Data Driven Decision Making

“*If you’re able to go into a meeting and other people have opinions, but you have data to support your arguments and your recommendations, you’re going to be influential.*”

Professor Jan Hammond, Harvard Business School

BUSINESS ANALYTICS (BA)

- ✓ BA is a set of disciplines and technologies for solving business problems using data analysis, statistical models and other quantitative methods.
- ✓ BA tools are used to visualize and explore the patterns & trends in the data to predict future business outcomes.
- ✓ BA makes extensive use of big data & advanced statistical analysis, mathematical modeling, and data mining to explore & understand the business performance.

BA & ITS IMPORTANCE IN BUSINESS

- ✓ *BA tries to answer the following questions:*
  - ➢ What is happening & why did something happen?
  - ➢ Will it happen again?
  - ➢ What will happen if we make changes to some of the inputs?
  - ➢ What the data is telling us that we were not able to see before?

- ✓ *BA is used to extract meaningful insights about consumer purchase pattern, attrition rate, etc.*
- ✓ *Finance analyzes data of millions of customers to assess risk & behavior*

TYPES OF BA

✓ *The BA area can be divided into 3 major categories*

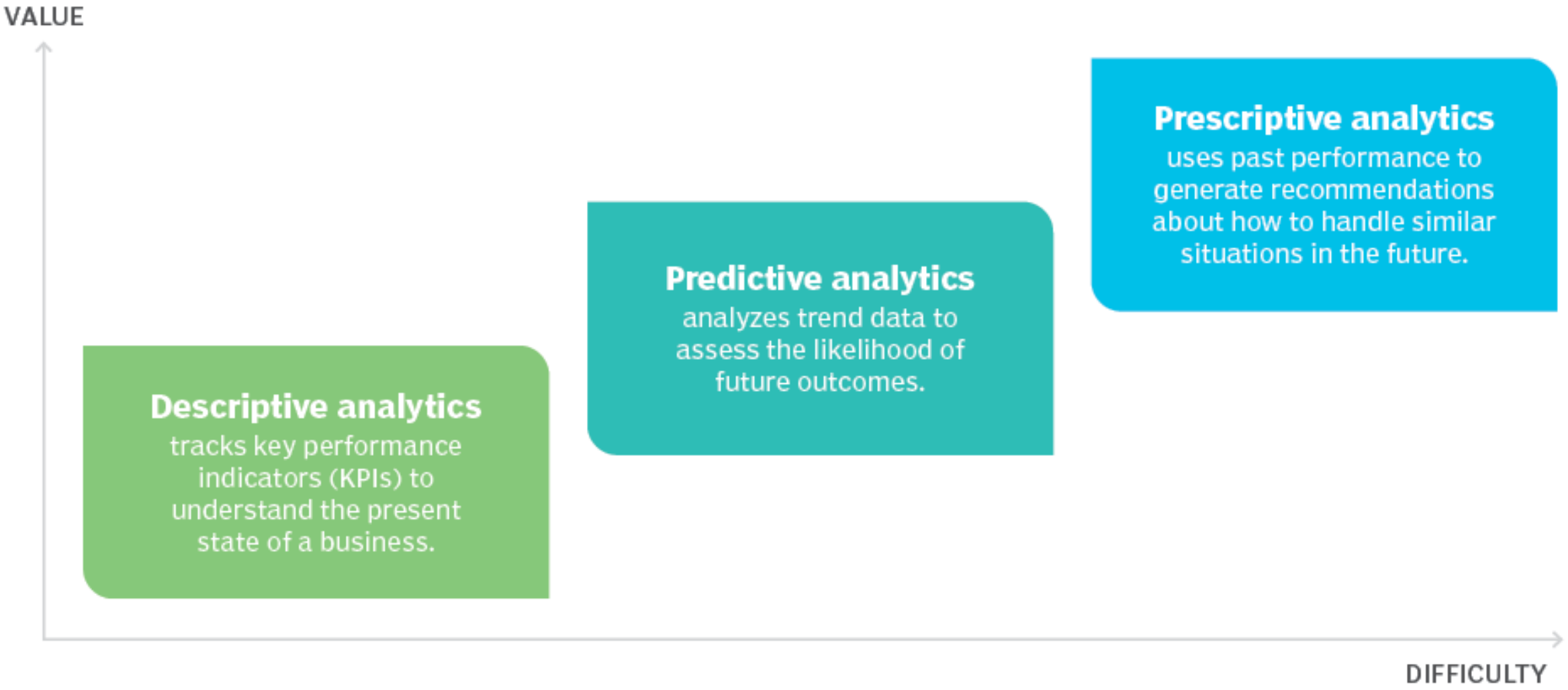
- Descriptive Analytics (answers the question, “What happened?”)
- Diagnostic Analytics (Answers “Why something has happened”)
- Predictive Analytics (answers the question, “What might happen in the future?”)
- Prescriptive Analytics (answers the question, “What should we do next?”)

TYPES OF BA

✓ *The BA area can be divided into 3 major categories*

- Descriptive Analytics (graphs, charts, dashboard, central tendency, dispersion, shape, var. etc.)
- Diagnostic Analytics (Drill down, Data discovery, data mining, root cause, correlation, etc.)
- Predictive Analytics (regression, data mining, forecasting model, etc.)
- Prescriptive Analytics (linear programming, optimization, simulation, decision analysis, etc.)

3 TYPES OF BA


VALUE
Descriptive analytics
tracks key performance indicators (KPIs) to understand the present state of a business.
Predictive analytics
analyzes trend data to assess the likelihood of future outcomes.
Prescriptive analytics
uses past performance to generate recommendations about how to handle similar situations in the future.
DIFFICULTY

COST OF BAD DATA

An over calculation of its commission cut caused Uber drivers to be underpaid. This incorrect formula circulating in its circulation cost tens of millions of dollars to Uber. They had to repay about **$900** per driver.

The Story of a Fruit Seller and Mr. X

A Fruit seller

Reasons for relocation:

- Less competition
- People are more affluent
- Customers do not haggle

He is a school drop out!

Customer profiling – expensive cars, big houses
Competitors analysis

Business Analytics

What Analytics Mao Zedong did on this bird?

Mao Zedong (1893-1976): Great Leap Forward(1958-62)

Four pests campaign 1958-1960 (rats, flies, mosquitoes, sparrows)

About 36 million odd people died of starvation (during 1959-62)

- Rather than being increased, rice yields after the campaign were substantially decreased.

- Extermination of sparrows upset the ecological balance, and insects destroyed crops as a result of the absence of natural predators.

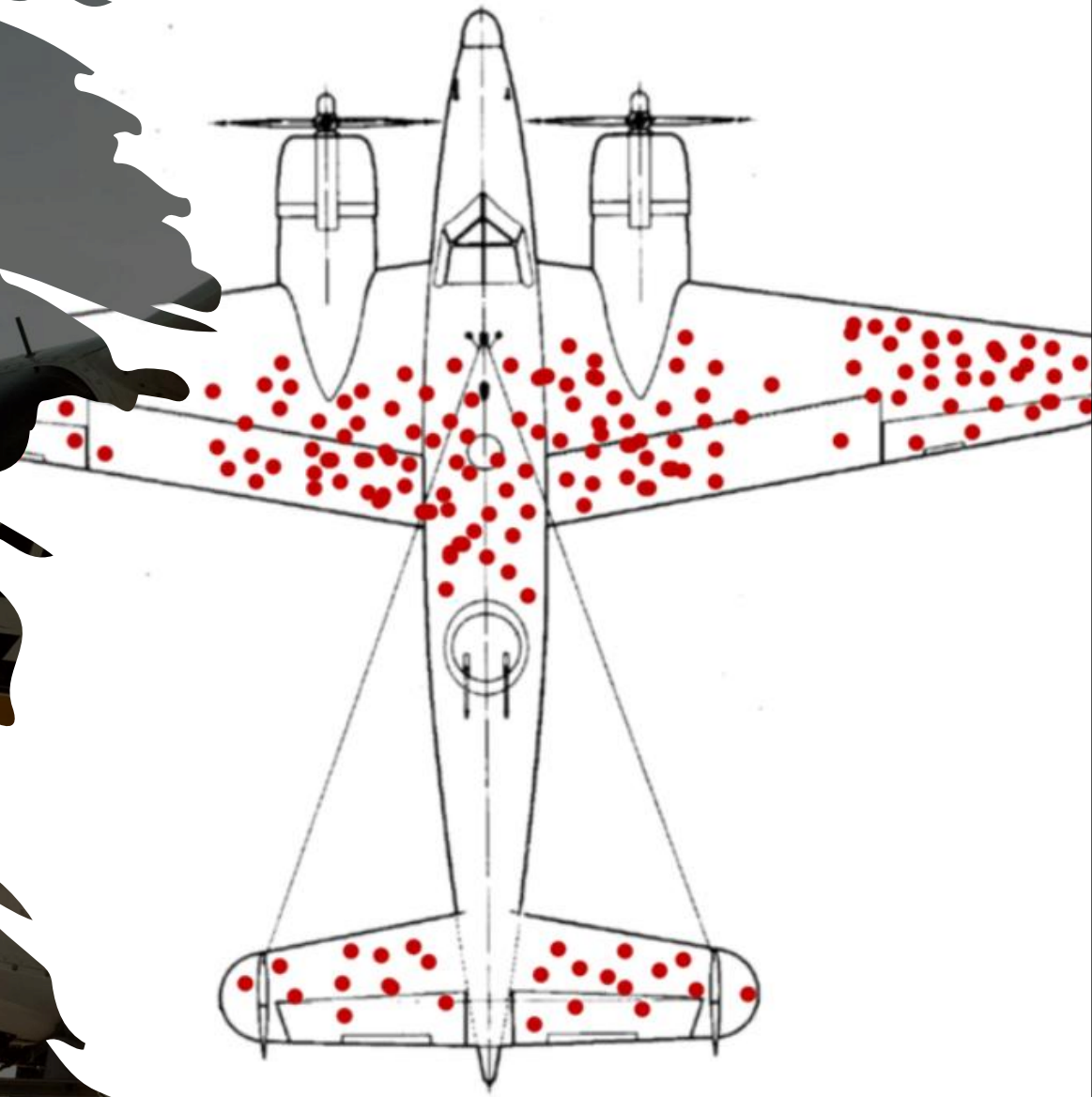
- Ecological imbalance is credited with exacerbating the Great Chinese Famine, in which 36 million people died of starvation.

- The Chinese government eventually resorted to importing 250,000 sparrows from the Soviet Union to replenish their population.

Who is smarter?

The fruit seller or Mao Zedong?

How WW2 aircraft can teach us about data driven insights today.



mage pattern on a WW2 bomber, dot pattern: Creative C

The world's most valuable resource is no longer oil, but data...

The Economist

Data Volume

Average Transaction size in Walmart – Day? Year?

10000+ stores across the world

Over 1 Million Transactions every hour

10,000 to 100,000 SKUs per store

Transaction :

In one year, more than 10 billion transaction!

Excel – Capacity : 1 million Row

Over 100M transactions a day

transactions a day

Offer protection on
over 2 billion accounts

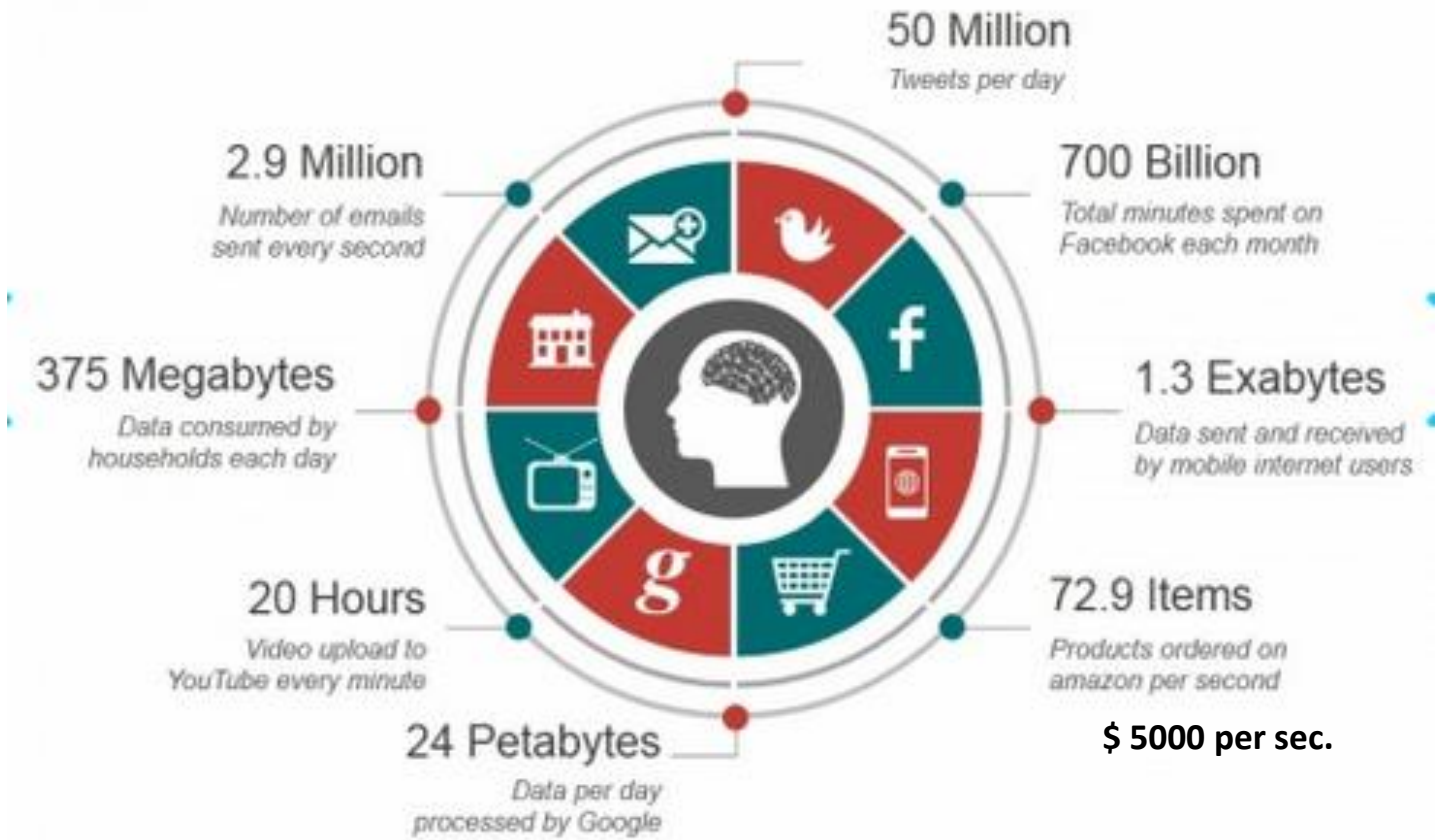
How Big is Big Data

No. of emails sent per day ____

Amazon --Orders ($) per day __

No of tweets per day _____

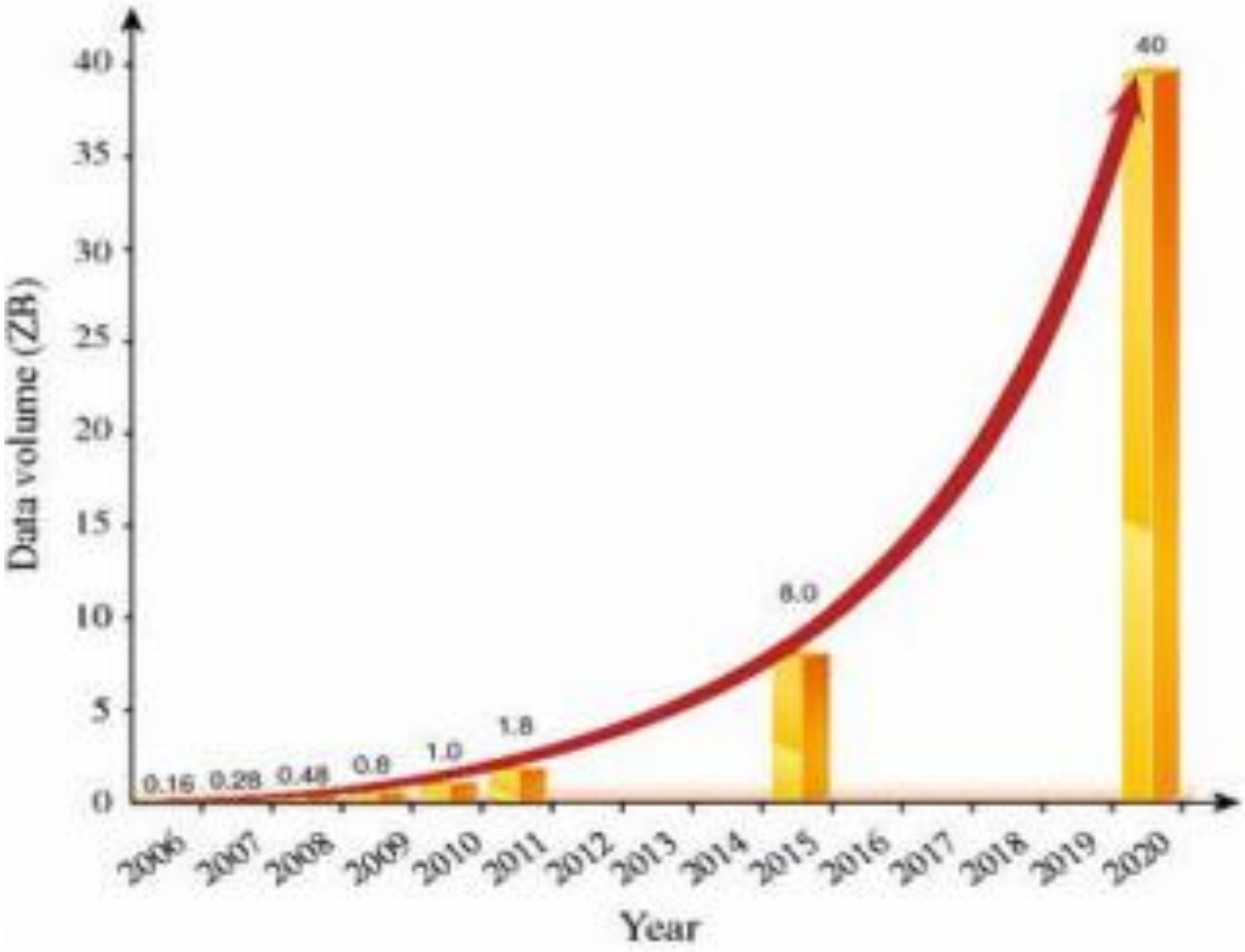
How Big Is Big Data



$ 5000 per sec.

Sources Of Big Data

- Fig: Global Growth Trend of Big Data Volume



What is Big Data?

- Information from multiple internal and external sources:
  - Transactions
  - Social media
  - Enterprise content
  - Sensors
  - Mobile devices

- Companies leverage data to adapt products and services to:
  - Meet customer needs
  - Optimize operations
  - Optimize infrastructure
  - Find new sources of revenue
  - Can reveal more patterns and anomalies

Big Data Characteristics

- **Volume:**

  Organizations collect data from a variety of sources, including business transactions, social media and information from sensor or machine-to-machine data.

  This makes most data sets too large to store and analyse using traditional database technology. New big data tools use distributed systems so that we can store and analyse data across databases that are dotted around anywhere in the world.

. New technologies (such as Hadoop) have made it easier to handle this sort of data.

- **Velocity:**

  Data streams in at an unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time.

  Just think of social media messages going viral in seconds.

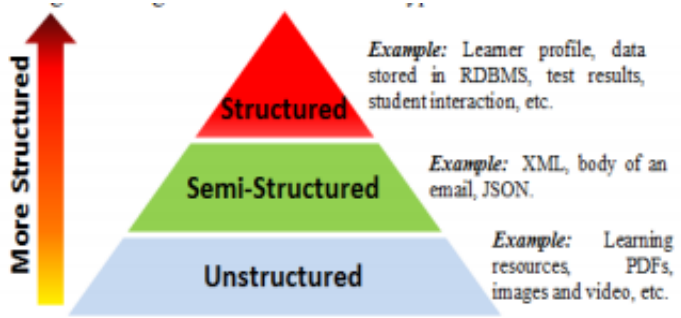
  Technology allows us now to analyze the data while it is being generated (sometimes referred to as in-memory analytics), without ever putting it into databases.

Big Data Characteristics

- **Variety:**

  Data comes in all types of formats – from structured, numeric data in traditional databases to unstructured text documents, email, video, audio, and financial transactions.

  In fact, 80% of the world's data is unstructured (text, images, video, voice, etc.)



- **Veracity**

  The fourth V is **veracity**, which in this context is equivalent to quality. We have all the data, but could we be missing something? Are the data "clean" and accurate? Do they really have something to offer?

- **Value**

  Finally, the V for **value** sits at the top of the **big data** pyramid. This refers to the ability to transform a tsunami of data into business.

A NOTE ABOUT "BUZZ WORDS"

- There are lots of buzz words that closely resemble business analytics:
  - Business intelligence
  - Decision science
  - Data science
  - *Blah, blah, blah…*
- Don't worry too much about the differences.
- They all support the same goal: "using tools and techniques to turn data…".

Descriptive Analytics

- Descriptive analytics, such as reporting/OLAP (online Analytical processing), dashboards, and data visualization, have been widely used for some time.
- They are the core of traditional BI.

Year 2000				
	Audio Division		Video Division	
Line Items	Budget	Actual	Budget	Actual
Cost of Goods Sold	$6,851,006.48	$7,132,961.38	$4,322,514.74	$4,526,954.71
Marketing Expense	$750,179.20	$756,596.17	$455,048.05	$462,815.40
Research and Development Expense	$538,243.39	$538,014.73	$329,890.95	$336,808.13
Selling Expense	$1,632,921.64	$1,579,790.18	$986,887.49	$927,970.90
Taxes	$314,659.05	$319,390.19	$202,636.67	$200,205.01

Year 2001				
	Audio Division		Video Division	
Line Items	Budget	Actual	Budget	Actual
Cost of Goods Sold	$2,664,556.31	$2,700,773.16	$1,726,031.16	$1,773,448.08
Marketing Expense	$294,766.22	$290,696.70	$187,757.29	$176,778.55
Research and Development Expense	$200,719.90	$193,236.83	$134,270.95	$125,725.88
Selling Expense	$620,427.30	$611,649.47	$405,092.93	$400,181.91
Taxes	$130,926.70	$122,526.31	$82,450.76	$80,671.87

Eg;
Total Stock in inventory
Y to Y change in sales
Attendance of Employees

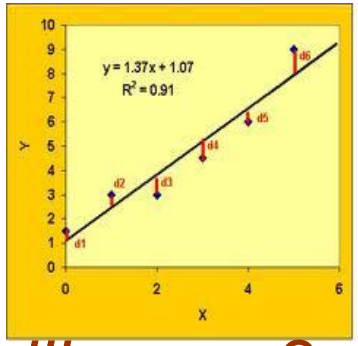
RMG Workers in Africa

What has occurred?

Descriptive analytics, such as data visualization, is important in helping users interpret the output from predictive and predictive analytics.

Predictive Analytics

- Predictive analytics has its roots in the ability to “Predict” what might happen. These analytics are about understanding the future. Predictive analytics provide estimates about the likelihood of a future outcome.
- No statistical algorithm can “predict” the future with 100% certainty.



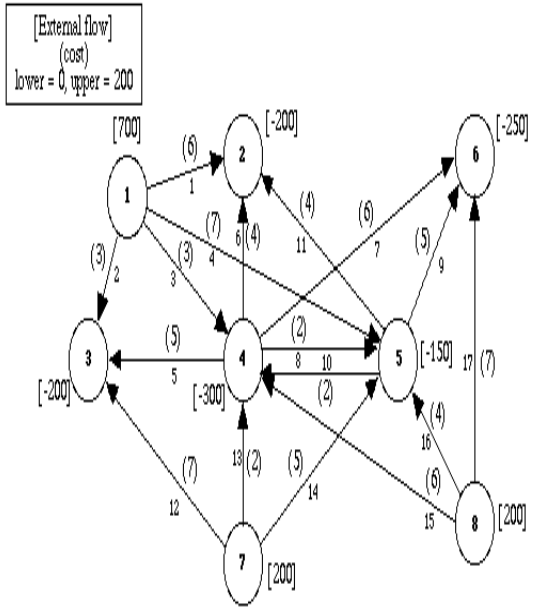
Time series Analysis

What will occur?

- Predictive analytics can be used throughout the organization, from forecasting customer behavior and purchasing patterns to identifying trends in sales activities. They also help forecast demand for inputs from the supply chain, operations and inventory.

Prescriptive Analytics

- The relatively new field of prescriptive analytics allows users to “prescribe” a number of different possible actions to and guide them towards a solution. These analytics are all about providing advice. Prescriptive analytics attempt to quantify the effect of future decisions in order to advise on possible outcomes before the decisions are actually made.



Optimization Model

- Larger companies are successfully using prescriptive analytics to optimize production, scheduling and inventory in the supply chain to make sure that are delivering the right products at the right time and optimizing the customer experience.

An Example LP Problem

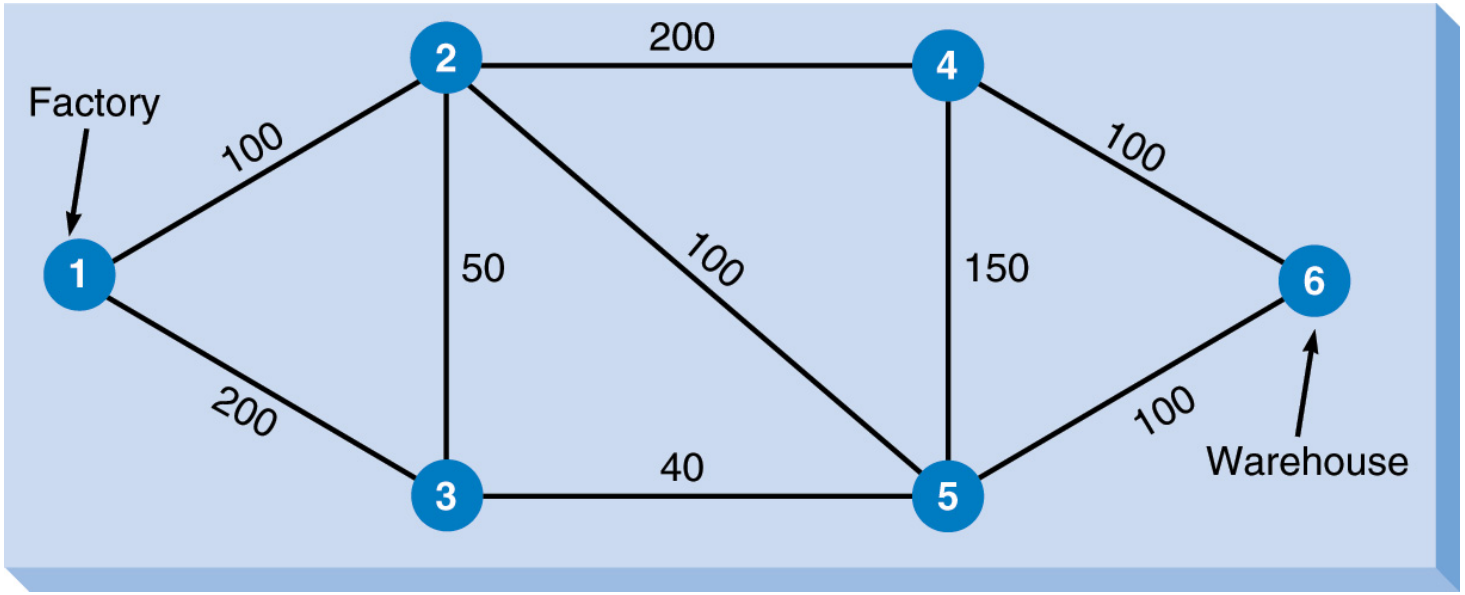
Blue Ridge Hot Tubs produces two types of hot tubs: Type X & Type Y.

	Type X	Type Y
Pumps	1	1
Labor	9 hours	6 hours
Tubing	12 feet	16 feet
Unit Profit	$350	$300

There are 200 pumps, 1566 hours of labor, and 2880 feet of tubing available.
How many type X and type Y tubs should be produced to maximize profit?

Shortest Path

- Want to find the shortest path from the factory to the warehouse



Example: Amazon.com

- Online retailers like Amazon.com and Overstock.com are high volume operations who rely on analytics to compete.
- When you enter their sites a cookie is placed on your PC and all clicks are recorded.
- Based on your clicks and any search terms, recommendation engines decide what products to display.
- After you purchase an item, they have additional information that is used in marketing campaigns.
- Customer segmentation analysis is used in deciding what promotions to send you.
- How profitable you are influences how the customer care center treats you.
- A pricing team helps set prices and decides what prices are needed to clear out merchandise.
- Forecasting models are used to decide how many items to order for inventory.
- Dashboards monitor all aspects of organizational performance

UPS

UPS has over 55,000+ delivery trucks in the US alone and around 106,000 drivers, globally. UPS delivers more than 20 million packages daily. When you consider the fact that every driver at UPS has millions of ways to run their delivery routes, the number of possibilities increases geometrically.

A reduction of one mile per driver per day translates to savings of up to $50 million a year. To focus on ways to reduce driving miles, UPS created the ORION (On-Road Integrated Optimization and Navigation) initiative.

- Capturing data through sensors, GPS, traffic and vehicles and using data mining techniques and analytics:

- UPS reduced 85 million miles driven per year. That equates to more than 8M fewer gallons of fuel used.

- Able to reduce engine idling time by 10M minutes. This led to savings in fuel consumption – around 650,000 gallons – and reduced carbon emissions by over 6,500 tons.
- Optimizing the supply chain using real-time actionable insights about customers, drivers, trucks, airplanes is the next frontier. Directed Actions from machine intelligence is the next frontier.

DATA CAPTURE & PREPARATION

WHAT IS DATA CAPTURE?

Data capture is the process of extracting information from any type of structured/unstructured document (paper or electronic) to transform it into a machine-readable digital format.

- ✓ Technological advancements in the field of Artificial Intelligence (AI) have taken data capture to new heights.

METHODS OF DATA CAPTURE

- ✓ Manual Data Capture
- ✓ Automated Data Capture
- ✓ Voice capture technology (Amazon’s Alexa, Microsoft’s Cortana, and Apple’s Siri)
- ✓ Video/Image Capture (security scanning at airports, theft prevention, etc.)

Data capture

Classical methods of data capture:

- Surveys
- Interviews
- Focus group discussions
- Panel data

Data capture

Digital Data Collection

- Transactional Tracking
- Online Tracking
- Online forms
- Social Media Monitoring

Data capture | Scale of measurement

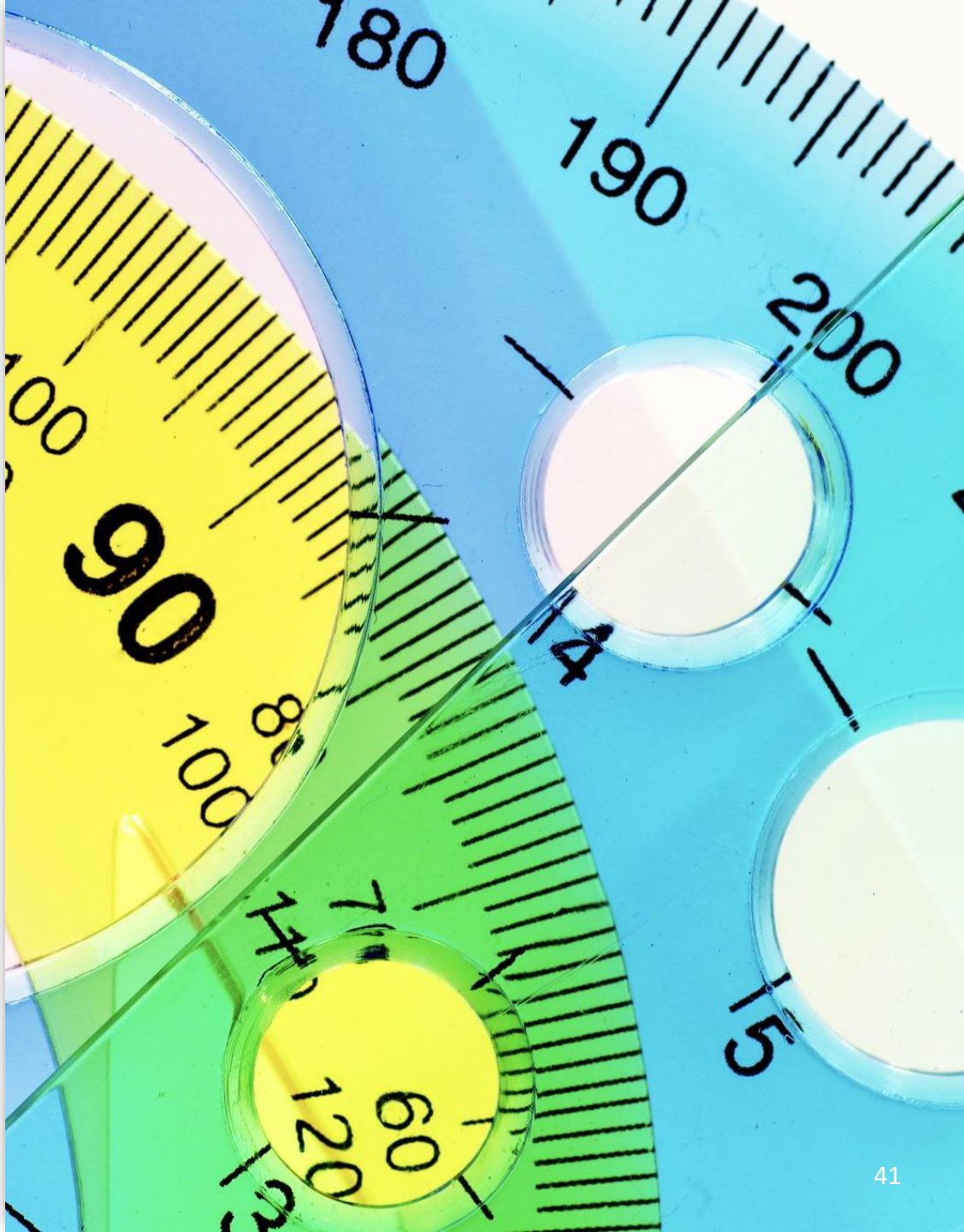
4 levels of measurement

1. Nominal
2. Ordinal
3. Interval
4. Ratio

Data capture | Scale of measurement

Nominal level of measurement

- Elements are represented as labels or names.
- Have no order.
- Can only be classified and counted.
- Example: Male / Female; Ethnicity: Bangladeshi / Non-Bangladeshi; Religion: I/ H/ B/ C/ J



Data capture | Scale of measurement

Ordinal level of measurement

- Based on a relative ranking or rating of items based on a defined attribute or qualitative variable
- Differences are not “mathematically” meaningful
- Example: Grades: A/B/C/D/E, Very Good ----- Good ----- Poor ----- Very Poor, Stage 3 vs Stage 4 Cancer

Data capture | Scale of measurement

Ordinal level of measurement

	Rating	Code	Frequency	%	Mean
Rate your experience with ACMP program at IBA, DU	Extremely poor	1	6	10%	2.62
	Poor	2	26	43%	
	Neither good nor poor	3	16	27%	
	Good	4	9	15%	
	Excellent	5	3	5%	

Data capture | Scale of measurement

Ordinal level of measurement

Economy	Global Rank
New Zealand	1
Singapore	2
Hong Kong SAR, China	3
Denmark	4
Korea, Rep.	5
United States	6
Georgia	7
United Kingdom	8
Norway	9
Sweden	10

Ease of Doing Business rankings 2019 Accessed from https://archive.doingbusiness.org/en/rankings#:~:text=A%20high%20ease%20of%20doing,equal%20weight%20to%20each%20topic.

Data capture | Scale of measurement

Interval level of measurement

- Ordered
- Interval or the distance between values is meaningful
  - Can state the difference between any two data values
  - Cant state the relative difference
- Based on a scale with a known unit of measurement
- Can + and -, but cannot X or ÷
- No “real zero”
- Example: Temperature (Zero doesn’t mean zero)

Data capture | Scale of measurement

Interval level of measurement

Temperature
05 degrees °F
10 degrees °F
20 degrees °F
40 degrees °F

Data capture | Scale of measurement

Ratio level of measurement

- The “highest” level of measurement
- Based on a scale with a known unit of measurement
- A meaningful interpretation of zero on the scale
- Example:
  - Age, Income/Salary, Marks, Weight, population

	Marks (out of 10)
Student 1	0
Student 2	10
Student 3	5

Data capture | Scale of measurement

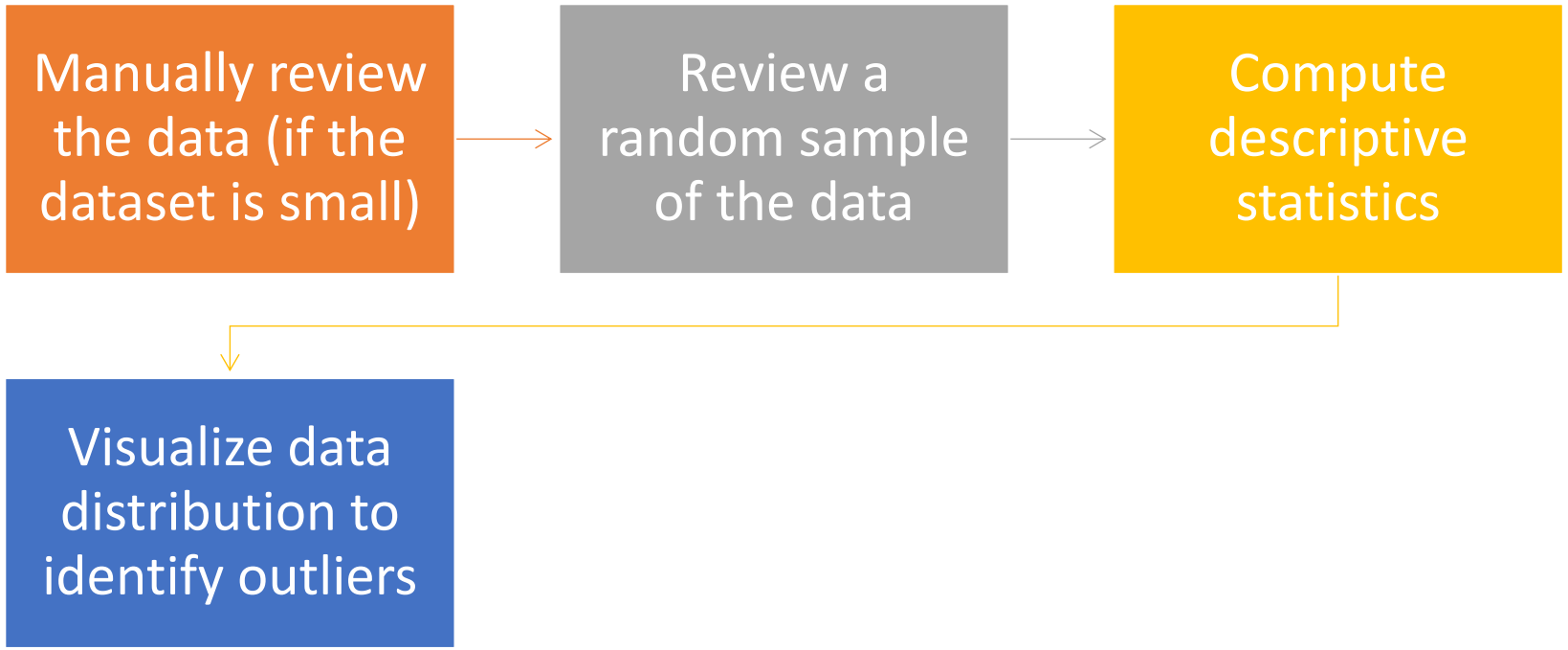
Lowest order *Highest order*

	Nominal	**Ordinal**	**Interval**	**Ratio**
Categorizes (no order/direction)				
Ordered categories (rank/order/scale)				
Has interval (no true zero)				
Has interval and true zero				

Data Preparation | Data Sanity Check

- Are the dates & time valid?
- Are the data in the right format?
- Are they on the right scale?
- Are there missing values?
- Are numerical values within range?
- Are there any outliers?
- Are there data duplications?
- Do we need to combine multiple data files? In that case, do we have an unique identification data?

Data Preparation | How to Check Data Sanity



Data Preparation | Handling Missing values

- *Missing values*: no data value is available for a variable in an observation

Data Preparation | Handling Missing values

Date	Sales
03/21/16	48
03/22/16	47
03/23/16	36
03/24/16	32
03/25/16	25
03/26/16	33
03/27/16	0
03/28/16	50
03/29/16	49
03/30/16	30
03/31/16	29
04/01/16	32
04/02/16	33
04/03/16	40

Data Preparation | Handling Missing values

- Remove the corresponding row/column
- Impute (“estimate”) a value
  - With zero
  - With average
  - With similar data points (“interpolation”)
- Make “missing” its own category

Data Preparation | Handling Missing values

Date	Sales
03/21/16	48
03/22/16	47
03/23/16	36
03/24/16	32
03/25/16	25
03/26/16	33
03/27/16	0
03/28/16	50
03/29/16	49
03/30/16	30
03/31/16	29
04/01/16	32
04/02/16	33
04/03/16	40

- Impute ("estimate") a value
- With zero
  - With average = 37.23
  - With similar data points
  - Using sales on Sunday, April 3 as an estimate = 40

Data Preparation | Handling Missing values

No single accepted solution.

Consider context.

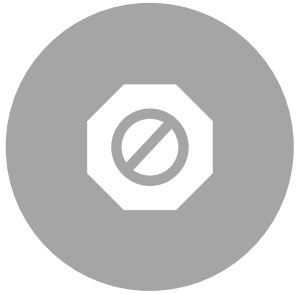
Try not to induce biases or distortions.

Need enough data to remain for meaningful analysis.

Pattern of missing values can itself carry information.

Data Preparation | Handling Outliers

Outlier: An observation “far away” from other observations.



No consensus definition of “far away”

More than 3 standard deviations away from the mean.

Other definitions based on statistical tests, nearest neighbors, quartile ranges.

Ethics in Analytics

- Advertising
- Privacy Issues