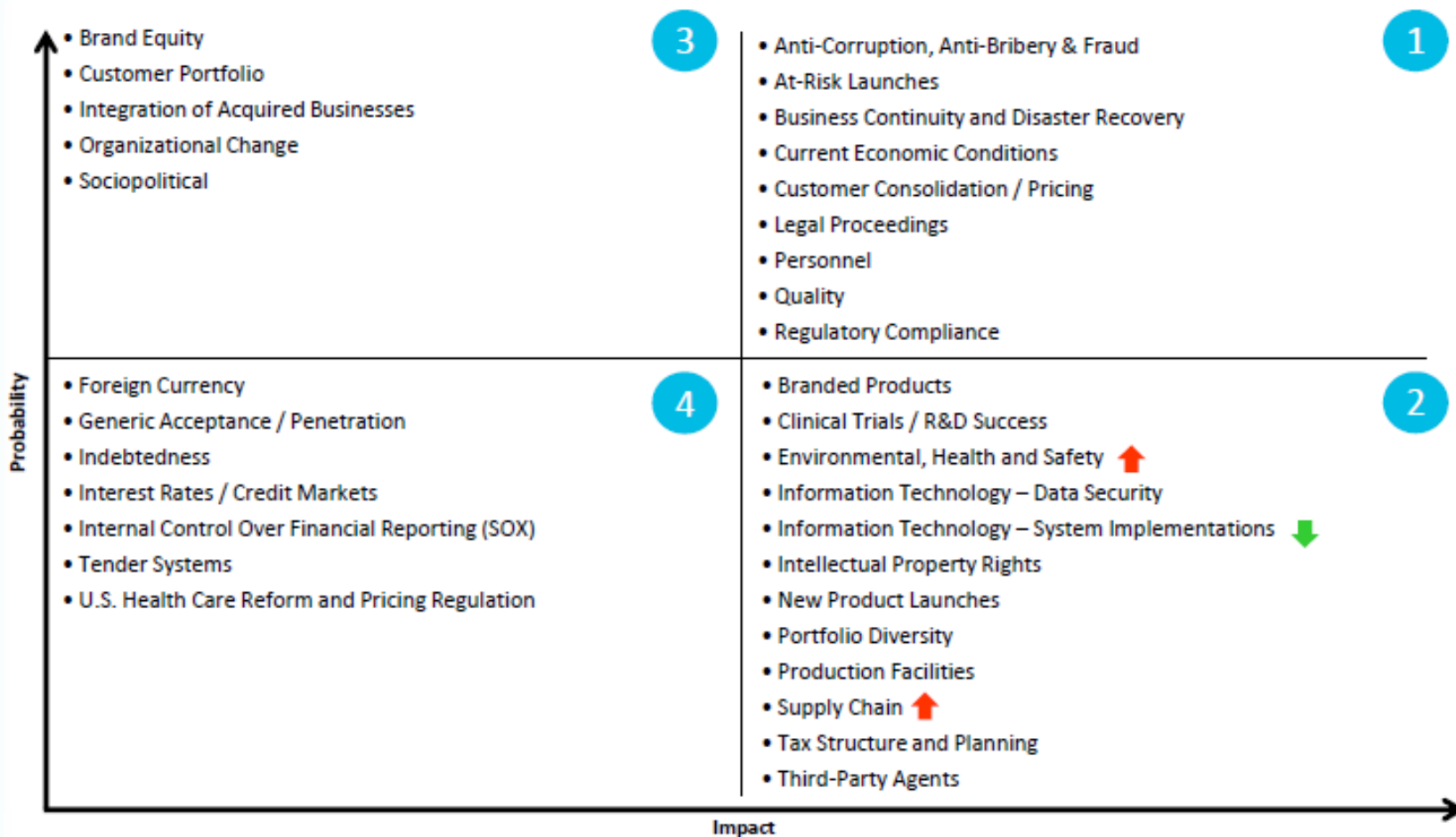
(1) Analytics Design



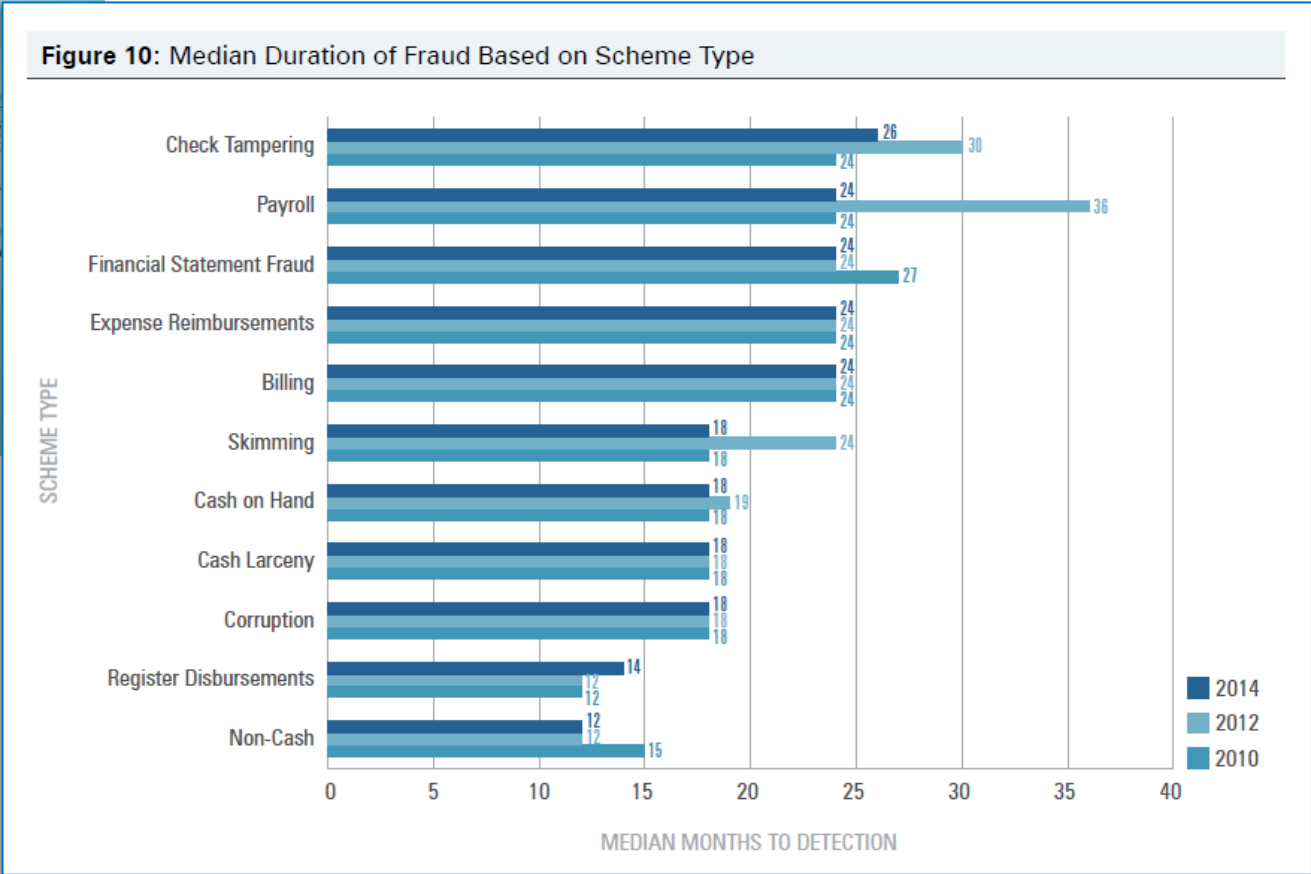
Risk quadrants

Macro view

Draft 2015 Risk Profile



ACFE 2014 Report to the Nations and fraud scheme discussion

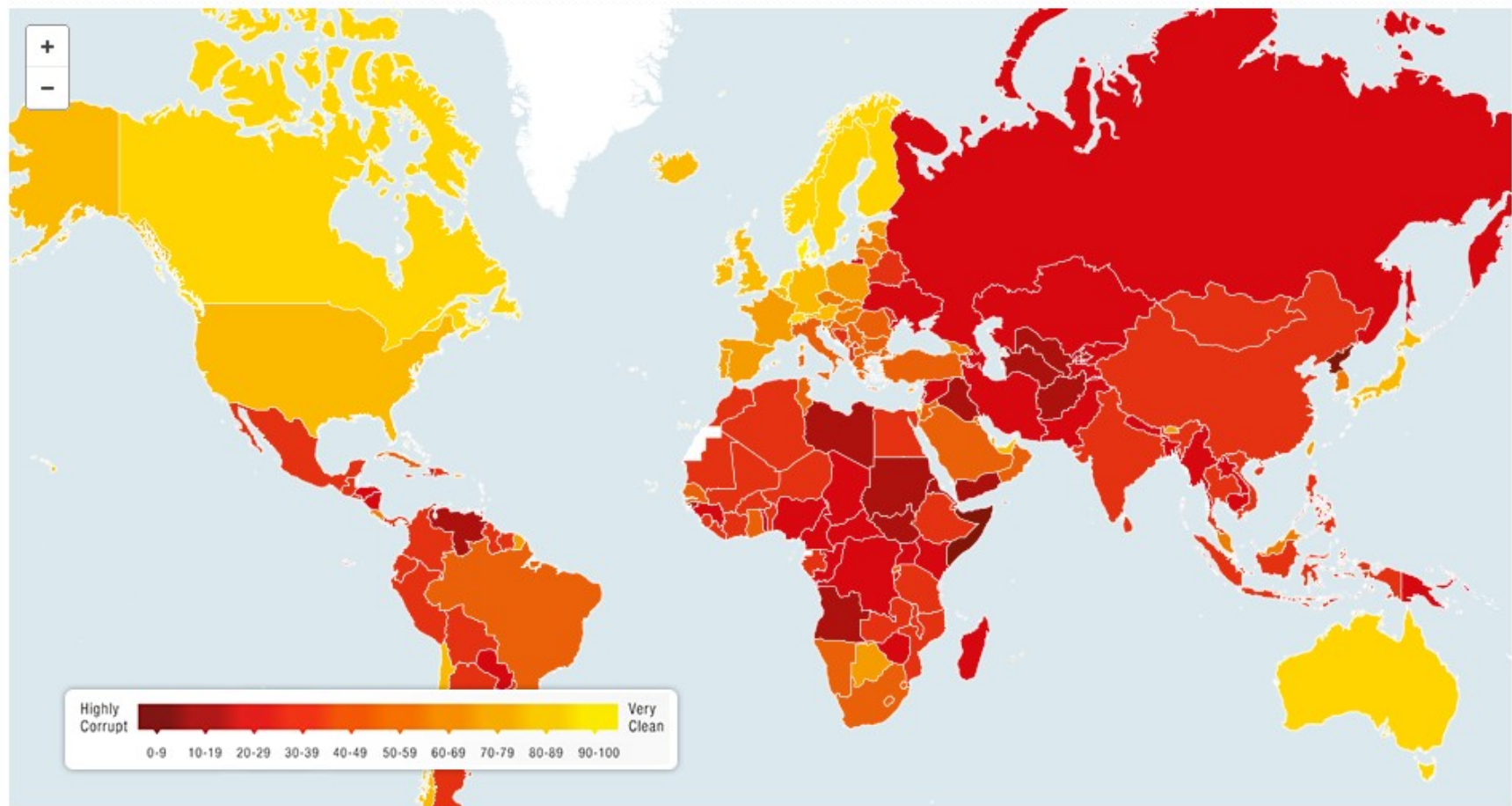


Transparency International 2014

Corruption Perceptions Index

CORRUPTION PERCEPTIONS INDEX 2014: RESULTS

[VIEW BROCHURE](#) [VIEW RESULTS TABLE](#)



Trends and leading practices in analytics¹

What we are seeing

- ▶ More precision in compliance program and operational assessments by sophisticated clients
- ▶ Increased operationalizing of compliance processes for specific risks
- ▶ Advanced data analytics to facilitate risk based continuous monitoring

Legal and regulatory risks

- ▶ Bribery and corruption
- ▶ Third-party integrity
- ▶ Export controls/ITAR

US Government business risks

- ▶ Cost accounting
- ▶ DFARS business systems
- ▶ Mandatory disclosure
- ▶ Cost structure simplification and cost reduction
- ▶ Claims and disputes
- ▶ Suppliers and pricing

Technology risks

- ▶ Data privacy
- ▶ Cyber security
- ▶ Cloud computing and big data
- ▶ Data management requirements for investigations

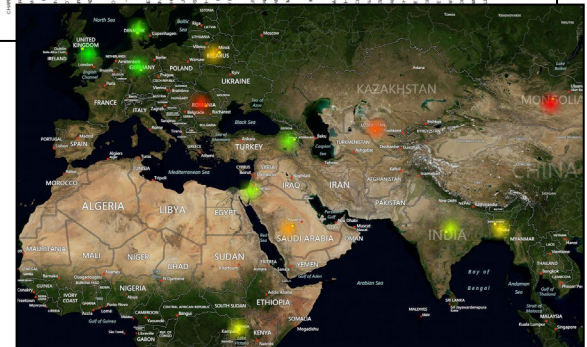
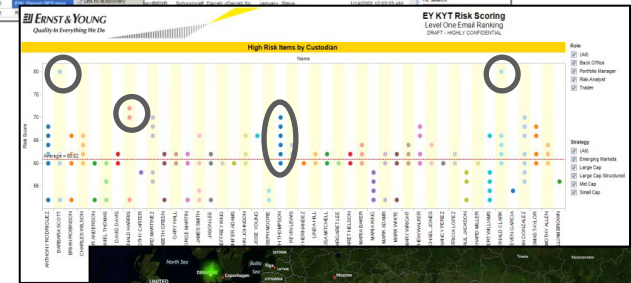
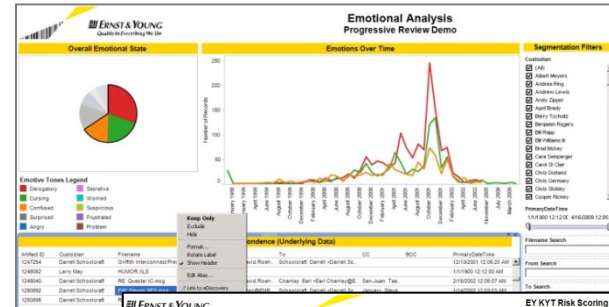
(2) Data Collection



Addressing risks by combining multiple data sources

2

Data Collection






Other external data sources

- **Global watch lists, state-owned enterprises and adverse media**

- Over 1.2 million named individuals and entities
- Over 160 sanction lists
- Enhanced country list addressing Presidential Executive Orders (e.g., state-owned entities, subsidiaries, vessels, etc.)
- Over 250 official lists (e.g., World Bank, Asian Development Bank, etc.)
- People who have been formally accused, arrested, or convicted in white-collar and terrorism crimes
- Relatives and known associates
- Entities in the news for adverse media


- **Social data**

Direct Access



and more

Managed Public API



and more

(2) Data Analysis





Applied Data Analytics

Purpose – Why?

3

Data
Analytics

- ▶ Assess and quantify potential areas of risk
- ▶ Analyze large volumes of data which would be time-consuming for manual review
- ▶ Identify exceptions and anomalies
- ▶ Find patterns and trends through periodic testing

Applied Data Analytics Approaches – What?

3

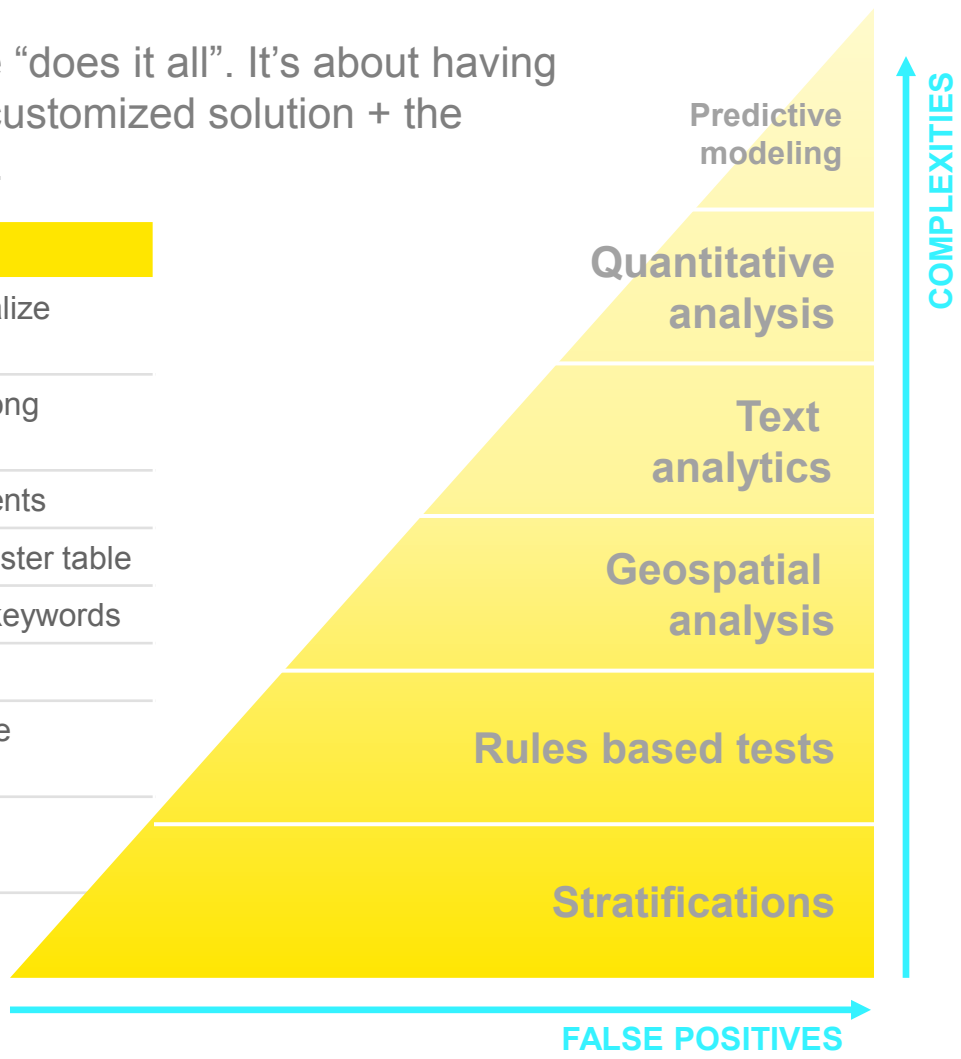
Data Analytics

Rules-based analytics	Testing data against known behaviors/patterns	<ul style="list-style-type: none">▶ Suited to well-defined processes▶ Ideal for anomaly testing▶ Used for identifying data not meeting expected behavior
Risk scoring	Using scoring models to identify and determine areas of higher risk	<ul style="list-style-type: none">▶ Using rules-based tests with weighted scores▶ Combining weighted scores from different datasets to determine high risk areas
Text analytics	Analyzing natural language text: <ul style="list-style-type: none">▶ Keyword searching▶ Concept analysis▶ Sentiment analysis	<ul style="list-style-type: none">▶ Search against keywords, e.g., fraud, bribery and corruption, black lists▶ Analyze text to determine concepts discussed▶ Analyze emails for emotional tone insight
Visual analytics	Using interactive dashboards for reporting and trend analysis	<ul style="list-style-type: none">▶ Visualizing facilitates understanding▶ Makes comparative analysis easier▶ Aids in spotting unusual trends

Applying the right techniques using the right tools

In our experience, no one tool or technique “does it all”. It’s about having the capacity and capabilities to develop a customized solution + the domain knowledge to contextualize results.

Techniques	Sample Counter Fraud Tests
Visualization	Stratify transactions by attributes to visualize trends and patterns
Rules-based	Identify split and duplicate expenses among employees
Rules-based	Identify anomalous mileage reimbursements
Rules-based	Reimbursements to employees not in master table
Text mining	Mine free-text fields for fraud/corruption keywords
Text mining	Term frequency analysis (unsupervised)
Data augmentation	Identify employees/attendees that may be sanctioned or politically exposed
Entity matching	Employees that share same attributes as vendors
Entity matching	Identify frequently entertained attendees



Applied Data Analytics Applications – Where?

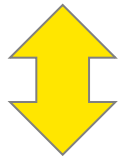
3

Data Analytics

Travel & Expense

Identify high risk employees through rules and text based analytics. Analyze findings in a dynamic risk scoring dashboard.

- ▶ Expense reports
- ▶ Meal attendee information
- ▶ Employee master



Conflict of Interest

- ▶ Employees Appearing in Vendor Master

Procurement

Identify high risk vendors through rules and text based analytics. Analyze findings in a dynamic risk scoring dashboard.

- ▶ Accounts payable ledger
- ▶ Vendor Master

Labor and Payroll

Identify anomalies and exceptions in labor charging and payroll.

- ▶ Time sheets
- ▶ Contract information
- ▶ Payroll

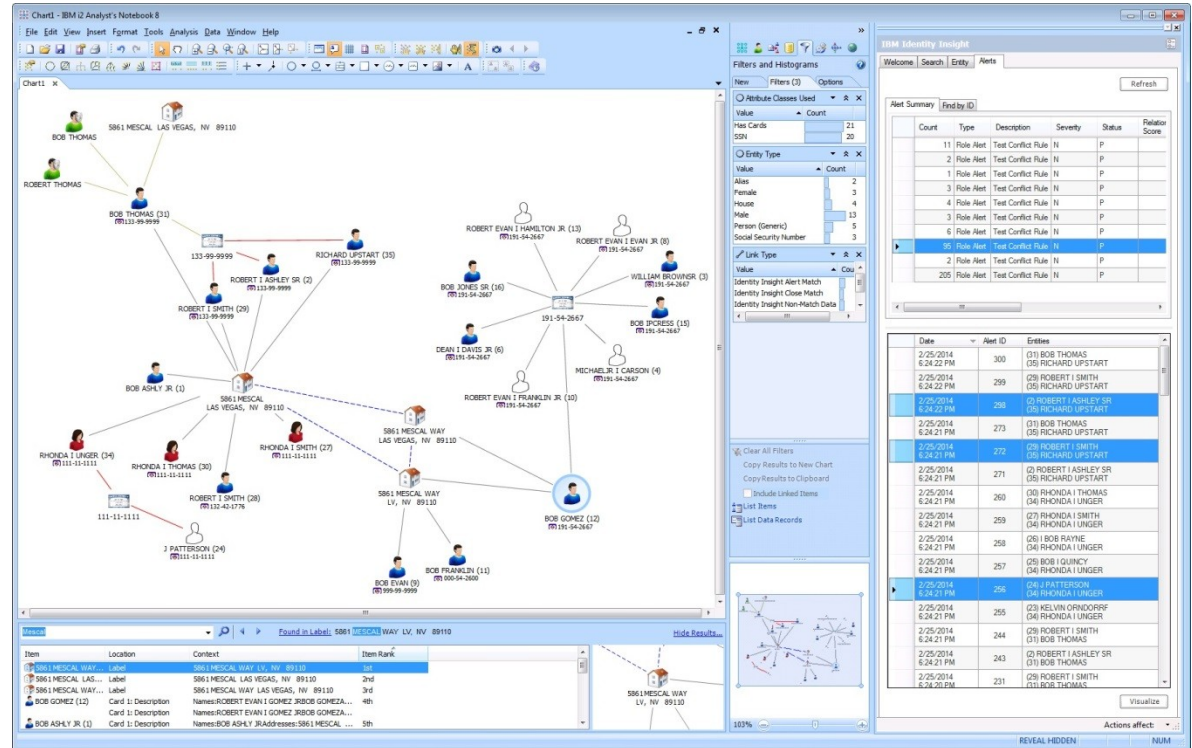
General Ledger / Accounting

Identify anomalous trends and patterns across time, geographies, and business units.

- ▶ General ledger
- ▶ Chart of accounts

Entity resolution and link analysis

- ▶ Automatically match and disambiguate individuals and entities in your master reference data.
- ▶ Vast improvements over fuzzy matching.
- ▶ Better conflict of interest tests, fewer false positives.
- ▶ More accurate network link analysis with deeper degrees of separation.



(4) Explore and Risk Rank



Per analytic:

$$\begin{array}{r} 1 \\ \times 5 \\ \hline 5 \end{array} \begin{array}{l} \text{(yes/no flag)} \\ \text{(1-5 metric weight multiplier)} \\ \text{(analytic score)} \end{array}$$

A fully loaded transaction score takes into account the sum of analytic scores that all hit on that transaction (e.g. 1-duplicate, 2-wkend, 3-round payment 5+5+5=15).

Using analytics to enable continuous monitoring

Continuous monitoring works when a defined, repeatable transactional tests are applied to a wide range of data. Risk scores are aggregated and compared to better focus resources based on evolving risks.

Aggregate risk scores as compared to previous analyses to help focus on higher risk areas

EY has partnered with IBM to leverage the power of big data platform

Summary of scope

EY Building a better working world
FORENSIC DATA ANALYTICS
POWERED BY IBM BigInsights

Data Analytics Dashboard
DRAFT - Confidential Information
as of January 30, 2014

The Data Analytics Dashboard integrates analytical results from multiple transactional data sets to present a unified view of regulatory, compliance, and operational risks. Changes in risk scores from prior analysis are also presented to show trends. The intent of the dashboard is to take a data-driven approach and help focus monitoring on areas that present highest risks.

Summary of Scope

Scope of data	Aug-1-2012 through Jul-31-2012
Total number of records analyzed	1 Million
Total volume of raw data	10GB
Number of raw transactional tables	40
Total number of vendor records	105,485
Total number of customers	894
Total number of employees records	3,164
Source Systems	SAP Dow Jones Watchlist

Accounts Payable		
Number of Records	Number of Vendors	Amount Disbursed
60,855	88	\$1,428,094,969.14

Third Party Payments		
Number of Vendors	Number of Records	Amount
88	53,165	\$1,428,094,969.14

Travel and Entertainment		
Employees	Number of Records	Amount Disbursed
50	7,690	\$762,438.36

Sanction and Political Exposure Risks		
Number of Records	Avg. Similarity	Amount USD
381	86.33%	\$2,811,067.04

General Ledger			
Number of Records	Number of Accounts	Amount in USD	
108,388	205	\$3,361,392,928.46	

General Ledger			
Number of Records	Number of Accounts	Debit Credit Indicator	Amount in USD
46,653	182	Credit	\$1,680,696,464.04
61,735	196	Debit	\$1,680,696,464.42

Sales		
Number of Records	Revenue	Counter-Revenue
98,301	\$13,537,401.34	(\$985,398.59)

Distributor Monitoring		
Customers	Revenue	Counter-Revenue
50	\$13,537,401.34	(\$985,398.59)

Unit Price/Margin Analysis		
Products	Average Unit Price	Average Margin
12,091	\$399.73	37%

Risk Score	Change
291	↓ -2.9%
612	↑ +4.6%
451	↑ +9.7%
406	↓ -1.2%
396	- ±0.0%
276	↓ -0.6%

Detail analyses of each domain

(5) Findings & Observations



Example #1

Summary analysis



Company XYZ
Labor Analytics and Trending
DRAFT - CONFIDENTIAL

Time Period

Work Date: 1/1/2013 to 12/31/2013

Summary

No. of Time Entries	No. of Employees	No. of Contracts	No. of Projects	Hours
20,198	258	77	7,726	67,969.80

Analytics

Over 24 Hours: (All)

No Sick or Vacation Days: (All)

Different Fiscal Year: (All)

Contract Not in Master: (All)

Employee Not in Master: (All)

Duplicate Employee ID: (All)

Duplicate Employee Name: (All)

Weekly Hour Variations: (All)

Signed off by: (All)

Statistic by Cost Center

Cost Center	Hours	Time Entries
Engineering	28,221.80	8,289
Manufacturing	13,241.40	4,194
Procurement	4,388.90	1,320
Program Support	3,631.10	1,196
IT	2,905.70	898
OSN	7,531.20	2,148
Maintenance	1,692.10	
Facilities	1,600.30	

Summary by Employee

Employee	Employee Direct/Indirect	Employee Location	Process Status	Approved by	Hour Type	No. of Time Entries	Hours		
Adrienne Klein	Direct	San Diego	Approved	Margit Minder	ATTENDANCE..	53	165.00		
					OVERTIME 1.5	4	9.00		
					SICK DAY	1	3.00		
					VACATION	2	16.00		
					Cancelled	Margit Minder	ATTENDANCE..	11	28.00
					OVERTIME 1.5	1	4.50		
Albert Jenkins	Direct	Seattle	Approved	Albert Jenkins	ATTENDANCE..	1	1.00		
					Georgene Gaver	ATTENDANCE..	53	163.00	
					INCENTIVE V..	1	8.00		
					NON-CONTRA..	1	8.00		
					SICK / PERSO..	1	8.00		

Custom Summary

Select Field: Contract Number

Selected Field	Direct or Indirect	Process Status	Hour Type	No. of Time Entries	Hours
0001384163L..	Direct	Approved	ATTENDANCE HOURS	2,684	8,026.60
			VACATION	5	18.90
			ABSENCE W/PERM AWP..	1	3.40
			LEAVE EARLY	1	2.90
			ABSENCE WITH PERM	1	2.20
			ABSENCE W/ PERM AWP..	2	7.00
			OVERTIME 1.5	55	114.70
			OVERTIME 2.0	8	21.20
	Cancelled		ATTENDANCE HOURS	342	1,088.30


Summary by Month

Year of Work Date	Month of Work Date	Employee	Cost Center Category	Cost Center	Process Status	No. of Time Entries	Hours
2013	January	Albert Jenkins	General Services	Rocket Testi..	Approved	1	0.40
			Home Office	Rocket Testi..	Approved	3	6.90
			Human Resources	Rocket Testi..	Approved	2	13.30
			IT	Rocket Testi..	Approved	5	7.40
			Maintainance	Rocket Testi..	Approved	3	8.80
			OSN	Planning Pro..	Approved	1	8.00
		Alberta Winslow	OSN	Structural Dynamics	Approved	1	1.50
				Cancelled	1	6.00	
				Approved	6	21.10	
		Alma Fitzsimmons	Engineering	Systems Engineering	Approved	6	21.10
				Cancelled	1	4.00	



Example #2

Social Media Analysis




Social Media Compliance Dashboard

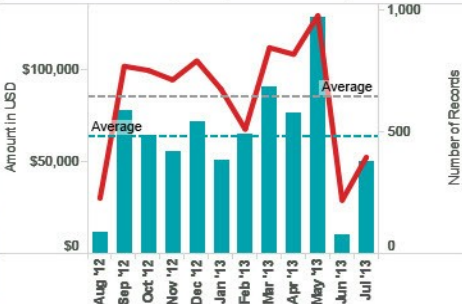
DRAFT - Confidential Information

Summary	
No of Factiva articles	No of Tweets
379	2,113


Media buzz by geography



Tweet frequency and stock price



Fraud risk terms



Date range
8/1/2012 7/31/2013

Brand names
 (All)
 Bizex
 Damart
 Dylaxilix
 Flonalvia

Known actors
(All)

On/Off label terms
(All)

Fraud risk terms
(All)

Sentiment
(All)

Stock price
(\$395.86) \$7,645.92

Custom search

Factiva feed

Headlines 1 - 20 of 2,137 Next 20 Total duplicates: 11

- Impax resubmits NDA for Parkinson's disease drug Rytry to US FDA**
Pharmaceutical Business Review · 0:24, 14 April 2014, 215 words, (English)
US-based Impax Pharmaceuticals, a division of Impax Laboratories, has resubmitted its new drug application (NDA) for Rytry (IPX066) to the US Food and Drug Administration (FDA), for the symptomatic treatment of Parkinson's disease (PD).
3 duplicate article(s) identified
- 4/14/14 - Kite Pharma Announces FDA Orphan Drug Designation for Anti-CD19 T Cell Cancer Immunotherapy Product**
Pharmacy Choice · 1:00, 14 April 2014, 335 words, (English)
By a News Reporter-Staff News Editor at Clinical Trials Week -Kite Pharma, Inc. (Kite), a clinical-stage biotechnology company focused on developing engineered autologous T cell therapy (eACT7) products for cancer, announced that the U.S. ...
1 duplicate article(s) identified
- 4/14/14 - 3rd Annual ISPE-FDA CGMP Conference to Address Latest Compliance Trends and Clarify Regulatory Expectations**
Pharmacy Choice · 1:00, 14 April 2014, 334 words, (English)
By a News Reporter-Staff News Editor at Pharma Business Week Top officials from the US FDA will meet with ISPE Members and other pharmaceutical industry leaders to discuss compliance trends and identify practical solutions for modern ...
- 4/14/14 - FDA Awards Fast Track Status to Tetraphase Pharmaceuticals for IV and Oral Formulations of Eravacycline**
Pharmacy Choice · 1:00, 14 April 2014, 587 words, (English)
By a News Reporter-Staff News Editor at Clinical Trials Week Tetraphase Pharmaceuticals, Inc. (NASDAQ:TTPH) announced that the U.S. Food & Drug

Twitter feed

MarketWatch @MarketWatch · Mar 31
Elianos **LifeSciences** gains 4% to lead S&P 500
Expand

ARMACAD @ARMACAD · Mar 31
International PhD Fellowship Program in **Life Sciences** And Their Ethical Consequences, Italy | .fb.me/2LgNBFfP
Expand

BioPortfolio @BioPortfolio · Mar 31
VRooms Powers Virtual Clinical Trials Bringing Security and Document Management Expertise to **Life Sciences** ow.ly/2FcVWgh
Expand


Pharmafocus @Pharmafocus · Mar 31
Life sciences boost from government. bit.ly/O83Fo #pharma #lifesciences
Expand

ThermoSciBio @ThermoSciBio · Feb 20
Have you taken our "Many Faces of Support Quiz"? Do it now and get your FREE Fun Facts on Molecular Cloning Poster! ow.ly/O83Fo
Promoted by ThermoSciBio
Expand

Followed by **Pharmafocus** and 1 other
Brett Wells @brettwells · Mar 31
Life sciences boost from government. bit.ly/O83Fo #pharma #lifesciences
Expand

Followed by **Glen Giovannetti**
Marcus Lippold @stclifeScience · Mar 31
Sanofi Appoints Anne Beal to Newly Created Position of Chief Patient Officer

Actors



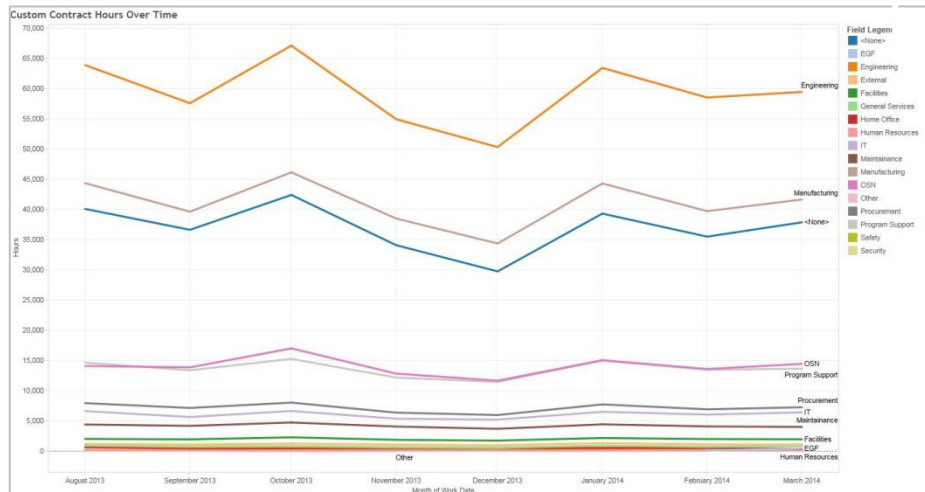
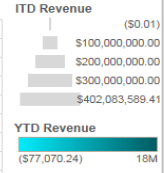
Example #3

Time charging and contract analytics

Contract analytics

- Identification of **high risk** contracts based on attribute-based risk ranking methodology
- Interactive tool allows for analysis and filtering by customer attributes including:
 - Contract
 - Contract type
 - Period of performance and percent complete
 - Customer
 - Financial performance (e.g. profit, billed accounts receivable, billings in excess of revenue, revenue in excess of funding)

Selected	Field	Number of C..	Current Contract ..	Funded Value	YTD Revenue	ITD Revenue
DFJ0000118709	1	\$7,313,267.00	\$7,313,267.00	\$40,704.65	\$7,240,602.25	
DFJ0000120428	1	\$7,152,282.00	\$7,152,282.00	\$0.00	\$7,152,282.00	
DFJ0000129016	1	\$9,000,000.00	\$2,053,411.00	\$0.00	\$2,026,696.60	
DFJ0000129946	1	\$42,837,413.00	\$34,180,420.00	\$3,290,060.82	\$24,544,238.28	
DFJ0000154999	1	\$6,312,584.72	\$2,151,942.00	\$0.00	\$2,151,942.00	
DFJ0000159931	1	\$28,788,881.00	\$28,788,881.00	\$0.00	\$28,788,881.00	
DFJ0000163670	1	\$20,597,402.00	\$5,952,480.52	\$0.00	\$5,948,793.46	
DFJ0001119323	1	\$26,535,657.00	\$2,164,555.00	\$0.00	\$2,164,555.00	
DFJ0001165399	1	\$7,100,000.00	\$7,100,000.00	\$3,426.24	\$7,042,430.48	
DFJ0001165511	1	\$20,174,000.00	\$8,906,783.34	(\$11,227.39)	\$8,551,459.69	
DFJ0001201166	1	\$14,465,089.41	\$14,465,089.41	\$0.00	\$14,447,413.38	
DFJ0001215376	1	\$8,247,770.00	\$8,247,770.00	\$0.00	\$8,242,316.22	
DFJ0001243632	1	\$7,265,433.00	\$306,106.00	\$0.00	\$303,604.54	
DFJ0001309412	1	\$180,632,584.88	\$56,386,899.64	\$10.22	\$56,206,216.86	
DFJ0001333217	1	\$11,521,360.53	\$4,211,808.00	\$0.00	\$4,046,518.27	
DFJ0001350232	1	\$6,986,244.00	\$6,986,244.00	(\$0.01)	\$6,943,476.82	
DFJ0001403419	1	\$8,621,257.32	\$718,940.69	\$0.00	\$119,647.15	
DFJ0001406782	1	\$21,280,262.00	\$21,280,262.00	\$0.00	\$21,280,262.00	
DFJ0001444981	1	\$6,851,628.00	\$6,756,628.00	\$0.00	\$6,542,337.52	
DFJ0001514302	1	\$14,047,060.00	\$14,047,060.00	\$0.00	\$13,890,456.00	
DFJ0001516097	1	\$70,179,871.69	\$3,761,711.69	\$0.00	\$3,119,293.00	
DFJ0001602020	1	\$170,064,114.00	\$35,811,874.00	\$6,128,935.10	\$11,249,227.79	
DFJ0001627931	1	\$8,600,000.00	\$8,600,000.00	\$10,957.12	\$8,599,668.35	
DFJ0001692378	1	\$35,137,703.00	\$7,068,700.00	\$0.25	\$7,061,524.52	
DFJ000170R275	1	\$12,785,189.75	\$11,747,327.13	\$0.00	\$11,181,350.42	



Time charging analytics

- Comparison of labor data to contract qualification requirements
- Identification of unusual trends in straight time and overtime labor charging
- Detection of charges made to multiple categories by the same employee
- Comparison of billed rates to contracted rates
- Trending of time charging practices by employee, contract, cost center and labor category

Contact Information

Ken Feinstein

Senior Manager, Forensic Technology & Discovery Services

203-674-3177

Kenneth.feinstein@ey.com