



# Profiling using Data and Predictive Analytics

Presented by: Manoj Chiba

Date: 22 February 2018

Insurance Crime Bureau Conference

Z

*Data is ~~[becoming]~~ the ~~[new]~~  
raw material of business*

~ Craig Mundie – Modified ~

Z



*If we have data, let's look at the data. If all we have are opinion's let's go with mine*

*~ Jim Barksdale~*

Z

# Agenda

- Why analytics and What is data?
- Data scientist and the Fraud Data Scientist
- Predictive Analytics
- What predictive analytics means for Fraud
- Use cases
- Fraud analytics process model



# Why Analytics and What is data?



# Why Analytics?

If the current rate of change and complexity were to remain constant, we would have experienced all the major milestones of the twentieth century – **in a single week in 2025!**

1. The creation of the automobile;
2. The first and second world war AND the Vietnam war;
3. Decoding of the DNA structure;
4. Nuclear energy;
5. Space travel;
6. The internet; and
7. Human genome sequencing

The challenge for organizations is: How to navigate this, build strategies that identify trends of the future: Analytics is postulated to be the answer! = IDENTIFICATION OF TRENDS: PRESENT AND FUTURE TRENDS



# Consider the following in a single day... online

1. Enough information is consumed to fill  $\pm 168$  Million DVDs
2.  $\pm 294$  Billion emails are sent
3.  $\pm 2$  Million blog posts are written
4.  $\pm 4.7$  Million minutes are spent on Facebook
5.  $\pm 864,000$  hours of video are uploaded on YouTube



For any analytics we need data...  
So what is data?

z



What many think data is...

**Gobbledygook**

*noun informal*

Language that is meaningless or is made unintelligible  
by excessive use of abstruse technical terms,  
nonsense

Synonyms: gibberish, claptrap, nonsense, balderdash, blather,  
garbage

# The often forgotten data

## Mining Text Data

---

*“The first forty years of life give us the text; the next thirty supply the commentary on it.”—Arthur Schopenhauer*

### 13.1 Introduction

---

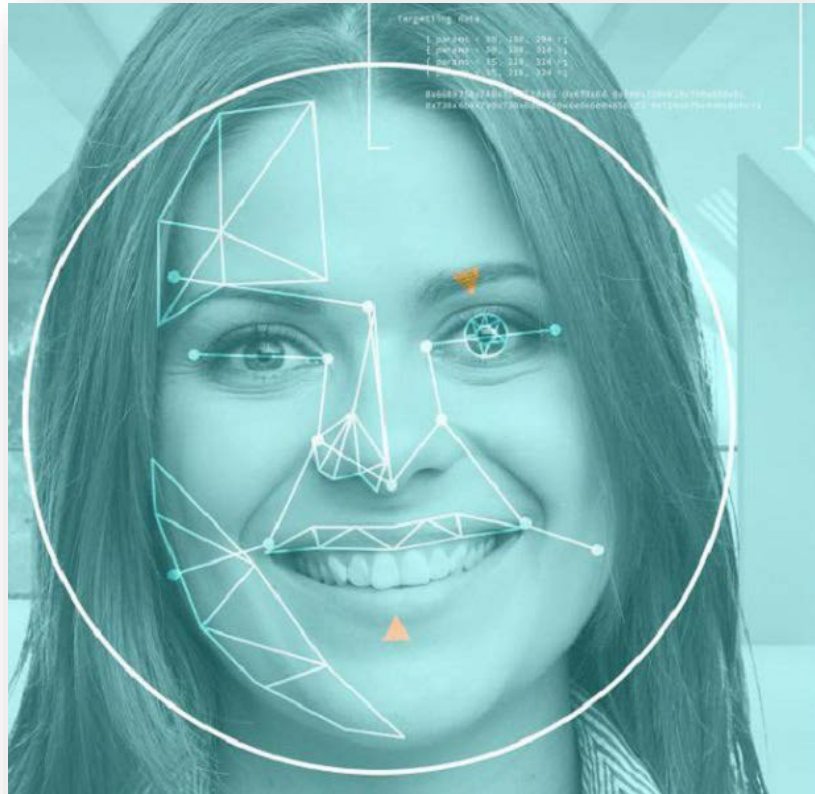
Text data are copiously found in many domains, such as the Web, social networks, newswire services, and libraries. With the increasing ease in archival of human speech and expression, the volume of text data will only increase over time. This trend is reinforced by the increasing digitization of libraries and the ubiquity of the Web and social networks. Some examples of relevant domains are as follows:

1. *Digital libraries:* A recent trend in article and book production is to rely on digitized versions, rather than hard copies. This has led to the proliferation of digital libraries in which effective document management becomes crucial. Furthermore mining tools are also used in some domains, such as biomedical literature, to glean useful insights.
2. *Web and Web-enabled applications:* The Web is a vast repository of documents that is further enriched with links and other types of side information. Web documents are also referred to as *hypertext*. The additional side information available with hypertext can be useful in the knowledge discovery process. In addition, many web-enabled applications, such as social networks, chat boards, and bulletin boards, are a significant source of text for analysis.
3. *Newswire services:* An increasing trend in recent years has been the de-emphasis of printed newspapers and a move toward electronic news dissemination. This trend creates a massive stream of news documents that can be analyzed for important events and insights.

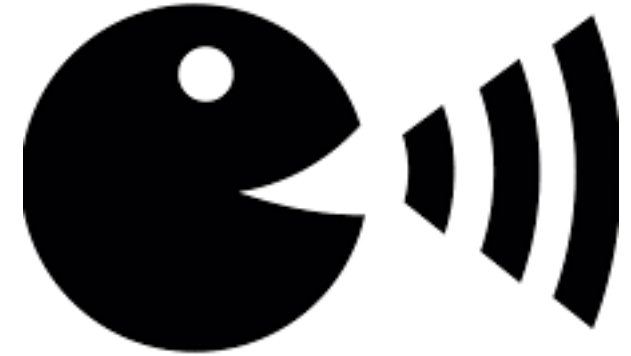




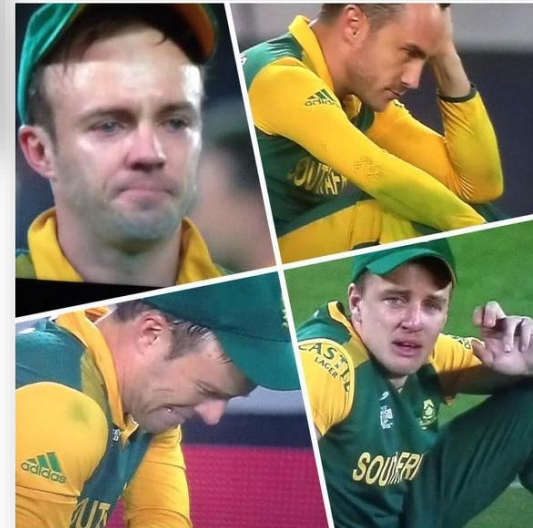
# The most often “forgotten” data



Macro-economic influences



Emotional Importance



Z

# Terminology

- Unstructured Data

Data that has no identifiable structure – for example, the text of email messages

- Structured Data

Data that is organised by a predetermined structure.

# The data problem?



Data exists but the problem is:

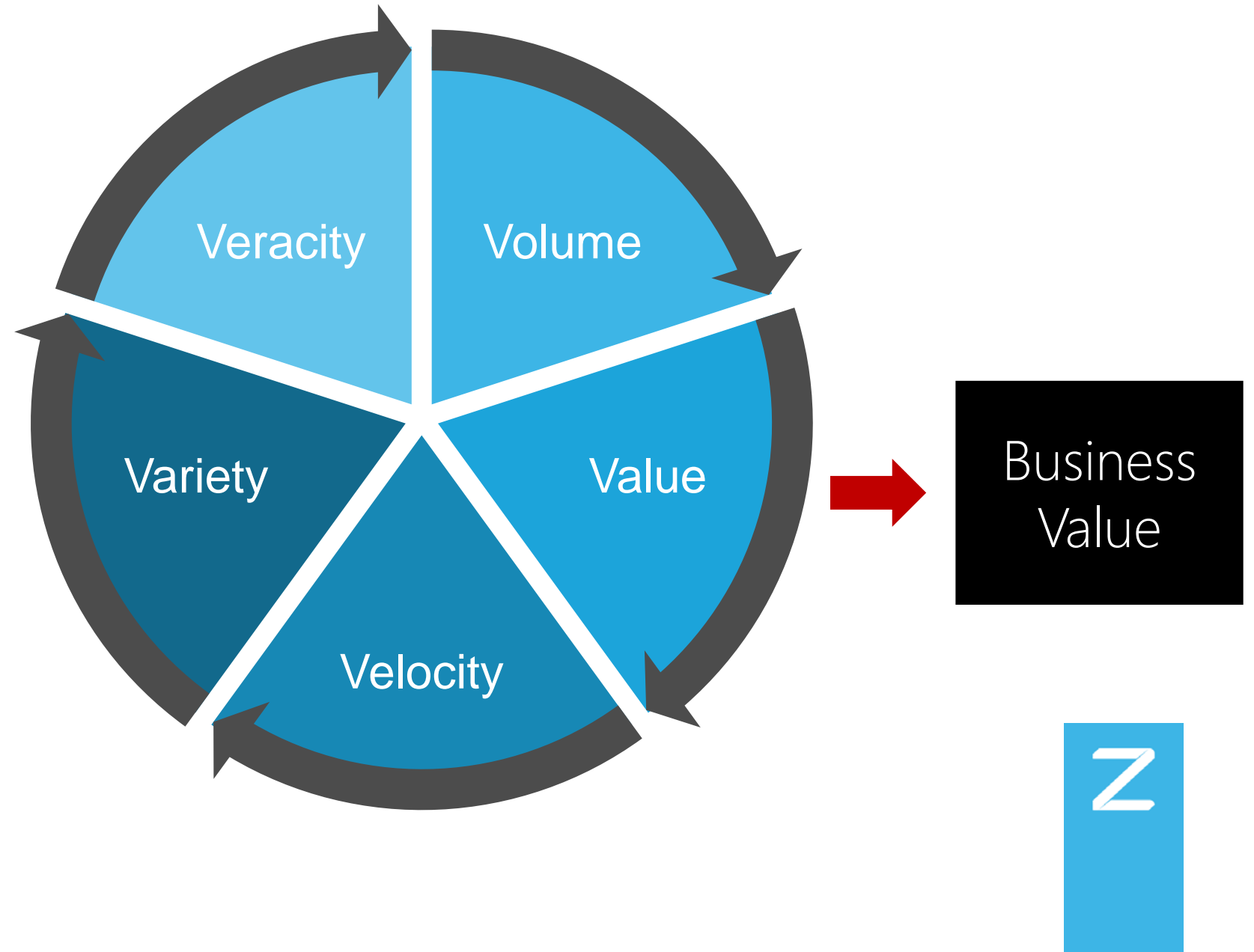
- Data Mining
- Data Analysis Skills
- Understanding **what it means for my business**

The question shifts from  
**what do we think,**  
to **what do we know?**



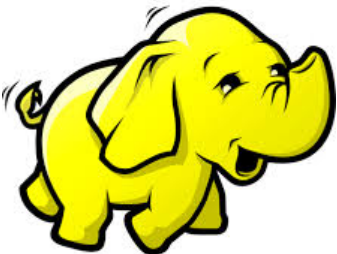
# BIG data

While “size” of data is traditionally the hallmark of big data, the term is poor, and may be better rooted in an understanding that Big Data is about capacity to **SEARCH, AGGREGATE** and **CROSS-REFERENCE** data sets





# But where are we???

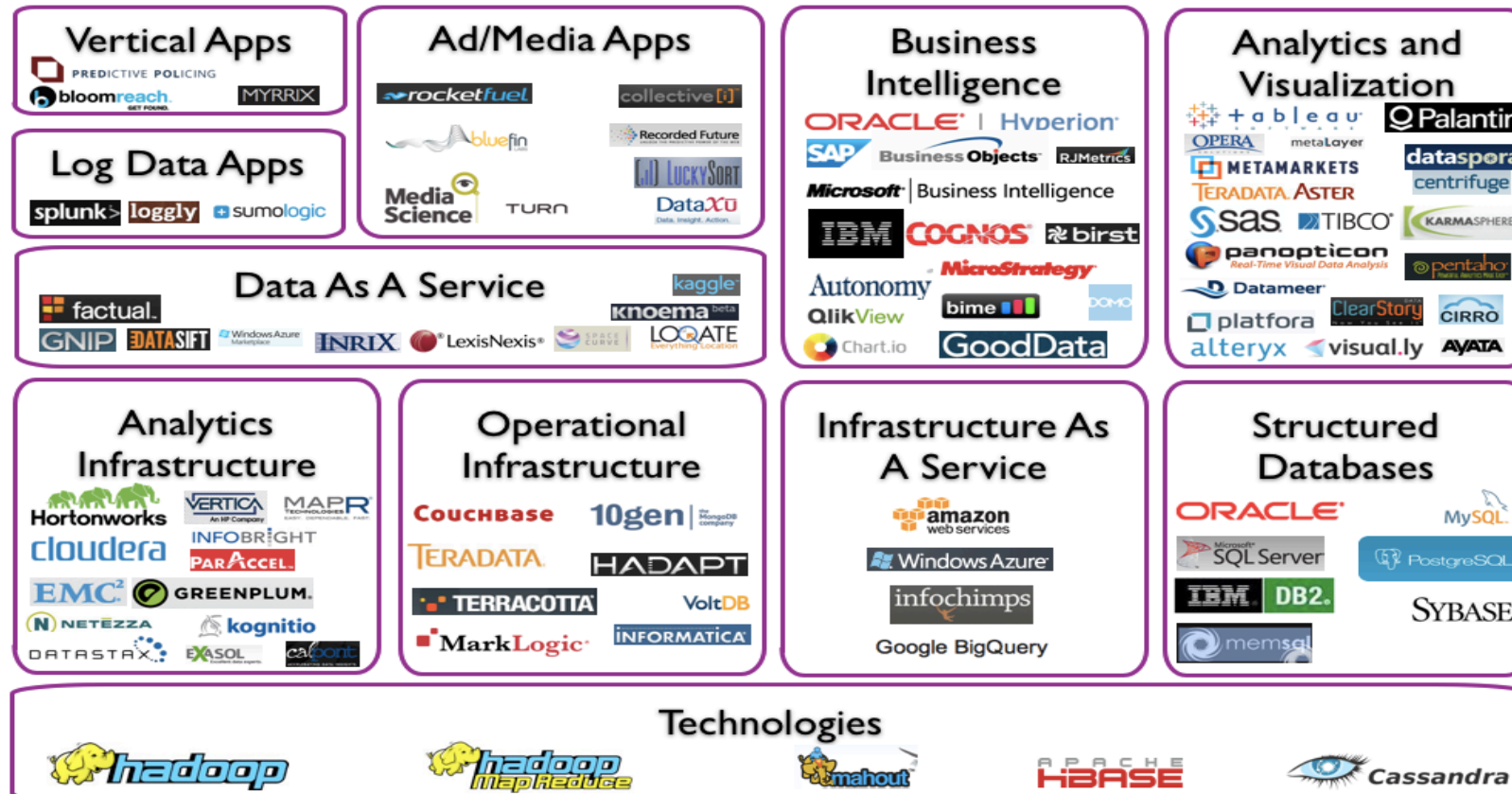


Hype of BD



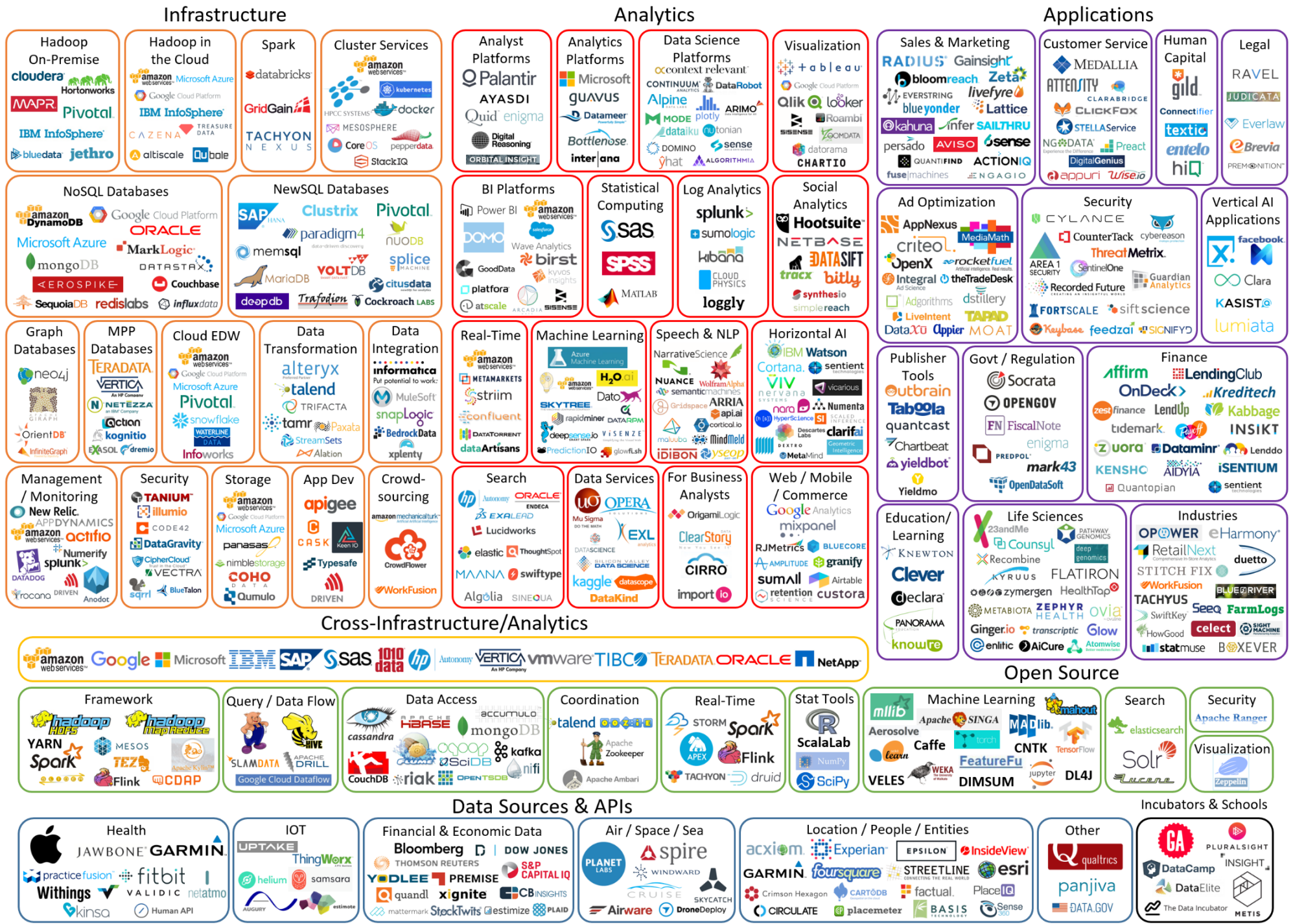
# How offerings have changed (2012)

## Big Data Landscape



# How the offerings have changed (2016)

Courtesy of Firstmark





Courtesy of Firstmark

[illegible]

# What does this mean?

- Maturity has been reached....
- Trend in:
- From Infrastructure (Developers and Engineers) to Analytics (Data Scientists and Analysts)
- From Analytics (Data Scientists and Analysts) to Application (Business users and consumers)- **In our context Fraud Detection!**



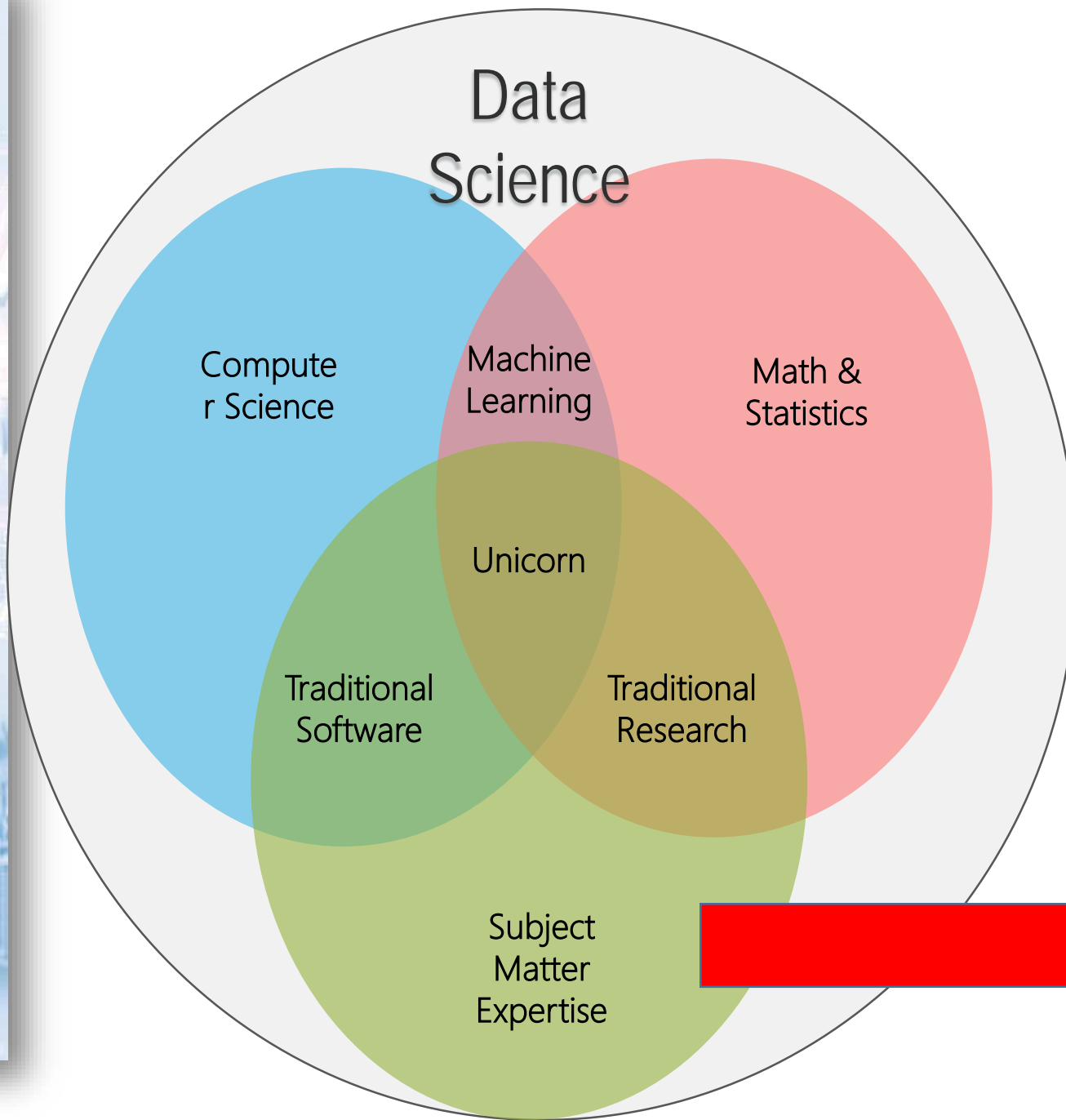
STOP! Who? The data scientist?

Z

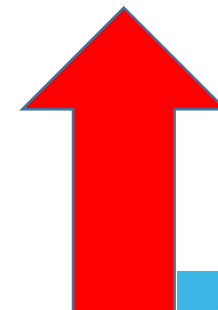


# DATA SCIENTIST

THE SEXIEST JOB OF THE 21ST CENTURY



Interesting data  
versus Actionable



Z

# Interesting vs Actionable

## Interesting:

- Nice to know
- Does **NOT** help you make informed decisions
- Does **NOT** provide insight: Why should we care?

## Actionable:

- Insights > Action
- Design Programmes
- Develop strategies
- Achieve goals

# Simple example of actionable data: My Fitbit

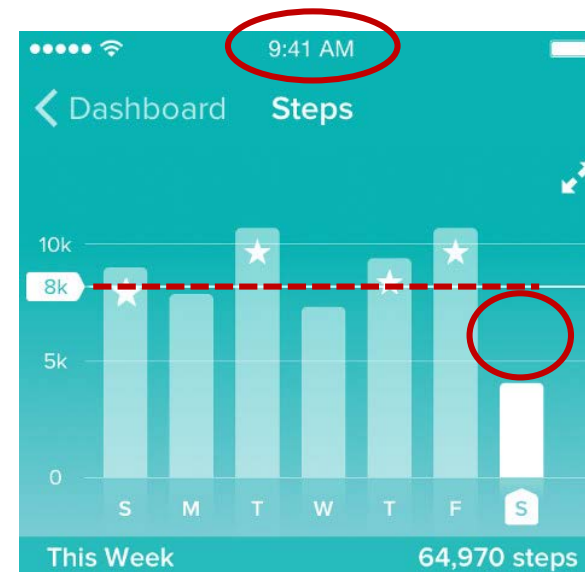
DATA → INFORMATION → INSIGHT → ACTION



2,259 steps



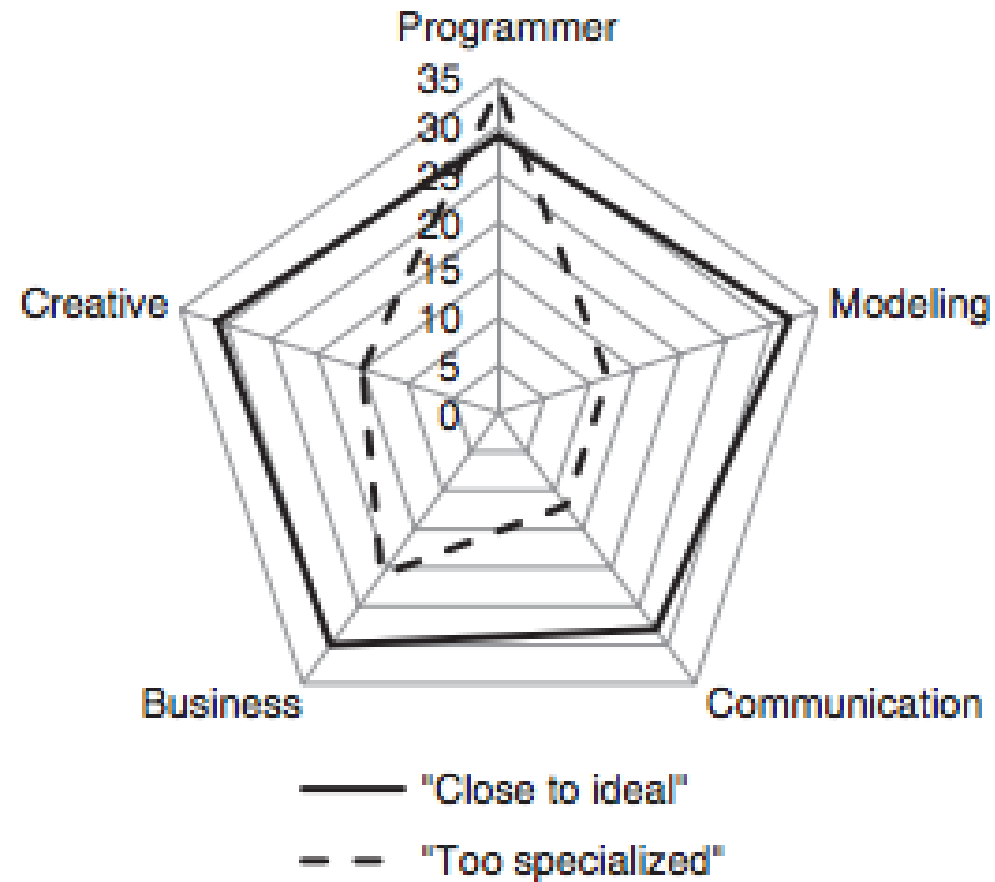
Fri	11,259 steps	★ >
Thu	9,320 steps	★ >
Wed	7,309 steps	>
Tue	11,172 steps	★ >



Fri	11,259 steps	★ >
Thu	9,320 steps	★ >
Wed	7,309 steps	>
Tue	11,172 steps	★ >



# The Fraud data scientist



# Predictive Analytics



# What is predictive analytics...Why Predictive Analytics matters...

**Predictive analytics** is the branch of the advanced analytics which is used to make predictions about unknown future events. **Predictive analytics** uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions about future.



## Descriptive Analytics



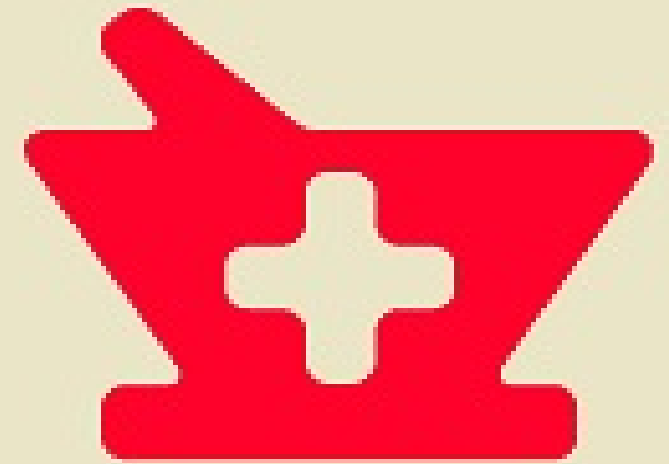
What has happened?

## Predictive Analytics



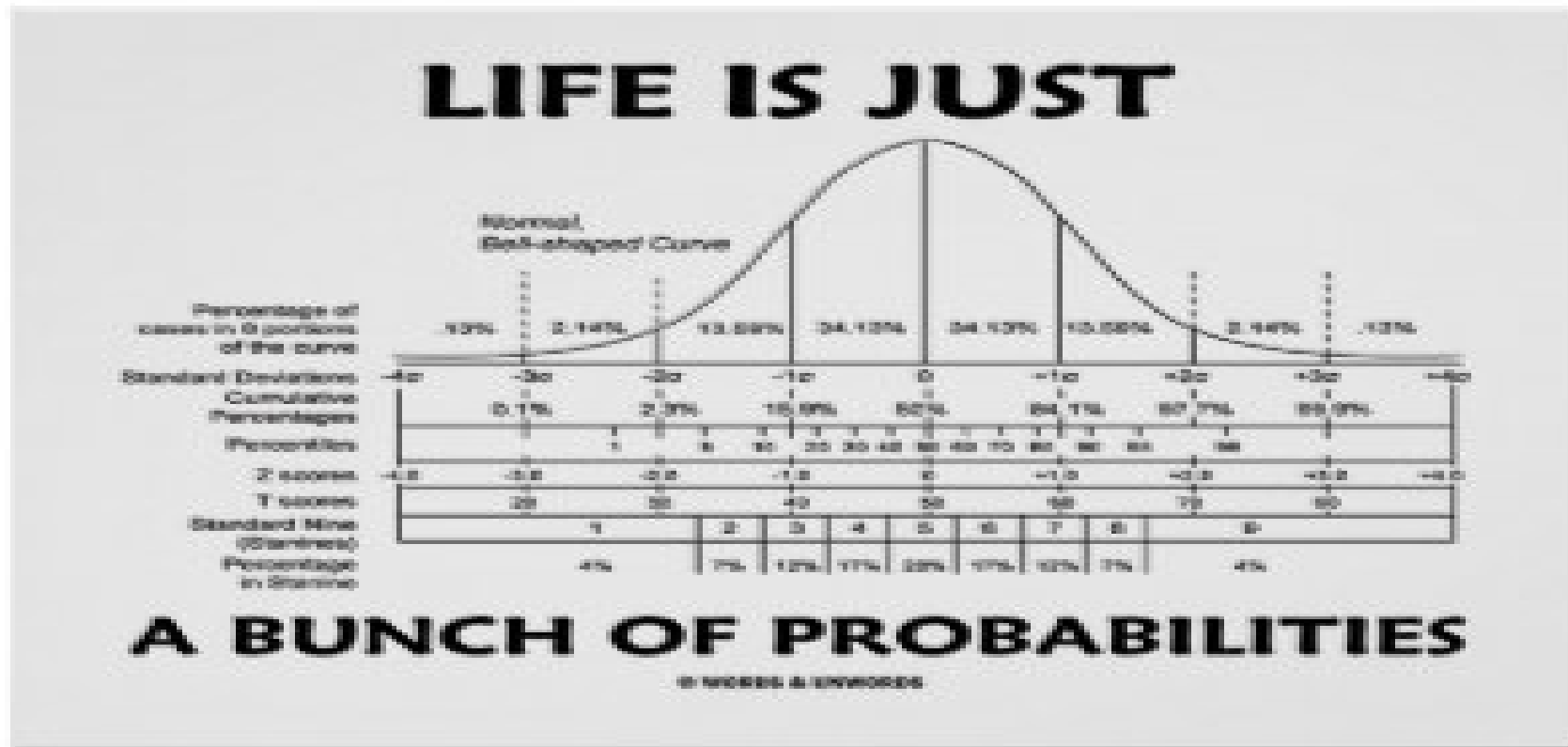
What could happen in the future based on previous trends and patterns?"

## Prescriptive Analytics



What should a business do?

Not what **will** happen... What **might** happen...

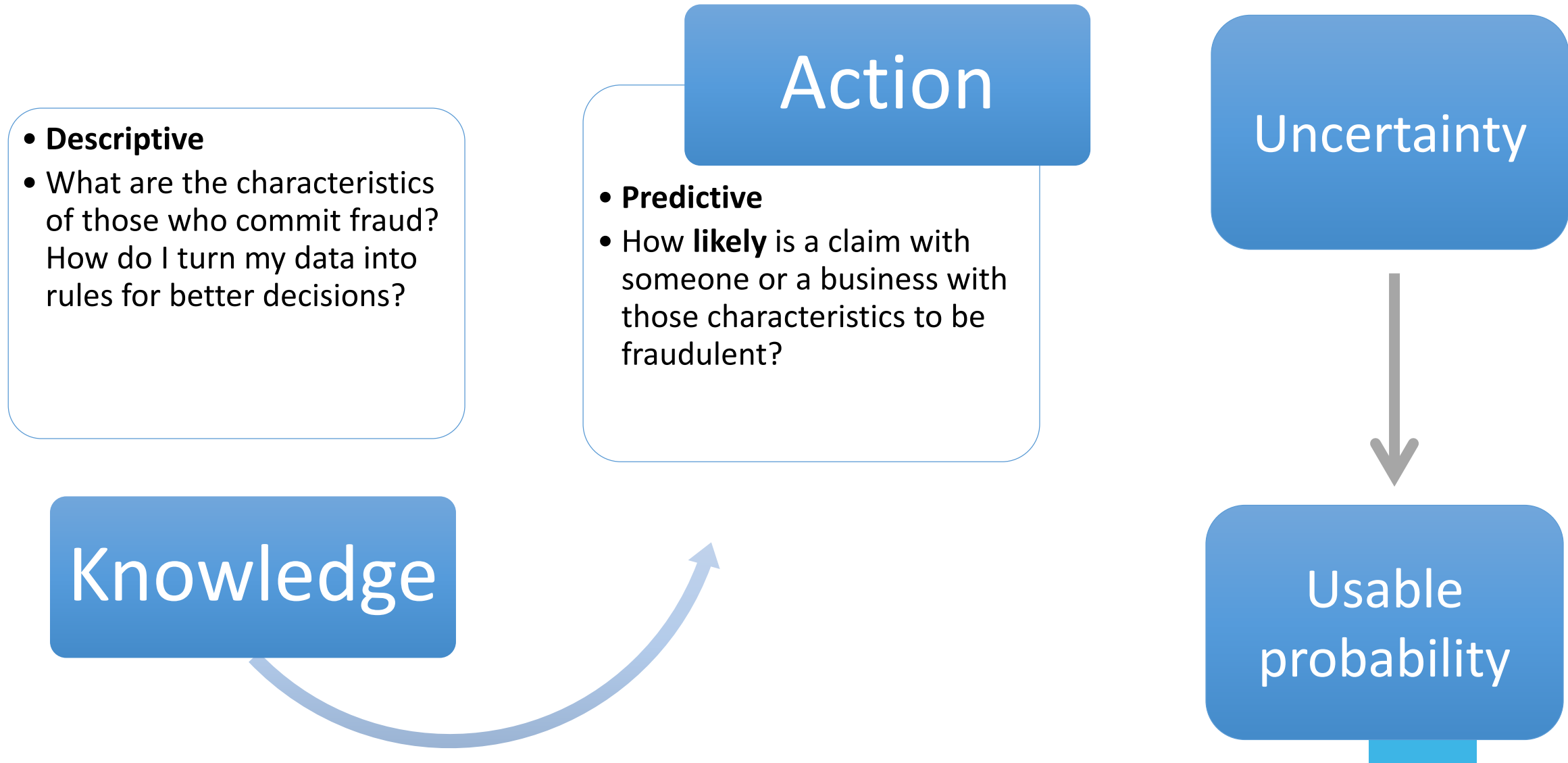


# Time for predictive analytics has come...

- Predictive Analytics like statistics has been around for a long time...
- So what has changed?
  1. Increase in volume and type of data
  2. Greater interest in data for insights
  3. Computing power, and “point and click”
  4. Tougher economic conditions and need for competitive differentiation:  
Business efficiency; ROI.....



# Why predictive analytics matters



What does this all mean for  
fraud detection?

z

# What does this mean for fraud detection and prevention

- Big Data and analytics provide powerful tools that *may* improve an organizations fraud detection systems
- COMPLIMENTARY to traditional expert-based fraud-detection approaches- DOES NOT REPLACE!!!

Fraud is an uncommon, well-considered, imperceptibly concealed, time-evolving and often carefully organized crime which appears in many types of forms.

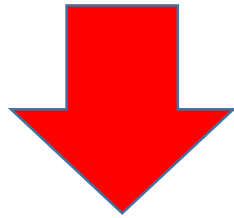
Social networks: That is, fraudulent companies are more connected to other fraudulent companies than to non-fraudulent companies.



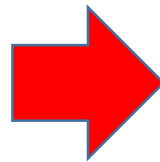


# What does this mean for fraud detection and prevention

**Fraud is an uncommon, well-considered, imperceptibly concealed, time-evolving and often carefully organized crime which appears in many types of forms.**



Social networks: That is, fraudulent companies are more connected to other fraudulent companies than to non-fraudulent companies.



Contextual information:  
Social Network Analysis

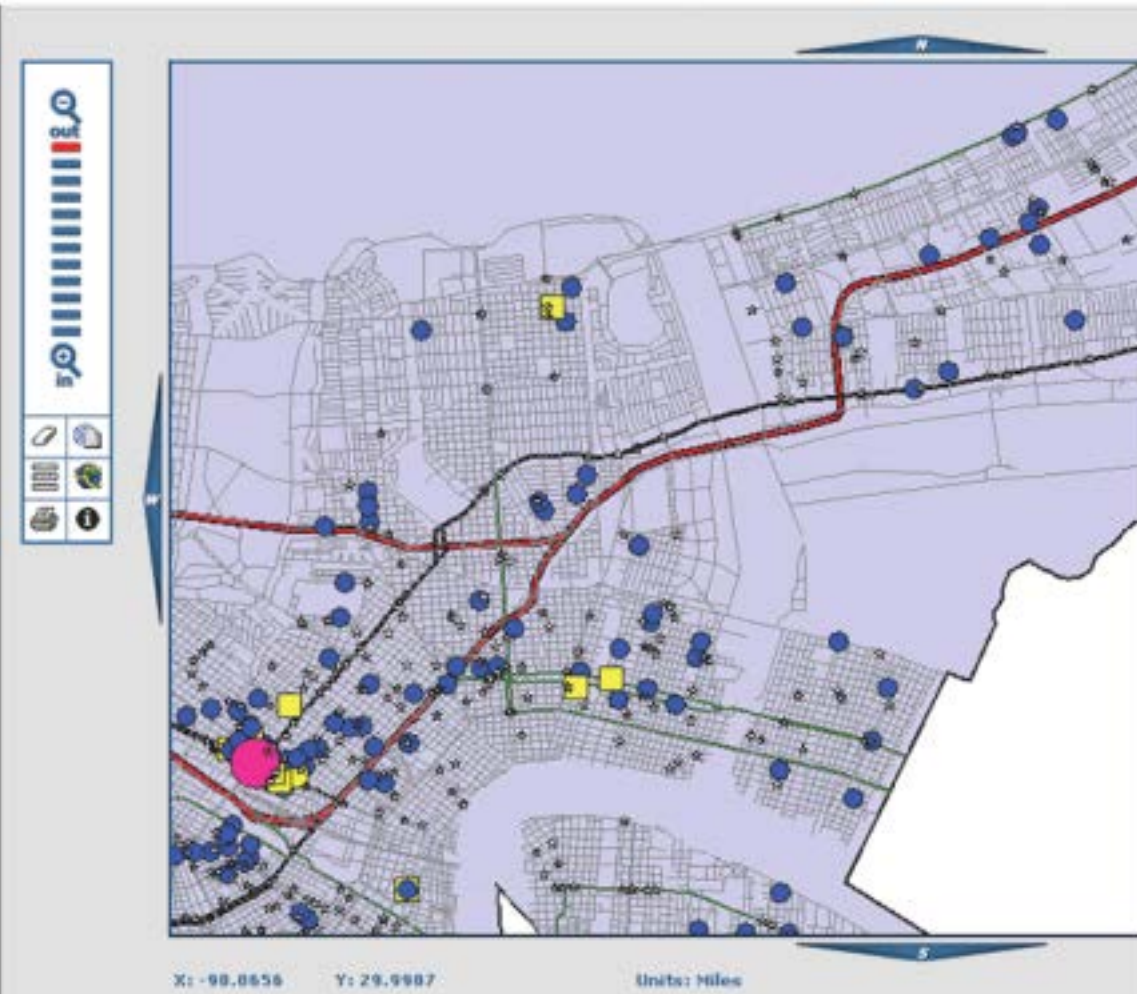
Z

# Use Case 1

Analytics Applied to Fraud Detection



# Use case of predictive analytics to detect fraud



- **Context:** Car insurance company in SA, operates globally
- Declining profits > increased premiums = fraudulent claims
- Historical claims data with known fraud outcomes to predict probability that new claims are fraudulent!
  - Understanding what has happened
- **Problem:** Repair shops that inflate estimates

# What we do using analytics... Geo-spatial data

- **Our problem:** Repair shops that inflate repair estimates

Use of Data:

- Claimants' address (Geocoded)
- Location of repair shops
- Average claim estimate for a particular problem

Analyzing the data:

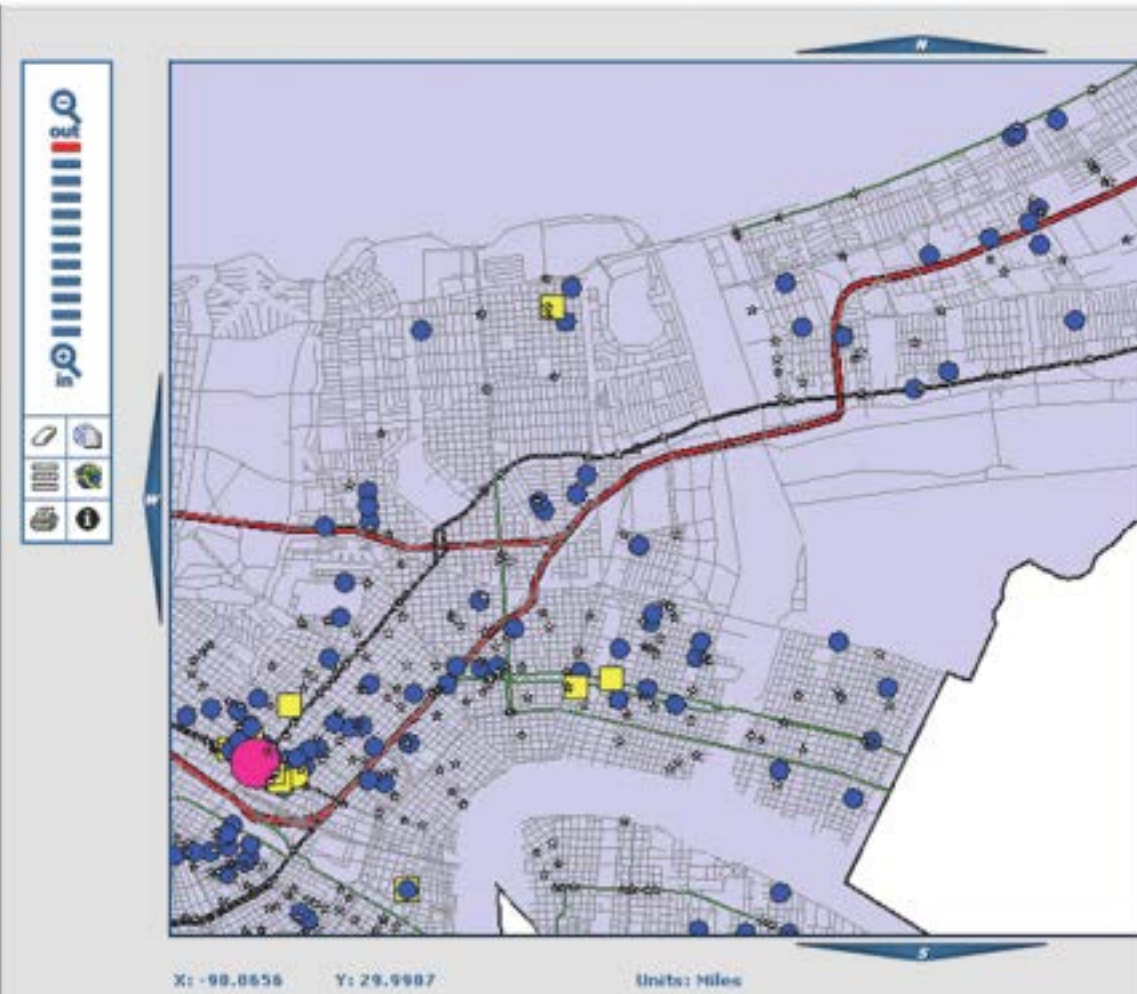
- Map areas where estimates are higher than the average
- Overlay claimants' address

Algorithm:

- Predict based on distance claimant travels to get a repair done > WHY travelling outside a radius?



# Use case of predictive analytics to detect fraud



- **Algorithm:** Claimants travelling a distance to get a repair done correlates with the repair shop providing over-estimates (above average)
- > inflated estimate > potential fraud
- **Outcome:**
  - Reduce time required to refer questionable claims for investigation by as much as 95%.
  - Success rate in pursuing fraudulent claims from 50% to 88%!
  - **Healthcare in Kenya!**

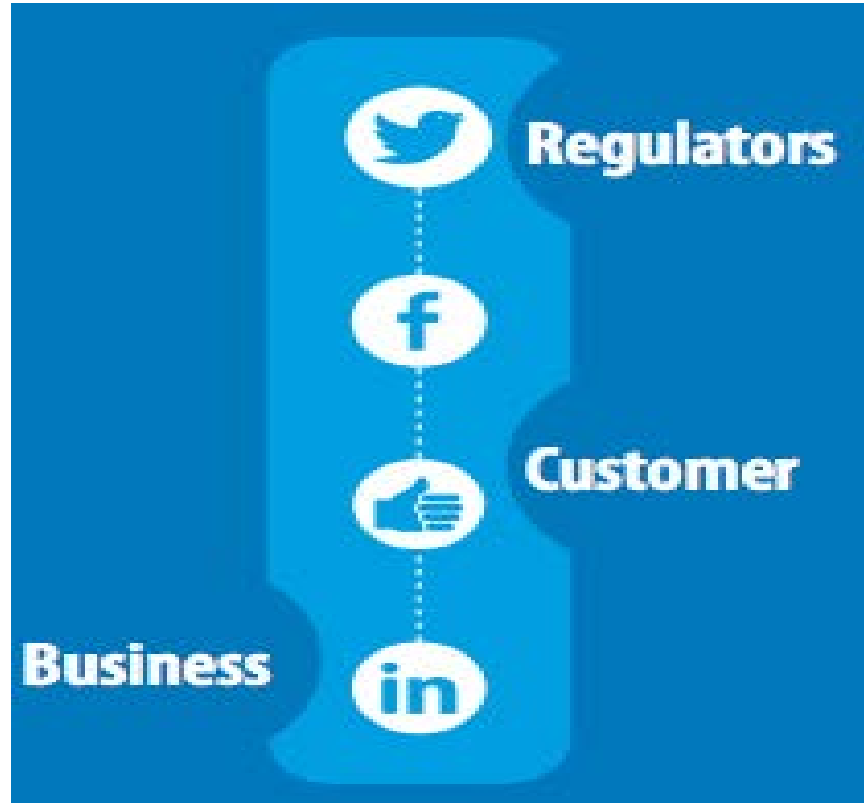


# Use Case 2

Analytics Applied to Fraud Detection



# Use case of predictive analytics to manage & prevent fraud



- Context: Insurance (Turkey)
- Mismatch between public and private profiles of individuals (narratives for claims) > Public data to serve as a reference for internal database records
- Relationship between customer profile and fraudulent claims
- Use of social media as a listening tool

# What we do using analytics... Social CRM

- **Our problem:** Characteristics (customer behavior and fraudulent claims)

## Use of Data:

- Consumers internal “known” data corroborated with external social data (e.g. check-in at “home” is 50km away from registered address)
- Using social analytics (text and images; check-in’s; likes etc.)

## Analyzing the data:

- Build behavioral profiles from social media data;
- Overlay behavioral data with known fraudulent claims

## Algorithm:

Predict based on behavioral data PROBABILITY of fraudulent claim (relationship between customer behavior and fraudulent claims)

Send for investigation: 86% accuracy. Social analytics is only an indicator > Investigators confirm independently

# Use Case 3

Analytics Applied to Fraud Detection



# Use case of predictive analytics to understand credit card fraud... Early adopters



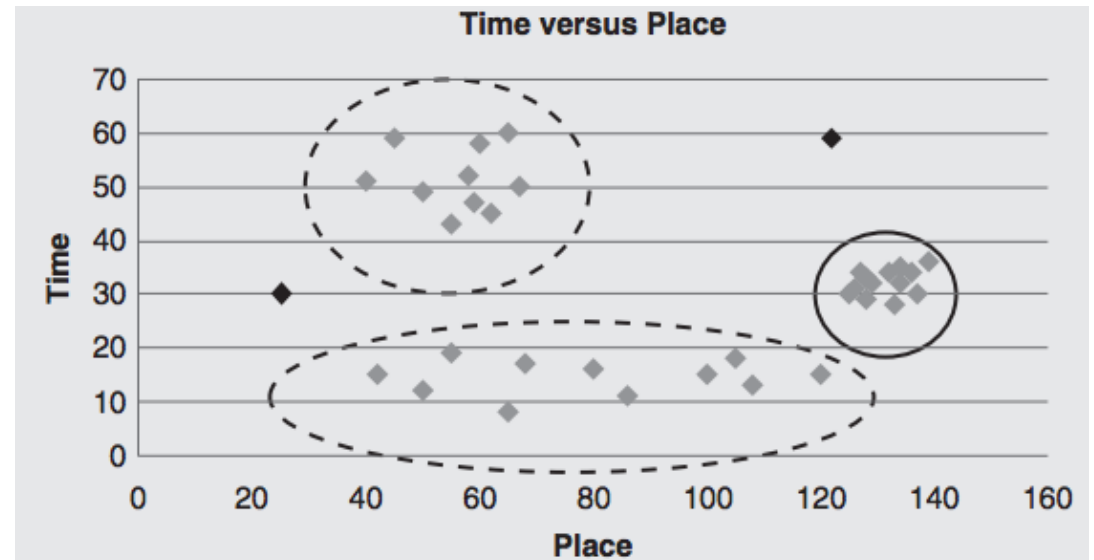
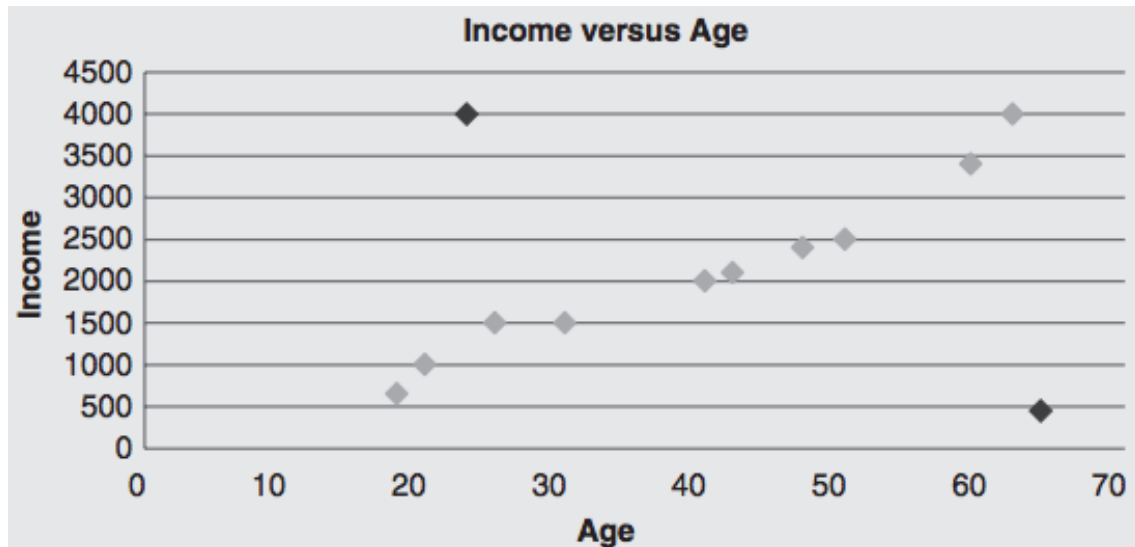
- **Context:** Financial institution (large impairments on CC fraud)
- **"Classic" symptoms:** Small purchase followed by a big one; large number of online purchases in a short period of time; spending as much as possible quickly; smaller amounts, spread across times
- **Problem:** "Normal" behaviour patterns of CC usage > outliers

# What we do using analytics... Supervised and unsupervised learning

- **Our problem:** Identify characteristics of transactions that deviate from the normal behavior

Use of Data:

- 2 million + CC holders





# Results

## Business Results

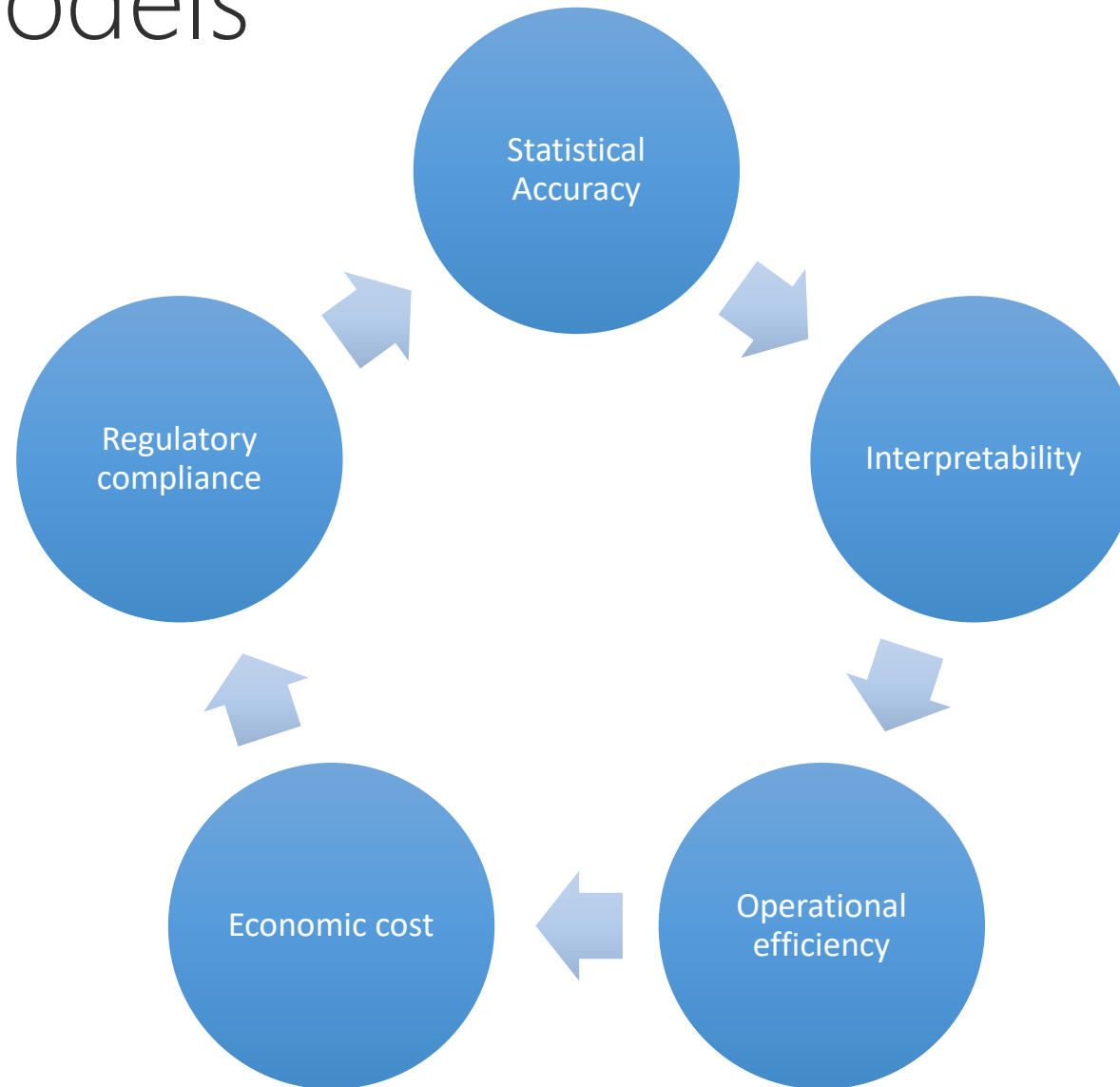
- 350 + hours of pure analysis
- 3 Months understanding
- Near-real time detection of fraudulent purchases and CC use
- 76% accuracy... > 85% once data issues fixed



# Fraud analytics process model



# Key characteristics of successful fraud analytics models



# With the right data

- Garbage data in > Garbage data out
- Master Data Management
  - Policies
  - Governance
  - Processes
  - Standards and Tools
  - Leads to increased accuracy of predictive models



# At the heart of predictive analytics



# So why are we NOT leveraging predictive analytics...

1. Data-driven **company culture**
2. What is the **value** (cost vs benefit)
3. **Innovation**: Saying no before trying – losing first mover advantage
4. **Leadership**: More data does not lead to success – making sense of the data with clear goals does!
5. **Talent management**: As data becomes cheaper, the complements become expensive
  1. **Data Scientists** with a business understanding become central – Do we have the skills? What skills do we need? What is a data scientist?
  2. **Problem solving skills**: logic and reasoning – the ability to know how non-traditional and traditional data sources can assist business derive and drive value





# Thank You

For more information, email: [ChibaM@gibs.co.za](mailto:ChibaM@gibs.co.za); [manojchiba@gmail.com](mailto:manojchiba@gmail.com)

A blue square logo with a white letter 'Z' inside, located in the bottom right corner of the slide.