



Simplify, Perfect, Innovate

The Future of Big Data

(good, bad, or ugly ?)

16-FUTBIGDATA-1A

Mark J. Kiemele, Ph.D.
President and Co-Founder
Air Academy Associates

Office: 719-531-0777
Cell: 719-337-0357
mkiemele@airacad.com
www.airacad.com

Motivation

**“It’s tough to make predictions –
especially about the future.”**

Yogi Berra



Pew Internet Project Survey*

- **Non-random sample of 1,021 internet experts and other internet users**
- **53% of the survey respondents agreed with a positive scenario presented to them about Big Data in 2020.**
- **39% agreed with the negative scenario presented.**

* source: Data Science Central

Fact

Big Data is here and will continue to get (much) bigger.

- 90% of world's data was created in the last 2 years.*
- 80% of the data is completely unstructured.*
- Facebook receives 100 TB of data daily (and growing daily).*
- Twitter has 400 Million tweets each day (and growing daily).*
- Every hour Walmart controls 1 million customer transactions.*

* source: statspotting.com/big-data-stats

Fact

An Explosion of Data in the New Era of Computing



1.3 Zettabytes

of annual Internet traffic
by 2016



> 24 Petabytes

of data processed by
Google in a single day



150 Exabytes

of Healthcare data

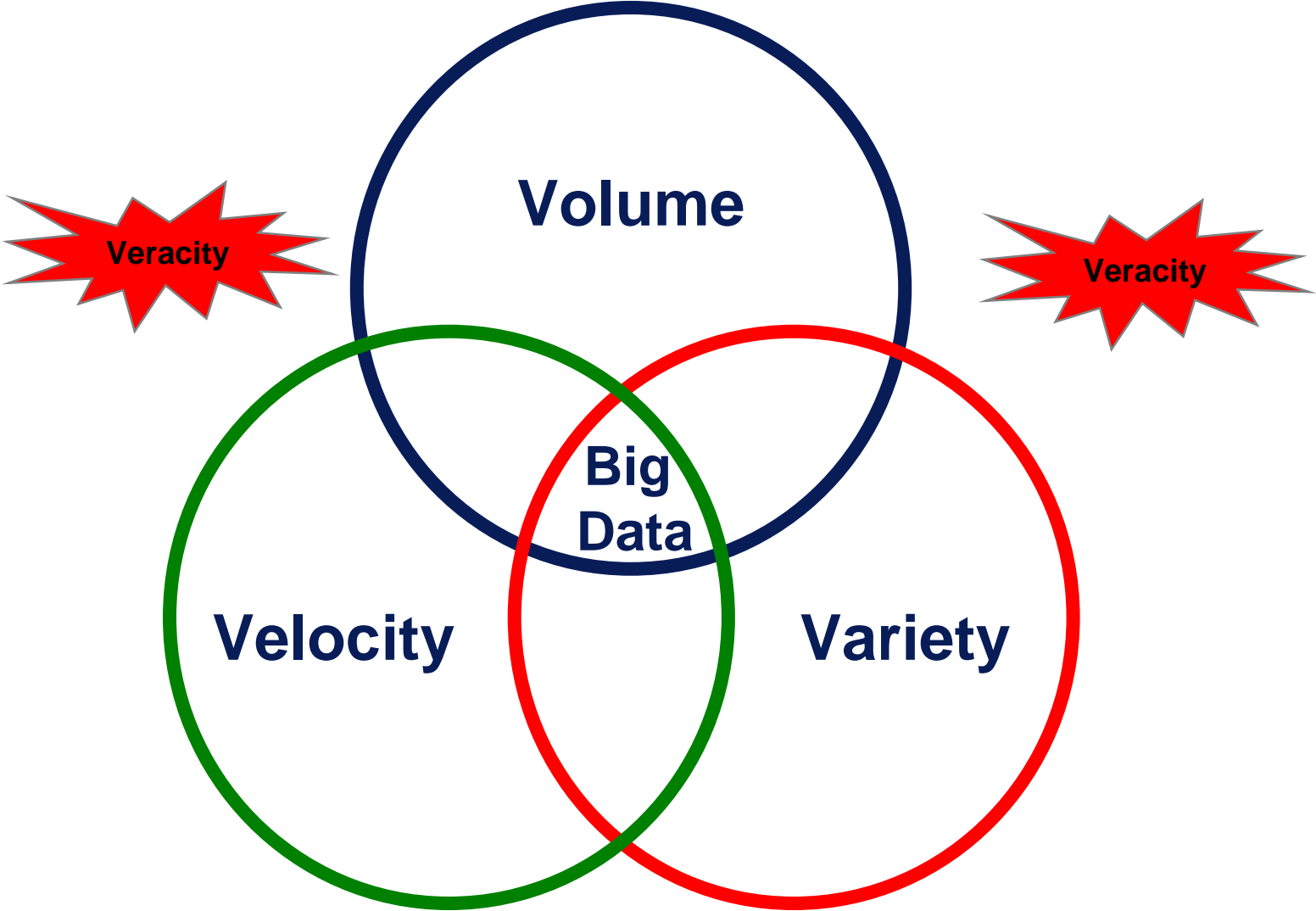


30 Petabytes

of data per day
transferred by AT&T
through its network

4V Definition of Big Data*

(*source: Doug Laney)



Definition

We have “Big” data when the volume/velocity/variety/veracity of the data becomes part of the problem we are trying to solve.

The Good

- Big Data is the “oil” of the future.
- It has provided the opportunity for many business to be successful and stay out of bankruptcy.
- Predictive analytics applied to Big Data can be a combination that provides extraordinary value.
- Better software and databases are on the rise to help.
- While a predictive analytics capability may be a competitive advantage today, it will be necessary for survival in the future.
- Provides great opportunity in every industry

Industries Where Predictive Models are Used

(Source: *Predictive Analytics* by Eric Siegel)

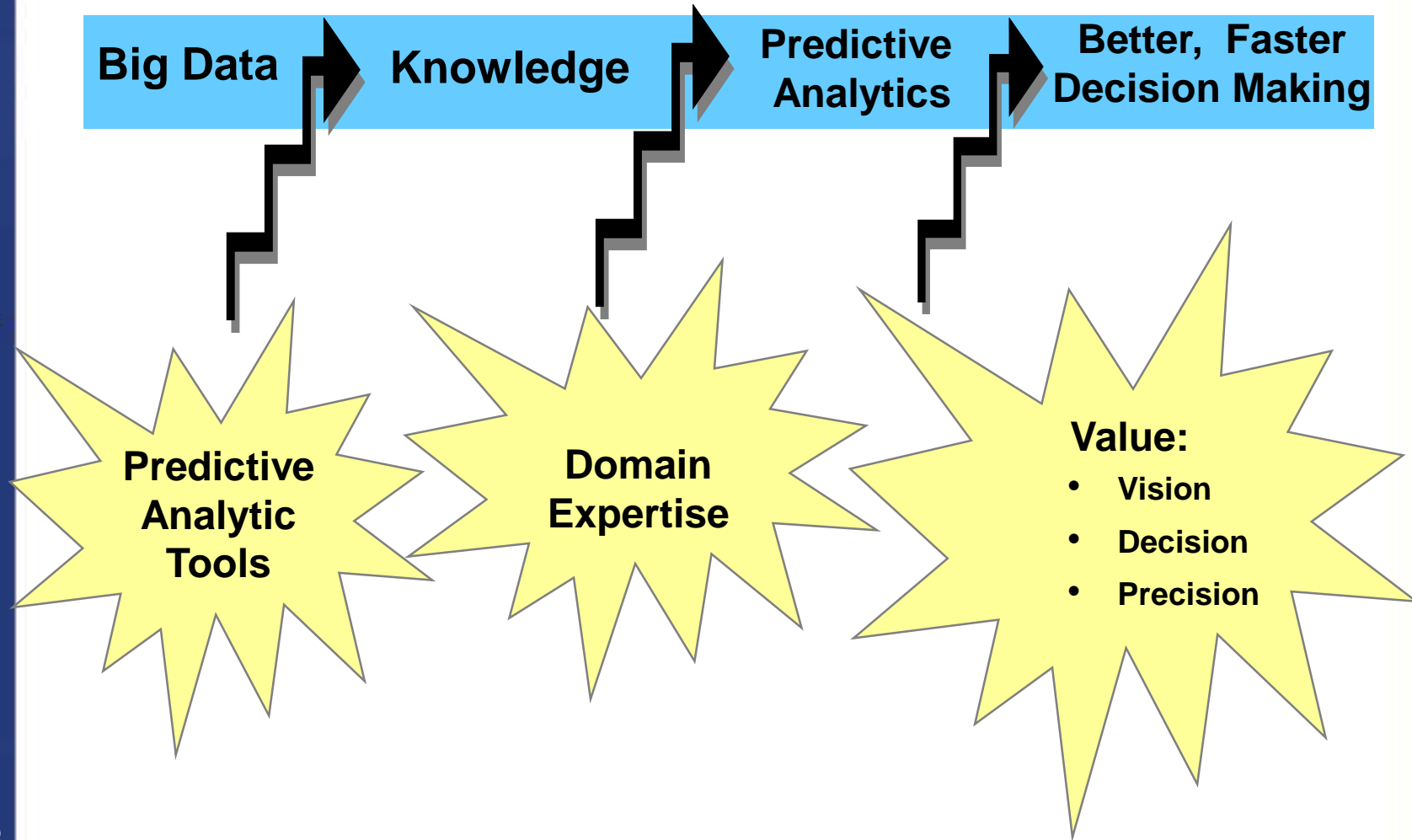
- Marketing, Advertising, and the Web
- Family and Personal Life
- Financial Risk and Insurance
- Healthcare, Medical, Pharmaceutical
- Crime Fighting and Fraud Detection
- Fault Detection for Safety and Efficiency
- Government, Politics, Nonprofit, and Education
- Human Language Understanding, Thought, and Psychology
- Staff and Employees – Human Resources
- Defense
- Engineering
- Oil, Gas, Energy

Examples of What is Being Predicted

(Source: *Predictive Analytics* by Eric Siegel)

- Likelihood of elderly dying within the next 18 months
- Target predicts customer pregnancy from shopping behavior
- Love (Match.com predicts interest in communicating)
- Breast cancer; premature births and other health related metrics
- Fraudulent tax returns, insurance and warranty claims
- Terrorist attacks
- Online activities that are malicious intrusions and attacks vs. legitimate activities
- Nuclear reactor failures (e.g., cracks in cooling pipes)
- Which voters are positively influenced by what type of contact and which voters will be affected adversely by contact
- Electricity demand growth to direct infrastructure development
- Lying and deception via eye movement and written statements
- Employee churn, job performance, and student dropouts

Big Data + Predictive Analytics = Value



The Value of Predictive Analytics Applied to Big Data

▪ Vision

- Provides the ability to identify trends, detect patterns, and uncover relationships between variables
- Lends direction in decision making

▪ Decision

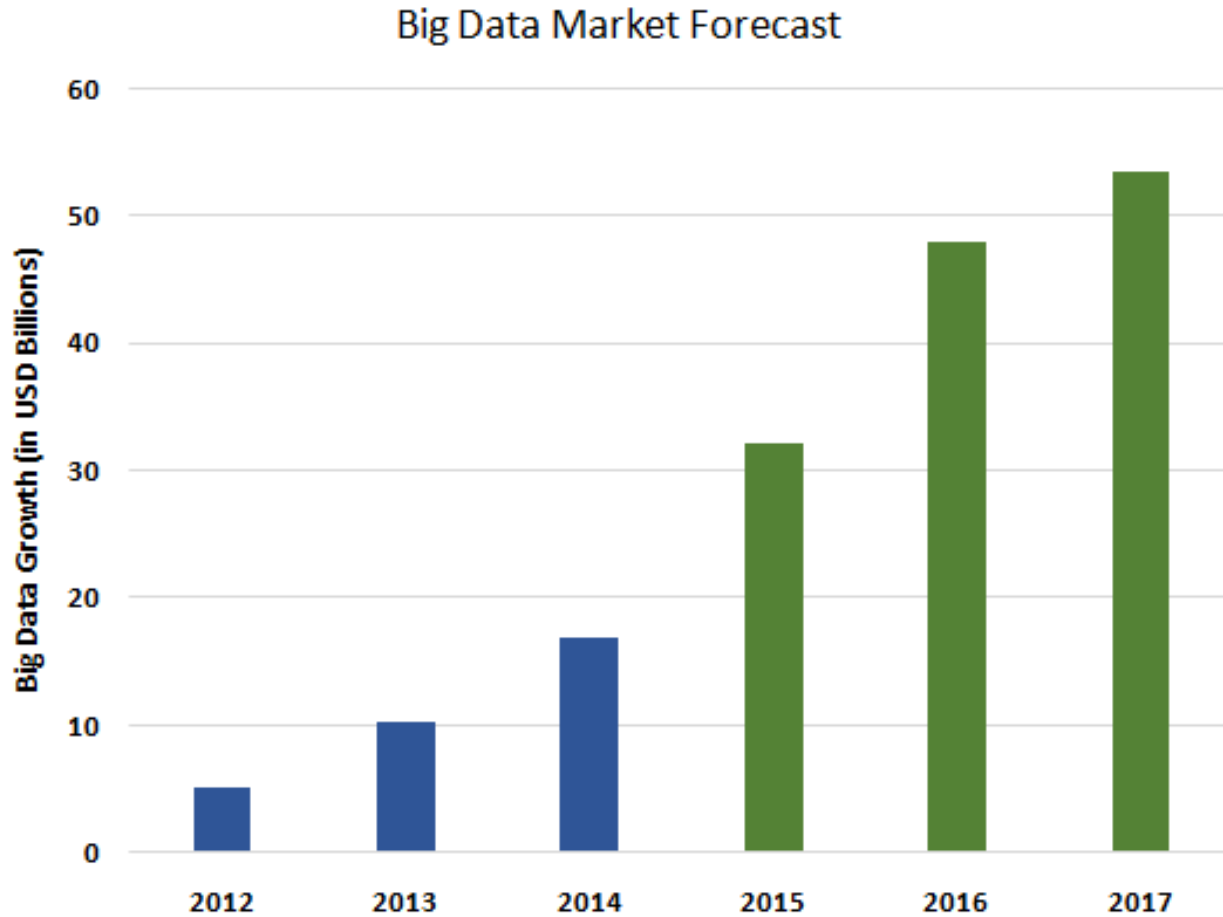
- Removes bias and emotion from decision making
- Provides consistent insights derived speedily to support the decision

▪ Precision

- Reduces risk in the decision making process
- Saves time and resources, reducing human error

Forecasting Big Data's Spending

(source: Gartner)

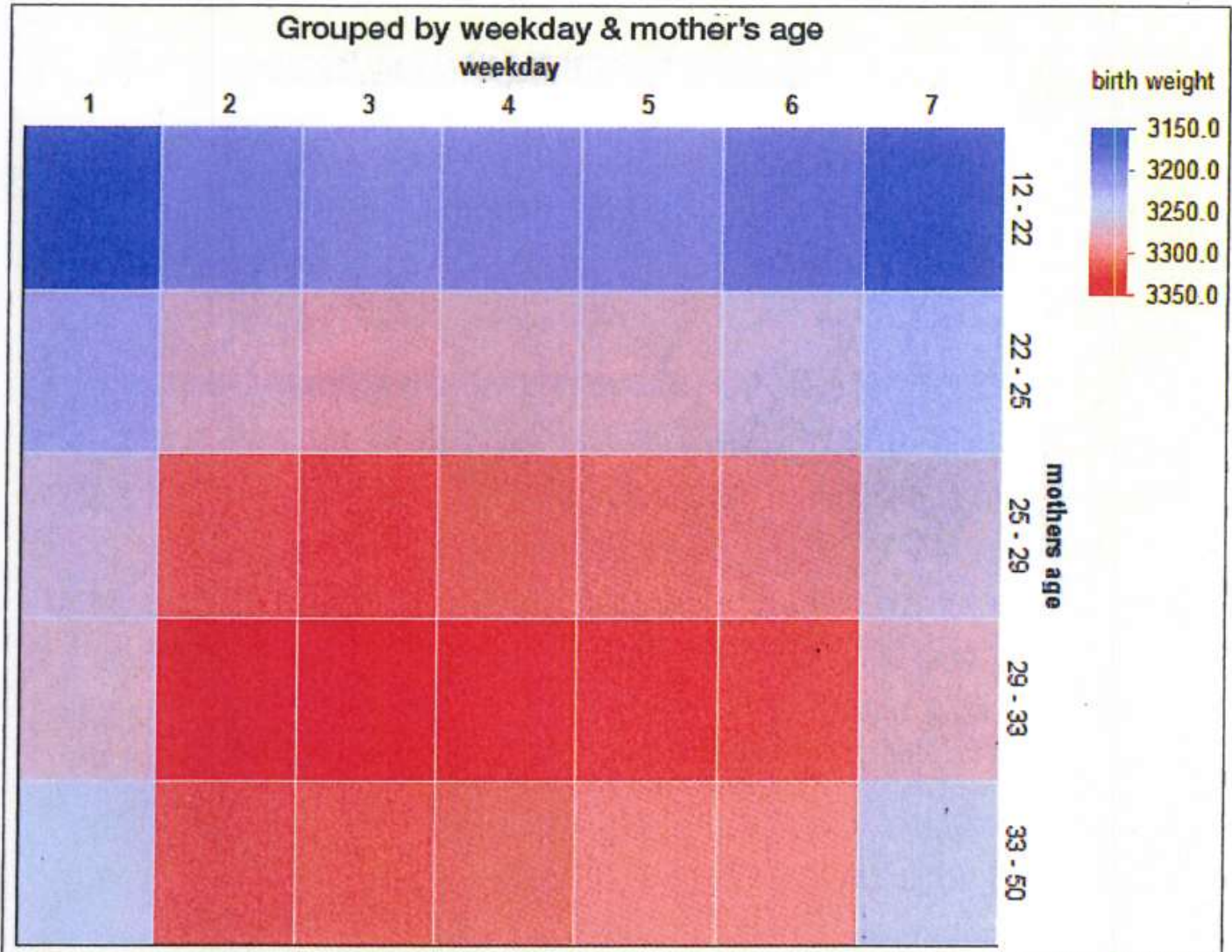


The Bad

- The myth that more data is always better.
- The myth that big data makes up for the lack of quality in the data.
- The low cost of data storage allows us to stash away troves of data, not knowing if we will ever need it or use it.
- 55% of Big Data projects have failed (Gartner).
- Your personal data footprint is large and is being sold as we speak.
- Anytime there is an opportunity to use a method for good, there will always be those who will use it for destructive/nefarious purposes.
- Predictive Analytics/Big Data tools are difficult to use.
- Predictive Analytics/Big Data will lead to an even greater divide between the rich and the poor (Gartner).
- There is much more noise in Big Data, making it more difficult to ferret out the true signal.
- While correlation may help us predict, correlation does not imply causality.

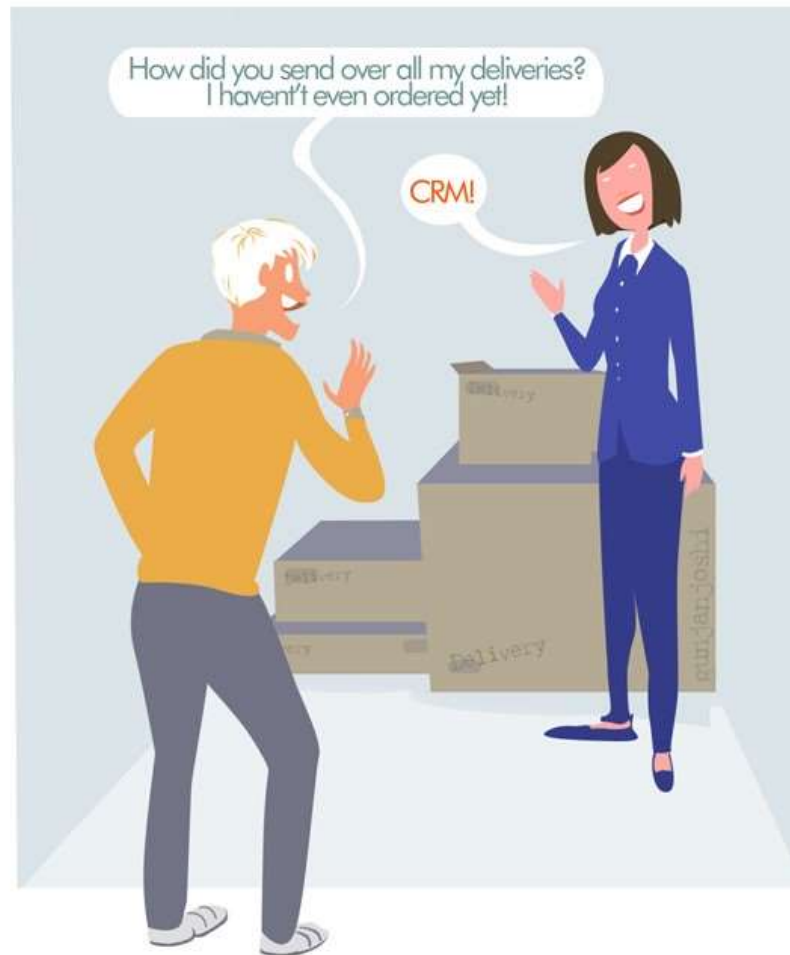
Example of a Predictive Analytics Heat Chart

(from more than 12 million records – these are average weights)



The Ugly

- Predictive Analytics with Big Data sometimes create a pathway where technology and privacy collide.



The Ugly

- Is Big Data Big Brother?
 - Privacy concerns
 - 51% of US 18-29 year olds prefer anti-terrorism surveillance to privacy.
 - 66% of US adults over 30 prefer anti-terrorism surveillance to privacy.
 - Everywhere we go, everything we touch is surveilled and contributes to our digital footprint.
 - Big Data never dies. It just keeps getting bigger.
 - We live in a world of scores.
 - **Job security score:** predicts future income and capacity to pay
 - **Churn score:** predicts when customers move their account
 - **Brand name medicine propensity score:** predicts if you will buy generics or brand name medications.
 - **Fraud score:** predicts if someone is not who they claim to be or is up to mischief.

Big Data Applied to SOS and NIE

- Big Data can be used to continuously monitor SOS and Network Integration throughout the lifecycle of a System.
- While we are familiar with DT&E and OT&E, we also need Continuous T&E (CT&E). That is, continuously collect data to prevent unwanted system states from occurring. This is where Big Data, along with the appropriate analytics, can help us.
- In complex System of systems, the factors or variables to be monitored or tested are no longer the individual components or parts of the System but rather the **interactions** of the entities within a complex System.
- To demonstrate the magnitude of number of interactions, let's take a simple scenario: suppose there are 10 entities that comprise a system.
 - How many two-way interactions are there? 45
 - How many three-way interactions are there? 120
 - How many four-way interactions are there? 210
- This will extend the concept of the scientific test and analysis tools that we use in DT&E and OT&E to a higher level of abstraction in CT&E.
- If implemented and used properly, Big Data analytics should be able to help us detect, predict, and ultimately prevent emergent behavior of a complex system.