Finance, Management, & Operations

Applications for Business Intelligence, Predictive Analytics and Big Data

Patrick Bogan, Chief Information Officer, Fuzion Analytics

Kyle Korzenowski, Chief Information Officer, Univita Health

ILTCI

14th Annual Intercompany Long Term Care Insurance Conference

Big Data: Just Another Buzzword?

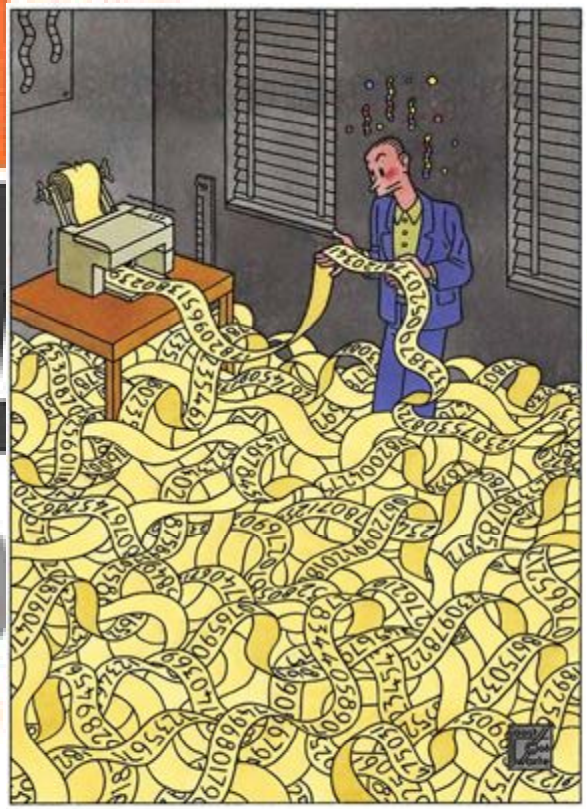
Source: Dilbert.com (http://dilbert.com/strips/comic/2013-01-09/)

What Is Big Data?

Just a whole lotta data?

Source: Spiral16

Source: Intrapromote

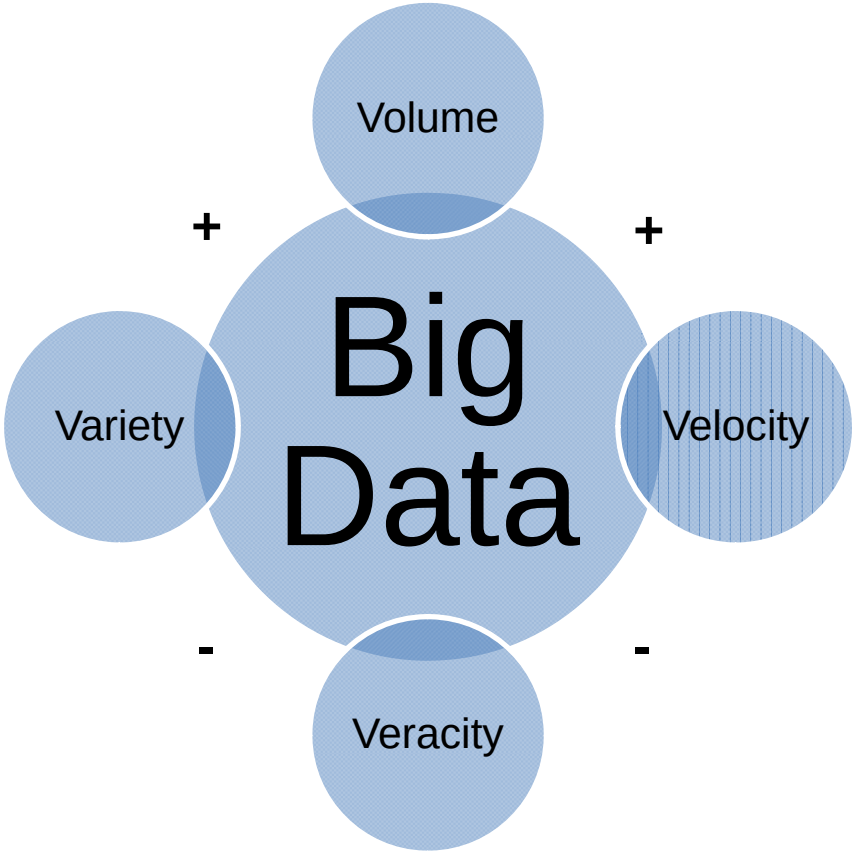


Source: Joost Swarte, The New Yorker

Definition of Big Data

IDC defines "big data" as follows:

"Big-data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and / or analysis."



Volume

How Big Is Big?	Byte	Examples
Byte	**1**	
Kilobyte (KB)	**10^3**	*1,000 bytes OR 10^3bytes* 2 Kilobytes: A Typewritten page.
Megabyte (MB)	**10^6**	1 Megabyte: A small novel OR a 3.5 inch floppy disk. 5 Megabytes: The complete works of Shakespeare. 10 Megabytes: A minute of high-fidelity sound. 100 Megabytes: 1 meter of shelved books. 500 Megabytes: A CD-ROM.
Gigabyte (GB)	**10^9**	1 Gigabyte: a pickup truck filled with books. 4.7 Gigabytes: DVD 20 Gigabytes: A good collection of the works of Beethoven. 100 Gigabytes: A library floor of academic journals.
Terabyte (TB)	**10^{12}**	1 Terabyte: 50000 trees made into paper and printed. 2 Terabytes: An academic research library. 10 Terabytes: The print collections of the U.S. Library of Congress. 400 Terabytes: National Climactic Data Center (NOAA) database.
Petabyte (PB)	**10^{15}**	1 Petabyte: 3 years of EOS data (2001). 2 Petabytes: All U.S. academic research libraries. 20 Petabytes: Production of hard-disk drives in 1995. 200 Petabytes: All printed material.
Exabyte (EB)	**10^{18}**	2 Exabytes: Total volume of information generated in 1999. 5 Exabytes: All words ever spoken by human beings.
Zettabyte (ZB)	**10^{21}**	

Source: Adapted from Roy Williams "Data Powers of Ten" web page at Caltech.

Volume

According to a 2010 IDC study, 3.4 exabytes are produced and replicated daily…that is 1.2 zettabytes annually.

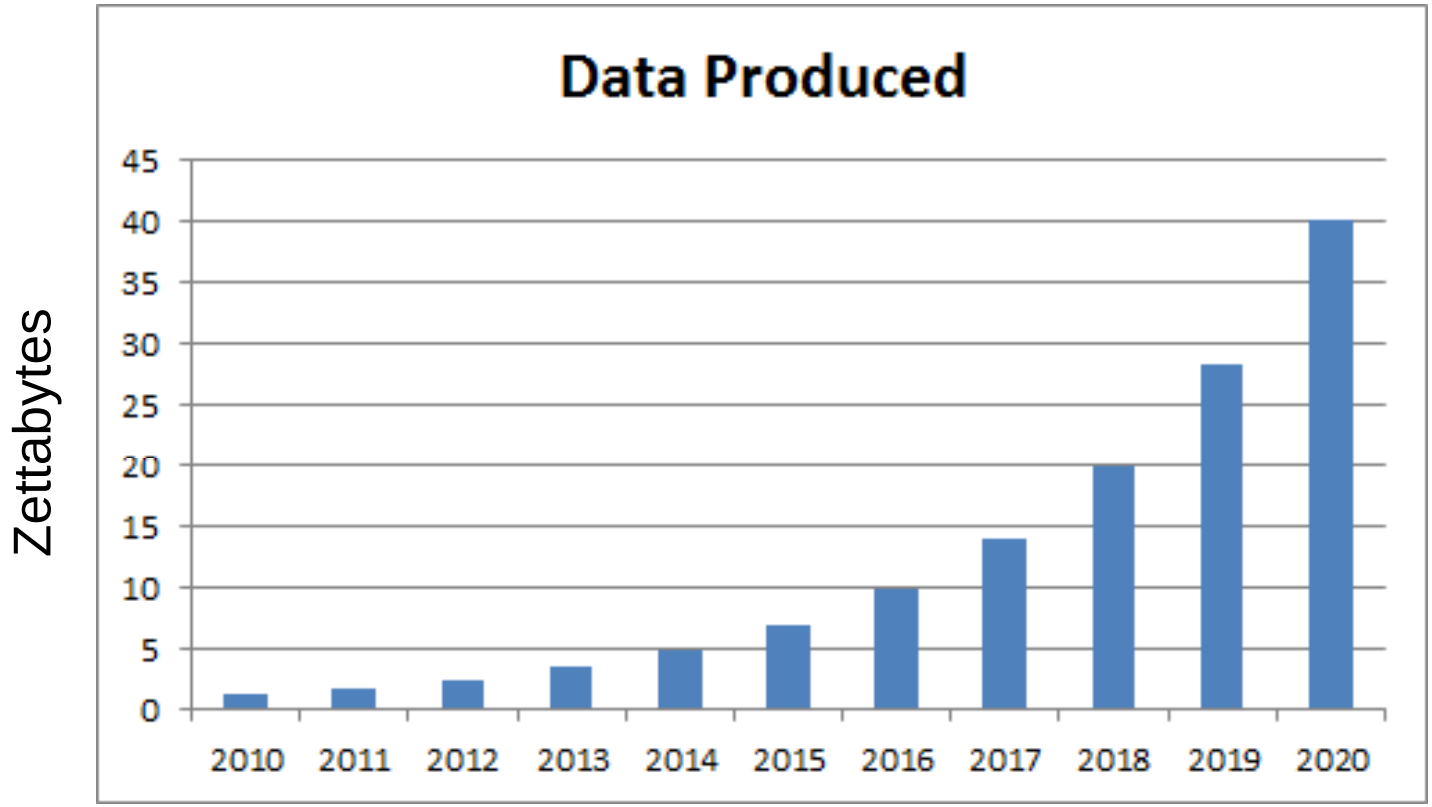
Equates to…

255 billion DVDs
or
600 quadrillion typewritten pages

…Daily!

Volume Growth

IDC estimates the volume of digital data will grow 40-50% per year through 2020.



Data about data, or metadata, is growing twice as fast as the digital universe as a whole.

Data Velocity

“Every two days now we create as much information as we did from the dawn of civilization up until 2003…The real issue is user-generated content.”

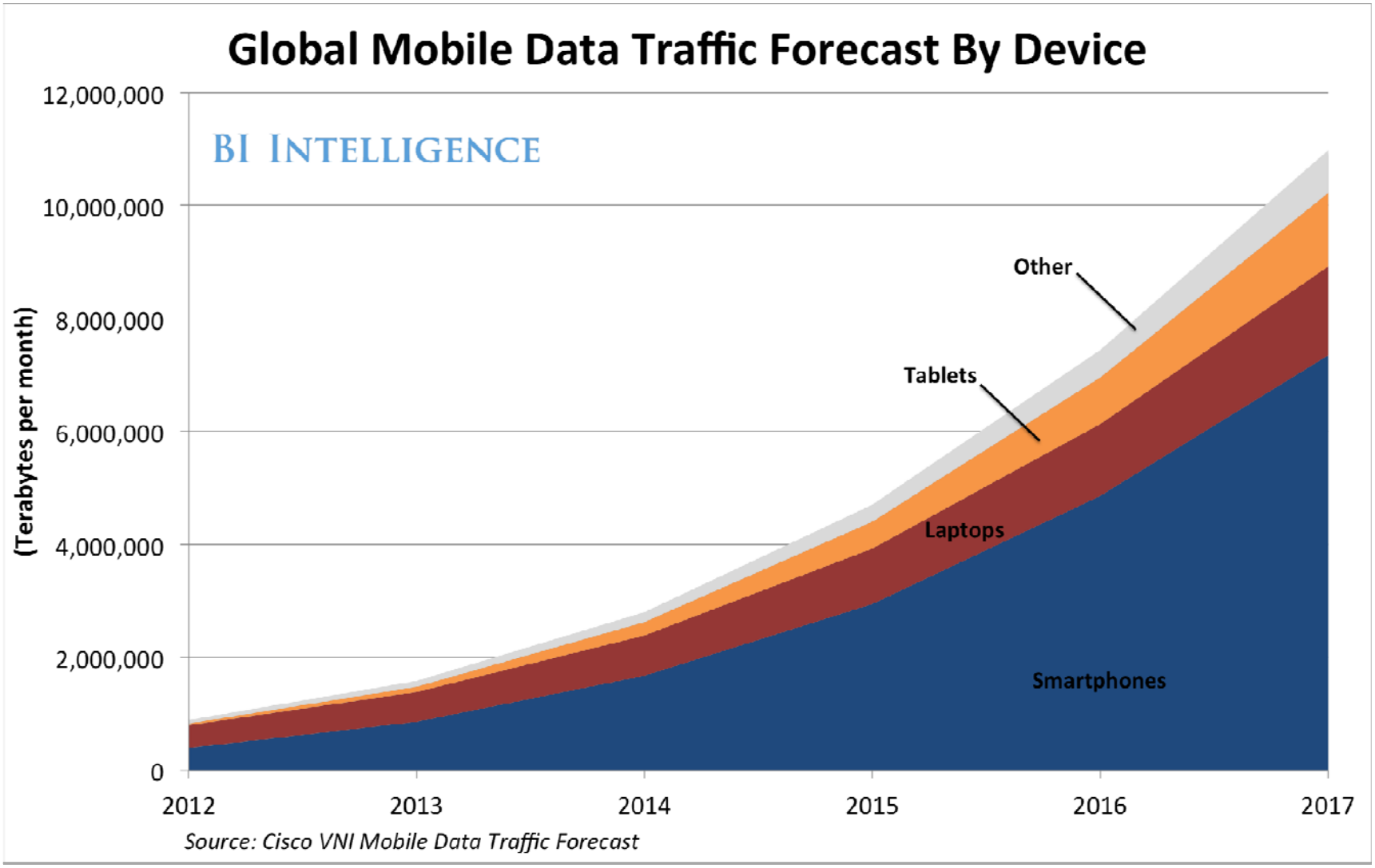
- Eric Schmidt, then-CEO, Google

A better estimate:

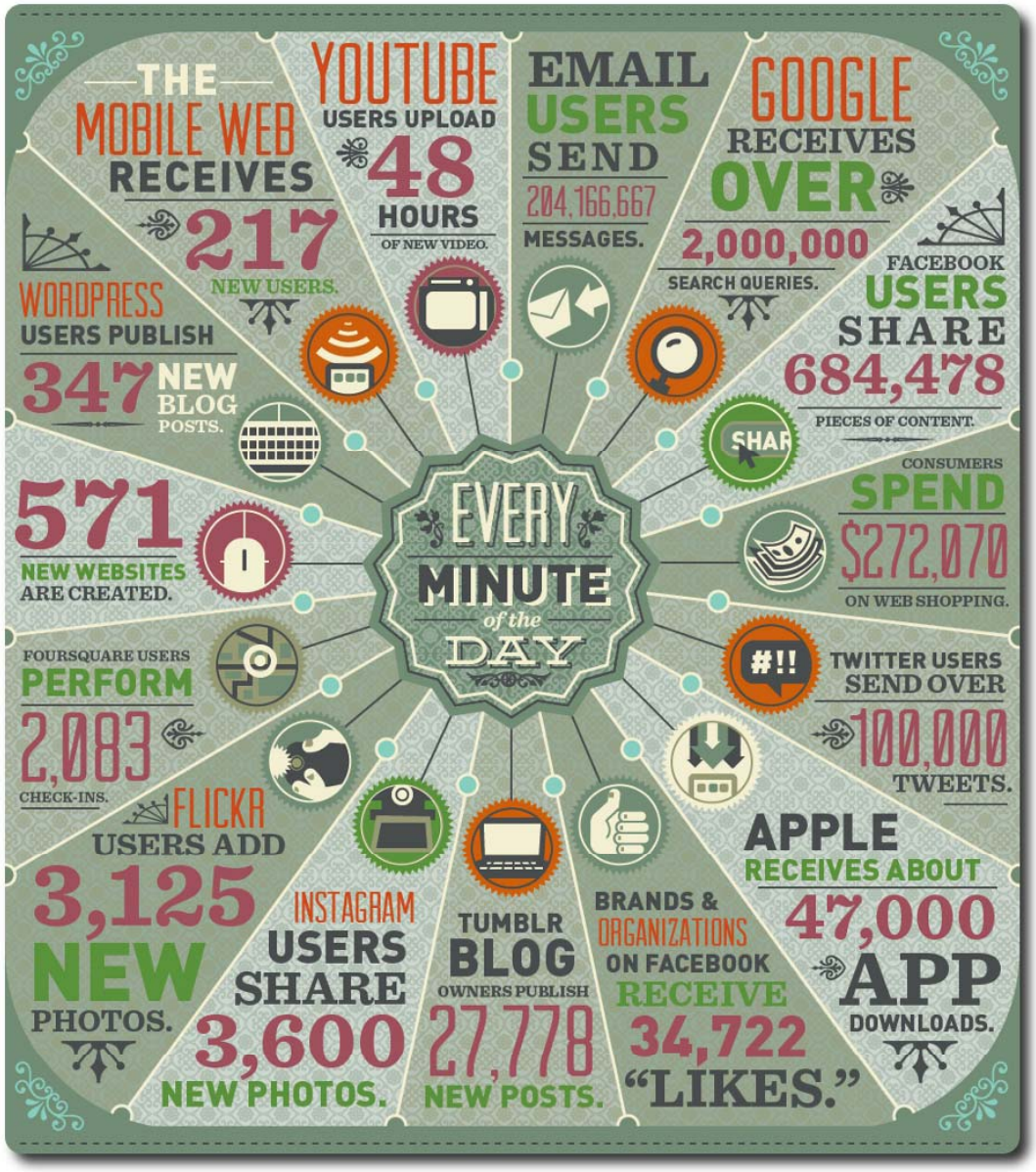
"23 Exabytes of information was recorded and replicated in 2002. We now record and transfer that much information every 7 days.”

- Robert J Moore (RJMetrics)

User-Generated Content



Data Variety



Source: DOMO (http://www.domo.com/blog/2012/06/how-much-data-is-created-every-minute/?dkw=socf3)

Some General Implications

Investment

- Spending on data infrastructure will grow (but at a slower rate)

Sources and Liability

- 68% of data is created and consumed by consumers — watching digital TV, interacting with social media, sending camera phone images and videos between devices and around the Internet, and so on
- But enterprises have liability or responsibility for nearly 80% of the information in the digital universe

Security

- The proportion of data in the digital universe that requires protection is growing faster than the digital universe itself, from less than a third in 2010 to more than 40% in 2020.

Analytic Value

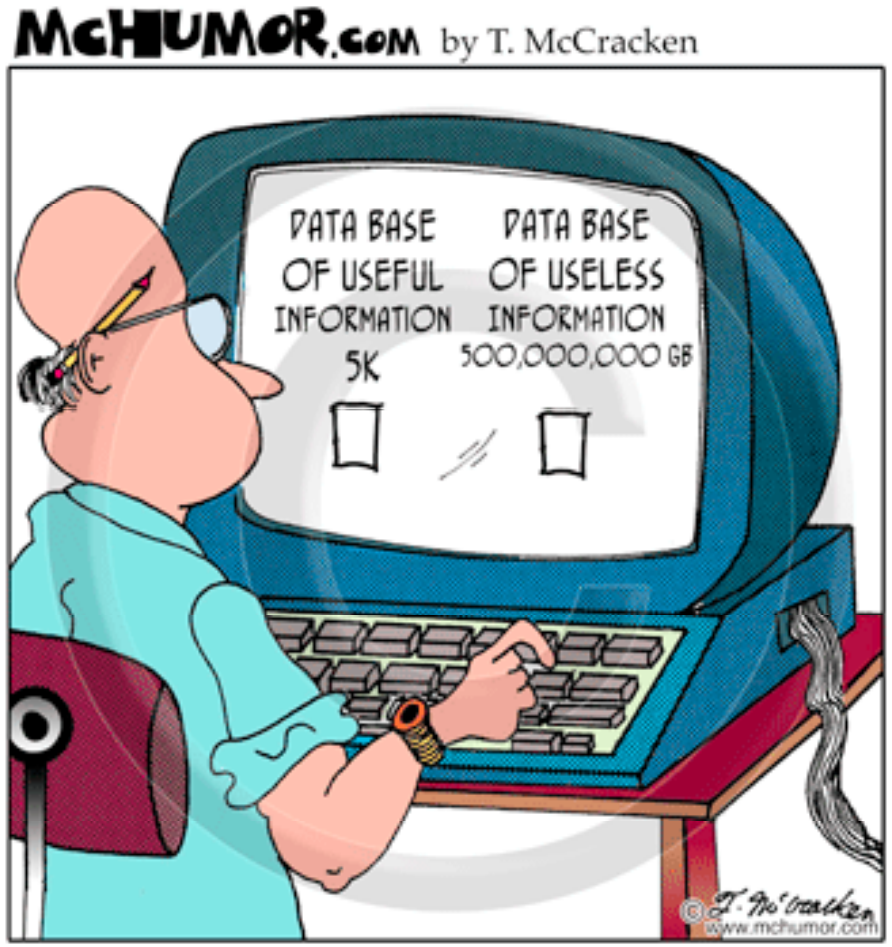
- Small fraction of digital universe has been explored for analytic value (especially in LTCI)

Source: IDC Analyst Perspectives: John Gantz and David Reinsel
The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East

Big-Data Challenges

- **Volume** – Infrastructure to process and store high volumes
- **Variety** – Designs to incorporate disparate sources, especially unstructured data
- **Velocity** – Collection mechanisms for high-velocity data
- **Veracity** – Expertise, creativity to design analytics and vet results

Is Big Data Useful for LTCI?



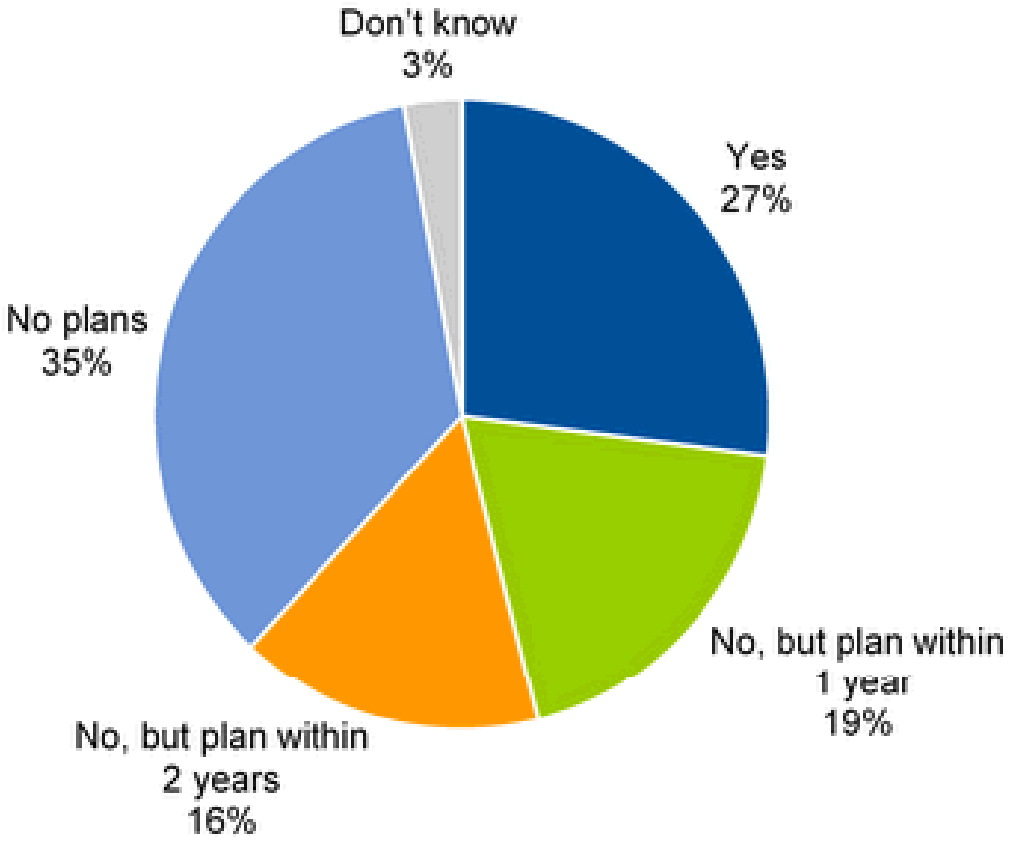
Value of Big Data

- **Low penetration so far** – 30% of organizations have invested in big data; only a quarter (8% of the total) have made it into production.
- **Big data investments in 2013 continue to rise** -- 64% of organizations investing or planning to invest in big-data technology (58% last year). Planned investments the next two years are highest for transportation, healthcare and **insurance.**
- **Enhanced customer experience** is the top big-data priority, with process efficiency close behind. Organizations struggle most with knowing how to get value from big data.
- **Big Data is touted as being about unconventional data sources** and the use of new and innovative technologies; this is not yet reflected in the chosen sources for first projects – transaction and log data still dominate the big data being analyzed.
- **Big-data technologies supplement** — but do not replace — existing information management and analytics. As a result, cloud adoption, with its supplementary nature, is the overriding technology that companies are using to derive value from big data.

Source: Gartner Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype; September 2013.

Big Data in Insurance

Has your organization already invested in technology specifically designed to address the big data challenge?



Source: Gartner (December 2013)

What Is Big Data in LTCI Context?

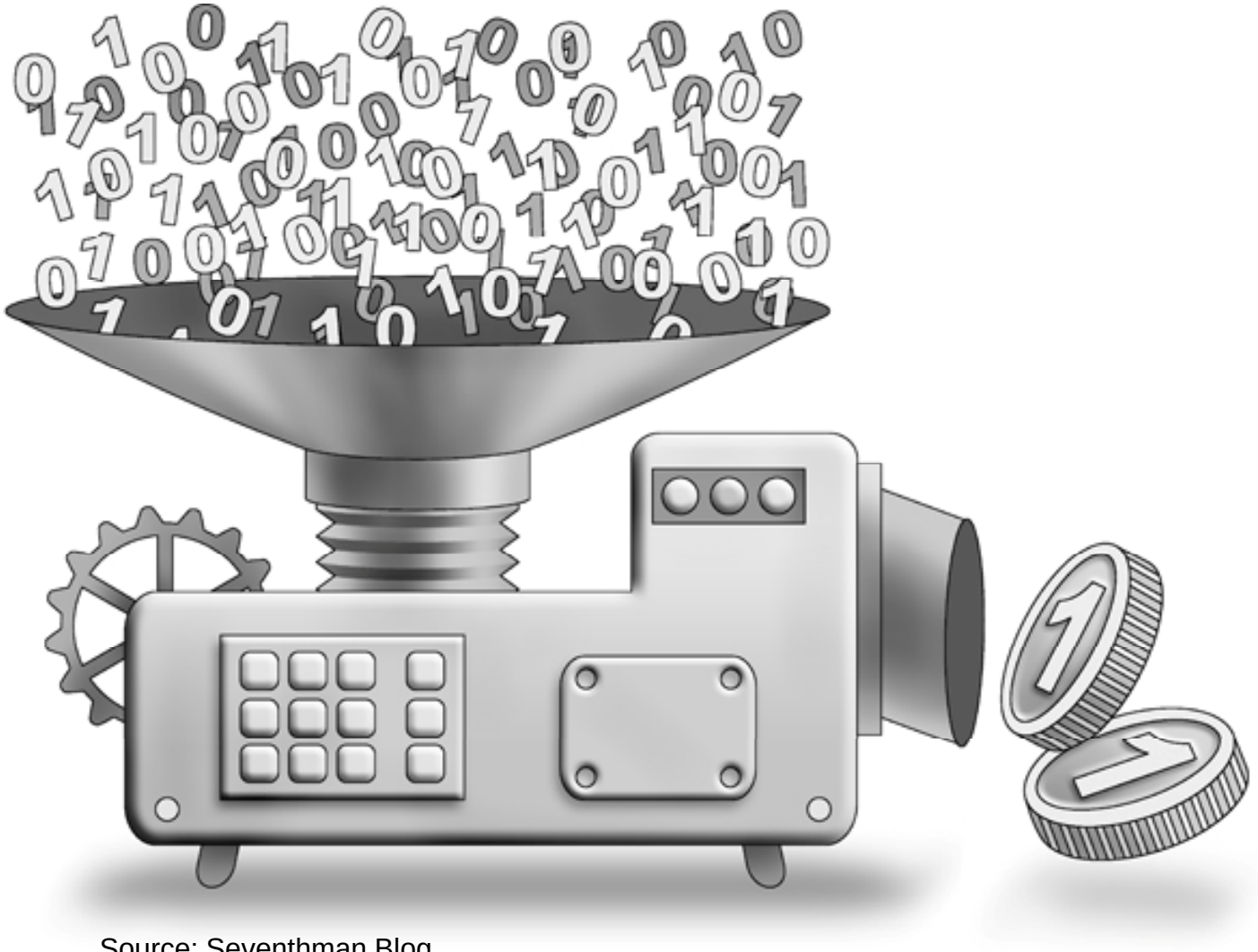
Taking clinical and unstructured observational information and connecting that with administrative/process information and social media

- Policyholders
  - Claim validation
  - Fraud detection
  - Marketing & sales
  - Underwriting
  - Predictive modeling
- Business Processes
- Providers
  - Provision of services
  - Observation
  - Electronic claims submission
- Market
  - Industry benchmarks
  - Trends and forecasts
- Healthcare convergence (integrated care delivery)

Additional Considerations for LTCI

- How to merge LTCI and healthcare data?
- How to combine unstructured, observational data with structured LTCI and medical data?
- A study of medical costs at end of life relative to LTCI coverage (CalPERS) – Dr. Stephen Holland presented on this yesterday – This is but one example of merging these two seemingly disparate data sources to improve our understanding of LTCI impact (to be published in *Population Health Management*)

Is There Value to be Gained?



Source: Seventhman Blog

How to Get Real

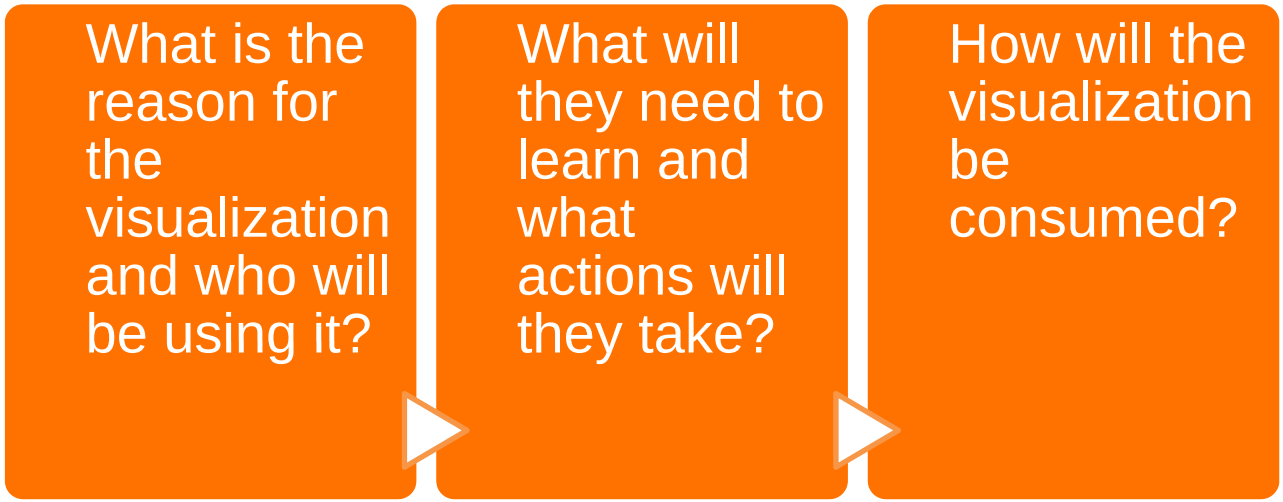
- Two-pronged approach
  - Define problem/need and value (top-down)
  - Study data to guide thinking (bottom-up)
- Start small and iterate!
- Ensure solid information-management platform and analytics
- Build upon traditional business-intelligence and analytic capabilities
- Explore ideas from outside the LTCI industry for interesting scenarios and applications

Dashboards for Business Intelligence

- Dashboards - Easy to read graphical representation of current status and historical trends of key performance indicators
- Actionable insight
- Static Dashboards
  - No analytic capability to explain results
  - Follow up is time consuming
- Interactive Dashboards

Creation of Interactive Dashboards

- Show only relevant, focused content
- Highlight interesting relationships in data
- Provide concise, relevant answers



LTC Interactive Dashboard Types

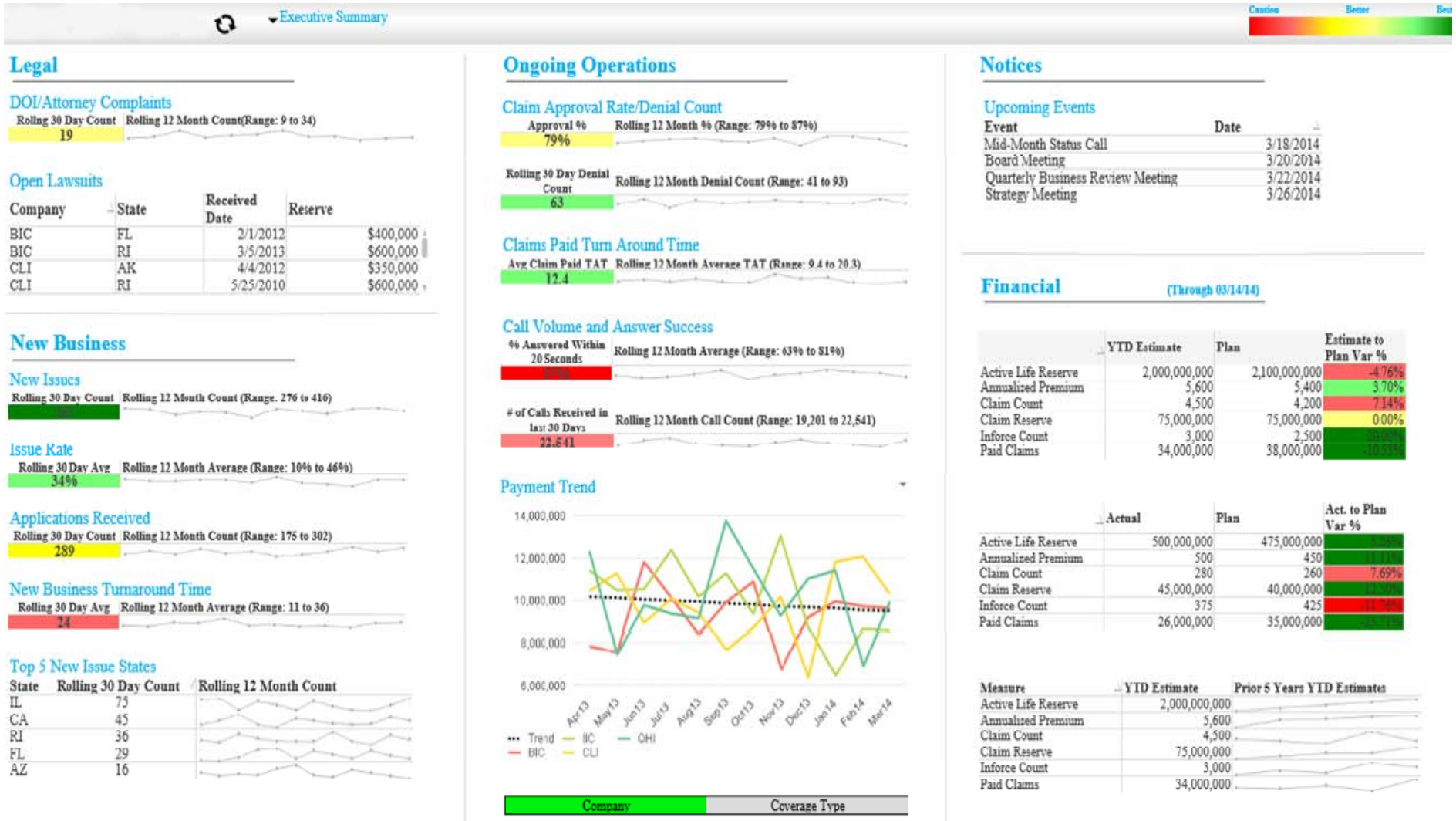
Executive Dashboards

Most relevant, actionable data at a glance for the executive team

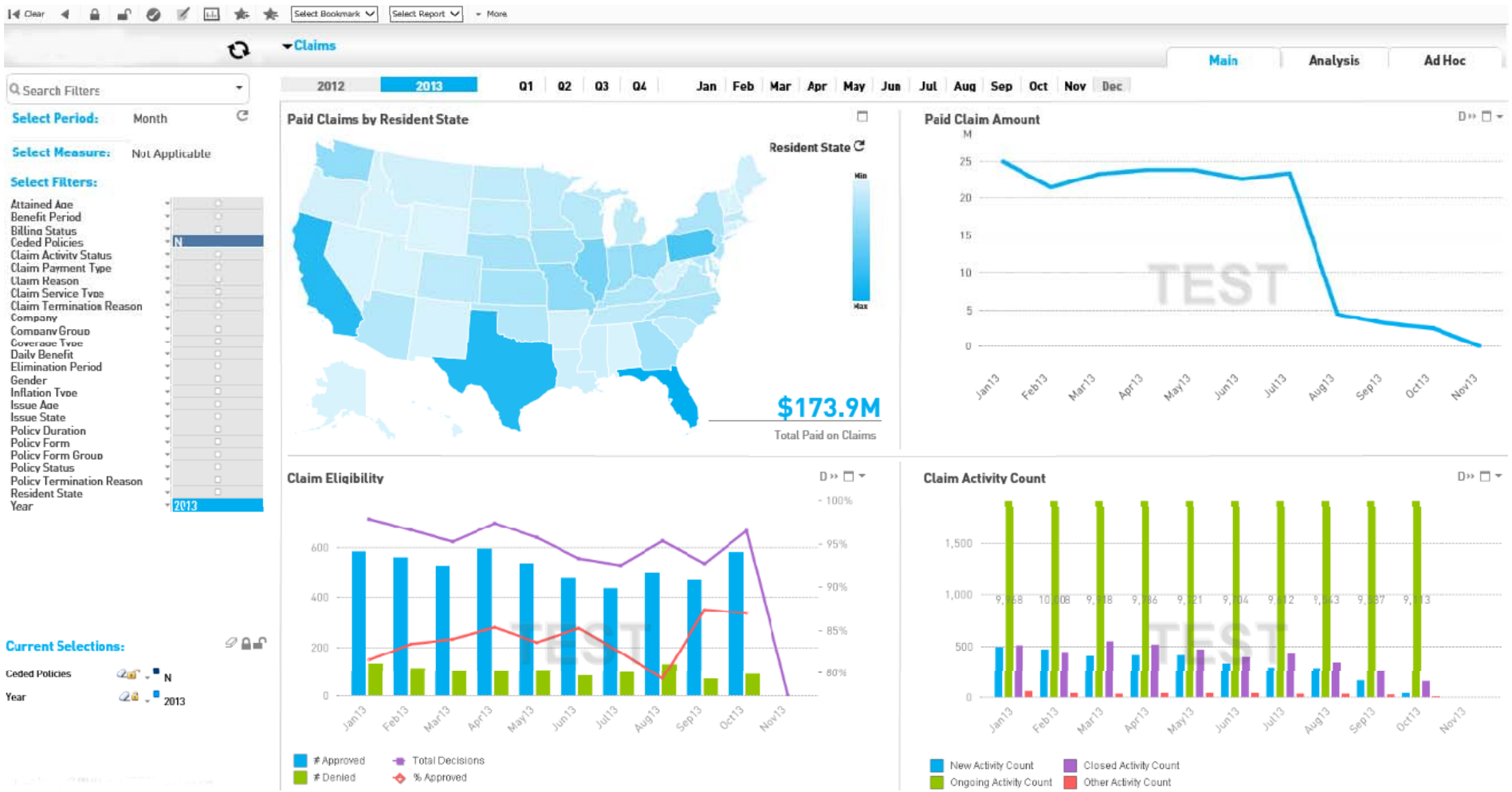
Financial and Actuarial	Legal and Regulatory	Ongoing Claims Operations	Ongoing Policyholder Operations	New Business
- Comparison to plan - Reforecasts - Loss analysis - Reserve metrics - Claim trends - Premium trends	- Litigation status - Complaints - Appeals	- Submission, approval and decision metrics - Cycle times - Service level metrics	- Member demographics - Billing metrics - Policy change metrics	- Underwriting metrics - Sales Analysis - Agent Analysis

Comparison to industry trends

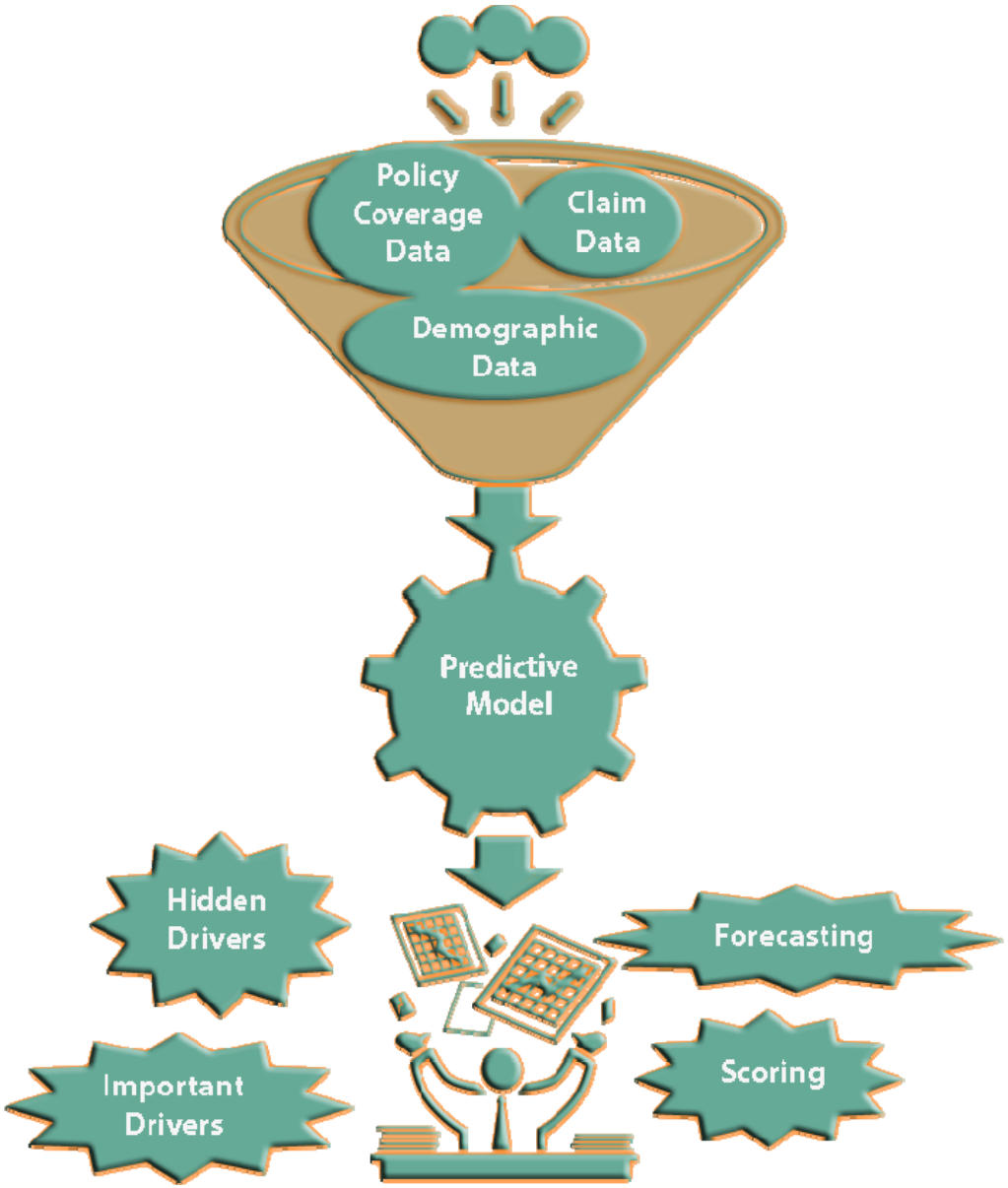
Executive Dashboard Sample



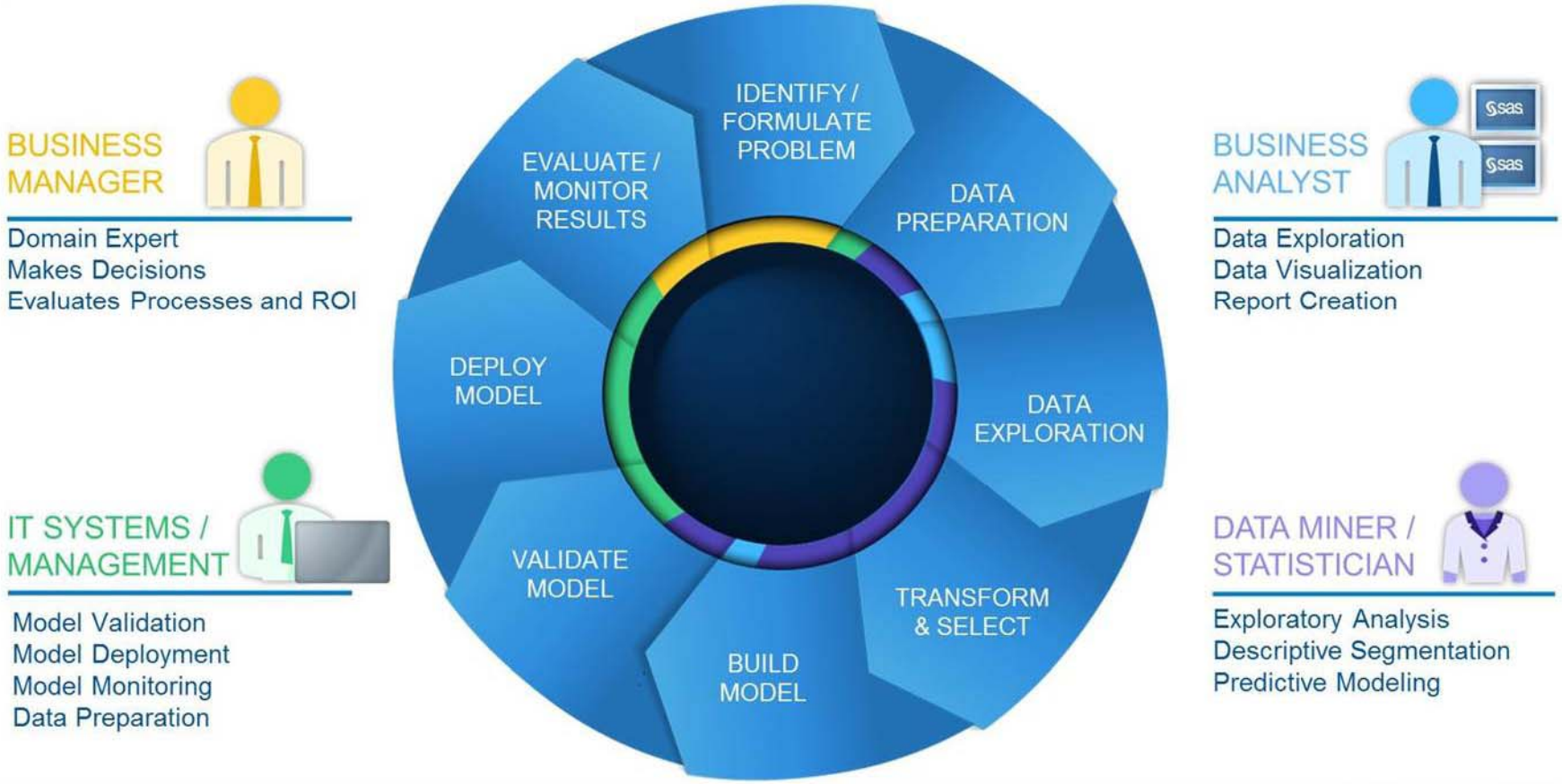
Claim Dashboard Example



Predictive Analytics – How it works

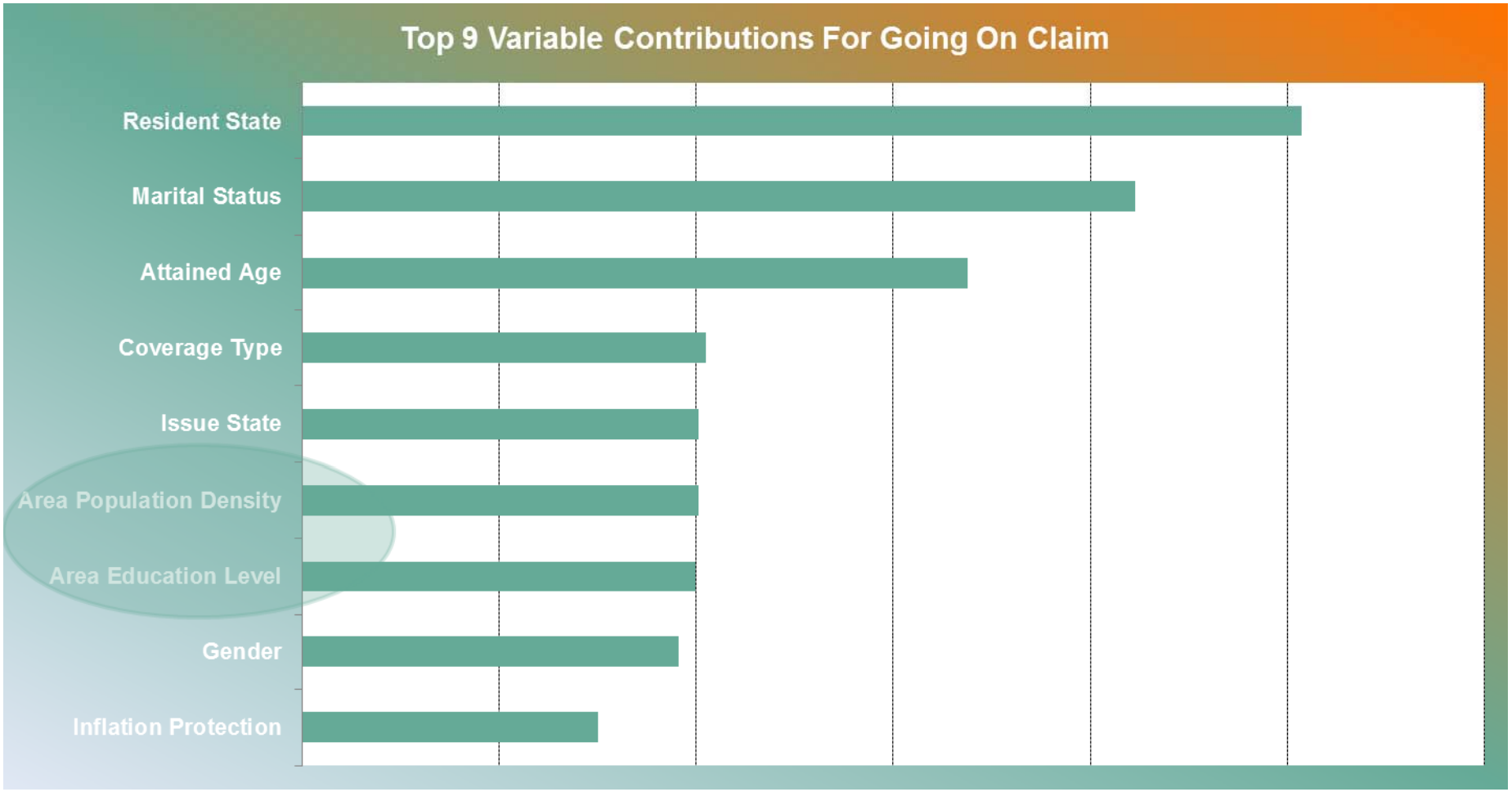


Predictive Analytics – How it works

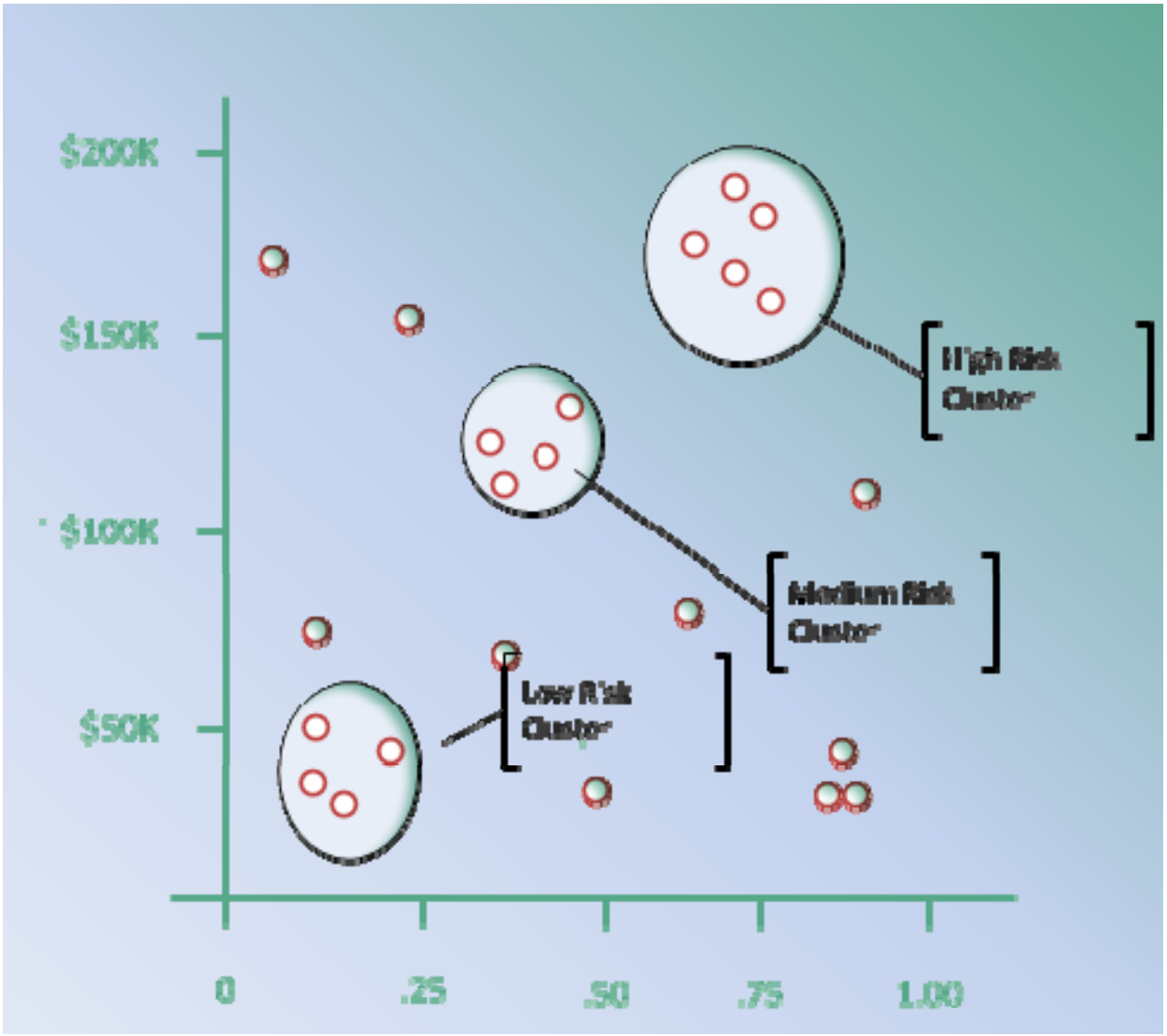


Source: http://blogs.sas.com/content/subconsciousmusings/2013/01/11/why-people-and-process-matter-in-addition-to-great-technology-in-predictive-analytics/

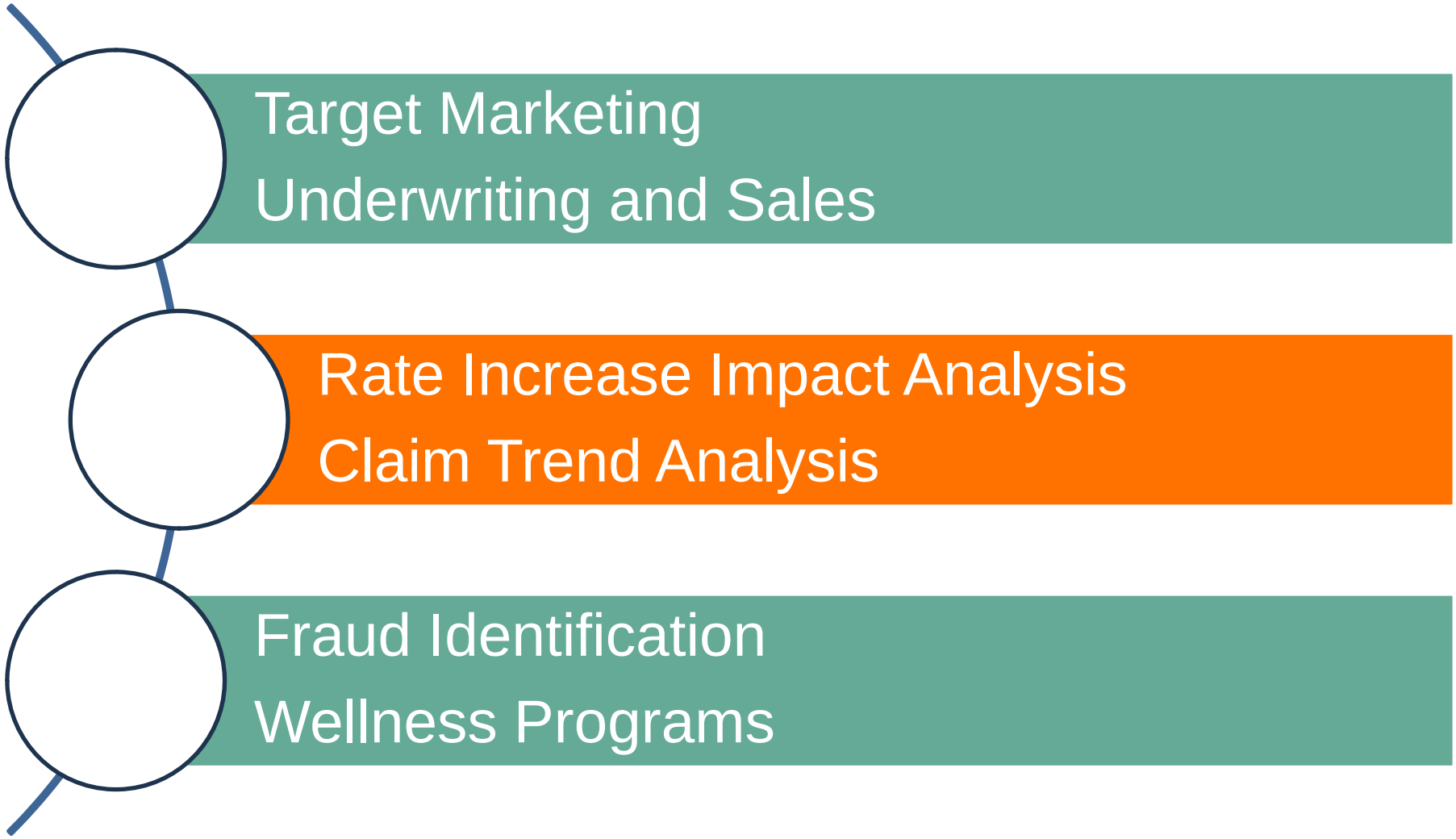
Predictive Analytics – Driver Identification



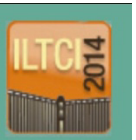
Predictive Analytics – Risk Management



Predictive Analytics - Model Types



Predictive Analytics – Fall Prevention Model



Select Model: Propensity to Fall

Select Filters:

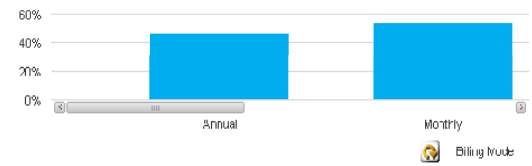
Billing Mode	
Billing Status	
Billing Type	
Ceded Policies	N
Coverage Type	
InflationType	
Issue Age	
Issue State	
Model Decision Reason 1	
Model Decision Reason 2	
Model Decision Reason 3	
Policy Form	
Policy Form Group	
Policy Status	
Policy Termination Reason	

Current Selections:

Top 30 At-Risk Policies — Show Top 10 20 30

Policy	Issue State	Issue Age	Model Score	Coverage Type
27	TN	60 - 64	98%	NH
49	TX	70 - 74	93%	NH
42	MO	65 - 69	92%	Shell
6	MI	60 - 64	91%	Shell
25	TX	60 - 64	90%	Comp-100
3	FL	60 - 64	89%	Comp-50
4	FL	55 - 59	89%	Shell
17	TX	0 - 54	89%	HHC
22	NC	65 - 69	87%	Comp-100
23	IN	0 - 54	87%	Comp-50
8	MI	60 - 64	82%	Comp-100
29	FL	75+	78%	HHC
41	VA	65 - 69	77%	NH

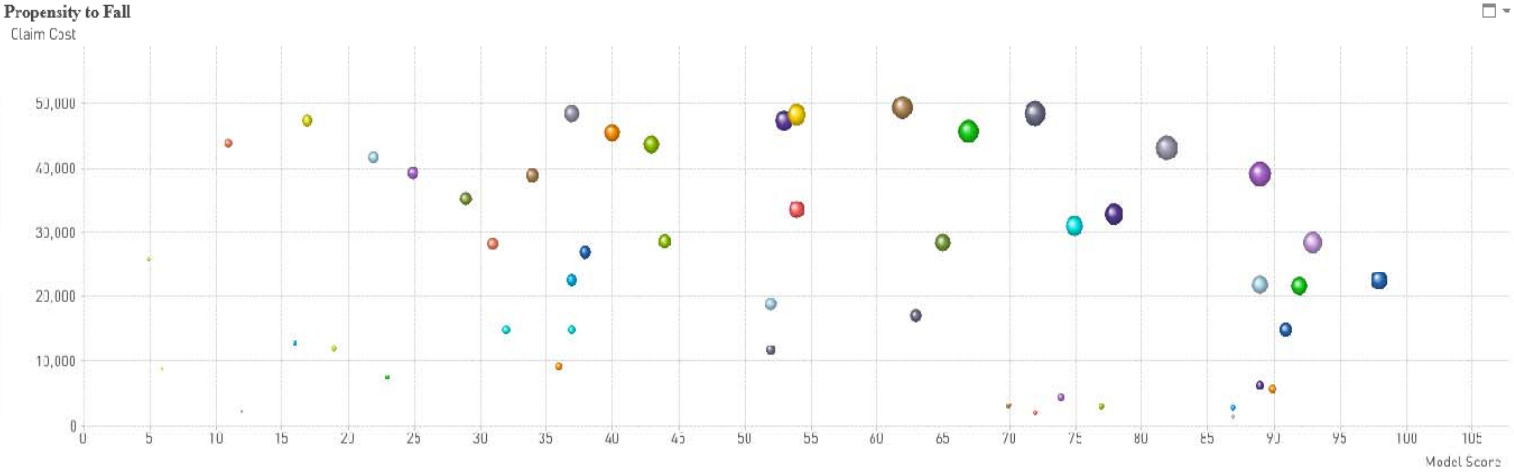
Avg. Propensity to Fall Score by Billing Mode



Propensity to Fall Details

Billing Mode	Billing Status	Billing Status Value	PolicyNumber	Percentage Score	Billing Type	Ceded Policies	Claim Cost	Coverage Type	InflationType
Semi-Annual	PPP	APP	27	98%	DRB	N	22,303	NH	Compound
Monthly	Term	TMV	49	93%	EFT	N	28,270	NH	Simple
Not Applicable	Term	TMD	42	92%	Not Applicable	N	21,367	Shell	Compound
Monthly	PPP	APP	6	91%	EFT	N	14,643	Shell	Compound
Monthly	PPP	APP	25	90%	EFT	N	5,553	Comp-100	None
Monthly	PPP	APP	3	89%	EFT	N	21,698	Comp-50	Compound
Monthly	CLW	CLW	4	89%	DRB	N	38,873	Shell	Compound
Not Applicable	Term	NTO	17	89%	DRB	N	6,113	HHC	None
Monthly	PPP	APP	22	87%	EFT	N	2,735	Comp-100	Compound
Not Applicable	Term	NTO	23	87%	DRB	N	1,405	Comp-50	None
Monthly	PPP	APP	8	82%	EFT	N	43,006	Comp-100	Compound
Monthly	Term	TMV	29	78%	DRB	N	32,704	HHC	None
Monthly	PPP	APP	41	77%	EFT	N	2,762	NH	Compound
Monthly	CLW	CLW	40	75%	DRB	N	30,877	Shell	Compound
Semi-Annual	PPP	APP	34	74%	EFT	N	4,250	Comp-50	Compound
Monthly	PPP	APP	19	72%	EFT	N	1,886	Comp-50	Compound
Monthly	Term	TMV	28	72%	DRB	N	48,222	NH	None
Monthly	PPP	APP	47	70%	EFT	N	3,017	Comp-50	None
Not Applicable	Term	NTO	16	67%	DRB	N	45,464	Comp-100	None
Monthly	PPP	APP	36	65%	EFT	N	28,330	NH	Compound
Annual	PPP	PPP	32	63%	DRB	N	16,807	Shell	Compound
Monthly	PPP	APP	11	62%	DRB	N	49,085	HHC	None
Monthly	NFO	NFO	33	54%	DRB	N	33,533	Comp-100	None
Annual	PPP	PPP	15	54%	DRB	N	48,143	Shell	None
Monthly	PPP	APP	5	53%	EFT	N	47,193	Comp-100	None
Not Applicable	Term	NTO	39	52%	DRB	N	11,552	Comp-100	None
Annual	Term	TNP	30	52%	DRB	N	18,783	HHC	None

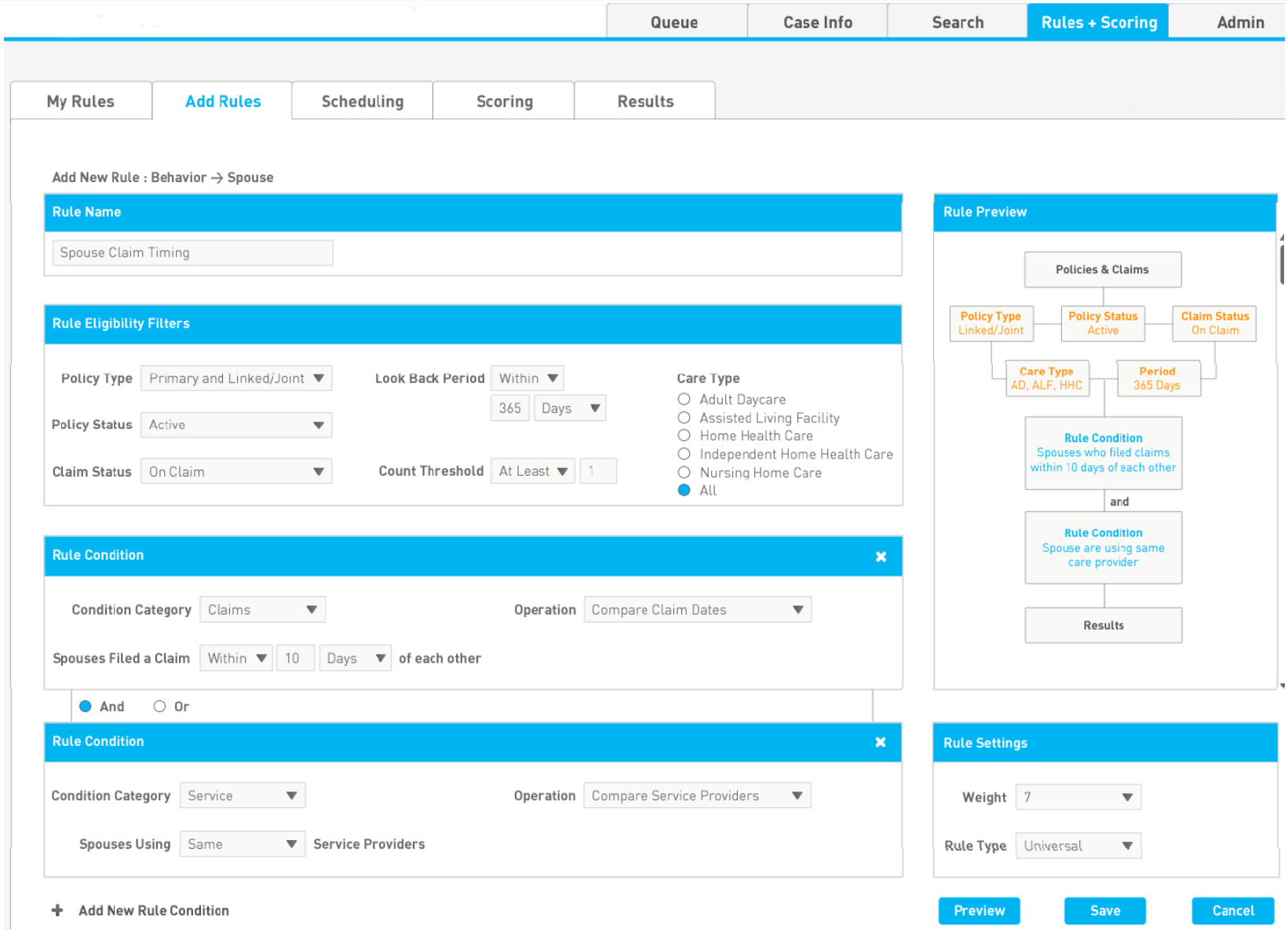
Propensity to Fall



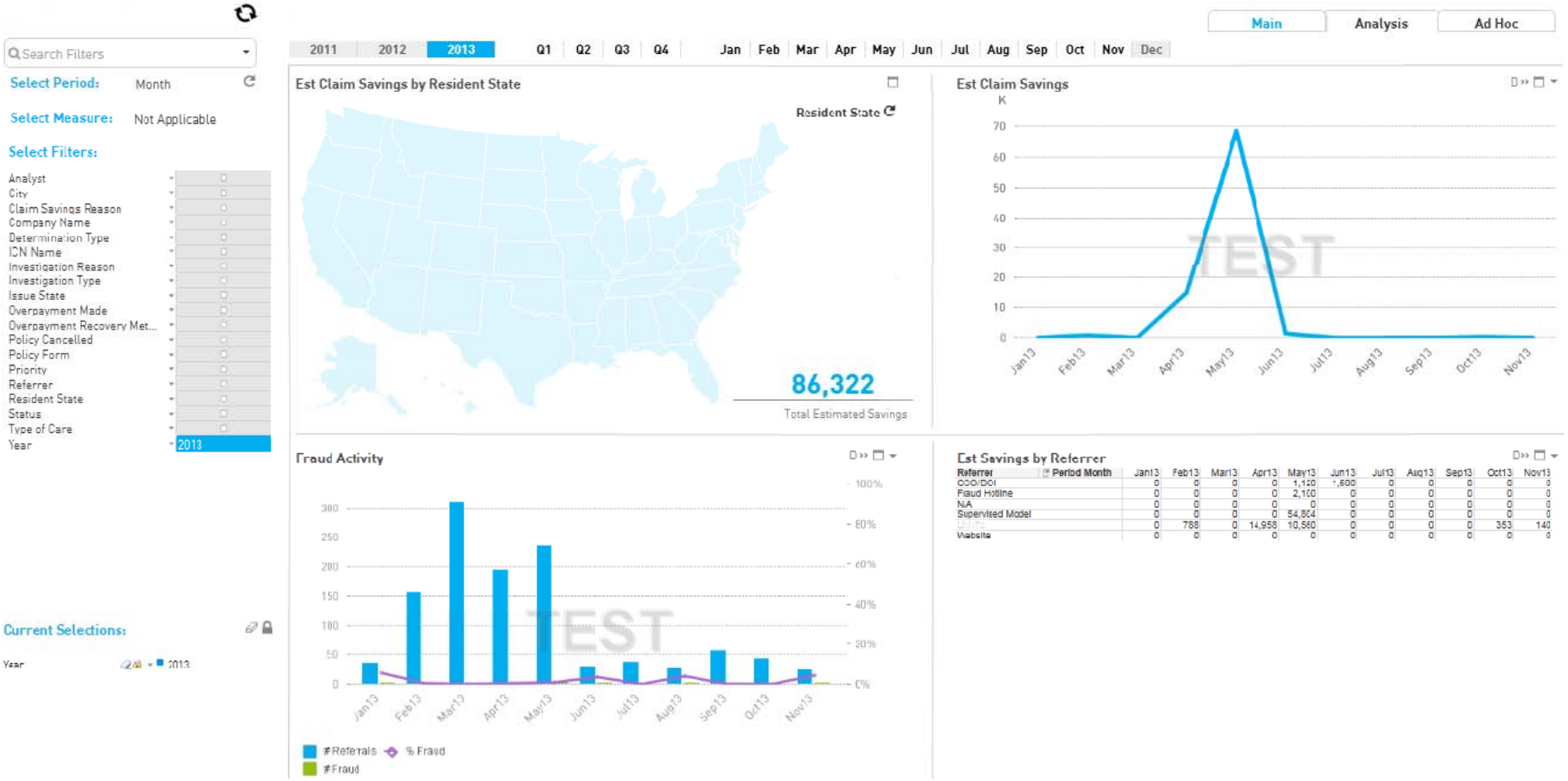
Fraud Detection - Overview

- Historically, a manual process
- Supervised models using rules based approach can improve results
- Iterative approach reduces false positives
- Prioritization of cases based upon fraud potential, risk, and recovery
- Identifies need for additional data
- Supplemental data
- Consortium models
  - Supervised
  - Unsupervised

Fraud Detection – Rules Engine



Fraud Detection - Reporting



Question and Answer from Audience

Any questions?