LOG 206

M8: Introduction to Big Data Analytics

Department of Logistics

Molde University College

Spring 2018

Høgskolen i Molde
Vitenskapelig høgskole i logistikk

Growing interest in Big data

Big Data: A Revolution That Will Transform How We Live, Work and Think – review

This informative introduction to the "datafication" of our lives looks at the benefits of big data in medicine, science and beyond

Thanks to the internet, social networking, smartphones and credit cards, more data is being collected and stored about us than ever before – a level of surveillance the Stasi could only dream about, say Mayer-Schönberger and Cukier in this informative introduction to the "datafication" of our lives. Big data analysis gives big business a competitive edge (all those Amazon recommendations), but governments have invested heavily in it, too. The risks to privacy and freedom are obvious, but the authors accentuate the positive. Big data has useful applications in medicine, science and "culturomics". Mayer-Schönberger and Cukier make interesting observations about data-crunching techniques and they also report that analysts have

Louis Columbus
Contributor
FOLLOW
I cover CRM, Cloud Computing, ERP and Enterprise Software
full bio →
Opinions expressed by Forbes Contributors are their own.

TECH 10/19/2014 @ 9:03PM | 18,453 views

84% Of Enterprises See Big Data Analytics Changing Their Industries' Competitive Landscapes In The Next Year

+ Comment Now + Follow Comments

87% of enterprises believe Big Data analytics will redefine the competitive landscape of their industries within the next three years. 89% believe that companies that do not adopt a Big Data analytics strategy in the next year risk losing market share and momentum.

These and other key findings are from an Accenture and General Electric study published this month on how the combination of Big Data analytics and the Internet of Things (IoT) are redefining the competitive landscape of entire industries. Accenture and GE define

DATA IS THE NEW OIL OF THE DIGITAL ECONOMY

Image: verifex/Flickr

Stordata revolusjonerer alt fra undervisning til politikk

Spranget fra undervisning som er skreddersydd din personlighetstype til politisk reklame som er spisset inn mot akkurat dine fordommer er ikke stort, i en verden hvor såkalte stordata skaper nye muligheter.

Ruth Lothe
kommunikasjonsrådgiver
NMBU - Norges miljø- og biovitenskapelige universitet

17

12.6 2017 04:00

– Men vi må huske at stordata også lett kan misbrukes, sier professor i biostatistikk Solve Sæbø.

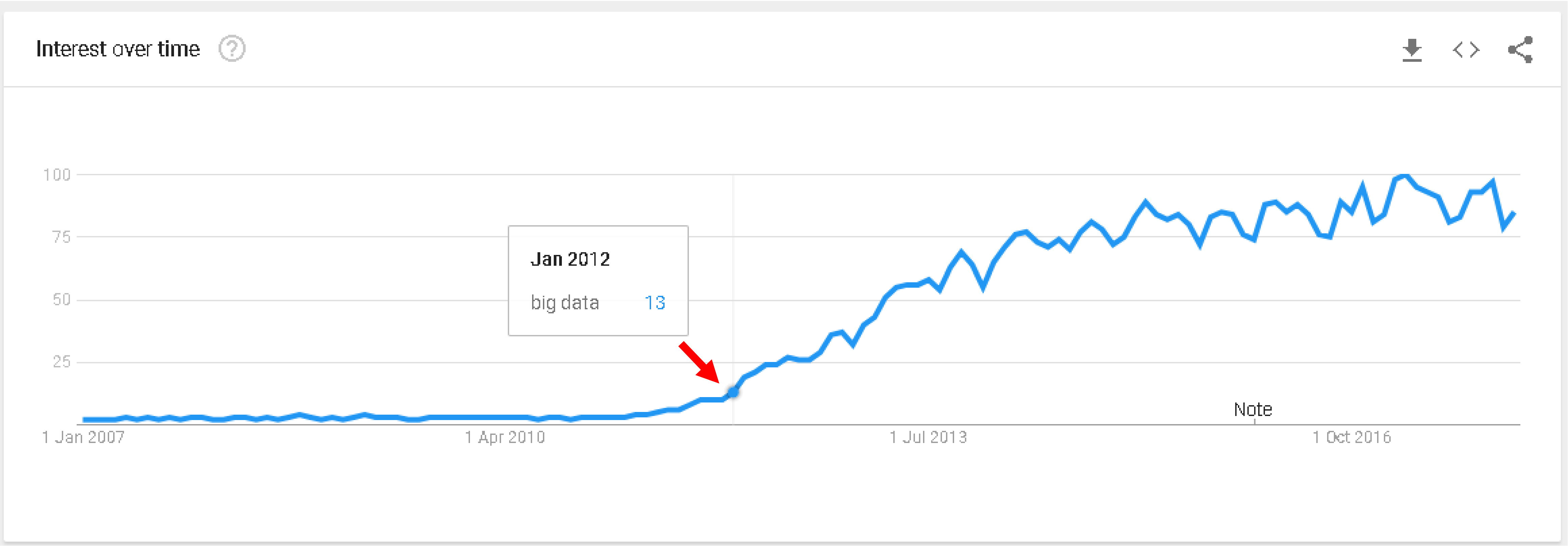
Som statistikkprofessor er han over gjennomsnittet interessert i muligheter og

Big data
Big Data er data som kan

29

DATA IN THE 21st Century is like Oil in the 18th Century: an immensely, untapped valuable asset. Like oil, for

Growing interest on Big data



Recall: Key strategic decisions faced by managers

- Decision 1: Digital business channel priorities
- Decision 2: Market and product development strategies
- Decision 3: Positioning and differentiation strategies
- Decision 4: Business, service and revenue models
- Decision 5: Marketplace restructuring
- Decision 6: Supply-chain management capabilities
- Decision 7: Internal knowledge management capabilities
- Decision 8: Organizational resourcing and capabilities

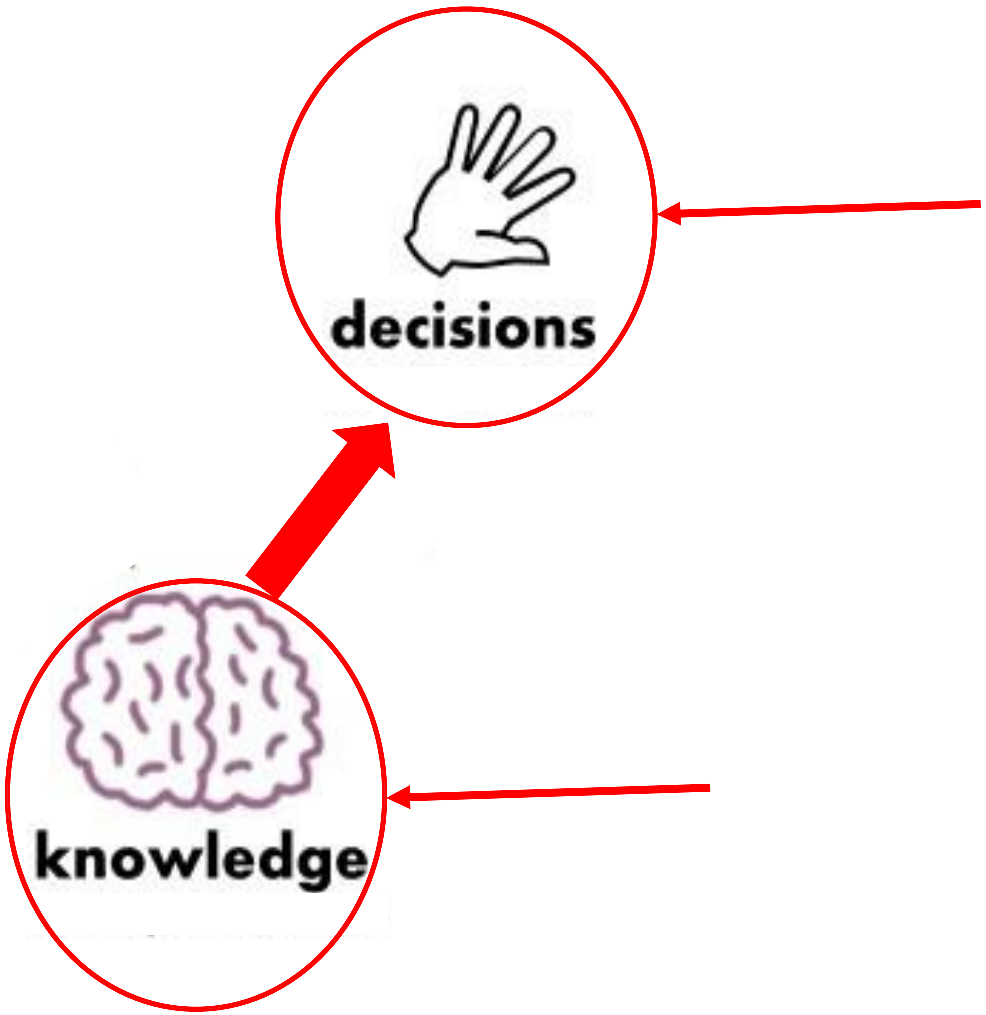
Making decisions

Role of knowledge

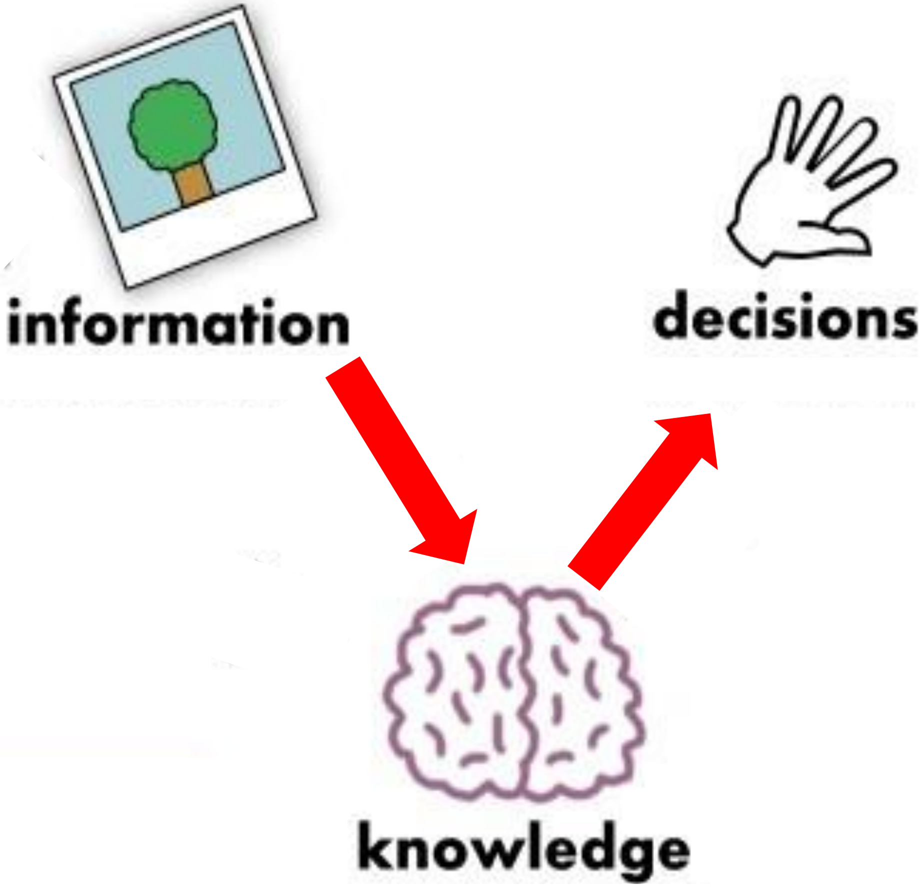
- Managers need knowledge to make good decisions and recognize business opportunities.
- The way a business gathers, shares and exploits knowledge can be central to its ability to develop successfully

Knowledge is a familiarity, awareness or understanding of someone or something, such as facts, information, descriptions, or skills, which is acquired through experience or education by perceiving, discovering, or learning.

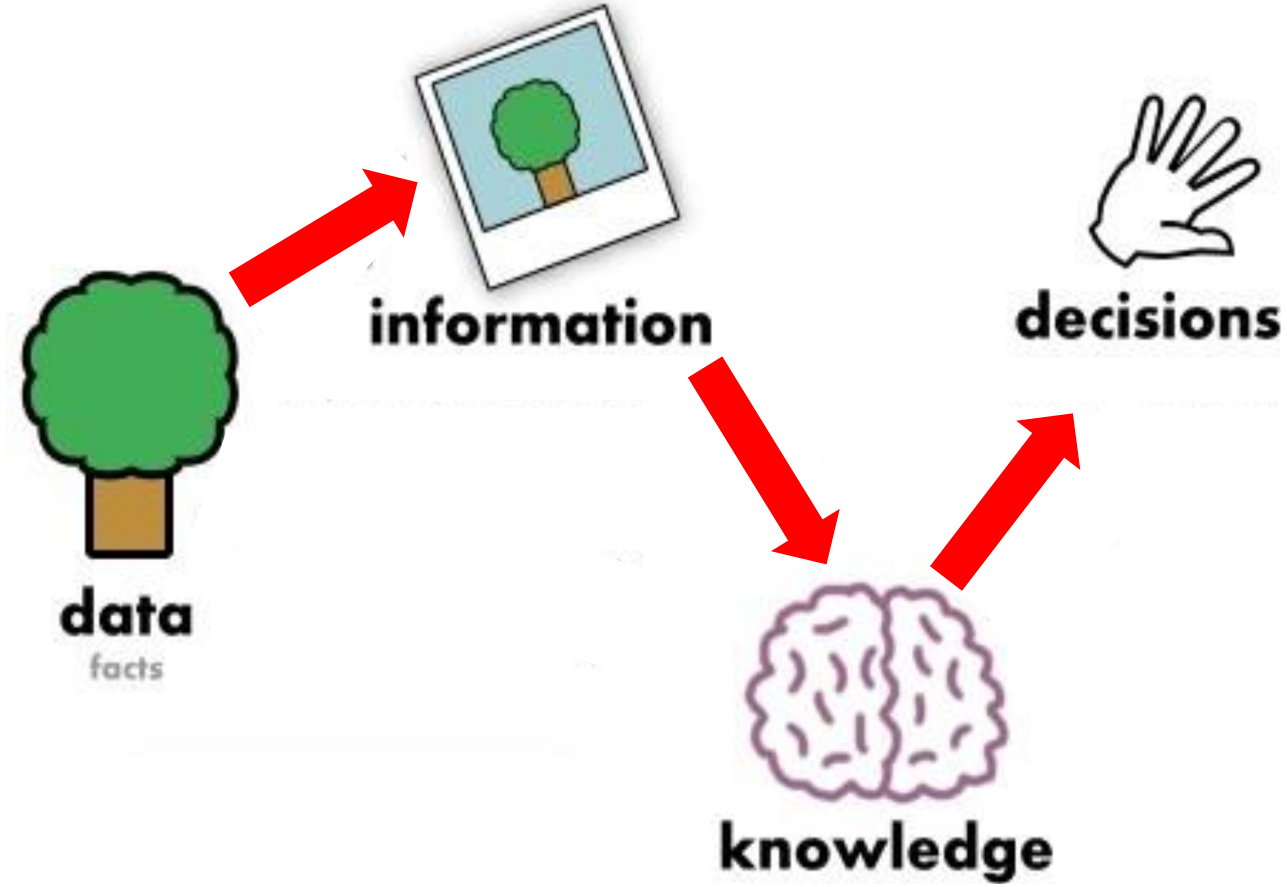
Knowledge and decision making



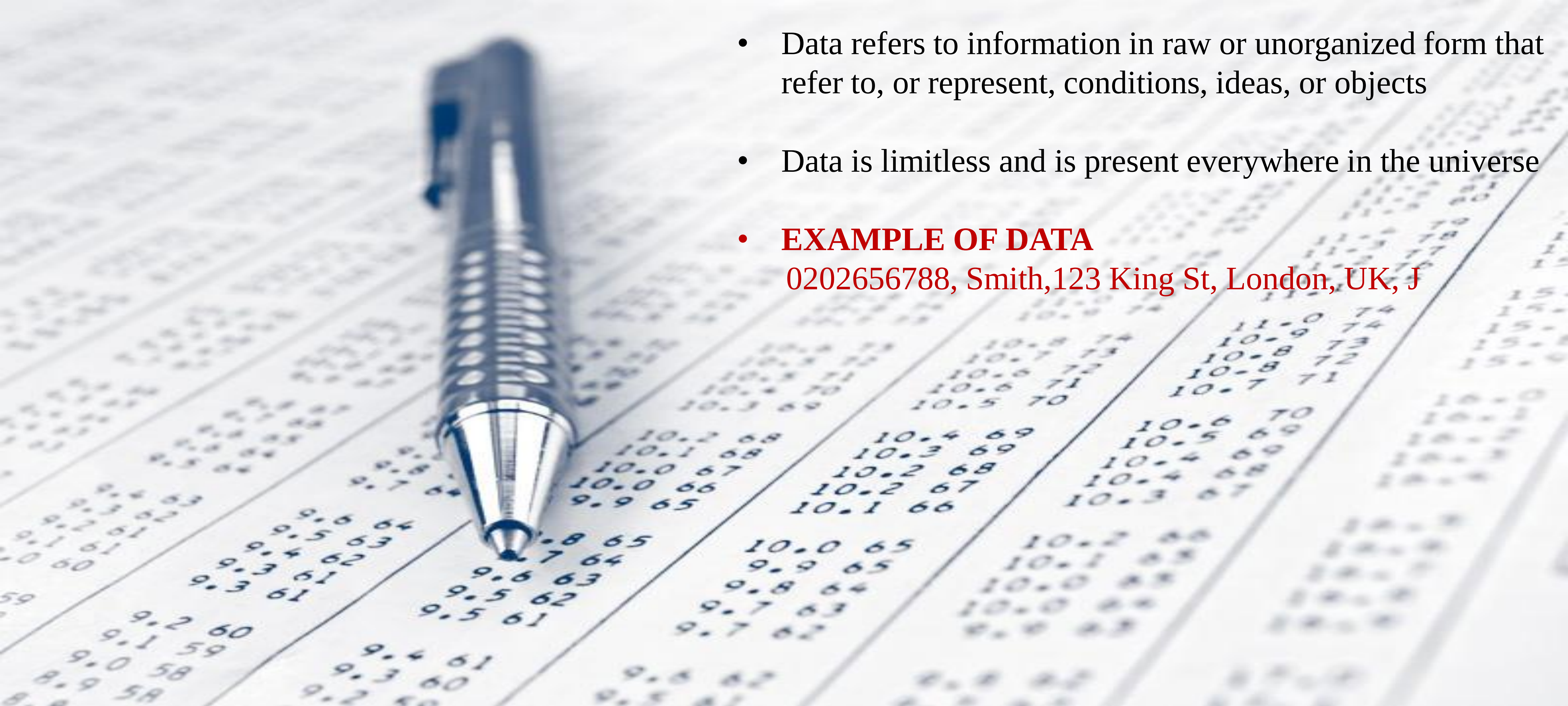
Where does knowledge come from?



Where does knowledge come from?



Data



- Data refers to information in raw or unorganized form that refer to, or represent, conditions, ideas, or objects

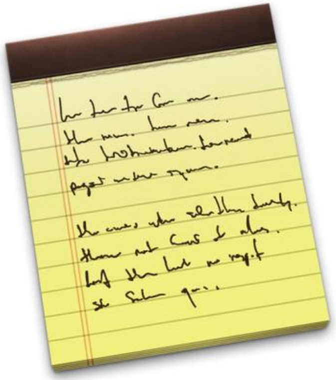
- Data is limitless and is present everywhere in the universe

- **EXAMPLE OF DATA**
 0202656788, Smith,123 King St, London, UK, J

Source of data for companies prior to the arrival of the mainstream internet

1. Interviews

Notes

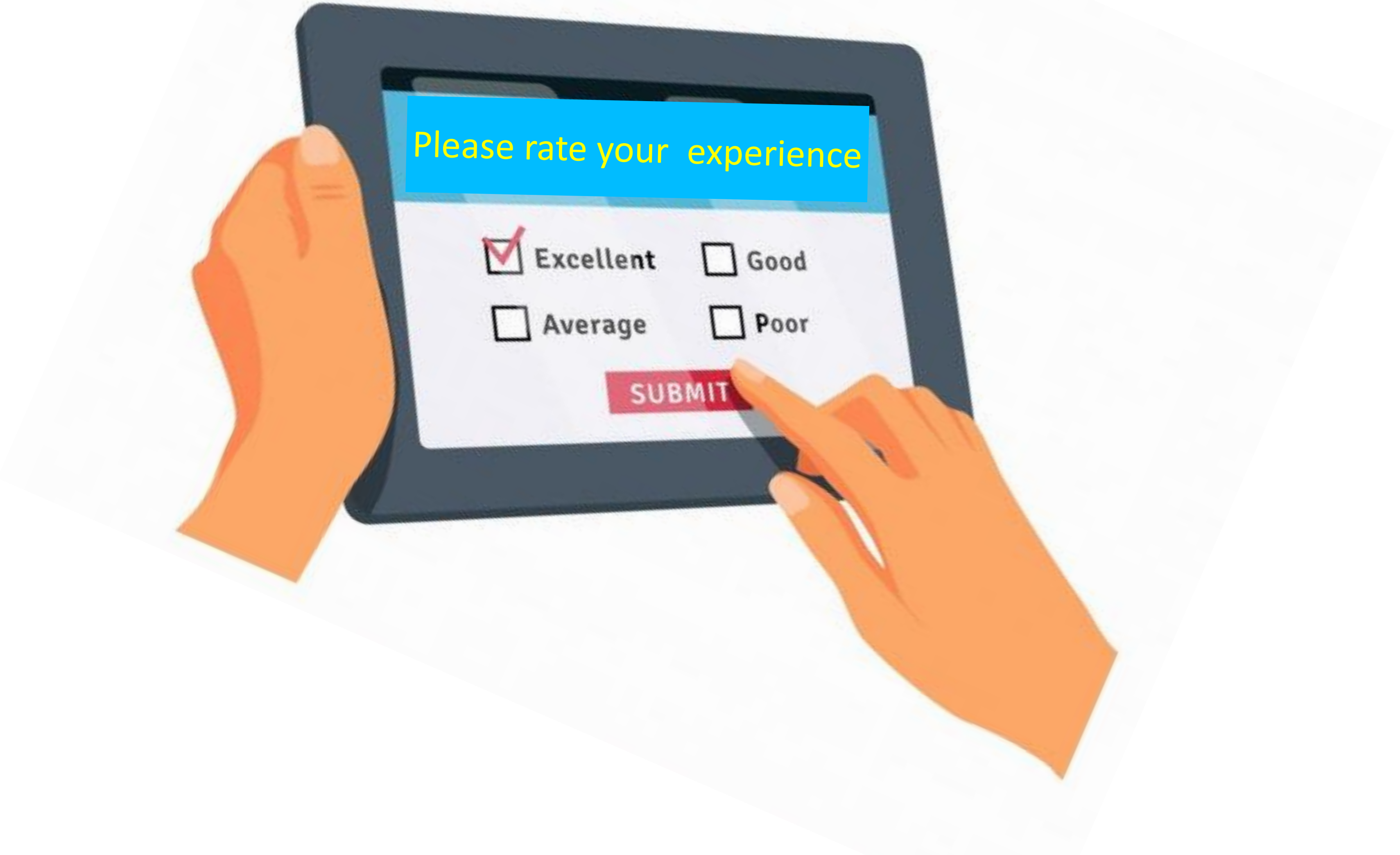


Voice recorder

2. Observations

Observer

3. Surveys


Please rate your experience
Excellent
Good
Average
Poor
SUBMIT

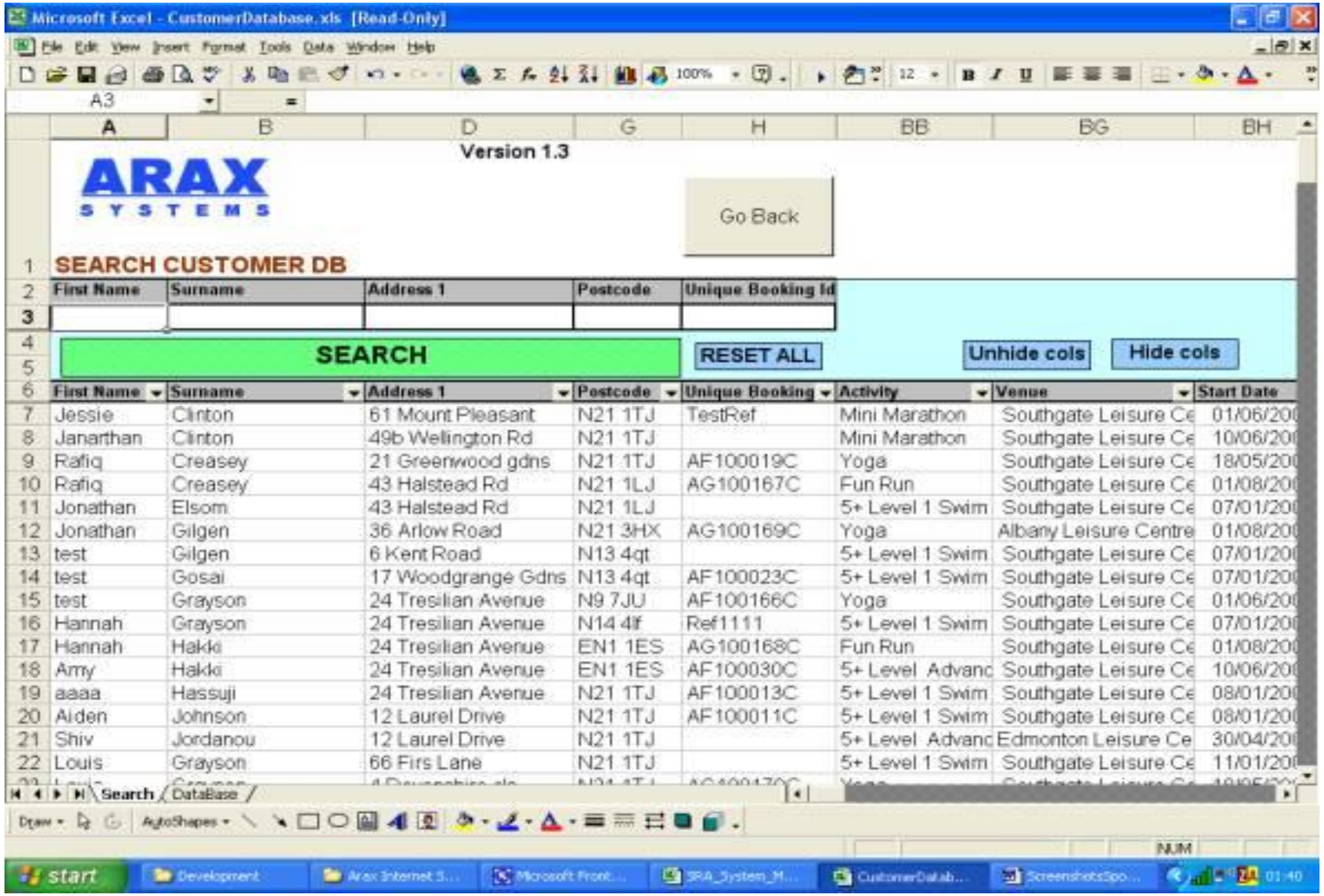
4. Focus groups

5. Secondary data

Example of sources of secondary data: Agencies such as industry bodies, government agencies, libraries and local councils

6. Retrieval from your own database

This is associated with business transactions like maintaining the general ledger (book keeping), payroll, billing, inventory management, etc.



First Name	Surname	Address 1	Postcode	Unique Booking	Activity	Venue	Start Date
Jessie	Clinton	61 Mount Pleasant	N21 1TJ	TestRef	Mini Marathon	Southgate Leisure Ce	01/06/20
Janarthan	Clinton	49b Wellington Rd	N21 1TJ		Mini Marathon	Southgate Leisure Ce	10/06/20
Rafiq	Creasey	21 Greenwood gdns	N21 1TJ	AF100019C	Yoga	Southgate Leisure Ce	18/05/20
Rafiq	Creasey	43 Halstead Rd	N21 1LJ	AG100167C	Fun Run	Southgate Leisure Ce	01/08/20
Jonathan	Elsom	43 Halstead Rd	N21 1LJ		5+ Level 1 Swim	Southgate Leisure Ce	07/01/20
Jonathan	Gilgen	36 Arlow Road	N21 3HX	AG100169C	Yoga	Albany Leisure Centre	01/08/20
test	Gilgen	6 Kent Road	N13 4qt		5+ Level 1 Swim	Southgate Leisure Ce	07/01/20
test	Gosai	17 Woodgrange Gdns	N13 4qt	AF100023C	5+ Level 1 Swim	Southgate Leisure Ce	07/01/20
test	Grayson	24 Tresilian Avenue	N9 7JU	AF100166C	Yoga	Southgate Leisure Ce	01/06/20
Hannah	Grayson	24 Tresilian Avenue	N14 4lf	Ref1111	5+ Level 1 Swim	Southgate Leisure Ce	07/01/20
Hannah	Hakki	24 Tresilian Avenue	EN1 1ES	AG100168C	Fun Run	Southgate Leisure Ce	01/08/20
Amy	Hakki	24 Tresilian Avenue	EN1 1ES	AF100030C	5+ Level Advanc	Southgate Leisure Ce	10/06/20
aaaa	Hassuji	24 Tresilian Avenue	N21 1TJ	AF100013C	5+ Level 1 Swim	Southgate Leisure Ce	08/01/20
Aiden	Johnson	12 Laurel Drive	N21 1TJ	AF100011C	5+ Level 1 Swim	Southgate Leisure Ce	08/01/20
Shiv	Jordanou	12 Laurel Drive	N21 1TJ		5+ Level Advanc	Edmonton Leisure Ce	30/04/20
Louis	Grayson	66 Firs Lane	N21 1TJ		5+ Level 1 Swim	Southgate Leisure Ce	11/01/20

Interview and Focus groups sdata

Interview Transcript

INTERVIEW
Audio File: Sample Interview

Transcribed: October 26, 2009

Interviewer: So is it generally people who know you already from meeting you at a fair?

Interviewee: It's either that or once in a while I'll have just somebody that searches for Astrology and I'll come up and call but most often it's either people that I know or people that they refer me to. I get a lot of referrals.

Interviewer: Do you keep any sort of regular schedule or do you see clients at certain intervals?

Interviewee: No!

Interviewer: No?

Interviewee: I have clients that I know I'm probably going to see once a month. I have this sweet lady from India that came here five years ago, that was escaping from an abusive husband and had never had a job in her life, had never made her own decisions. She's now Vice President of one of the University Federal Credit Unions.

Interviewer: Oh, really?

Interviewee: And it's just that every month, "You can do this, you can do this! Come on, you can do this!" Because I see it here in the cards and she's just amazing. I'm just so proud because she came out under [Inaudible 9:56] and never had any options and a friend got her away from this guy she was married to and sent her over here and she just landed. I mean she just landed. (Laughs)

Interviewer: Wow!

Interviewee: Yeah! She's just so bright and she comes from that background. We have monthly sessions but she calls me when she's ready.

Interviewer: Yeah! Do you schedule appointments days in advance?

Interviewee: Yes, I have a schedule. I'm going to be in a fair in Fort Worth this weekend but I have a schedule already for the next Saturday here at an acupuncture clinic.

Interviewer: Oh!

Interviewee: Yeah at West Gate.

Survey data

Illustrative+survey+for+research+design+seminar%2C+2017_September+27%2C+2017_11.25.xlsx - Excel

	StartDate	EndDate	Status	IPAddress	LocationLatitude	LocationLongitude	Distributi	UserLanguage	A1_1	A1_2	A1_3	A1_4	A1_5	A1_6	A1_7	A1_8	A2_1	A2_2	A2_3	A2_4	A3_1	A3_2	A3_3	A3_4
2	25.09.2017 05:36	25.09.2017 05:39	0	158.38.82.192	62.737503051758	7.1591033935547	email	EN	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
3	25.09.2017 06:20	25.09.2017 06:24	0	158.38.157.27	62.737503051758	7.1591033935547	email	EN	4	1	1	7	4	7	1	7	7	7	7	1	7	7	5	
4	25.09.2017 06:30	25.09.2017 06:36	0	158.38.157.2	62.737503051758	7.1591033935547	email	EN	5	3	6	4	3	7	5	5	6	5	3	3	2	2	2	
5	25.09.2017 06:39	25.09.2017 06:43	0	158.38.157.102	62.737503051758	7.1591033935547	email	EN	1	2	2	5	2	5	2	6	7	2	6	2	5	2	3	
6	25.09.2017 07:06	25.09.2017 07:09	0	158.38.157.42	62.737503051758	7.1591033935547	email	EN	4	4	5	5	3	5	4	4	5	6	4	3	5	3	5	
7	25.09.2017 08:15	25.09.2017 08:24	0	158.38.68.124	62.733306884766	7.1833038330078	email	EN	5	4	3	4	5	6	4	5	3	4	4	5	5	4	3	
8	25.09.2017 08:38	25.09.2017 08:42	0	158.38.157.130	62.737503051758	7.1591033935547	email	EN	5	4	2	5	1	6	2	6	7	4	6	4	2	2	2	
9	25.09.2017 08:40	25.09.2017 08:43	0	176.11.143.179	59.949996948242	10.75	email	EN	5	2	2	6	4	6	2	6	6	2	5	1	4	4	2	
10	25.09.2017 08:37	25.09.2017 08:51	0	158.38.159.136	62.737503051758	7.1591033935547	email	EN	5	2	4	6	6	6	4	6	5	4	4	4	4	5	4	
11	25.09.2017 08:49	25.09.2017 08:56	0	158.38.158.63	62.737503051758	7.1591033935547	email	EN	6	5	7	3	5	6	3	5	5	6	5	3	7	1	5	
12	25.09.2017 09:42	25.09.2017 09:48	0	92.220.204.9	62.730697631836	7.3274993896484	email	EN	5	2	5	5	5	5	3	4	5	7	3	3	5	5	3	
13	25.09.2017 09:59	25.09.2017 10:04	0	176.11.205.76	59.949996948242	10.75	email	EN	7	2	2	7	4	5	1	7	5	4	5	1	6	7	6	
14	25.09.2017 10:13	25.09.2017 10:17	0	158.38.157.96	62.737503051758	7.1591033935547	email	EN	5	2	2	6	3	5	3	6	6	4	7	3	5	3	3	
15	25.09.2017 06:31	25.09.2017 10:21	0	92.220.197.154	62.737503051758	7.1591033935547	email	EN	4	2	2	5	7	6	2	7	6	3	6	2	4	4	4	
16	25.09.2017 10:15	25.09.2017 10:28	0	158.38.156.163	62.737503051758	7.1591033935547	email	EN	2	1	3	5	2	7	3	6	6	4	5	3	3	2	2	
17	25.09.2017 11:26	25.09.2017 11:31	0	158.38.88.9	62.737503051758	7.1591033935547	email	EN	3	2	4	5	6	5	4	5	5	4	5	3	5	5	5	
18	25.09.2017 12:02	25.09.2017 12:12	0	77.18.54.90	59.949996948242	10.75	email	EN	2	2	4	5	2	5	3	6	7	2	6	2	5	6	3	
19	25.09.2017 12:33	25.09.2017 12:39	0	92.220.200.158	62.737503051758	7.1591033935547	email	EN	7	7	7	7	7	7	1	7	7	5	5	5	5	3	3	
20	25.09.2017 12:11	25.09.2017 12:48	0	158.38.88.127	62.737503051758	7.1591033935547	email	EN	7	5	6	7	5	7	2	6	6	3	5	3	6	5	5	
21	26.09.2017 01:24	26.09.2017 01:31	0	158.38.157.150	62.737503051758	7.1591033935547	email	EN	2	1	4	7	3	6	3	6	6	5	5	4	5	5	5	
22	26.09.2017 02:19	26.09.2017 02:25	0	158.38.158.47	62.737503051758	7.1591033935547	email	EN	3	2	2	5	2	2	2	7	7	3	4	4	3	4	2	
23	26.09.2017 02:22	26.09.2017 02:26	0	158.38.157.215	63.064407348633	7.5012969970703	email	EN	2	2	2	5	5	7	2	6	6	3	5	7	6	5	5	
24	26.09.2017 03:24	26.09.2017 03:37	0	158.38.157.181	62.737503051758	7.1591033935547	email	EN	6	3	5	7	7	7	6	5	6	2	6	7	3	7	5	
25	26.09.2017 01:19	26.09.2017 03:37	0	158.38.158.154	62.737503051758	7.1591033935547	email	EN	2	2	2	5	2	2	2	6	5	5	5	2	3	3	3	
26	26.09.2017 05:20	26.09.2017 05:24	0	158.38.157.182	62.737503051758	7.1591033935547	email	EN	5	2	2	6	3	6	2	6	6	4	4	5	7	6	5	
27	26.09.2017 05:22	26.09.2017 05:27	0	158.38.157.80	62.737503051758	7.1591033935547	email	EN	3	2	5	3	3	3	2	6	6	2	6	3	5	5	3	
28	26.09.2017 05:57	26.09.2017 06:19	0	158.38.158.62	62.737503051758	7.1591033935547	email	EN	5	3	6	6	5	4	5	5	6	2	6	3	1	3	3	
29	26.09.2017 06:14	26.09.2017 06:25	0	158.38.159.50	62.737503051758	7.1591033935547	email	EN	6	2	4	5	7	4	4	5	6	2	5	5	5	4	3	
30	26.09.2017 01:20	26.09.2017 14:58	0	158.38.88.89	62.737503051758	7.1591033935547	email	EN	2	2	6	7	7	2	2	7	5	5	6	1	5	2	6	

Customer database

A2 fx 2

	A	B	C	D	F	L	M
1	ID	Customers.Company	LastName	FirstName	JobTitle	City	StateProvince
2	2	Customer A	Bedecs	Anna	Owner	Seattle	WA
3	3	Customer B	Gratacos Solsona	Antonio	Owner	Boston	MA
4	4	Customer C	Axen	Thomas	Purchasing Representative	Los Angeles	CA
5	5	Customer D	Lee	Christina	Purchasing Manager	New York	NY
6	6	Customer E	O'Donnell	Martin	Owner	Minneapolis	MN
7	7	Customer F	Pérez-Olaeta	Francisco	Purchasing Manager	Milwaukee	WI
8	8	Customer G	Xie	Ming-Yang	Owner	Boise	ID
9	9	Customer H	Andersen	Elizabeth	Purchasing Representative	Portland	OR
10	10	Customer I	Mortensen	Sven	Purchasing Manager	Salt Lake City	UT
11	11	Customer J	Wacker	Roland	Purchasing Manager	Chicago	IL
12	12	Customer K	Krschne	Peter	Purchasing Manager	Miami	FL
13	13	Customer L	Edwards	John	Purchasing Manager	Las Vegas	NV
14	14	Customer M	Ludick	Andre	Purchasing Representative	Memphis	TN
15	15	Customer N	Grilo	Carlos	Purchasing Representative	Denver	CO
16	16	Customer O	Kupkova	Helena	Purchasing Manager	Honolulu	HI
17	17	Customer P	Goldschmidt	Daniel	Purchasing Representative	San Francisco	CA
18	18	Customer Q	Bagel	Jean Philippe	Owner	Seattle	WA

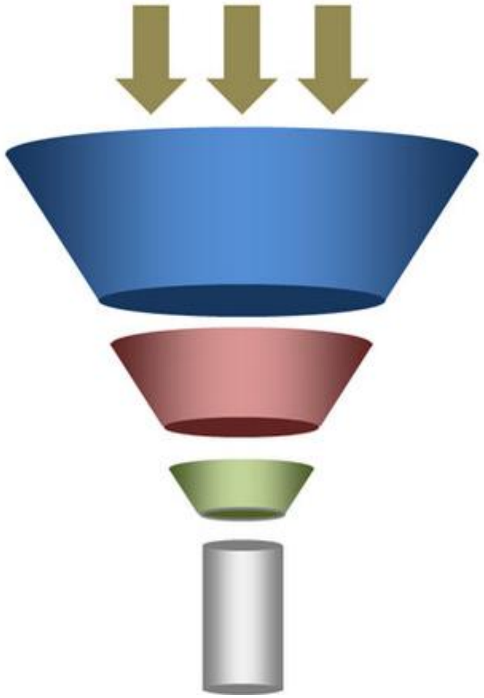
Information

- Data that is (1) accurate and timely, (2) specific and organized for a purpose, (3) presented within a context that gives it meaning and relevance, and (4) can lead to an increase in understanding and decrease in uncertainty.

- Information isn't just data that's been neatly filed away, it has to be ordered in a way that gives meaning and context.

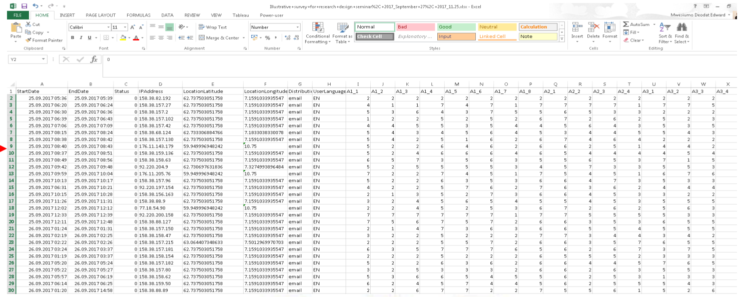
- Information provides answers to "who", "what", "where", and "when" questions

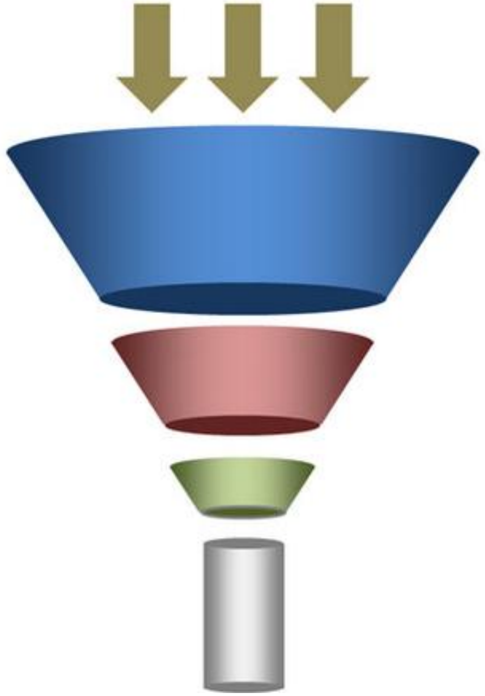
DATA → 0202656788, Smith,123 King St, London, UK, J

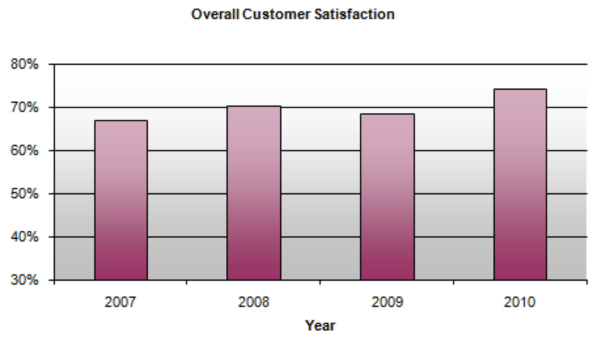


INFORMATION → John Smith
123 King Street
London, United Kingdom
(020) 2656788

DATA

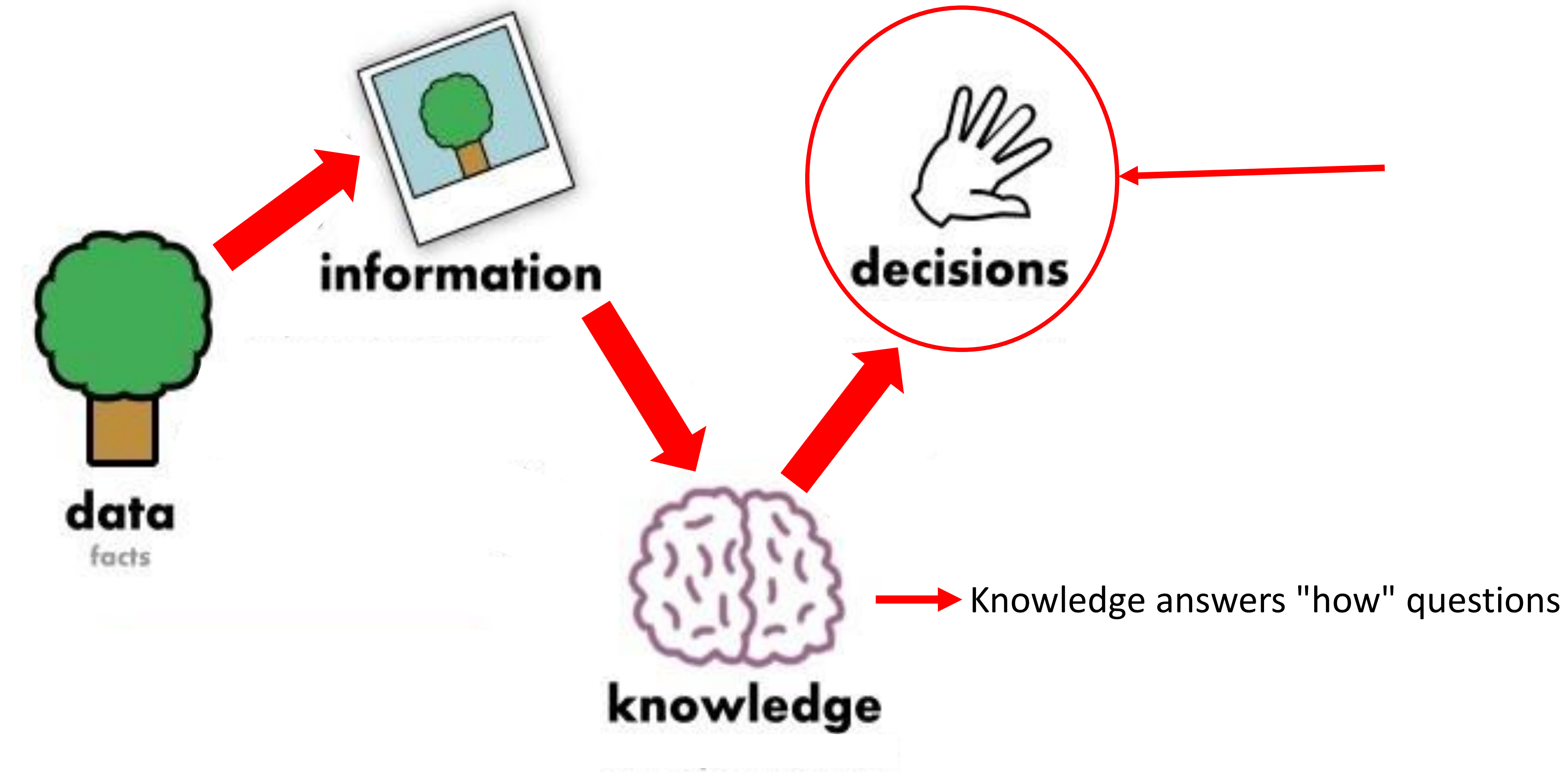






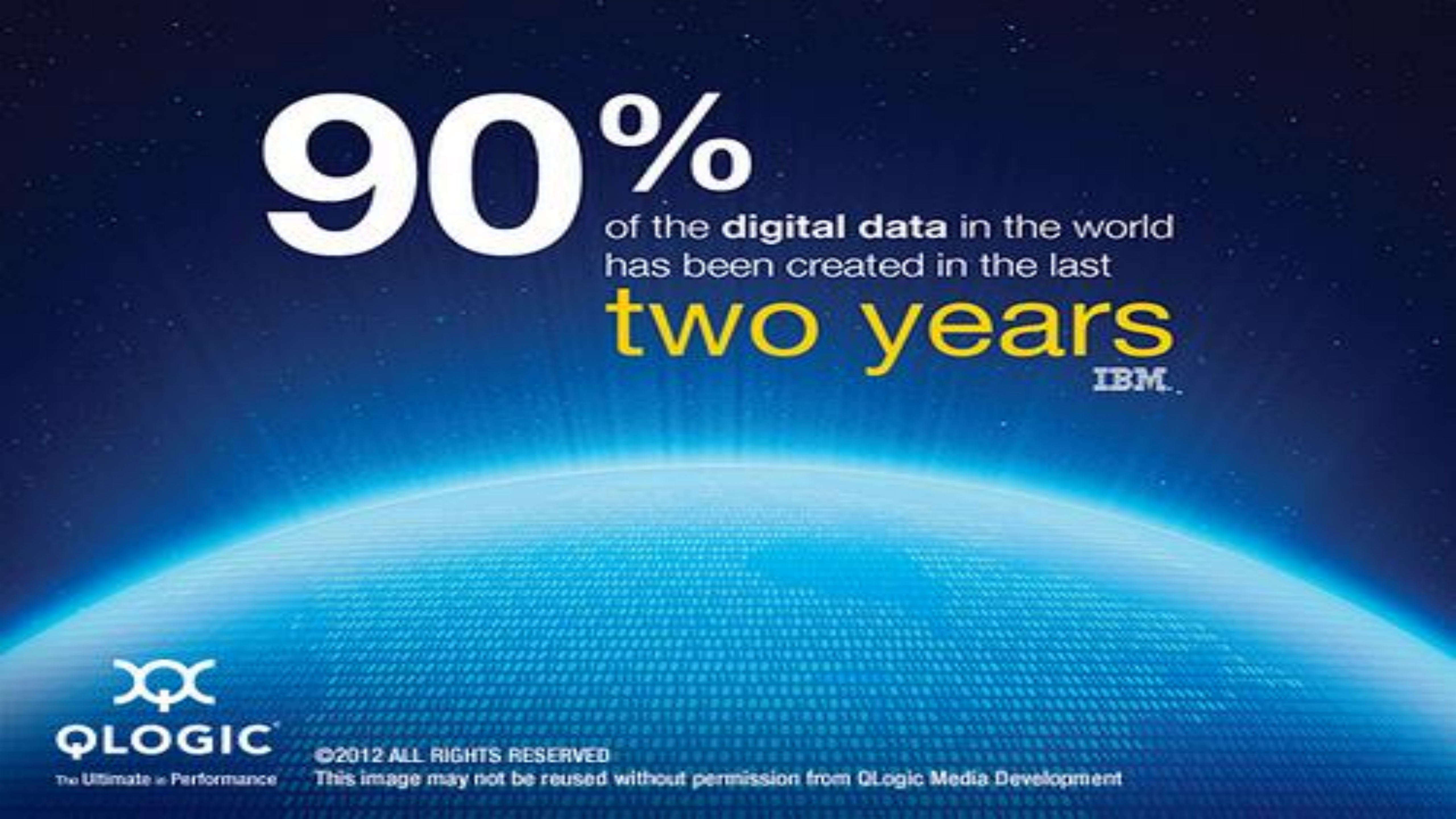
INFORMATION

Where does knowledge come from?



Data explosion

The arrival of the mainstream internet in the 1990s
expanded business capabilities and the role of



90%
of the digital data in the world
has been created in the last
two years
IBM
QLOGIC
The Ultimate in Performance
©2012 ALL RIGHTS RESERVED
This image may not be reused without permission from QLogic Media Development

DOMO

DATA NEVER SLEEPS 5.0

How much data is generated *every minute?*

90% of all data today was created in the last two years—that's 2.5 quintillion bytes of data per day. In our 5th edition of Data Never Sleeps, we bring you the latest stats on just how much data is being created in the digital sphere—and the numbers are staggering.

Easily available data.........

Annals of Tourism Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/annals

ELSEVIER

Research Note

Effect of component failure on tour package evaluation

D. Mwesiumo[a,*], N. Halpern[b]

[a] Department for Logistics, Molde University College – Specialized University in Logistics, 6402 Molde, Norway
[b] Department for Management and Organisation, Kristiania University College, 0107 Oslo, Norway

ARTICLE INFO

Associate Editor: ShiNa Li

Introduction

Tour packages comprise two or more service components sold as a bundle, and multiple actors are often used to deliver individual components, thus forming a value chain. Several studies have attempted to examine the contribution of individual components on overall tourist satisfaction and find that some contribute more to satisfaction than others (e.g. see Chan, Hsu, & Baum, 2015; Räikkönen & Honkanen, 2013). However, there also seems to be a consensus amongst researchers that tourists tend to consider tour packages as unified products, and therefore evaluate their experiences holistically rather than separately (Zach & Racherla, 2011).

Previous studies on tour package satisfaction are based largely on the one-way quality model, which examines the impact of the presence of a quality element on satisfaction, and treats the relationship between tourist satisfaction and the evaluation of individual components as linear. However, Kano's two-dimensional model suggests that quality attributes and customer satisfaction have an asymmetric and nonlinear relationship, which makes it important to consider one-dimensional quality attributes along with attributes whose presence does not have the same effect as their absence. In other words, Kano's model recognises that certain factors or elements can engender satisfaction but their absence does not necessarily lead to dissatisfaction (Alegre & Garau, 2010). This view is upheld by the findings of several studies such as Cadotte and Turgeon (1988) and Zhu and Tsai (2010). In light of Kano's model, an intriguing question in the context of tour package operations is whether individual service component failure can affect tourist evaluation of the entire bundle. That is the focus of this research note.

Methodology

The analysis is based on 286 customer reviews of tour packages offered by a North American tour operator that uses independent local actors at the destination to provide service components such as accommodation and tour guiding. Reviews were obtained from an independent online platform that provides space for visitors to write reviews about their experiences with a company as well as to give an overall rating. Using qualitative research software NVivo, reviews were coded according to five components of the customer experience; customer service (Cuss), information provided (Info), trip coordination (Coord), accommodation (Accom), and tour guiding (Tgd). Each of these was coded with 1 if positively evaluated, 2 if negatively evaluated, or 3 if not mentioned/neutral. The rating given by the customer (from 1 star being worst to 5 stars being best) was treated as a proxy for overall evaluation of the package, and was coded at three levels: low (1–2 stars), moderate (3 stars) and high (4–5 stars). The five components were used as independent variables and the corresponding overall evaluation as the dependent variable. A standard multinomial logit model is used to determine the odds of a customer giving a moderate or high overall evaluation of the package when a particular component is

$$\ln\left[\frac{P_{ij}}{P_{im}}\right] = \eta_{ij} + \eta_{im} = \beta\left[X_{ij} - X_{im}\right] \tag{4}$$

where P_{im} is equal to $1 - P_{ij}$, thus, odds ratios can be derived from the estimated coefficients.

Results

Table 1 presents results of the model estimation. The model fit was evaluated using Nagelkerke's rho-squared, which has a value of 72 per cent, indicating that the independent variables explain substantially the variations in the dependent variable. For each independent variable, a probability is reported quantifying the odds that the overall evaluation of the package by a customer is moderate (3 stars) or high (4 or 5 stars), given that the customer had a negative experience of the individual component.

The results show that when a customer has a negative experience with any of the individual components then the probability of that customer giving a high overall evaluation becomes significantly small (ranging from 0.008 to 0.07). Results also show that the odds of the customer giving a moderate overall evaluation increases for two components; accommodation and trip coordination (0.26 and 0.37, respectively) but they remain significantly low for the other components, especially when the customer experiences poor customer service ($P_2 = 0.07$). This means that if a customer experiences poor customer service, the chance that they will give a 3 star overall evaluation is 0.07 while giving a 4 or 5 star evaluation is 0.009, when other aspects of the package are not depicted.

Conclusion

This study finds that when a component of a value chain fails, all actors in the chain are likely to suffer as a result of the poor aggregate evaluation. This is important in an era when consumers increasingly rely on evaluations of previous customers (e.g. see Casado-Díaz, Pérez-Naranjo, & Sellers-Rubio, 2017). Theoretically, the finding supports the contention that actors in tourism value chains should work collectively rather than individually (Song, 2012; Zhang, Song, & Huang, 2009).

To the managers of tourism firms involved in bundled products, the results suggest that there is no room to engage in

Table 1
Estimation results.

	3 star rating			4 or 5 star rating		
	Coefficient	p-Value	P_2	Coefficient	p-Value	P_3
(Intercept)	1.988	0.005**		5.548	0.000**	
NegCuss	−2.520	0.003**	0.070	−4.218	0.000**	0.009
NegInfo	−1.508	0.010**	0.180	−3.749	0.000**	0.020
NegCoord	−0.438	0.430 ns	0.370	−2.333	0.001**	0.060
NegAccom	−0.923	0.187 ns	0.260	−2.249	0.011**	0.070
NegTgd	−1.680	0.005**	0.160	−4.772	0.000**	0.008
−2 log-likelihood:	73.4					
Nagelkerke rho-squared (ρ^2)	72.2%					
Number of observations:	286					

Notes:
**Significant at $P < 0.01$; ns not significant.
P_2: Probability that when a given variable is negatively evaluated in a review, the overall evaluation of a package becomes moderate (3 stars) rather than low (1 or 2 stars).
P_3: Probability that when a given variable is negatively evaluated in a review, the overall evaluation of a package becomes high (4 or 5 stars) rather than low (1 or 2 stars).

Analysis done by using customer reviews scraped from a review platform

Customer reviews scrapped from a website

DATA

★★★★★ *This review is featured by Tours4Fun*

Carmen of Fitchburg, MA ✔ Verified Reviewer $ Verified Buyer

Original review: Sept. 22, 2017

My kids, grandkids and I went to Niagara Falls Fireworks Cruise through Tours4Fun and it was an amazing experience. They had so much fun. I saw Tours4Fun when I was searching for a hotel. I was shown a tool where I could get videos, places, and things to do in places. I like that the price was affordable and I purchased a trip online. The only person who had a problem was my daughter because we bought the tickets same day and when we went to Canada for the fireworks, she had a

★★★★★

James of Temperance, MI ✔ Verified Reviewer $ Verified Buyer

Original review: March 28, 2018

Tours4Fun had some good specials. They had little packages for things to go see. Basic[illegible] only did the plane ride. But for some reason, I had a hard time purchasing online. I could si[illegible] deals, but it took me a while to search for the same one. It had taken a little time. N[illegible]eless, liked the Grand Canyon West Rim, the Scenic Airplane Tour. The experience was r[illegible]eat. We planning to go back someday, and when we go back, we're looking at those helicopte[illegible] get

★★★☆☆

J. of Ca, CA ✔ Verified Reviewer $ Verified Buyer

Original review: March 25, 2018

I booked a tour last year from Tours4Fun. I looked at the detail and it matched what we were looking for. Our Upper Antelope Canyon and Horseshoe Bend Tour were really good. I really enjoyed it. The local guide was cool. But the guy that picked us up at the hotel, the one who's driving us all day, did not give enough informative detail about when we're gonna arrive at the destination or the destination detail. But it was not a bad experience.

★☆☆☆☆

Julie of Waregem, Other ✔ Verified Reviewer $ Verified Buyer

Original review: March 24, 2018

Do not book with them!!! It's a ripoff!!! First there were issues with booking my trip. Then they forgot to p[illegible]k me up. I phoned the emergency number, a Chinese woman with bad English picks up. After [illegible] she sends me to a different spot in the city to meet the tour guide. Turns out he was the g[illegible] for [illegible]NG tour! He sent me by Uber to meet a driver! No idea if it was the correct bus se[illegible]on[illegible] spoke ENGLISH!!! After that I was transferred to a different bus, with no lunch sto[illegible]e at 7 am and got to my location at 11 30 pm starving and in a bad mood! No guide to meet

★☆☆☆☆

Harshidha of San Diego, CA ✔ Verified Reviewer $ Verified Buyer

Original review: March 12, 2018

I did online searches for Grand Canyon trip and I purchased a helicopter trip that would pick us up. But I might have misread it because on the day, it was just a bus pick-up. It would have been okay if

★★★★☆

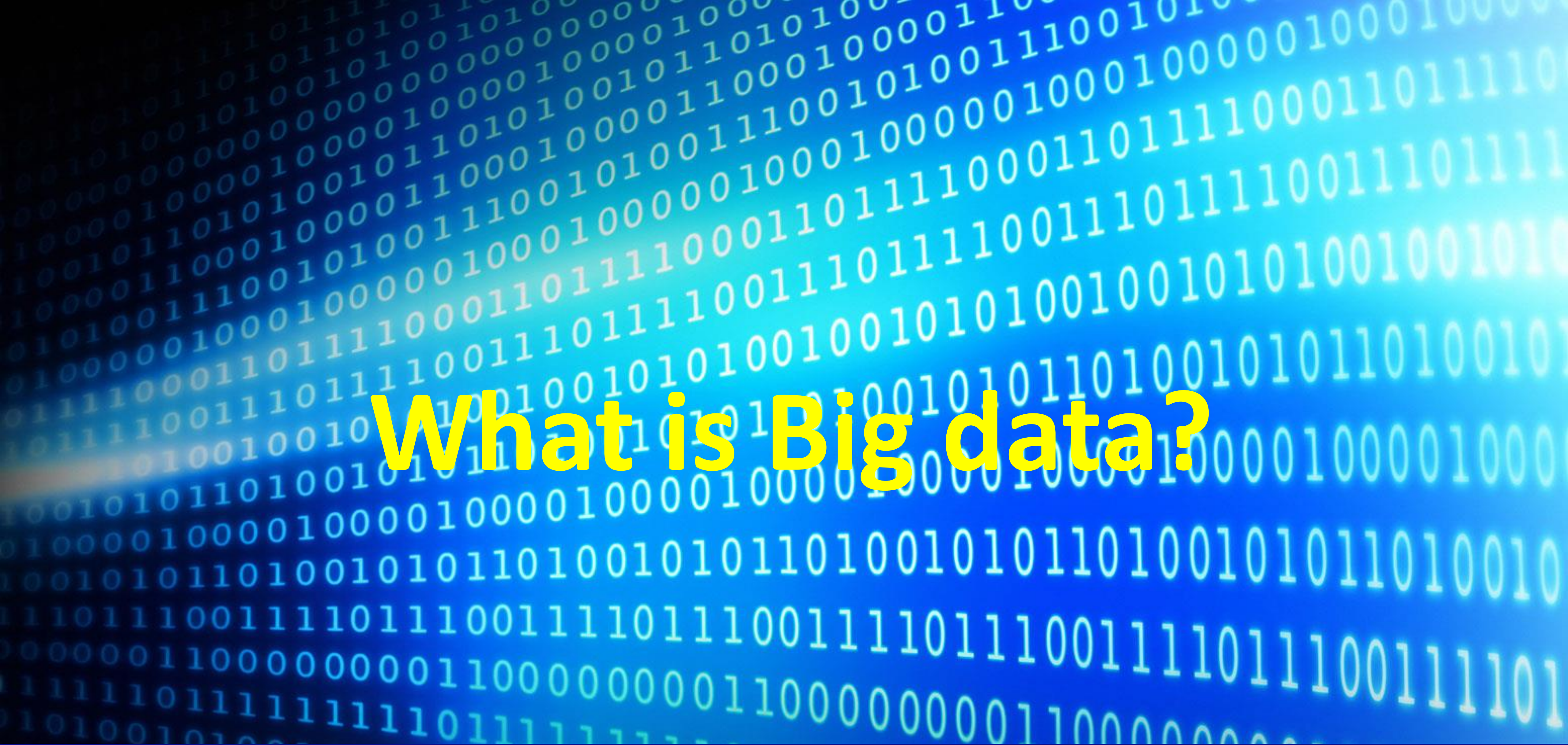
Harold of Aurora, CO ✔ Verified Reviewer $ Verified Buyer

Original review: March 11, 2018

I just Googled tours and I looked at the different ones that seemed to have the best deal. I got Tours4Fun and it was really easy to navigate their website. Tours4Fun was okay for the price and what we were trying to get out of it. But I have a lot of trouble walking. I have a bad back. So I should have done something more to facilitate so I could've gone do a little more. I probably have to ask for that and I didn't ask anything. Or if there's a thing that would say that we have handicapped-accessible things that'll probably be good. But the bus driver was really a nice guy. The bus was clean and there were very nice people that were on the bus with us. We would probably look at them

BIG DATA

What is Big data?

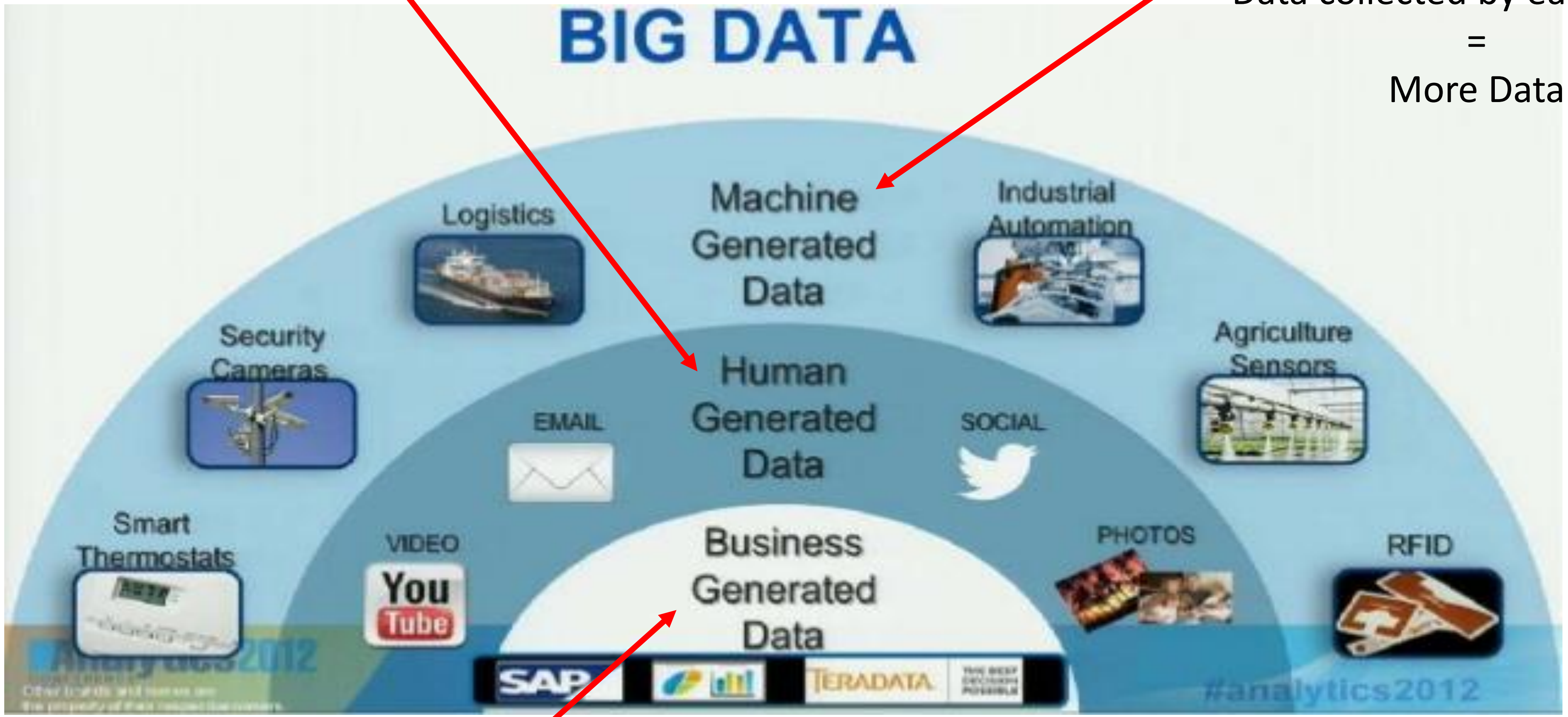


Big data is a buzzword, or catch-phrase, used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process using traditional database and software techniques.

SOURCES OF BIG DATA

People Data generated by people for example on social media, in-app activity, etc

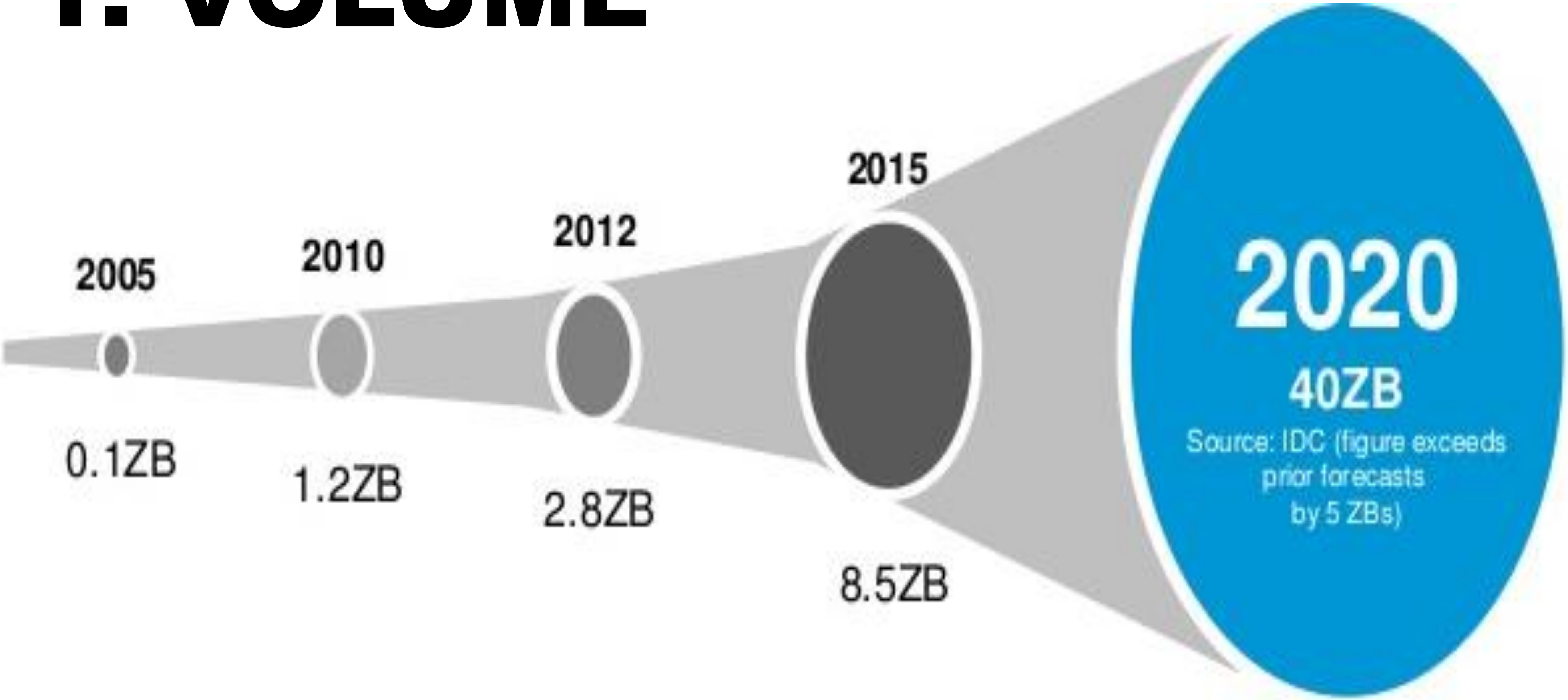
Increasing number of machines that sense
X
Data collected by each device
=
More Data



Highly-structured organizational data (ERP System data etc.)

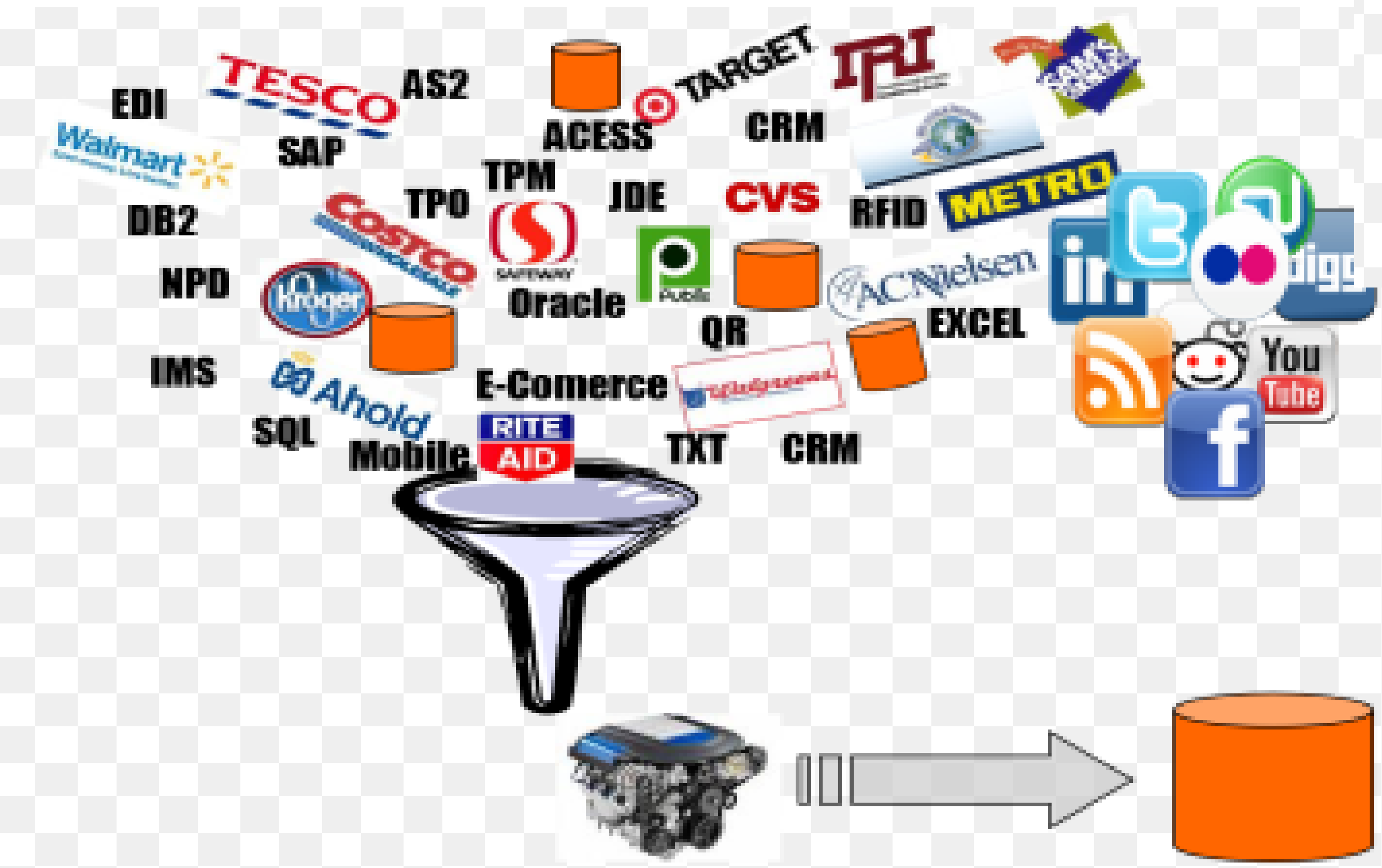
Main Characteristics of Big data

1. VOLUME

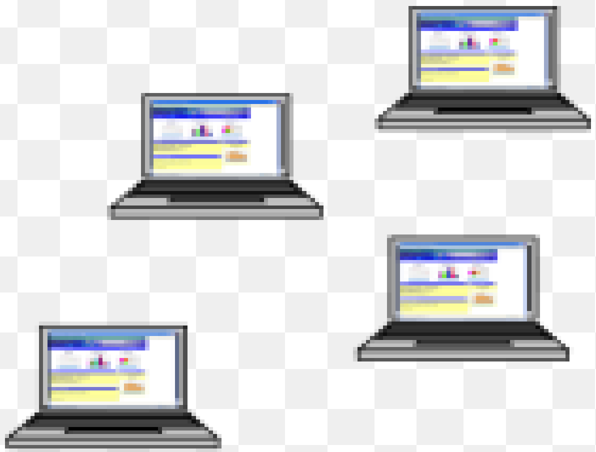


- Every 1.2 years, the volume of business data globally doubles, MIT Sloan reported.
- Looking at the above figures one can easily understand why the name 'Big Data' is given and imagine the challenges involved in its storage and processing.

2. VARIETY

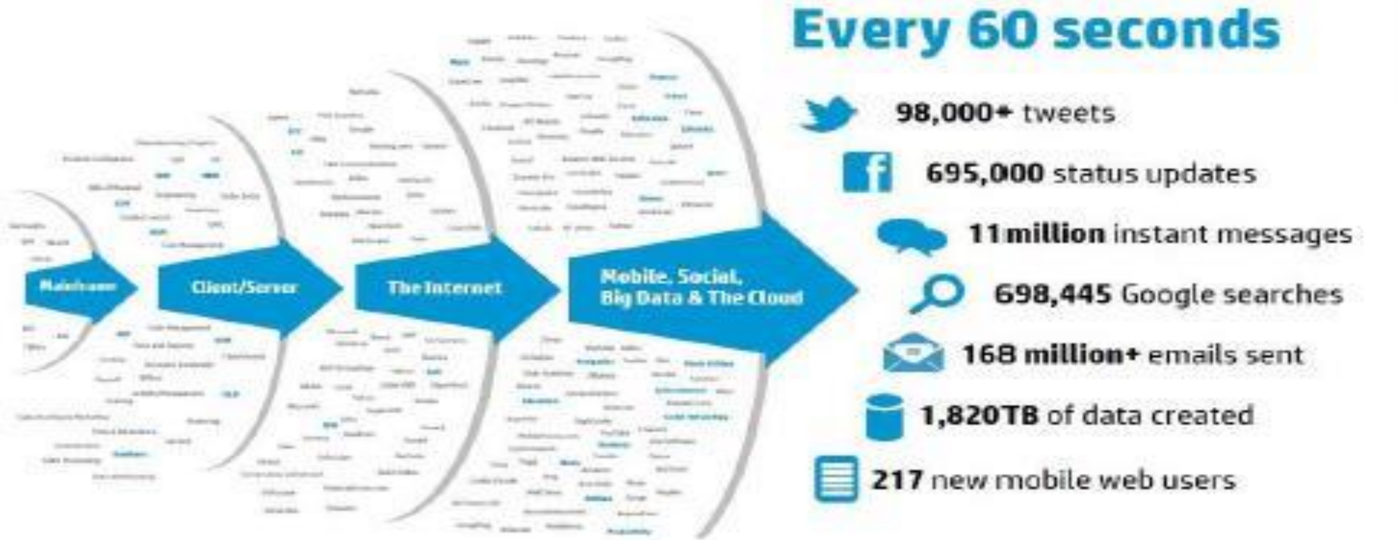


- Various formats, types, and structures
- Text, numerical, images, audio, video,time series, etc



3. VELOCITY

- Increasing speed at which big data is created
- This means high speed of processing is required
- Slow processing means missed opportunities

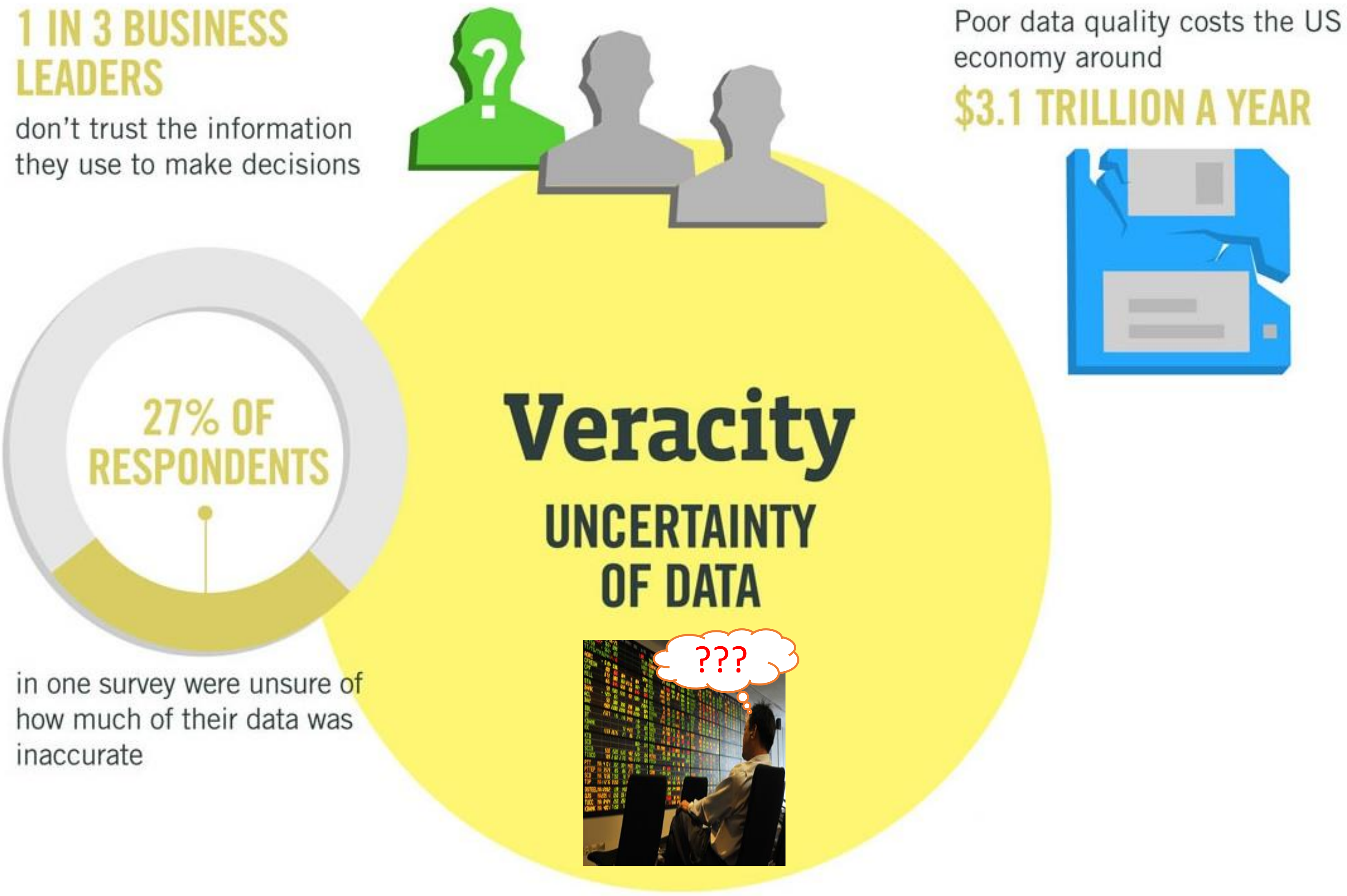


4. VERACITY

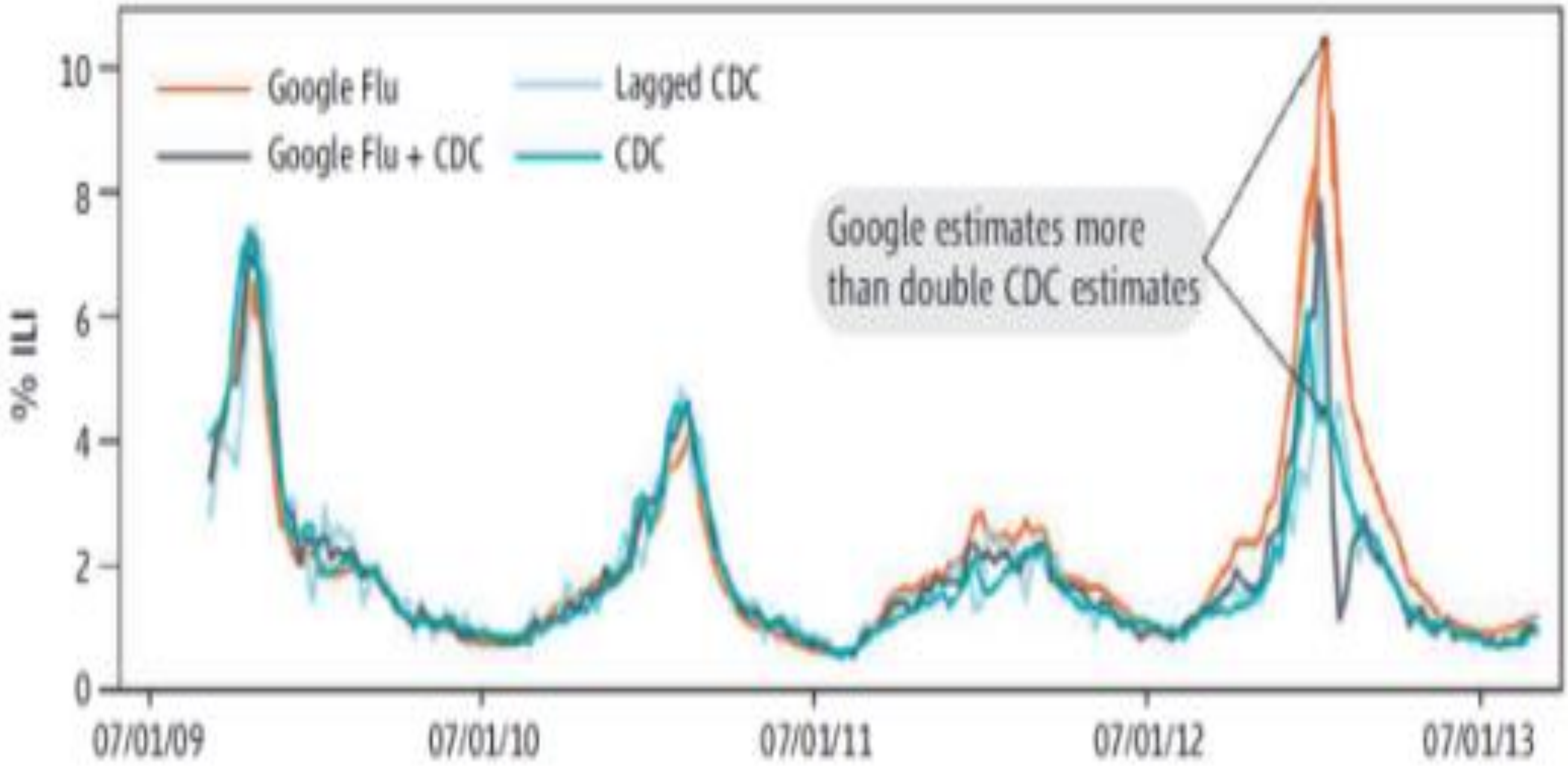
- This is concerned with uncertainty (provenance and reliability) related to big data

- As any volume, variety and velocity increase, the veracity (confidence or trust in the data) drops

- It is important to ensure accuracy, reliability of the data source, and consider context
- within analysis, and how meaningful it is to the analysis based on it



Google Flu Trends: A case of Big Data gone bad



5. VALENCE

- Simply put Valence refers to Connectedness.

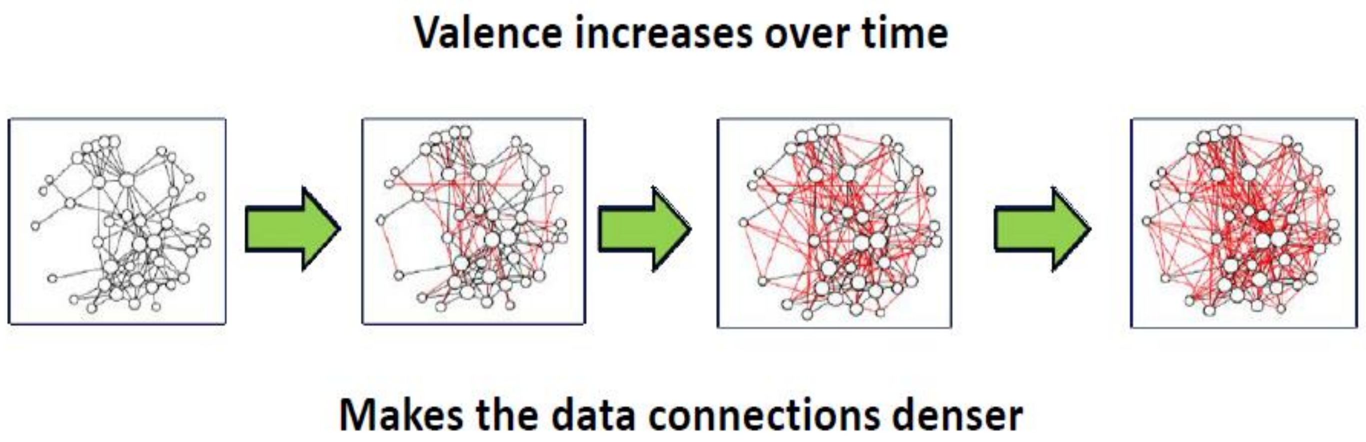
- Data items are often directly connected to one another.
  - A city is connected to the country it belongs to.

  - Two Facebook users are connected because they are friends.

  - An employee is connected to his work place.

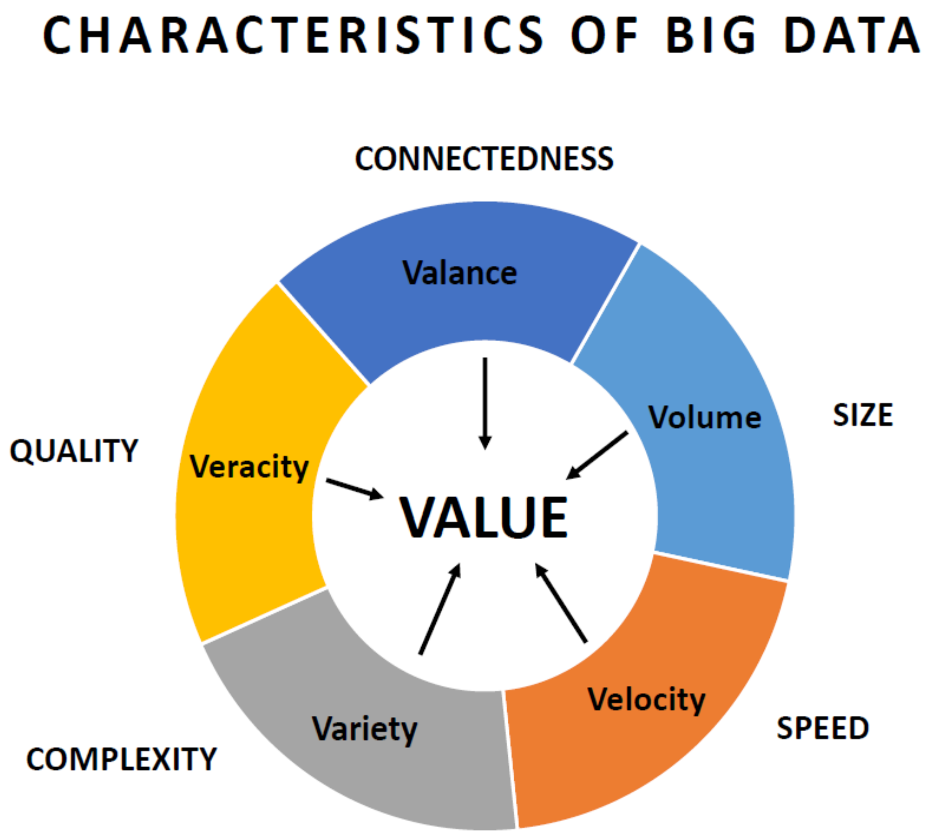
- Data could also be indirectly connected

- For a data collection valence measures the ratio of actually connected data items to the possible number of connections that could occur within the collection.



FOCUS ON VALUE

- The five Vs present a challenging dimensions of big data namely, size, complexity, speed, quality, and connectedness.
- Processing big data must bring about value from insights gained.
- Value is the ability to convert Big Data information into financial reward.
- For example, if you find a relationship between two products at a point of sale, you can recommend them to customers at a website or put the products next to each in a store.



Categories Of 'Big Data'

1. Structured data

Data that is organized in a table format

Employee_ID	Employee_Name	Gender	Department	Salary_In_lacs
2365	Rajesh Kulkarni	Male	Finance	650000
3398	Pratibha Joshi	Female	Admin	650000
7465	Shushil Roy	Male	Admin	500000
7500	Shubhojit Das	Male	Finance	500000
7699	Priya Sane	Female	Finance	550000

Categories Of 'Big Data'

2. Semi-structured

Semi-structured data has some structured form but it is not organized in a table format.

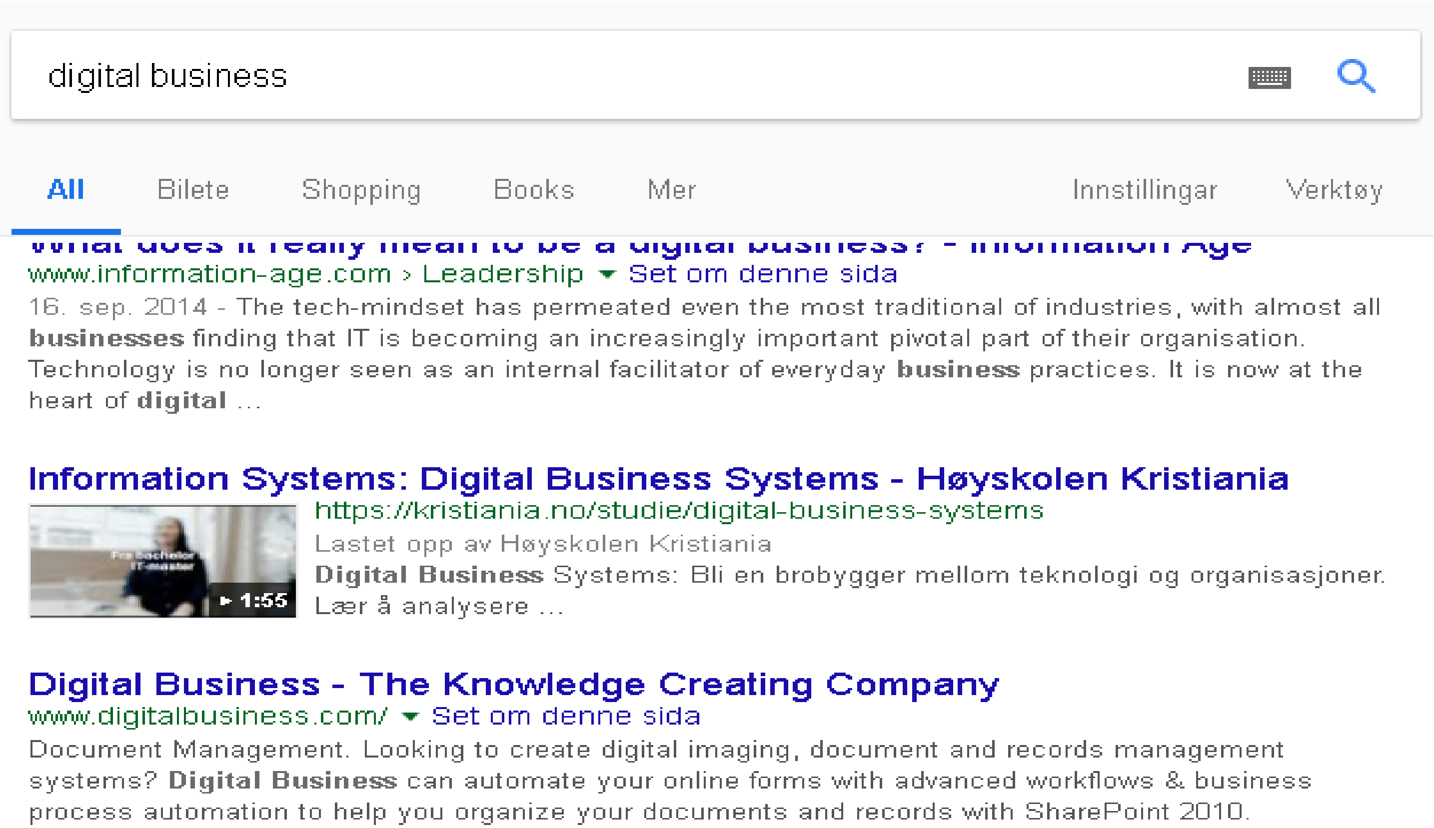
Personal data stored in a XML file-

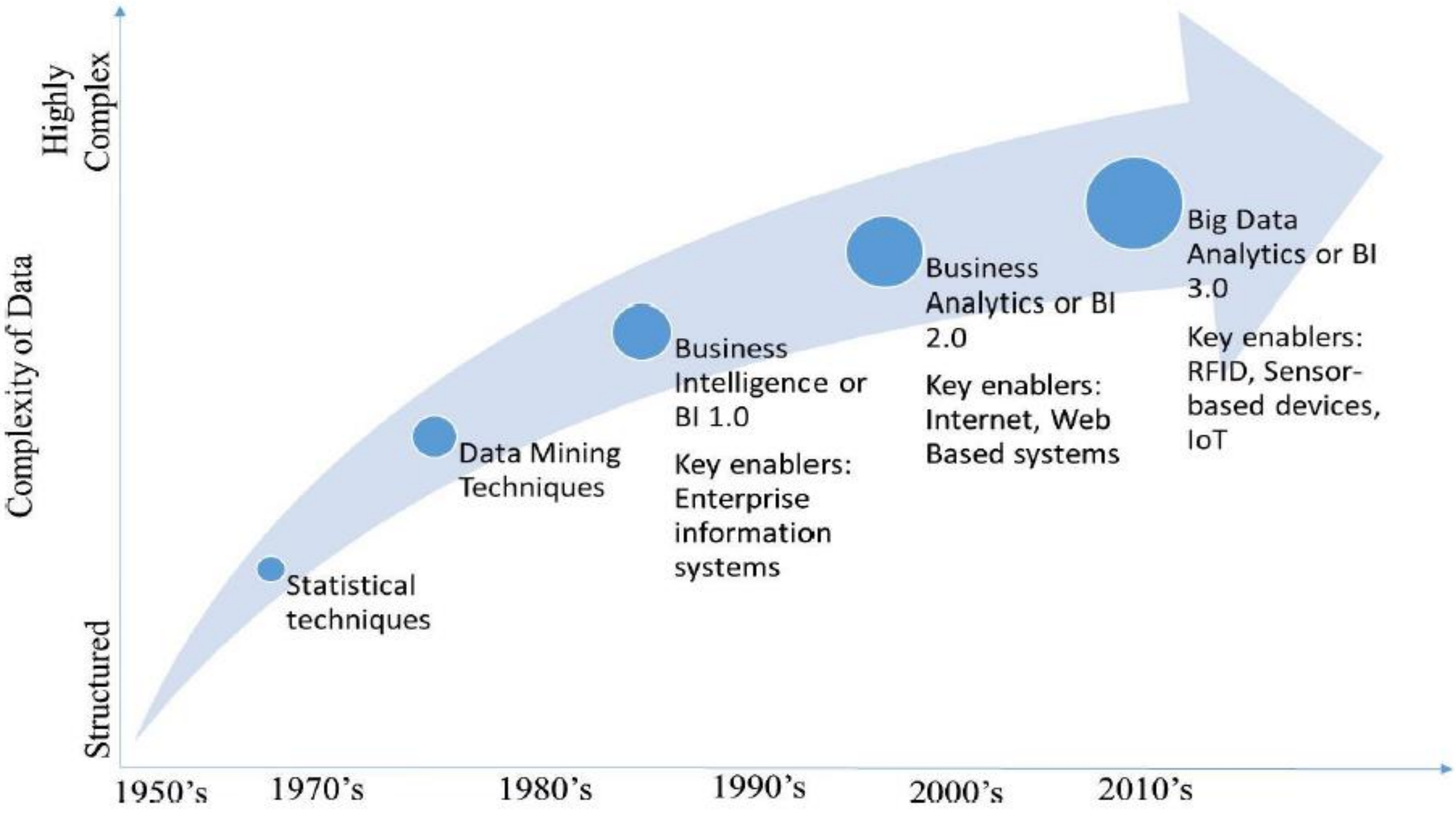
```
<rec><name>Prashant Rao</name><sex>Male</sex><age>35</age></rec>
<rec><name>Seema R.</name><sex>Female</sex><age>41</age></rec>
<rec><name>Satish Mane</name><sex>Male</sex><age>29</age></rec>
<rec><name>Subrato Roy</name><sex>Male</sex><age>26</age></rec>
<rec><name>Jeremiah J.</name><sex>Male</sex><age>35</age></rec>
```

Categories Of 'Big Data'

3. Unstructured

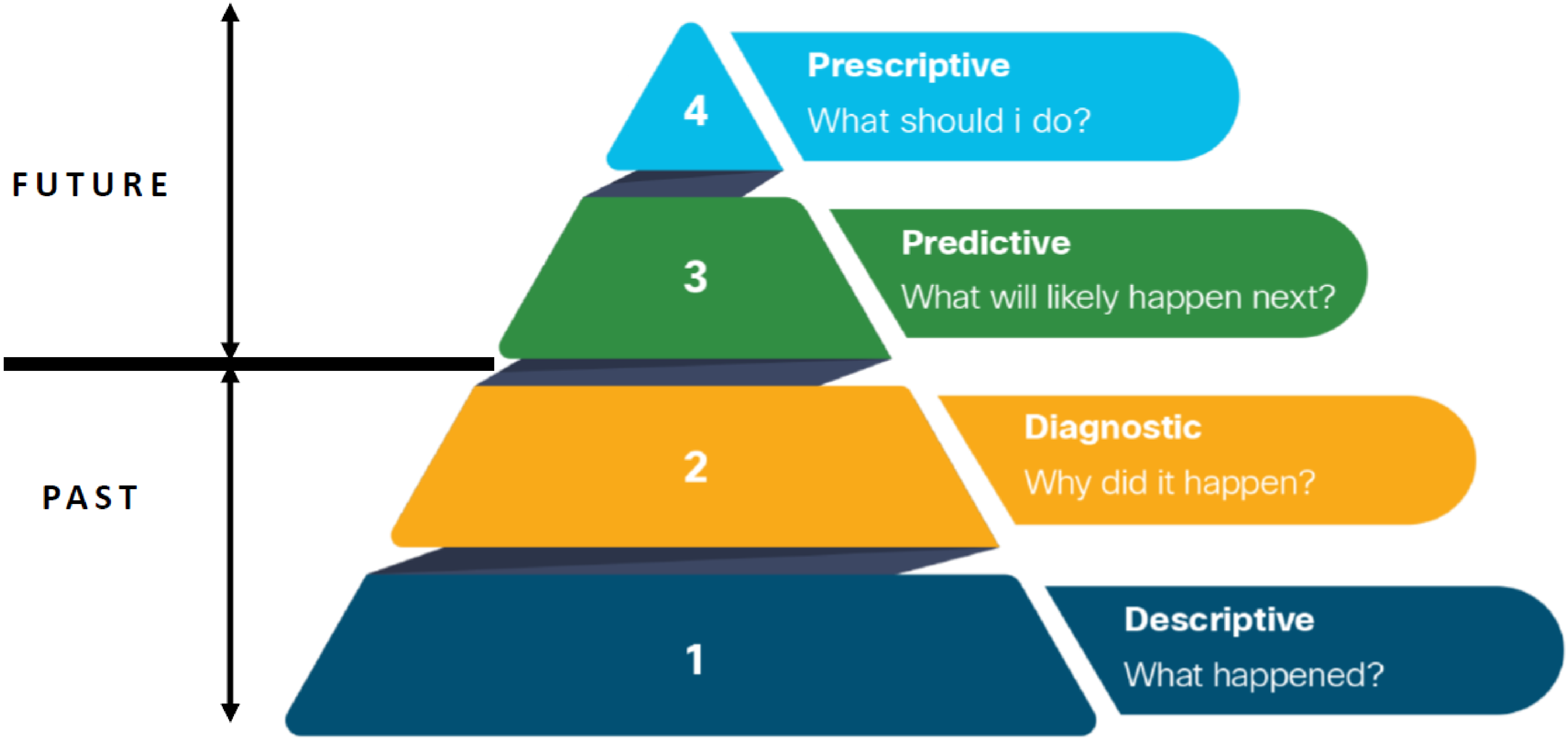
Data that doesn’t fit neatly in a table format





As data complexity increases, more advanced tools and software are required to access information

TYPES OF ANALYTICS



Descriptive analytics

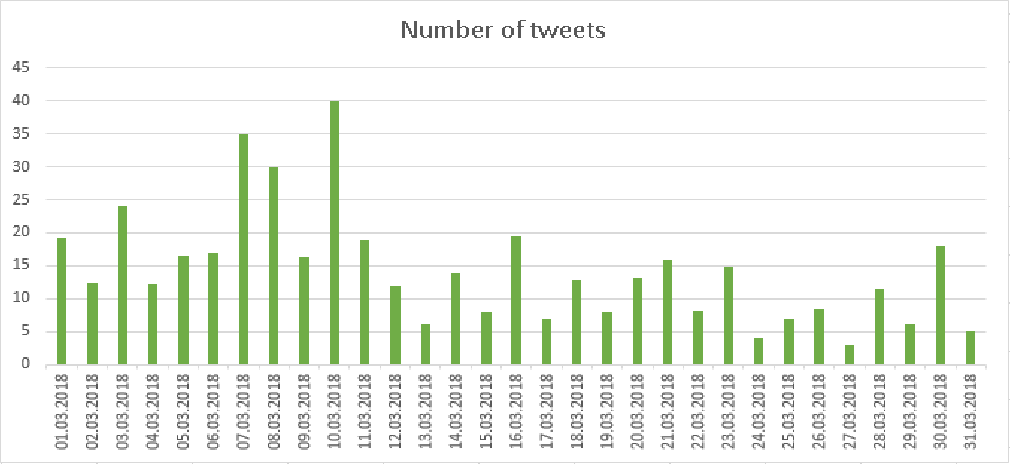
- Descriptive analytics summarize raw data and make it something that is interpretable by humans.

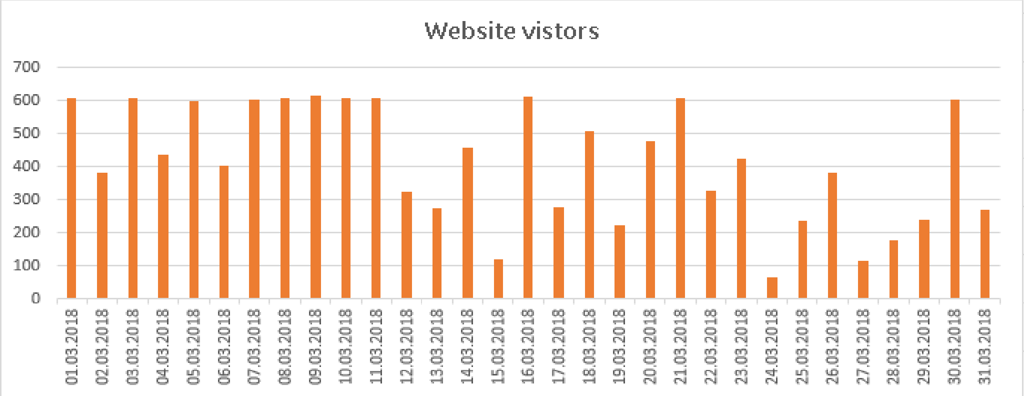
- They are analytics that describe the past- any point of time that an event has occurred, whether it is one minute ago, or one year ago.

e.g. Number of visitors on your website each day in the last three months

Descriptive analytics

Day	Number of tweets	Day	Website vistors
01.03.2018	19	01.03.2018	609
02.03.2018	12	02.03.2018	381
03.03.2018	24	03.03.2018	609
04.03.2018	12	04.03.2018	436
05.03.2018	17	05.03.2018	600
06.03.2018	17	06.03.2018	405
07.03.2018	35	07.03.2018	603
08.03.2018	30	08.03.2018	609
09.03.2018	16	09.03.2018	615
10.03.2018	40	10.03.2018	606
11.03.2018	19	11.03.2018	609
12.03.2018	12	12.03.2018	324
13.03.2018	6	13.03.2018	275
14.03.2018	14	14.03.2018	457
15.03.2018	8	15.03.2018	120
16.03.2018	19	16.03.2018	612
17.03.2018	7	17.03.2018	279
18.03.2018	13	18.03.2018	510
19.03.2018	8	19.03.2018	224
20.03.2018	13	20.03.2018	480





Diagnostic Analytics

- Diagnostic Analytics is a form of advanced analytics which examines data or content to answer the question "Why did it happen?"

- Diagnostic analytics takes a deeper look at data to attempt to understand the causes of events and behaviors.

e.g. Checking whether your social media activity relates with the number of visitors on your website each day in the last three months

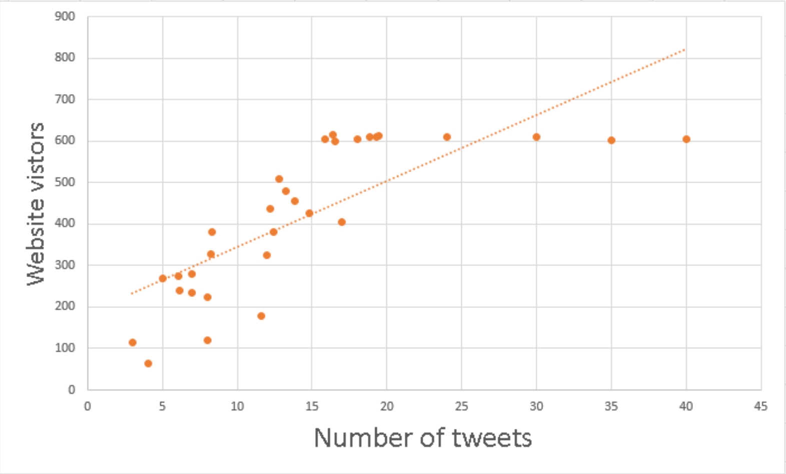
Social media activity

?

Number of visitors on your website

Diagnostic Analytics

Day	Number of tweets	Website vistors
01.03.2018	19	609
02.03.2018	12	381
03.03.2018	24	609
04.03.2018	12	436
05.03.2018	17	600
06.03.2018	17	405
07.03.2018	35	603
08.03.2018	30	609
09.03.2018	16	615
10.03.2018	40	606
11.03.2018	19	609
12.03.2018	12	324
13.03.2018	6	275
14.03.2018	14	457
15.03.2018	8	120
16.03.2018	19	612
17.03.2018	7	279
18.03.2018	13	510
19.03.2018	8	224
20.03.2018	13	480



Predictive analytics

- Predictive analytics tells what is likely to happen.

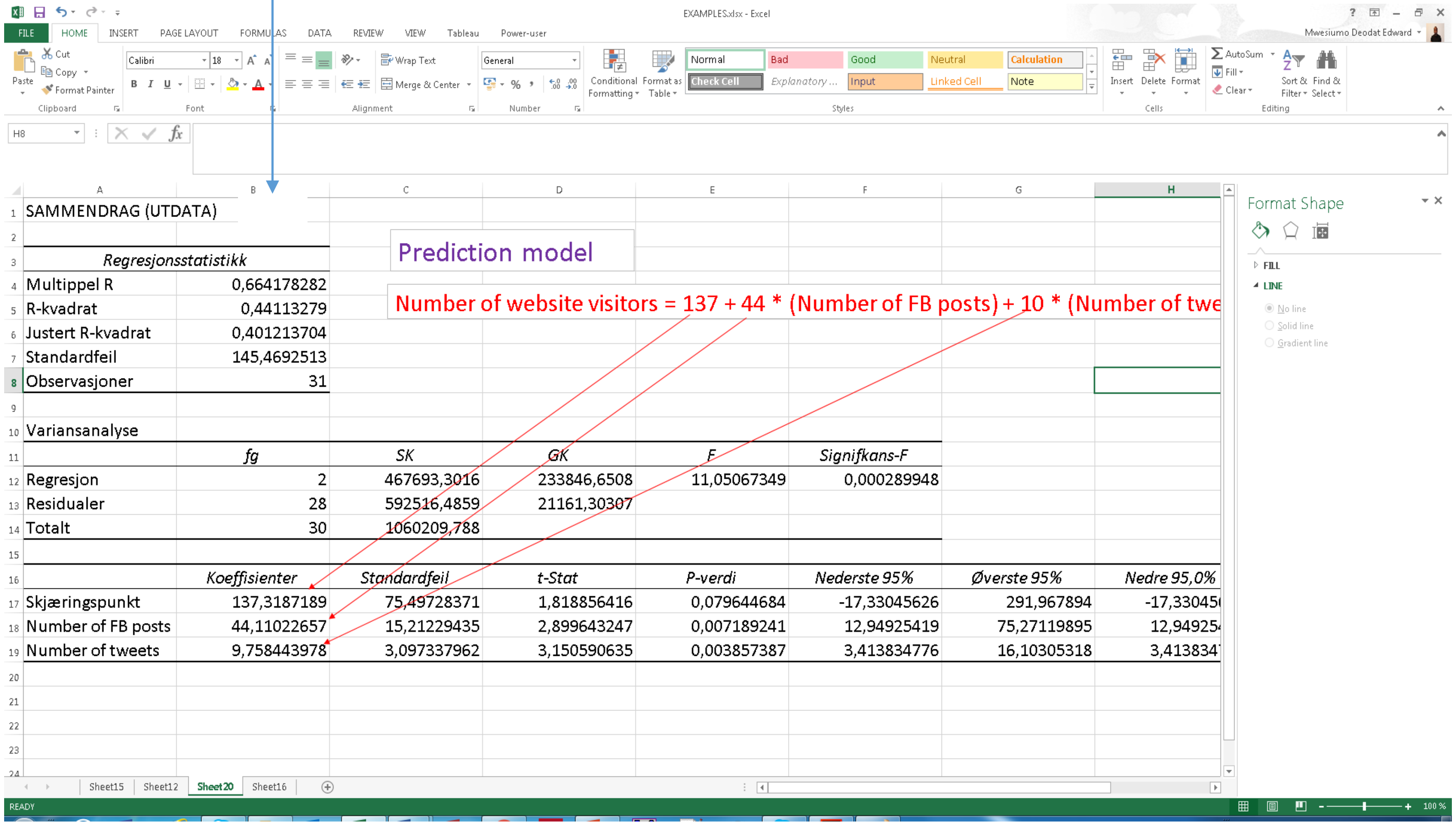
- Predictive analytics provide estimates about the likelihood of a future outcome.

- It is important to remember that no statistical algorithm can "predict" the future with 100% certainty.

e.g. Predictive analytics can help a telecom company, for instance, to identify the subscribers who are most likely to reduce their spend

Predictive analytics

Day	Number of FB posts	Number of tweets	Website vistors
01.03.2018	2,00	19	609
02.03.2018	7,00	12	800
03.03.2018	4,00	24	609
04.03.2018	4,00	12	436
05.03.2018	3,00	17	600
06.03.2018	4,00	17	405
07.03.2018	8,00	35	603
08.03.2018	4,00	30	609
09.03.2018	6,00	16	615
10.03.2018	4,00	40	606
11.03.2018	2,00	19	609
12.03.2018	4,00	12	324
13.03.2018	4,00	6	275
14.03.2018	4,00	14	457
15.03.2018	4,00	8	120
16.03.2018	4,00	19	612
17.03.2018	5,00	7	700



SAMMENDRAG (UTDATA)

Regresjonsstatistikk	
Multippel R	0,664178282
R-kvadrat	0,44113279
Justert R-kvadrat	0,401213704
Standardfeil	145,4692513
Observasjoner	31

Variansanalyse

	fg	*SK*	*GK*	*F*	*Signifkans-F*
Regresjon	2	467693,3016	233846,6508	11,05067349	0,000289948
Residualer	28	592516,4859	21161,30307		
Totalt	30	1060209,788			

	Koeffisienter	*Standardfeil*	*t-Stat*	*P-verdi*	*Nederste 95%*	*Øverste 95%*	*Nedre 95,0%*
Skjæringspunkt	137,3187189	75,49728371	1,818856416	0,079644684	-17,33045626	291,967894	-17,33045
Number of FB posts	44,11022657	15,21229435	2,899643247	0,007189241	12,94925419	75,27119895	12,94925
Number of tweets	9,758443978	3,097337962	3,150590635	0,003857387	3,413834776	16,10305318	3,413834

Predictive models

- **Predictive modeling** is a commonly used statistical technique to predict future behavior.

- Predictive models are models of the relation between the specific performance of a unit in a sample and one or more known attributes or features of the unit.

- The objective of the model is to assess the likelihood that a similar unit in a different sample will exhibit the specific performance.

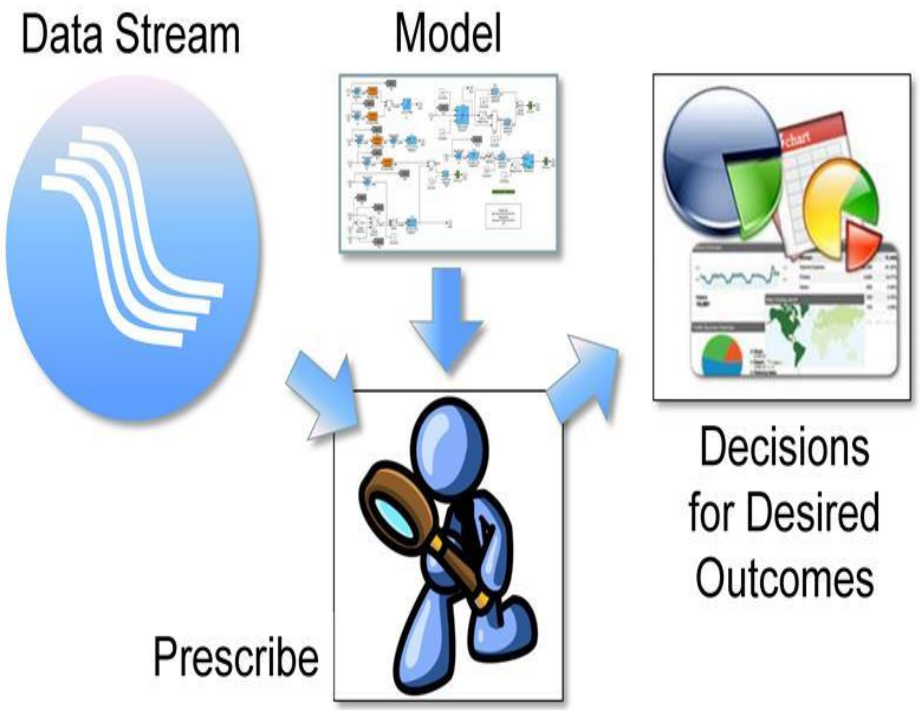
Prescriptive analytics

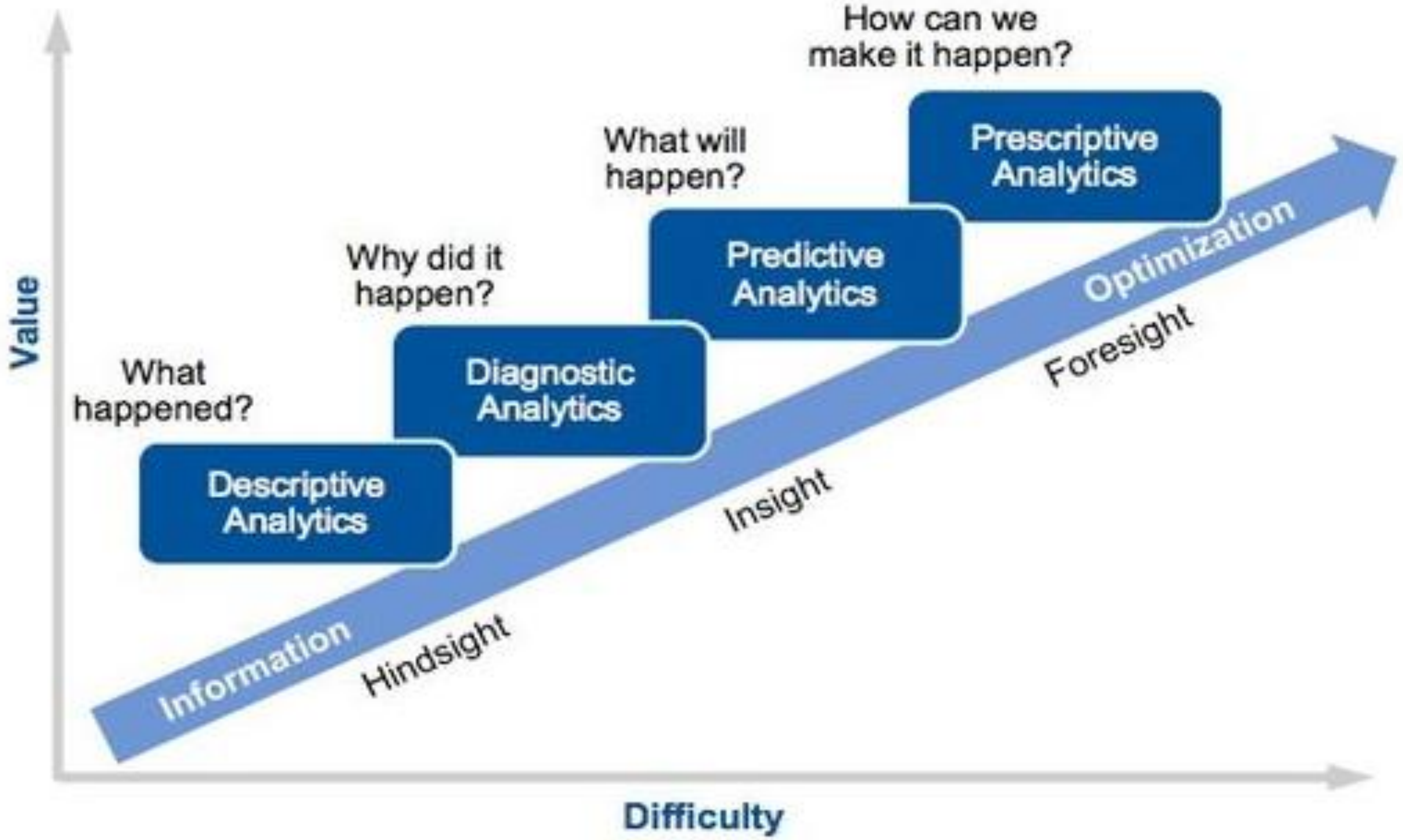
- Prescriptive analytics is the area of business analytics dedicated to finding the best course of action for a given situation.

- Prescriptive analytics goes beyond predicting future outcomes by also suggesting actions to benefit from the predictions and showing the implications of each decision option

- Prescriptive analytics incorporates both structured and unstructured data, and uses a combination of advanced analytic techniques

Google's [self-driving car] is an example of prescriptive analytics in action. The vehicle makes millions of calculations on every trip that help the car decide when and where to turn, whether to slow down or speed up, and when to change lanes



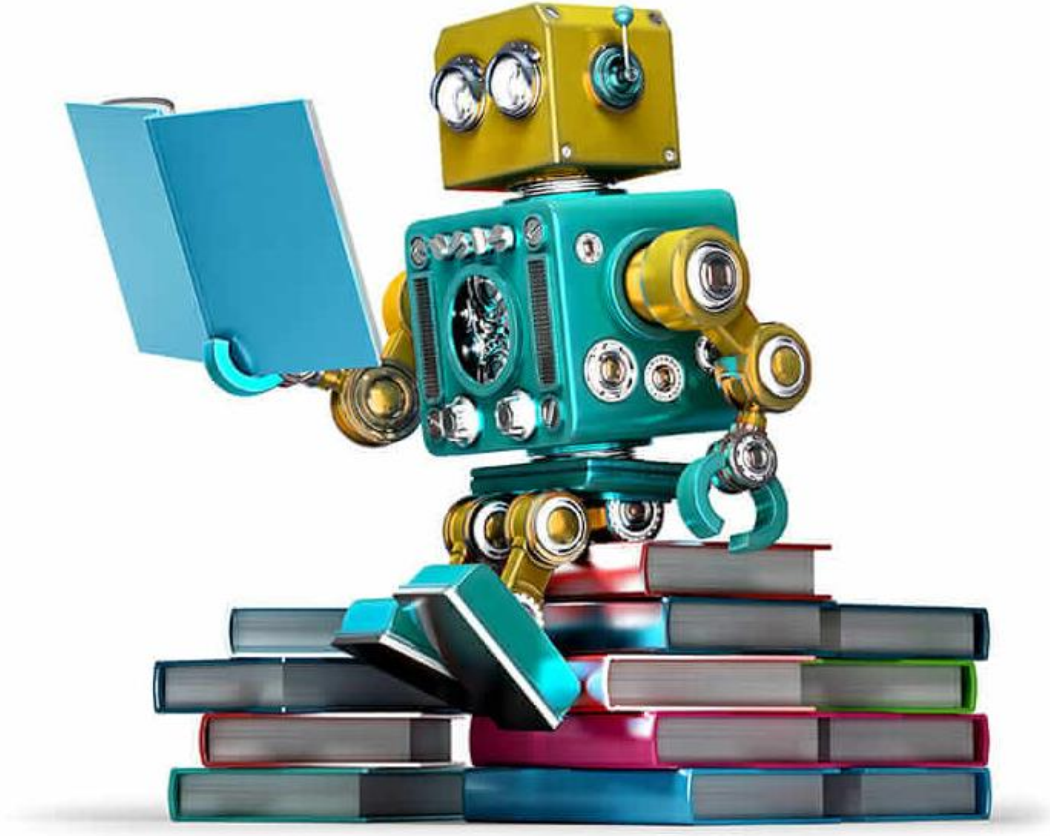

How can we make it happen?
Prescriptive Analytics
What will happen?
Predictive Analytics
Why did it happen?
Diagnostic Analytics
What happened?
Descriptive Analytics
Value
Information
Hindsight
Insight
Foresight
Optimization
Difficulty

Machine learning

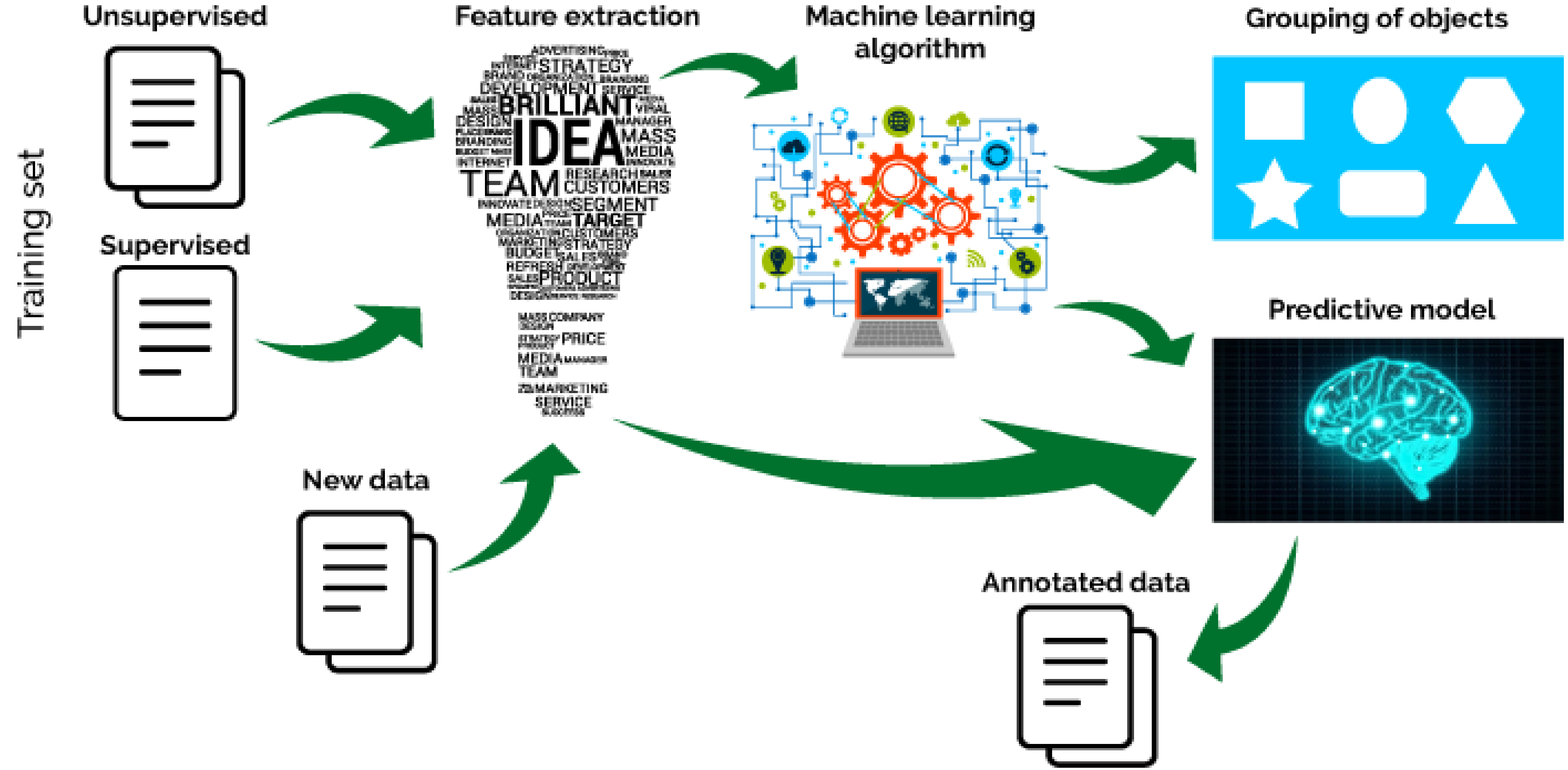
- Machine learning enables building of analytic models that learn from data.

- Machine learning in the world of big data, huge data sets, needs the ability to apply complex algorithms in a cost effective way.

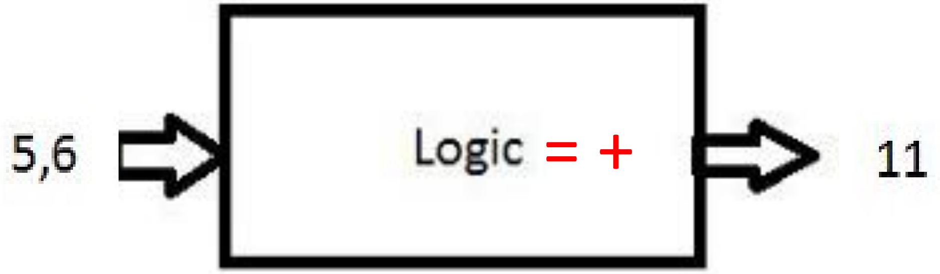
- It helps with the development of fast and efficient algorithms for real-time processing of data with as a main goal to deliver accurate predictions of various kinds.



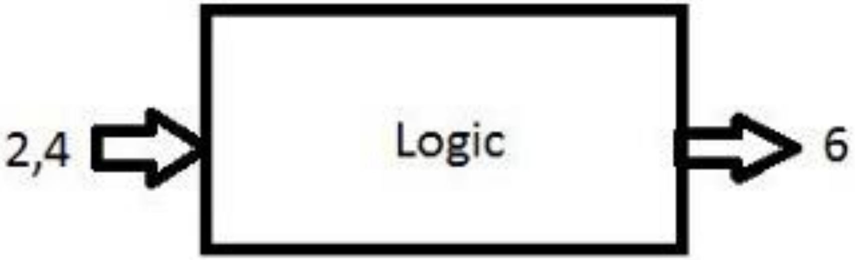
Machine Learning



Supervised learning

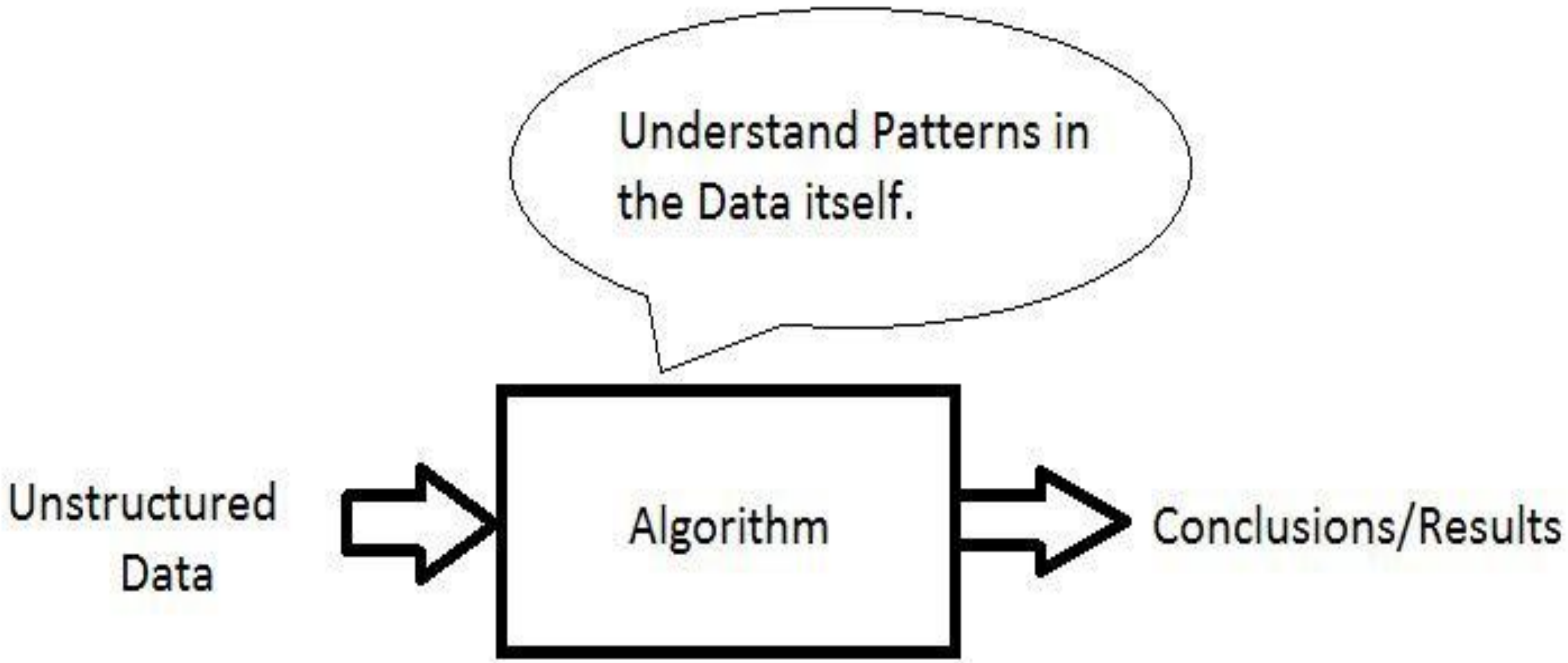


Training with Training data



Predicting with new data

Unsupervised learning



Dashboarding

- Dashboarding is an act of consolidating and arranging numbers, metrics and key performance indicators (KPIs) on a single screen.

- A dashboard is a page which provides a single view of a component or complete business.

- With interactive visualisations that link to each other, users are able to uncover behaviours and insights in their data.

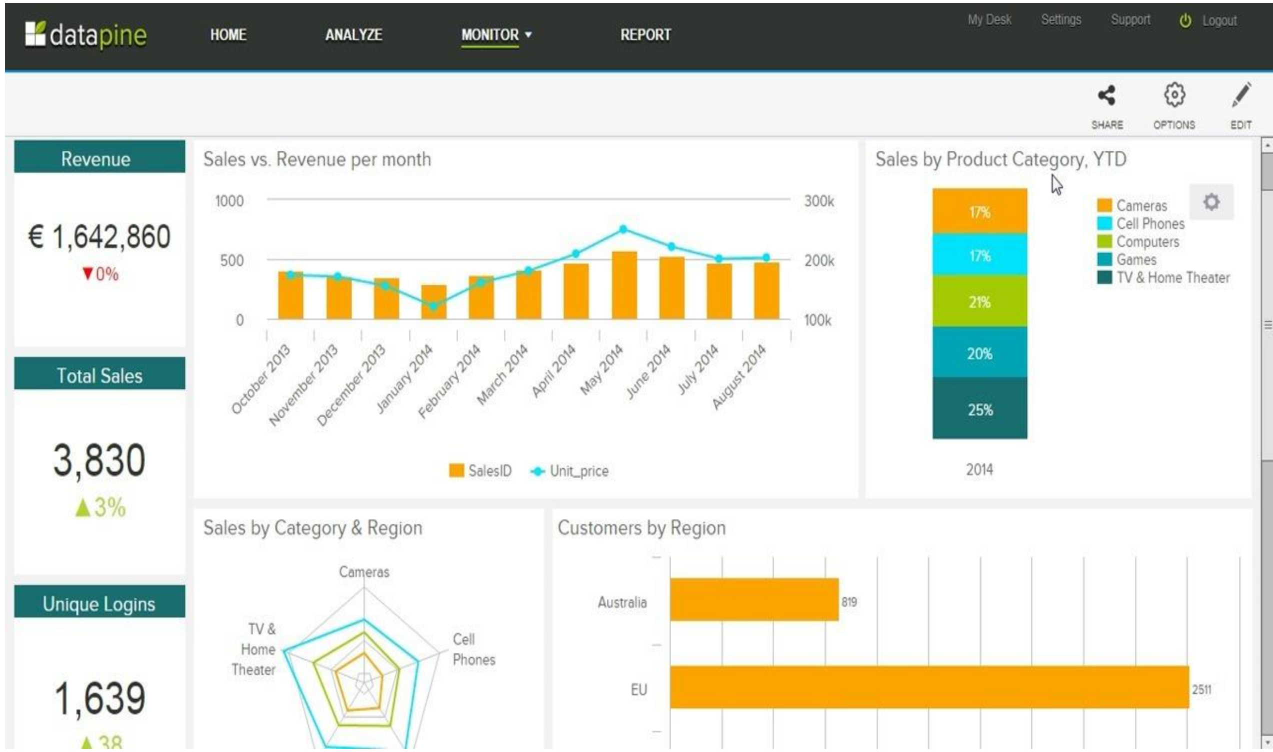

BOARD - Capsule
Capsule
100000
10000
1000
100
10
100000
10000
1000
100
10
Sparte IST % VI YY
Mountain Bikes Bikes 9.393 42,0 % 0 24,0 %
Road Bikes Bikes 5.417 24,2 % 0 11,2 %
Helmets Accessories 3.334 14,9 % 0 13,9 %
Jerseys Clothing 2.280 10,2 % 0 8,2 %
Road Frames Components 730 3,3 % 0 5,0 %
Mountain Frames Components 671 3,0 % 0 6,0 %
Socks Clothing 293 1,3 % 0 1,3 %
Caps Clothing 248 1,1 % 0 1,1 %
Handlebars Components 143 0,0 % 0 0,0 %
100000
10000
1000
100
10
2012
2013
2014
7%
3%
20%
10%
60%

Advantages of Dashboards

- Gives executives, managers and analysts convenient immediate access to key performance metrics

- On demand, accurate and relevant information in line with business priorities

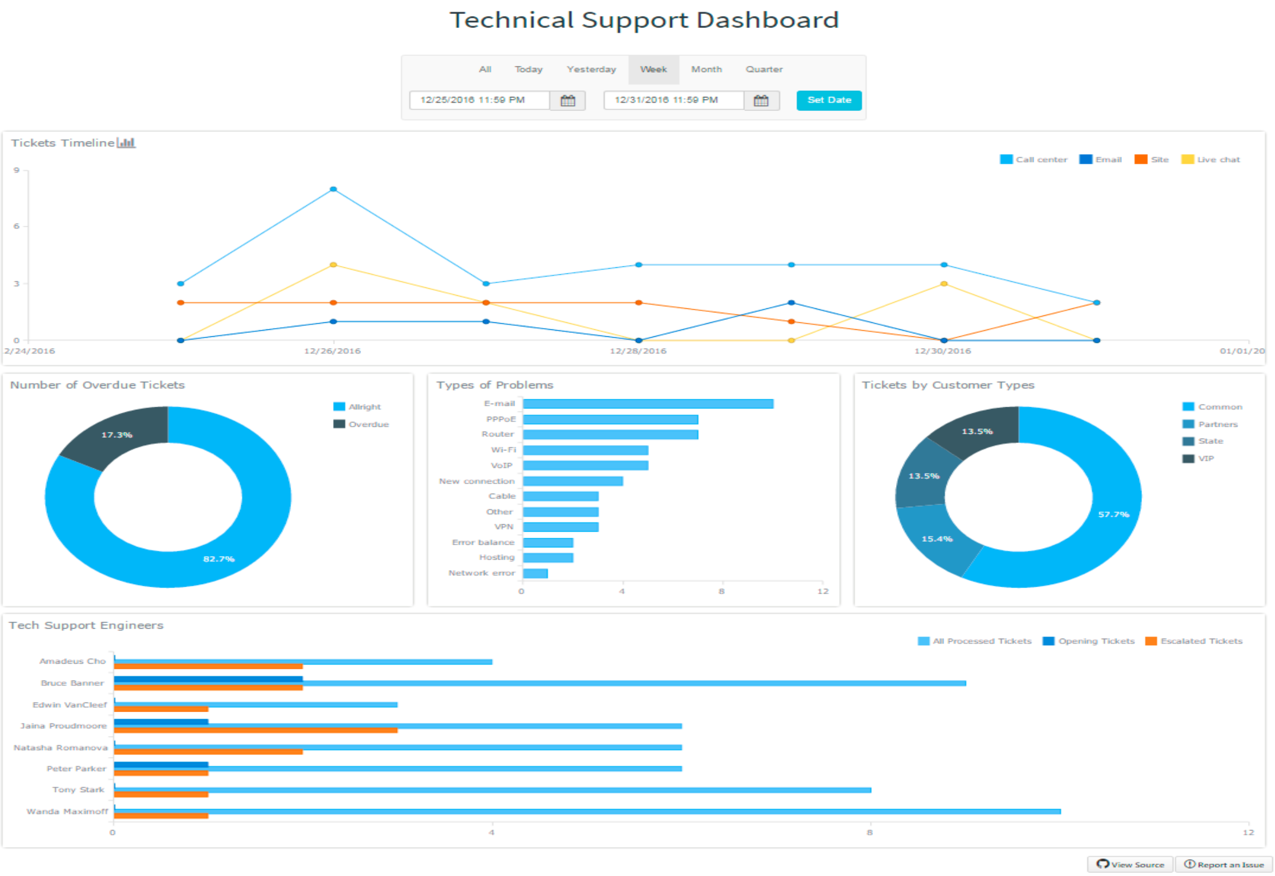
- Focused identification of problems, inefficiencies or negative trends for immediate action and improved performance



Best Practices for Building Effective Dashboards

1. Thoughtful Planning

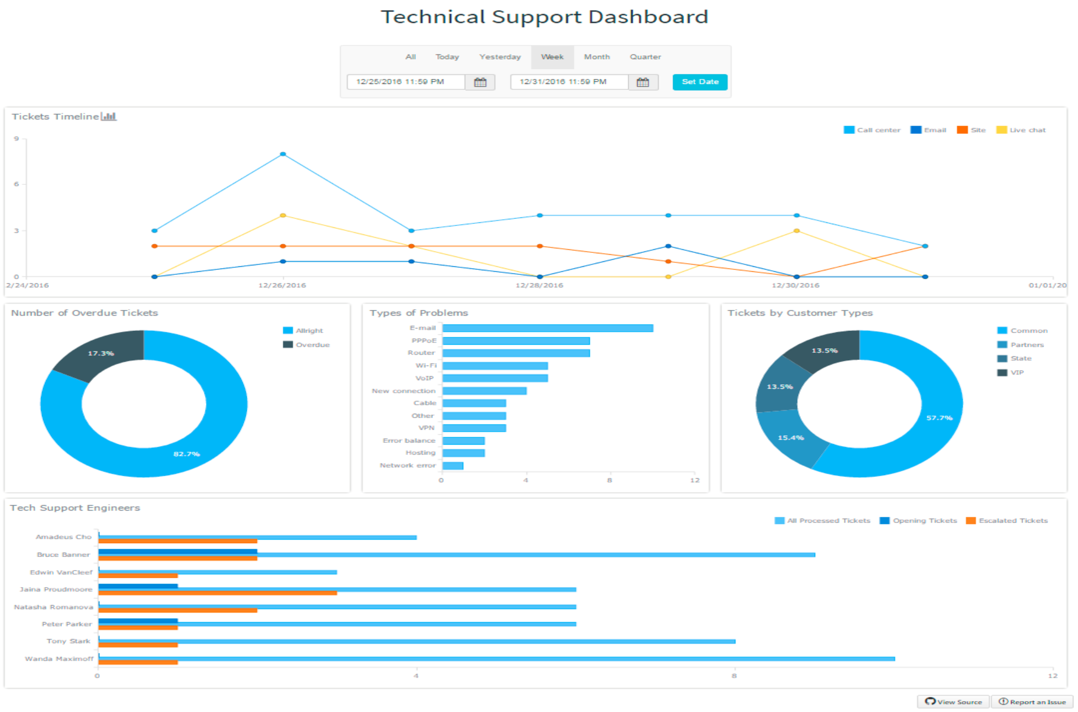
- ➢ Know your audience
- ➢ Consider display size
- ➢ Plan for fast load times



Best Practices for Building Effective Dashboards

2. Informed Design

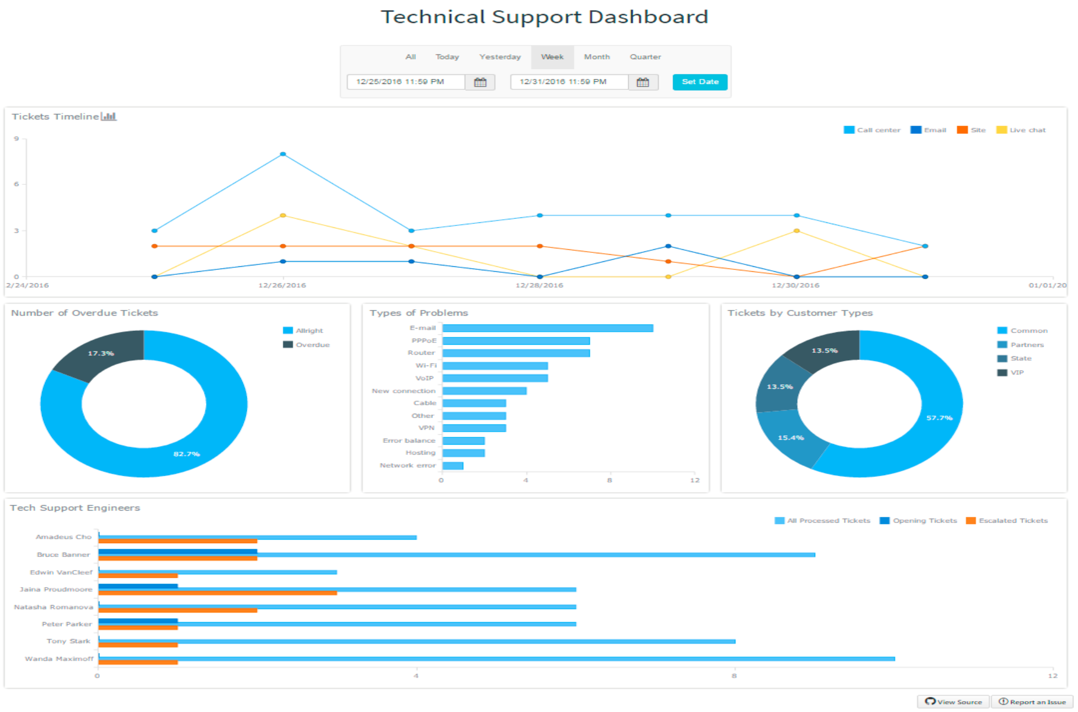
- ➢ Leverage the sweet spot
- ➢ Limit the number of views & colors
- ➢ Add interactivity to encourage exploration
- ➢ Eliminate clutter



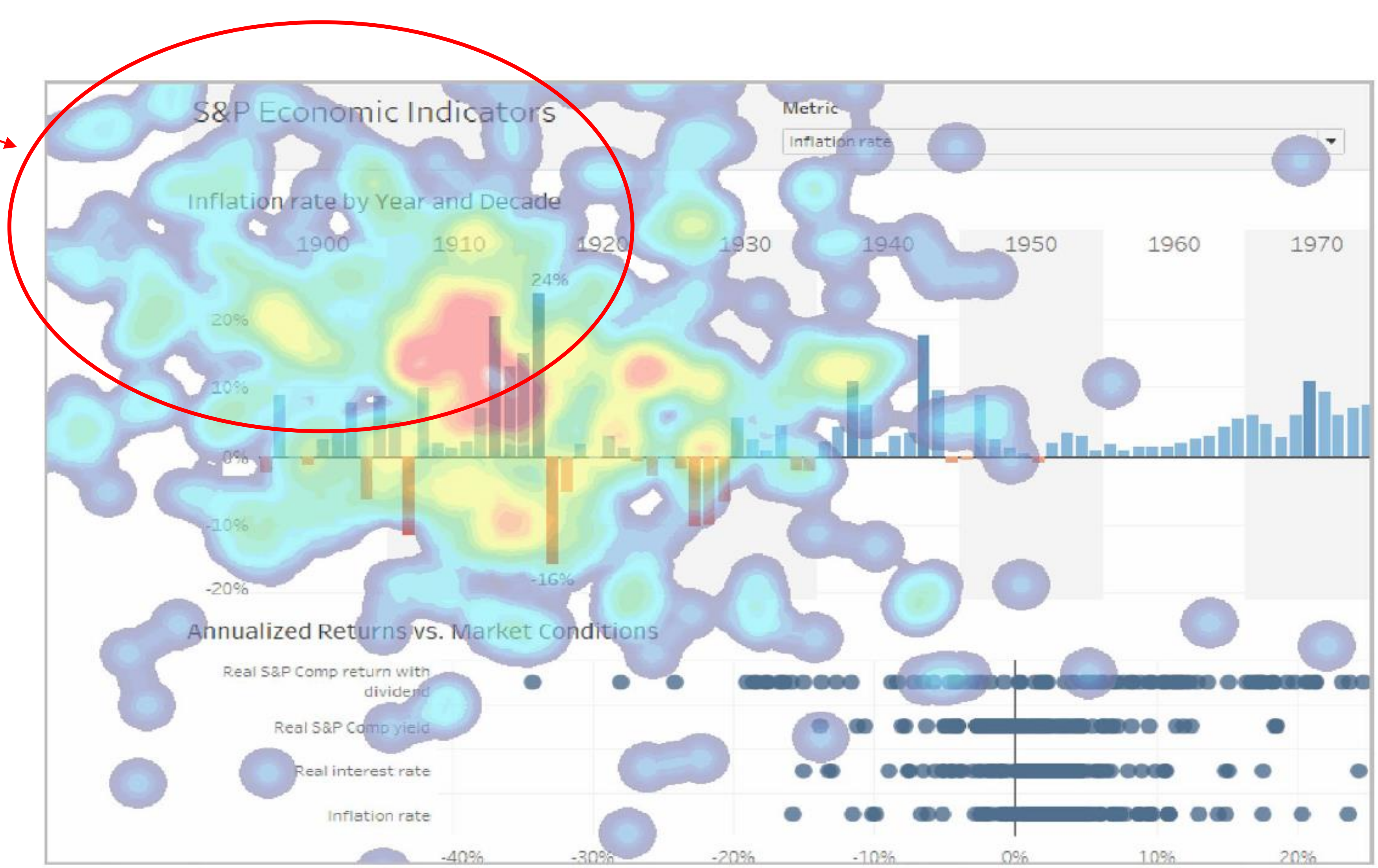
Best Practices for Building Effective Dashboards

2. Informed Design

- ➢ Leverage the sweet spot
- ➢ Limit the number of views & colors
- ➢ Add interactivity to encourage exploration
- ➢ Eliminate clutter



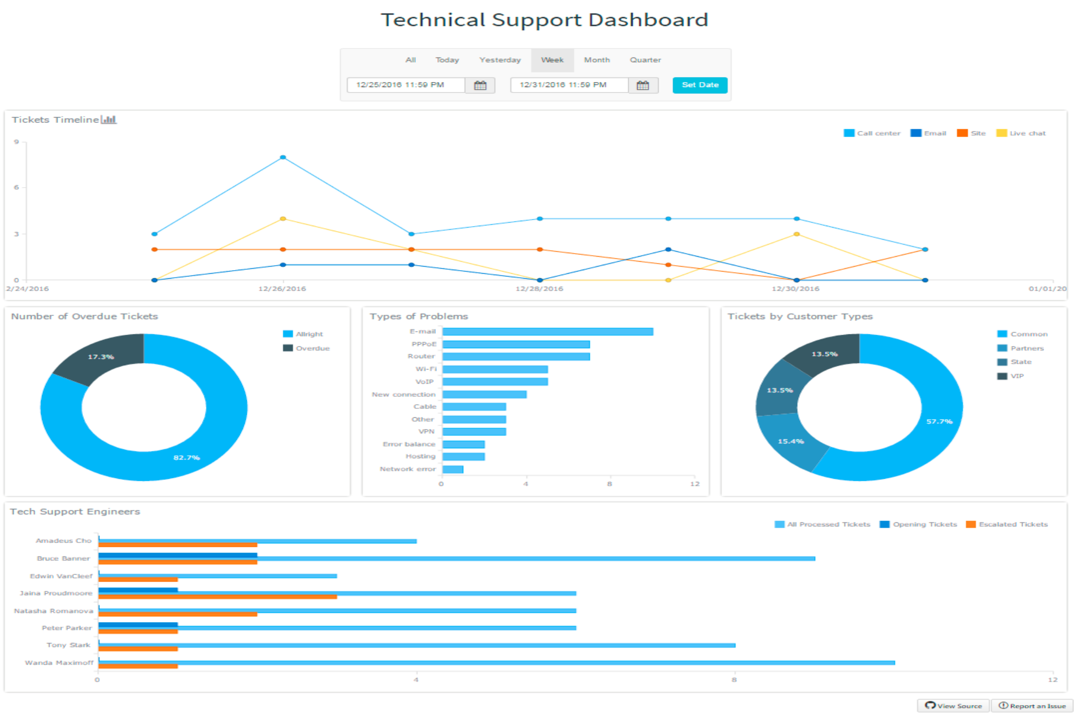
Leverage the sweet spot



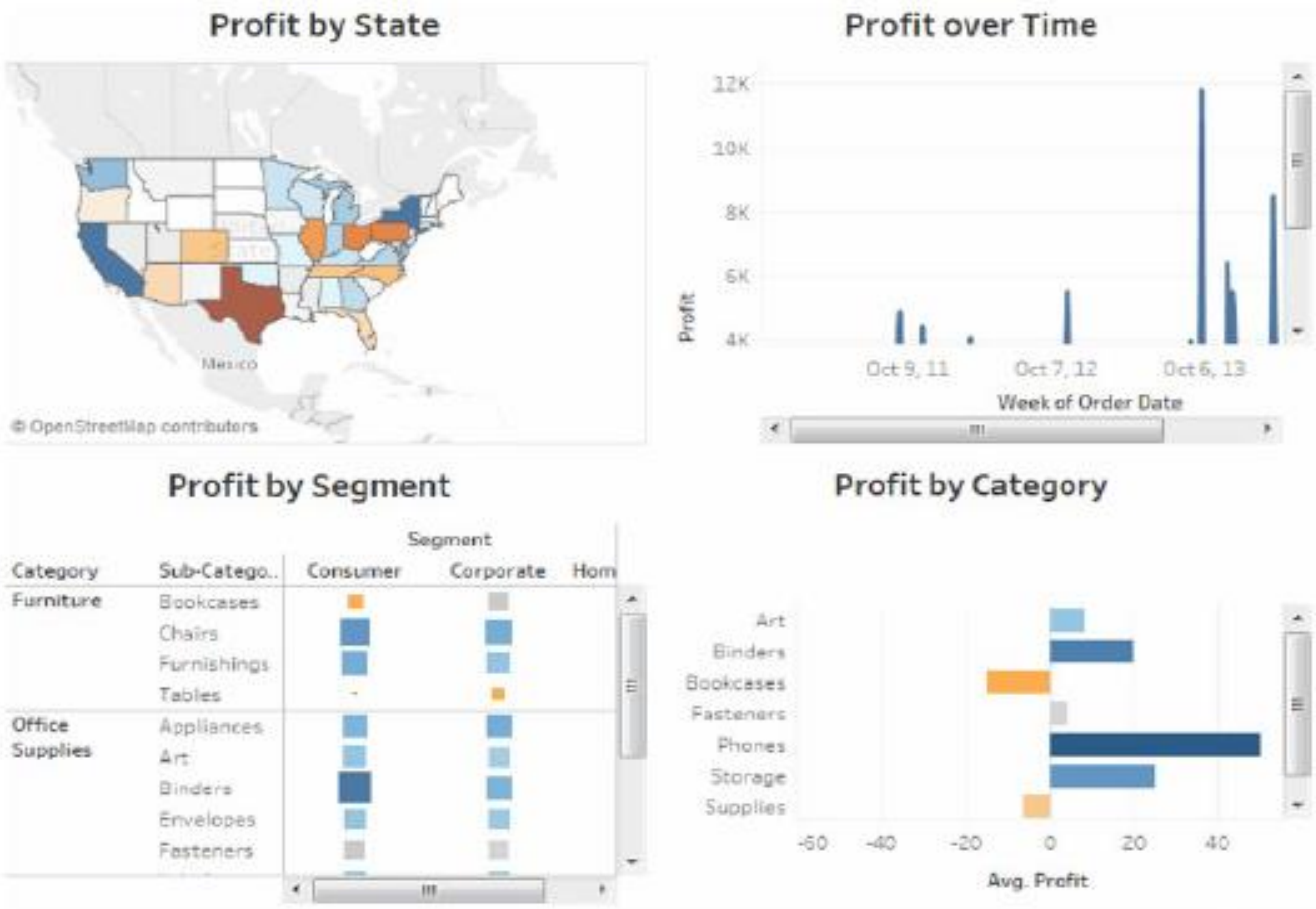
Best Practices for Building Effective Dashboards

2. Informed Design

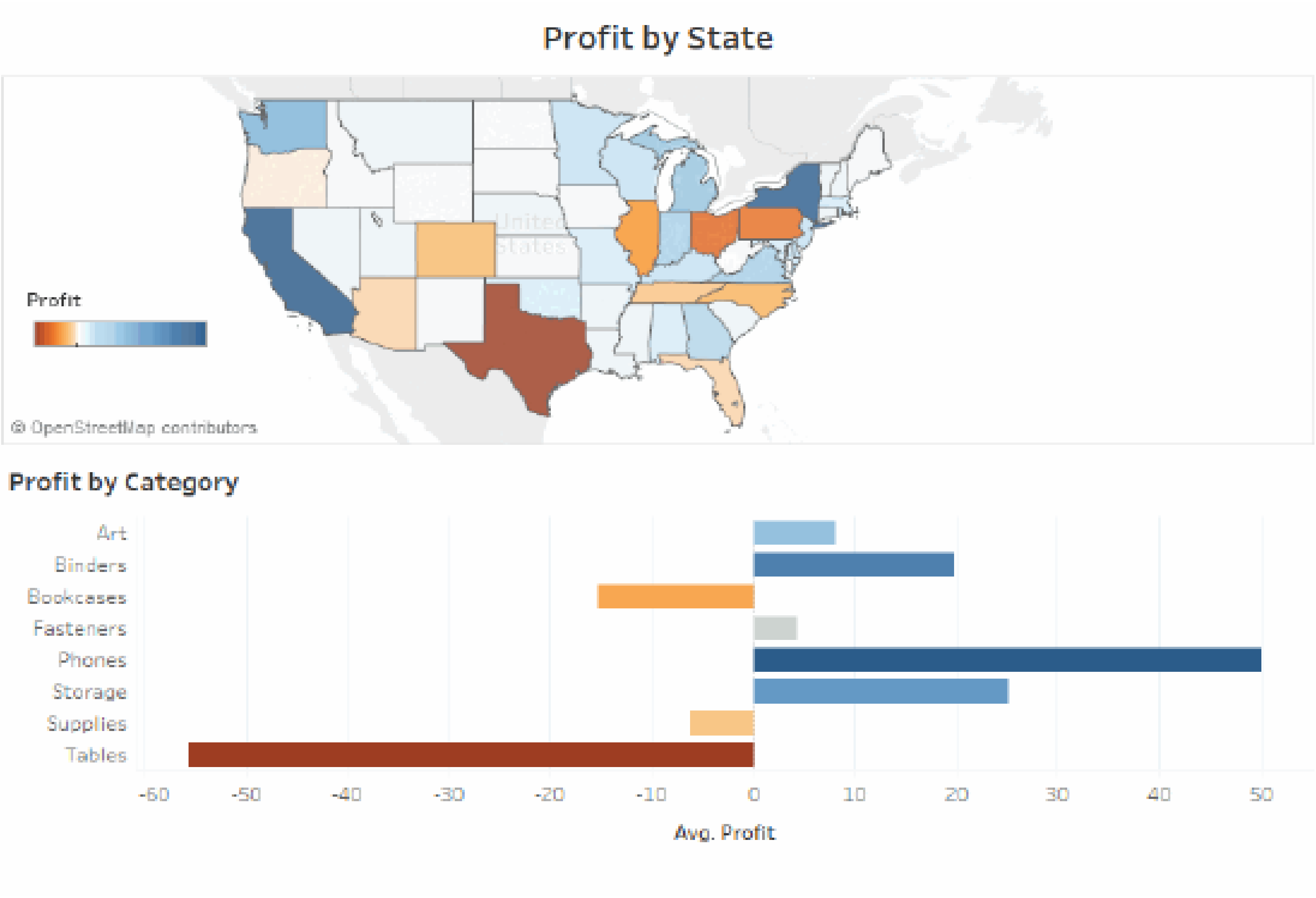
- Leverage the sweet spot
- Limit the number of views & colors
- Add interactivity to encourage exploration
- Eliminate clutter



Too many views are not recommended



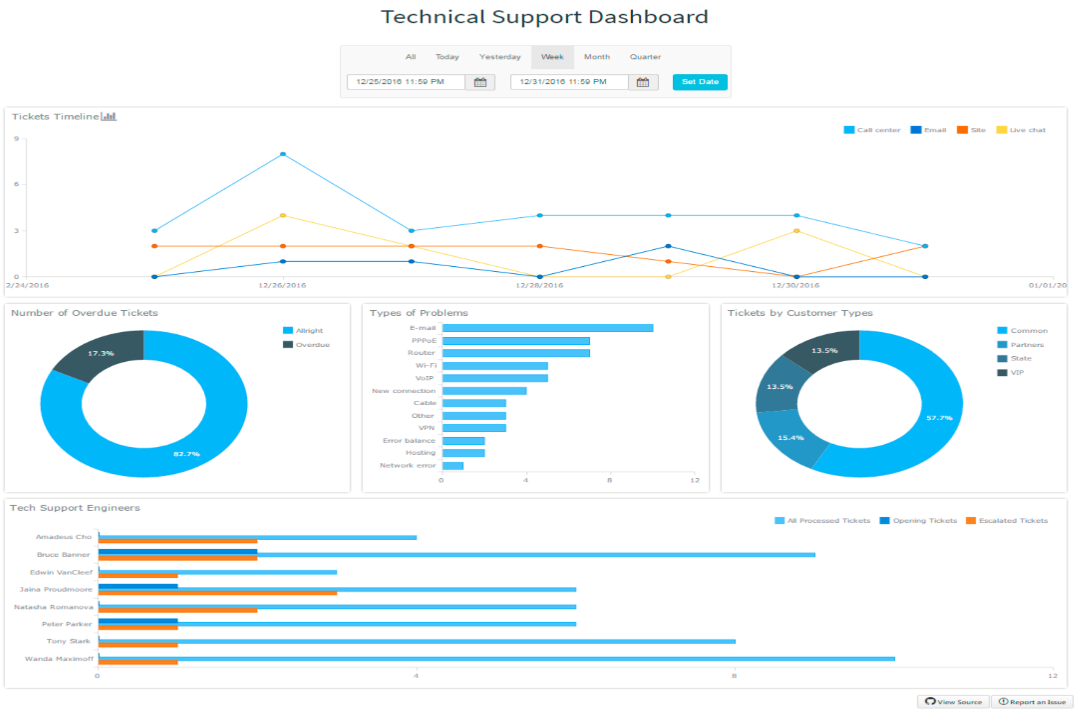
Recommended: two to three views



Best Practices for Building Effective Dashboards

2. Informed Design

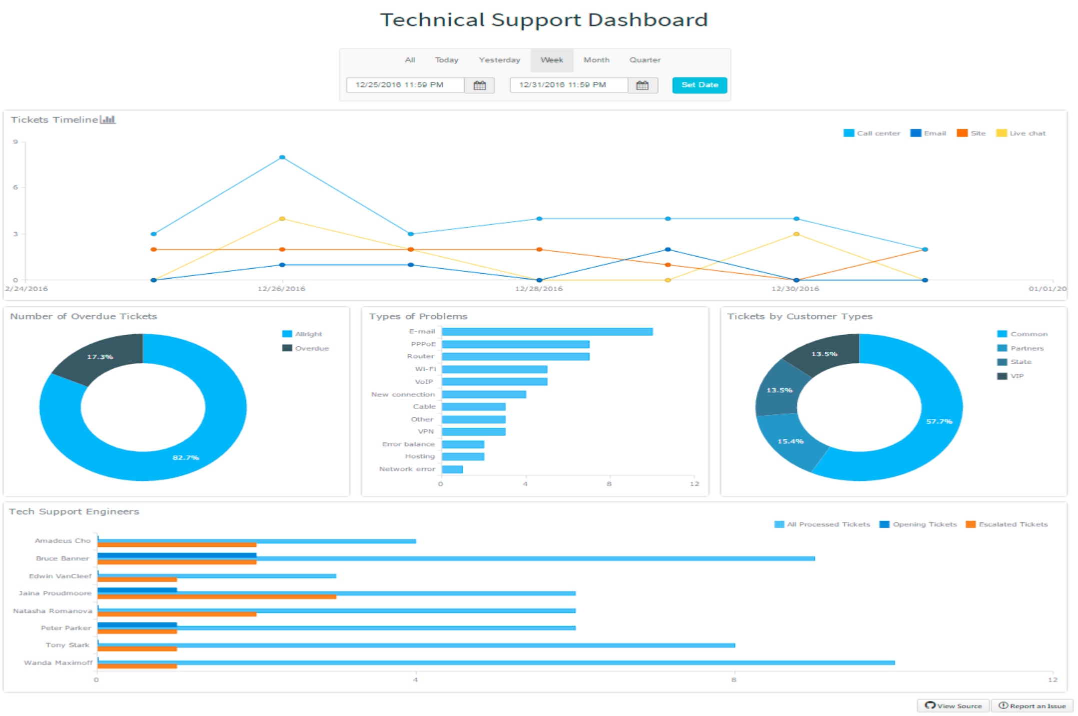
- Leverage the sweet spot
- Limit the number of views & colors
- Add interactivity to encourage exploration
- Eliminate clutter



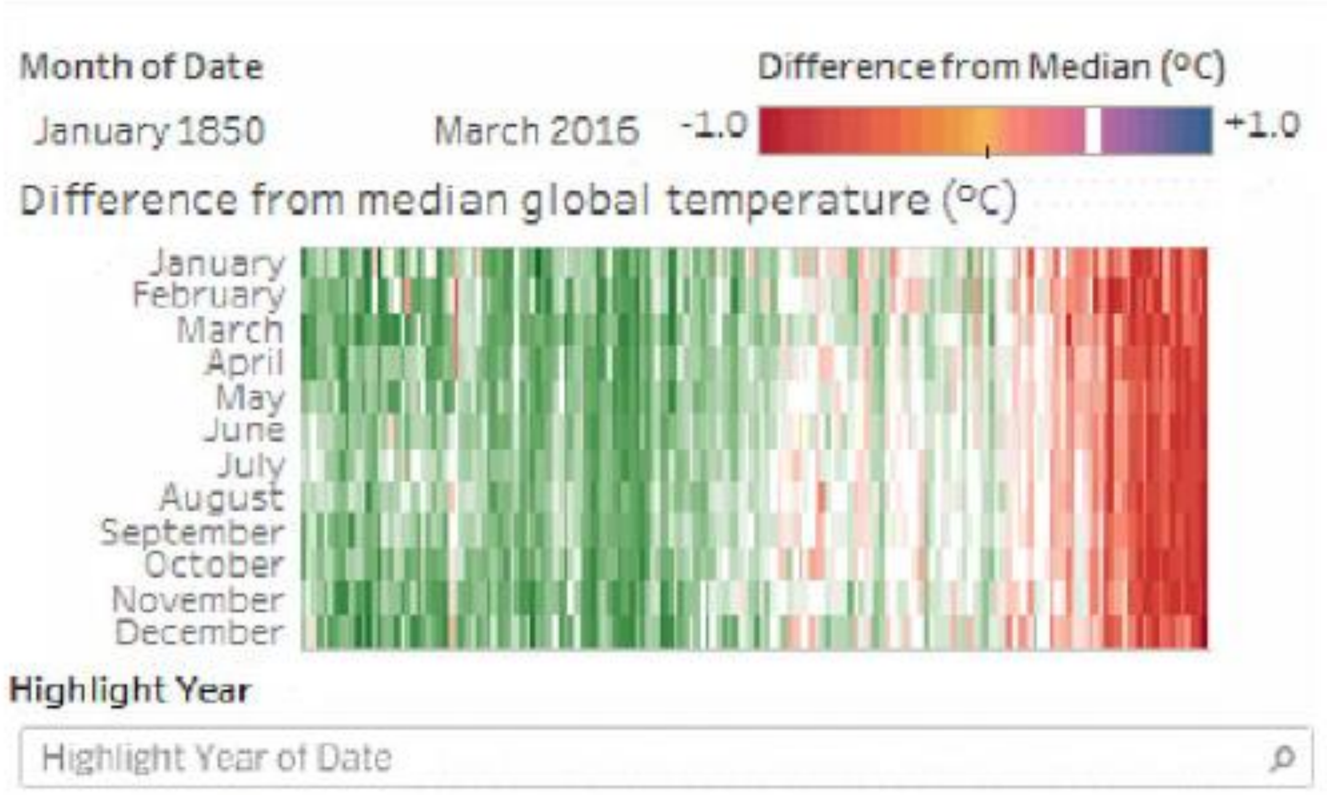
Best Practices for Building Effective Dashboards

2. Informed Design

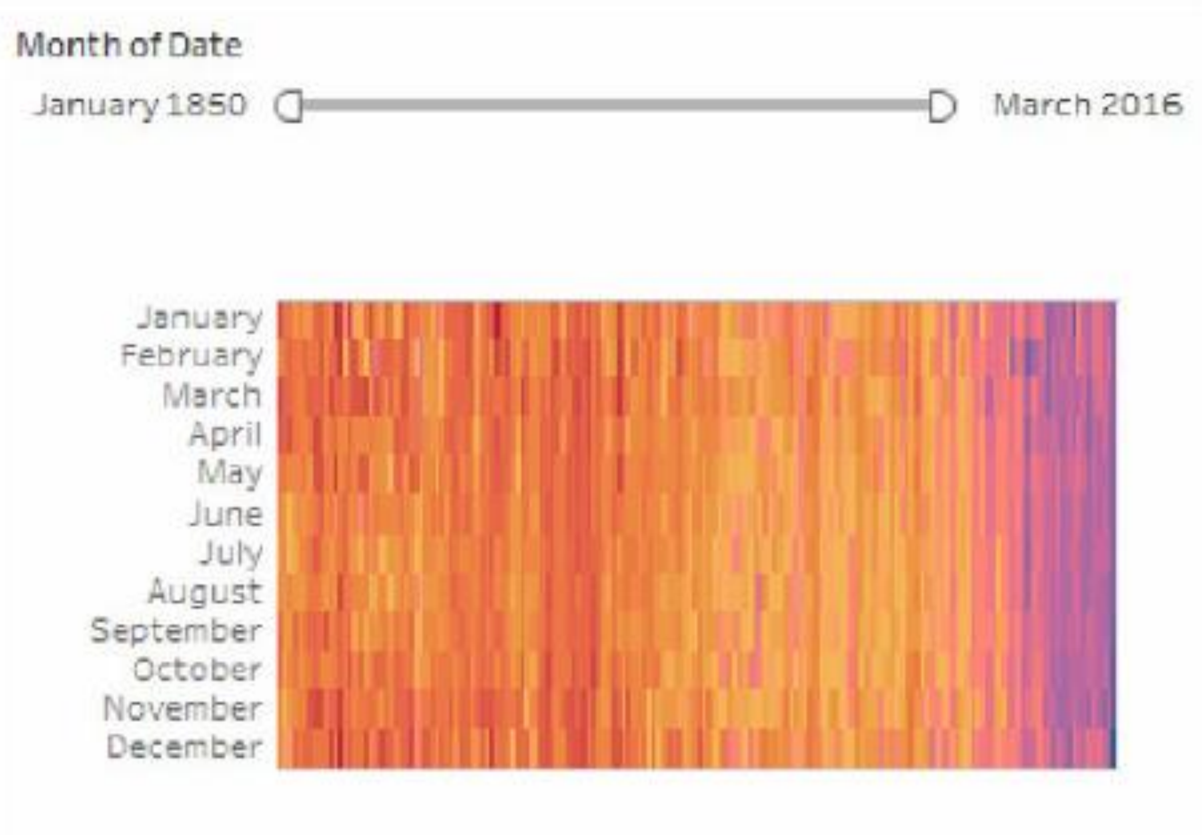
- ➢ Leverage the sweet spot
- ➢ Limit the number of views & colors
- ➢ Add interactivity to encourage exploration
- ➢ Eliminate clutter



Not recommended



Recommended



Role of Big data in decision making

- Because of big data, managers can measure, and hence know, radically more about their businesses.
- This knowledge can directly be translated into improved decision making and performance.
- Big Data is the next frontier for innovation, competition, and productivity.

EXAMPLES OF WAYS COMPANIES CAN BENEFIT FROM BIG DATA

1. **Targeting Customers**

- Big data is used to better understand potential customers and their behaviors and preferences.

- Expanding traditional data sets with social media data, browser logs as well as text analytics and sensor data provides a more complete picture of customers.

EXAMPLES OF WAYS COMPANIES CAN BENEFIT FROM BIG DATA

2. Product improvement

Big Data can also help you understand how consumers perceive your products so that you can improve them.
For example: Analysis of unstructured social media text allows you to uncover the perceptions of your customers and even segment those in different geographical locations or among different demographic groups.

EXAMPLES OF WAYS COMPANIES CAN BENEFIT FROM BIG DATA

4. New products and services

Perhaps the most interesting use of big data analytics is to create new products and services for customers. Online companies have done this for a decade or so, but now predominantly offline firms are doing it too. GE, for example, has made a major investment in new service models for its industrial products using big data analytics.

Big data Use SUCCESS: Netflix

- Being able to visualise data and gather valuable insights is key to Netflix's success

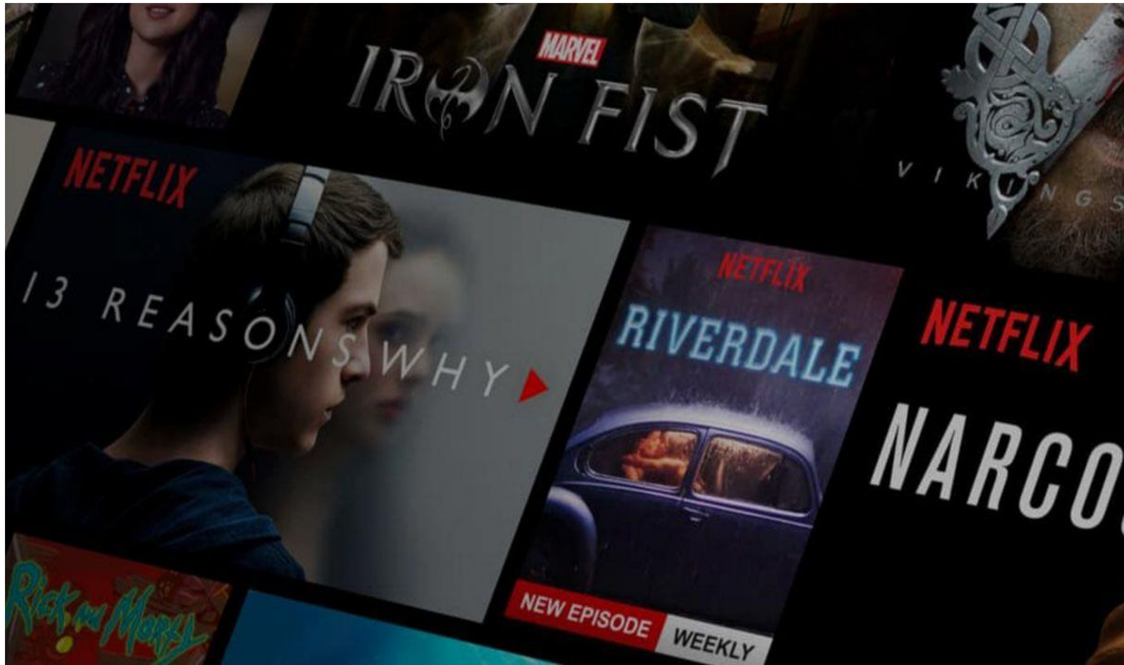
https://medium.com/swlh/how-netflix-uses-big-data-20b5419c1edf

- They are able to adjust algorithms, address insights and solve problems easily and efficiently

- Data on viewing habits: are pivotal data points for predicting consumer behaviour.
    - time of day a movie or TV show was watched,
    - time spent selecting movies, and even
    - how often playback was stopped.

- They started with analysing consumers and provide them with relevant and personalised content

- They eventually started using data to create shows: e.g. House of Cards.



EXAMPLES OF WAYS COMPANIES CAN BENEFIT FROM BIG DATA

5. Business environment scanning

Success not only depends on how you run your company. Social and economic factors are crucial for your accomplishments as well. Predictive analytics, fueled by Big Data allows you to scan and analyze newspaper reports or social media feeds so that you permanently keep up to speed on the latest developments in your industry and its environment.

EXAMPLES OF WAYS COMPANIES CAN BENEFIT FROM BIG DATA

6. Predicting customer attrition

- Customer attrition (customer churn) is the loss of customers.
- Through Big Data companies can create predictive models to understand who will attrite to competitors and who will remain loyal
- Example:
  By looking at customer banking behavior and integrating 100+ variables, American Express has come up with predictive models to understand who will attrite to competitors and who will remain loyal.

EXAMPLES OF WAYS COMPANIES CAN BENEFIT FROM BIG DATA

7. Real time solutions

Many large organizations are seeking both faster and better decisions with big data, and they're finding them.

Example

Caesars, a leading gaming company has embraced big data analytics for faster decisions. For example, Caesars has found that if a new customer to its loyalty program has a run of bad luck at the slots, it's likely that customer will never come back. But if it can present, say, a free meal coupon to that customer while he's still at the slot machine, he is much more likely to return to the casino later. The key, however, is to do the necessary analysis in real time and present the offer before the customer turns away in disgust with his luck and the machines at which he's been playing.

Faster and better decisions