



1ST EDITION

Data Literacy in Practice

A complete guide to data literacy and making smarter decisions with data through intelligent actions



ANGELIKA KLIDAS | KEVIN HANEGAN

Preface

Data is more than a commodity in our digital world. It is the ebb and flow of our modern existence. Individuals, teams, and enterprises that work confidently with data unlock a new realm of possibilities. The resultant agility, growth, and inevitable success have one origin: data literacy.

Data Literacy in Practice is a comprehensive guide that will build your understanding of data literacy basics, and accelerate your journey to independently uncovering insights with best practices, practical models, and real-world examples.

Discover the four-pillar model that underpins all data and analytics. Explore concepts such as measuring data quality, setting up a pragmatic data management environment, choosing the right graphs for your readers, and questioning your insights.

This guide is written by two data literacy pioneers, each with a thorough footprint within the data and analytics commercial world, including their lectures at top universities in the US and the Netherlands.

By the end of the book, you'll be equipped with a combination of skills and mindsets, along with tools and frameworks, that allow you to find insights and meaning within your data to enable effective and efficient data-informed decision-making.

Who this audiobook is for

This book is for data analysts, data professionals, and data teams starting or wanting to accelerate their data literacy journey. Discover the skills and mindset you need to work independently with data, along with the tools and frameworks to build a solid knowledge base, and start making your data work for you today.

What this audiobook covers

Chapter 1, The Beginning – The Flow of Data, covers the process of going from data to insights and action and shows how it is a multi-step process. Understanding this process is critical for anyone who is leveraging data to make decisions. This chapter will introduce the flow of data through this process, as well as common pitfalls that can get in the way at each step.

Chapter 2, Unfolding Your Data Journey, shows how, to be able to properly turn data into actionable insights, individuals need to be able to leverage multiple steps in analytics maturity: descriptive, diagnostic, predictive, prescriptive, and semantic. This chapter will introduce those steps with practical examples of what insights you can get from each step in the process.

Chapter 3, Understanding the Four-Pillar Model, looks at the four elementary pillars of data and analytics that we need to address in our businesses. Everybody knows and understands what data or a dashboard is. From that point of view, we see more demand and acceptance for data and analytics projects and the need for data literacy knowledge.

Chapter 4, Implementing Organizational Data Literacy, focuses on best practices related to organizational strategy and culture to support data literacy and data-informed decision-making. For individuals and organizations to be able to elicit insights and value from their data, there needs to be widespread adoption of data-informed decision-making. Despite many organizations having tools, technologies, and technical abilities, they are often unable to become data-informed due to their lack of a data literacy culture.

Chapter 5, Managing Your Data Environment, looks at how low-code/no-code solutions are maturing in an interesting way, giving all the benefits to their users in building rapid data lakes, data warehouses, and data pipelines. If we compare this technology against the more traditional solutions, we notice that we are able to get a better “race pace” in developing a data and analytics fundament. Due to the enormous growth (1.7 Mb of data is created every second for every person on earth) and complexity of data and data environments, a good and solid data strategy and taking care of a shared data vision was never as important as it is now. But in the last 2 years, there has been a shift occurring, and the necessity of a managed data environment has become more important.

Chapter 6, Aligning with Organizational Goals, explains how **Key Performance Indicators (KPIs)** are extremely vital in helping organizations understand how well they are performing in relation to their strategic goals and objectives. However, understanding what a KPI truly is versus what is just a measurement or a metric is important, along with understanding the right types of KPIs to track, including leading and lagging indicators.

Chapter 7, Designing Dashboards and Reports, talks about how visualizations provide a vital function in helping to describe situations. Visualizations can be used for both

finding insights and also for communicating those insights to others. Choosing the right visualization depends on both the data you are using and what you are trying to show. This chapter will focus on choosing the right chart type, as well as designing charts to make it easier for people to interpret relevant parts.

Chapter 8, Questioning the Data, covers learning to ask questions, analyze outliers (supporting story by Dr. Snow – *Death in the Pit*), exclude bias, and so on so that you will be able to ask the right questions and develop your curiosity. You will understand the difference between correlation and causation. By addressing those topics, you will be able to understand what signals and noise are, and how to analyze the outliers by addressing hypothetical questions. You will be able to recognize the good, the bad, and the ugly insights.

Chapter 9, Handling Data Responsibly, explains how ethics is a science in which people try to qualify certain actions as right or wrong. However, there are no unequivocal answers to ethical questions because they are often very personal. Today, data and analytics are everywhere, touching every waking moment of our lives. Data and analytics, therefore, play an enormous role in our daily lives – for example, Amazon knows what we buy and suggests other articles that we may be interested in; applications show us how we will look when we are older, and Netflix and Spotify know what we watch or listen to and give us suggestions of what else to watch or listen to.

Chapter 10, Turning Insights into Decisions, explores how many individuals and organizations come up with insights from their data. However, the process of turning insights into decisions and acting on them is much more difficult. This chapter focuses on what is required to support this step in the process, including introducing a six-step framework, which is both systemic and systematic. The chapter also includes how you can manage the change related to your decisions and how you can communicate effectively to all stakeholders via storytelling with data.

Chapter 11, Defining a Data Literacy Competency Framework, explains how the first step to increasing your own data literacy via education is to learn what exactly are the competencies that support data literacy. This chapter describes a data literacy competency framework, which includes the right hard skills, soft skills, and mindsets for data literacy. It also discusses how competencies have various levels, and you can progress up the levels as you become more experienced with data literacy. This chapter also focuses on best practices for getting started learning these competencies.

Chapter 12, Assessing Your Data Literacy Maturity, introduces how you can assess your own data literacy skills and then explains how to interpret the results of the assessment to personalize your educational journey. Before you begin your educational journey for data literacy, you should start by assessing your current level, and then using that assessment to understand what competencies to focus on next.

Chapter 13, Managing Data and Analytics Projects, explains the ways you can approach a data and analytics project and how you can manage it as a project leader and keep track of the business case and the value that it can bring. It all starts with the development of a data and analytics business case, in which you define the project scope, goals, and risks but also the beneficial value that it can bring to your organization. Data and analytics projects are often across organizations, departments, and processes of business units. They mostly contain a mix of strategic goals or have high political content and hidden stakeholders and have specific data and analytics risks that you should take care of.

Chapter 14, Appendix A – Templates, provides the materials to help you get started on your data literacy journey. All materials are also available on www.kevinhanegan.com.

Chapter 15, Appendix B – References, provides a summary of the references, books, and articles that we've read over the years. All of them inspired us and helped us to teach and write.

Supplementary materials

Packt audiobooks have been selected for a seamless audio experience. Some topics, however, do come with elements like images that aren't natural for this medium. We've adapted the content of the audiobooks so that you can listen to the audio without needing to refer to these visual elements unless necessary.

To give you the choice between listening to just the audio and listening to the audio while referring to the visual elements, we've created this PDF that contains all the elements that cannot translate to the audio. All references to images, tables, links, and every other visual element in the audiobook can be found within this PDF.

Get in touch

Feedback from our readers is always welcome.

- **General feedback:** If you have questions about any aspect of this audiobook, email us at audio@packt.com and mention the audiobook title in the subject of your message.
- **Errata:** Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you have found a mistake in this audiobook, we would be grateful if you would report this to us at the email mentioned above.
- **Piracy:** If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at copyright@packt.com with a link to the material.
- **If you are interested in becoming an author:** If there is a topic that you have expertise in and you are interested in either writing or contributing to an audiobook, please visit authors.packtpub.com.

Chapter 1

DATA IN OUR DAILY LIVES

When I wake up...



Figure 1.1 – Data in our daily lives

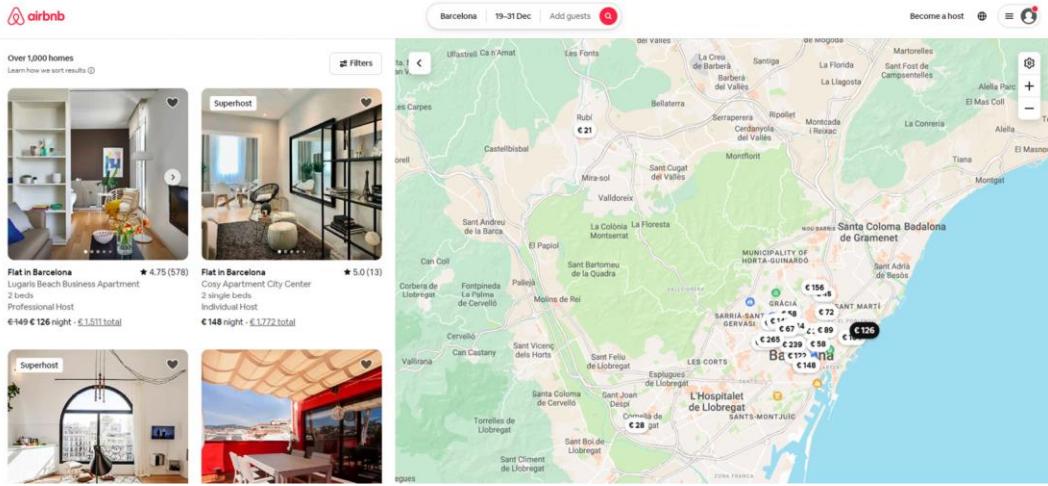


Figure 1.2 – DIDM for the summer vacation using Airbnb

★ 4.62 · 71 reviews

Cleanliness	<div style="width: 4.7/5 * 100%;"> </div>	4.7	Accuracy	<div style="width: 4.6/5 * 100%;"> </div>	4.6
Communication	<div style="width: 4.7/5 * 100%;"> </div>	4.7	Location	<div style="width: 4.8/5 * 100%;"> </div>	4.8
Check-in	<div style="width: 4.7/5 * 100%;"> </div>	4.7	Value	<div style="width: 4.4/5 * 100%;"> </div>	4.4



Cody

March 2022

Very nice place to stay in Barcelona with great restaurants and attractions near by!



Carlos

February 2022

This spot is located in a wonderful area of Barcelona that is safe, close to restaurants, and makes traveling to the nearby attractions easy. The metro line is nearby, there's a great bounty of restaurants to choose from and everything inside the room itself was as...

Show more >

Figure 1.3 – DIDM reviews to make a choice



Figure 1.4 – How we use data during our days

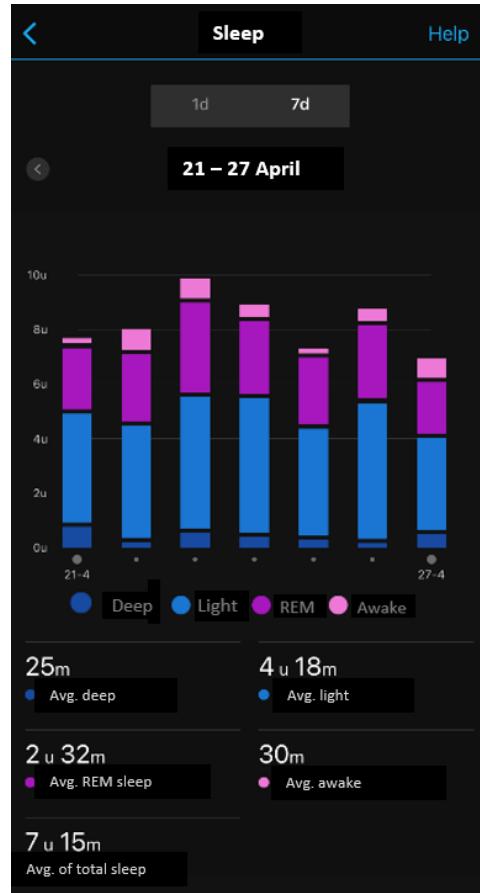


Figure 1.5 – Tracking your sleep

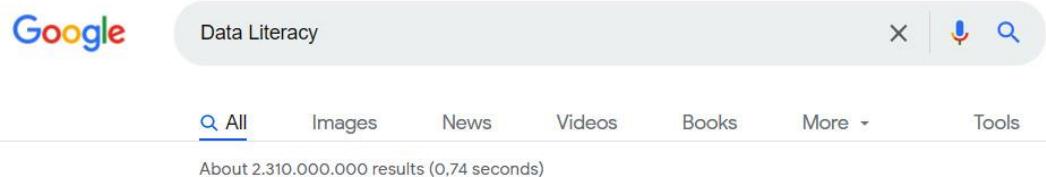


Figure 1.6 – Google search for data literacy



Figure 1.7 – COVID-19 headlines

Tuesday
March, 2nd
2021

Amsterdam News

Latest news & updates

Peter Woodward

Staggering increase in infections continues, the number of new corona patients in hospitals fluctuates!

The dramatic increase in infections continues, and the number of new corona patients in hospitals fluctuates, the number of new corona cases continues to rise dramatically now that the Omicron variant is most common.

Between Tuesday morning and Wednesday morning, 32,149 positive tests were reported. That is the highest number ever for a Wednesday. The number of new admissions of corona patients to the nursing wards has fluctuated in recent days. The experts say that most likely one of the reasons the numbers are rising this much is related to the increased testing capacity and the number of tests taken.



In the past seven days, the National Institute for Public Health was informed that the corona virus has been diagnosed 209,073 times. That equates to an average of 29,868 positive tests per day, the highest level ever. They even found in some areas a dramatic rise of 45%. This average is rising for the fifteenth day in a row during the COVID pandemic.

Figure 1.8 – News article about the COVID-19 pandemic

Tuesday
March, 2nd
2021

Amsterdam News

Latest news & updates

xx

Peter Woodward

Staggering _____ in continues, the number of new corona patients in hospitals fluctuates!

The dramatic _____ in infections continues, and _____ of new corona patients in hospitals fluctuates, _____. _____ of new corona cases continues _____ dramatically now that the Omicron variant is most common.

Between Tuesday morning and Wednesday morning, _____ positive tests were _____. That is the _____ ever for a Wednesday. _____ of new admissions of corona patients to the nursing wards has _____ in recent days. The experts say that most likely one of the reasons _____ this much is related to _____ testing capacity and _____ of tests taken.



In the past _____ days, the National Institute for Public Health was informed that the corona virus has been diagnosed _____. That _____ positive tests per day, _____ level ever. They even found in some areas a dramatic _____. This _____ is rising for the _____ during the COVID pandemic.

Figure 1.9 – News article removing data elements

Tuesday
March, 2nd
2021

Amsterdam News

Latest news & updates

Because more people are tested? Peter A. Voskamp

Staggering increase in infections continues, corona patients in hospitals fluctuates!

The dramatic increase in infections continues, and the number of new corona patients in hospitals fluctuates, the number of new corona cases continues to rise dramatically now that the Omicron variant is most common.

Between Tuesday morning and Wednesday morning, 32,149 positive tests were reported. That is the highest number ever for a Wednesday. The number of new admissions of corona patients to the nursing wards has fluctuated in recent days. The experts say that most likely one of the reasons the numbers are rising this much is related to the increased testing capacity and the number of tests taken.

Data literacy skills helps us to slow down and find out what we are reading.

45% increase in some areas? Where is that? Which areas?

In the past seven days, the National Institute for Public Health was informed that the corona virus has been diagnosed 399,073 times. That equates to an average of 29,968 positive tests per day, the highest level ever. They even found in some areas a dramatic rise of 45%. This average is rising for the fifteenth day in a row during the COVID pandemic.

Figure 1.10 – Questioning the news article to understand the message better

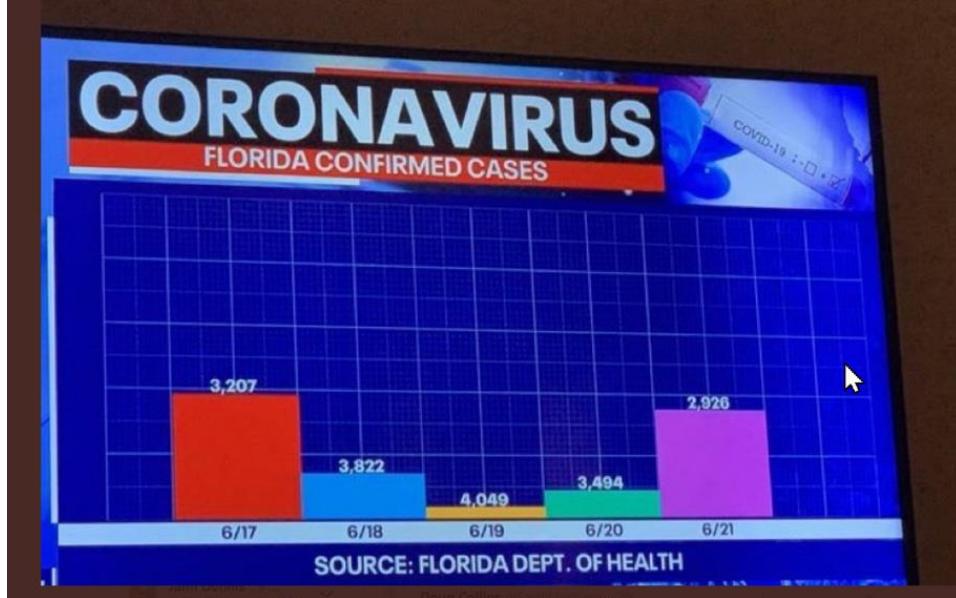


Figure 1.11 – Misleading graphs on TV

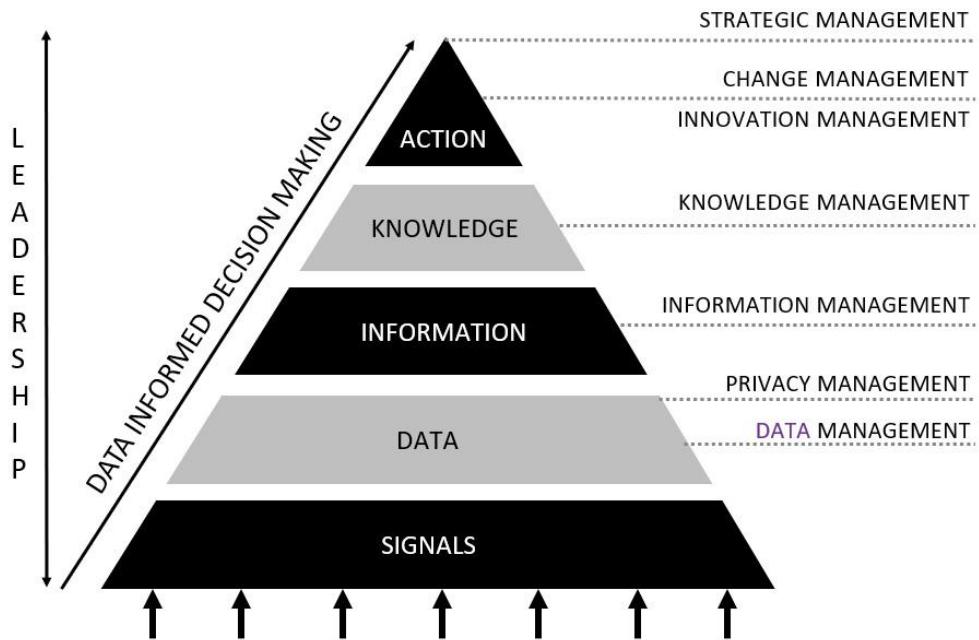


Figure 1.12 – The flow of data

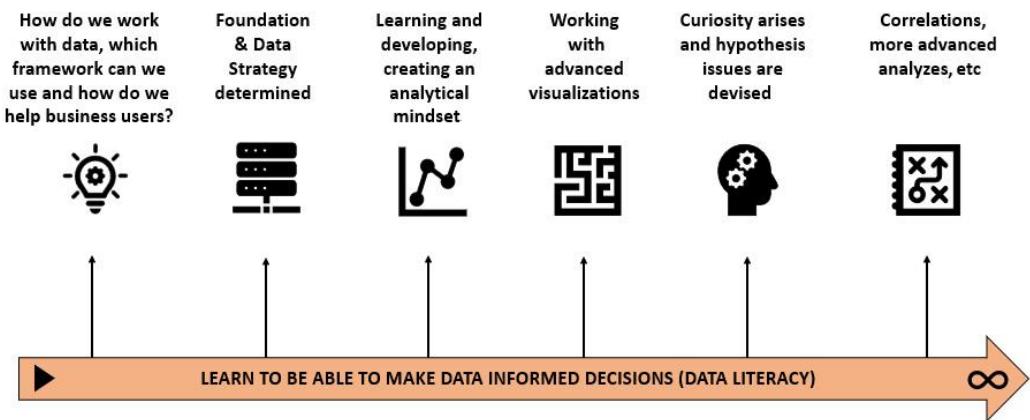


Figure 1.13 – The data-informed decision-making journey

Chapter 2

Name	Value in bytes
Kilobyte (KB)	1,000
Megabyte (MB)	1,000,000
Gigabyte (GB)	1,000,000,000
Terabyte (TB)	1,000,000,000,000
Petabyte (PT)	1,000,000,000,000,000
Exabyte (EB)	1,000,000,000,000,000,000
Zettabyte (ZB)	1,000,000,000,000,000,000,000
Yottabyte (YB)	1,000,000,000,000,000,000,000,000

Figure 2.1 Understanding data growth

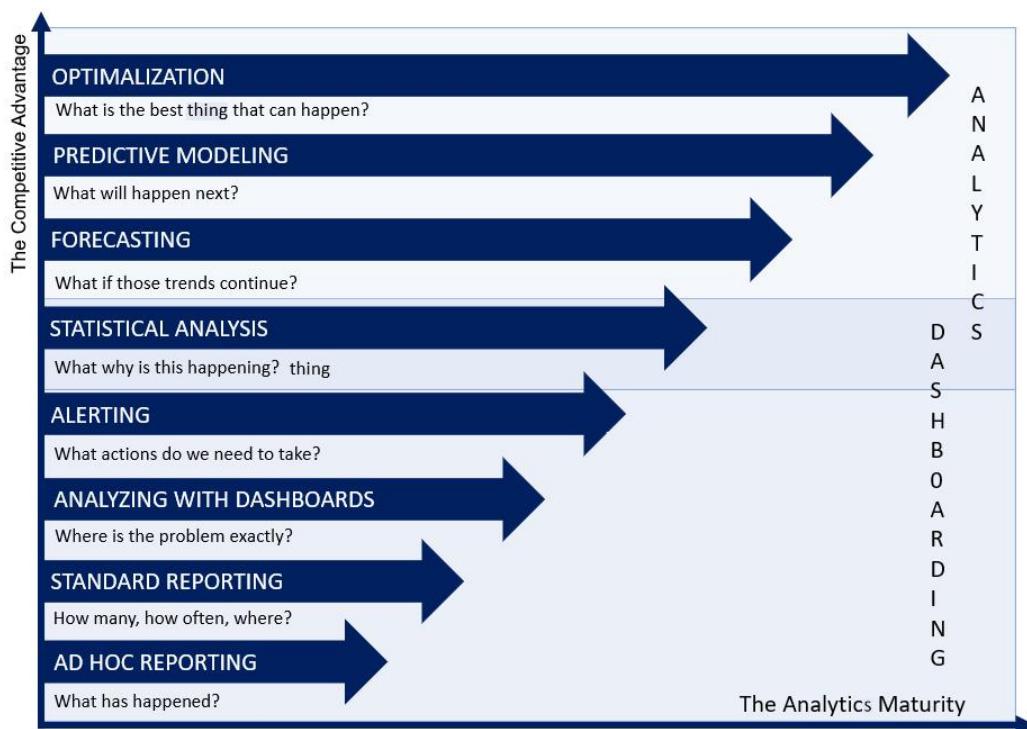


Figure 2.2, Introducing analytical maturity

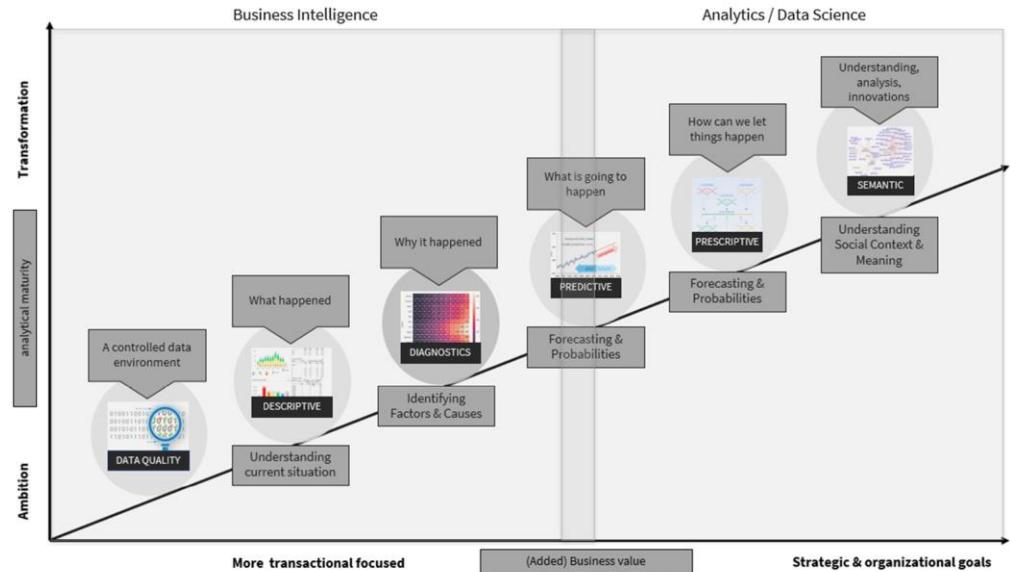


Figure 2.3, The different phases in the world of data and analytics

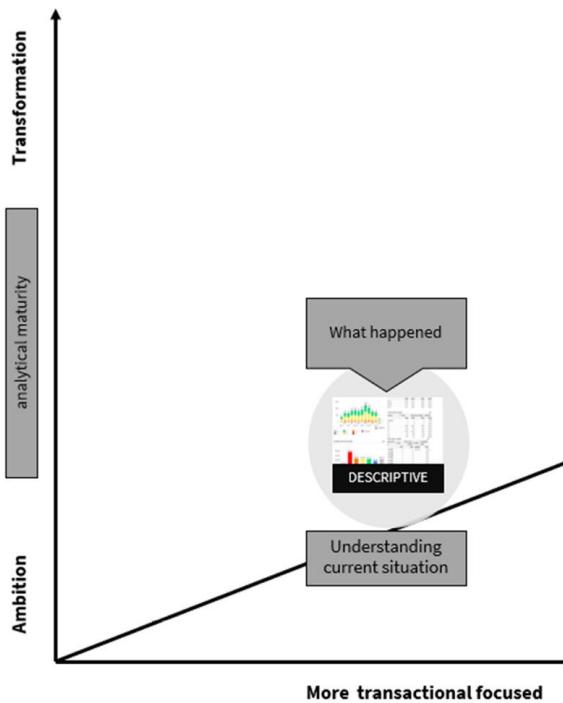


Figure 2.4, The first step of the data journey: the phase of descriptive analytics

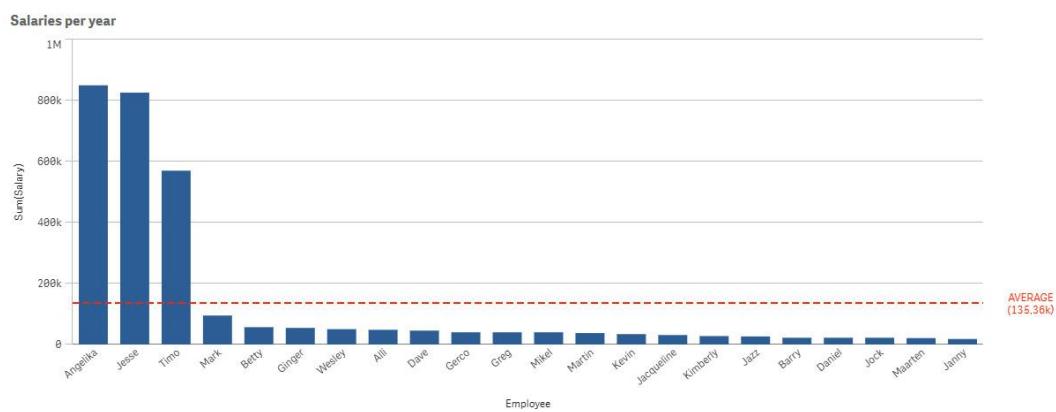


Figure 2.5, Describing the aggregation: average

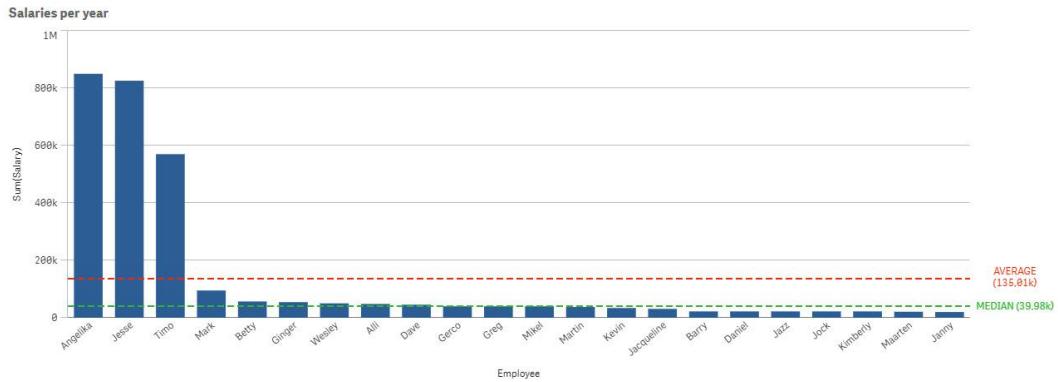


Figure 2.6, Describing the aggregation: median

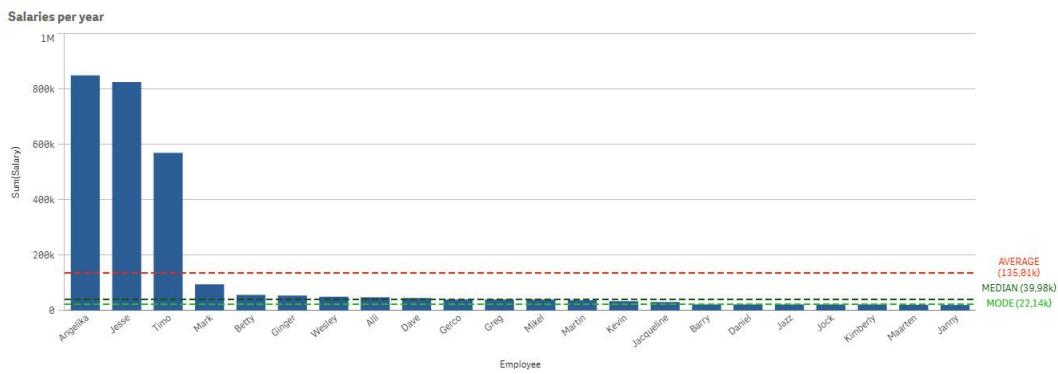


Figure 2.7, Describing the aggregation: mode



Figure 2.8, The usage of a distribution plot to find outliers

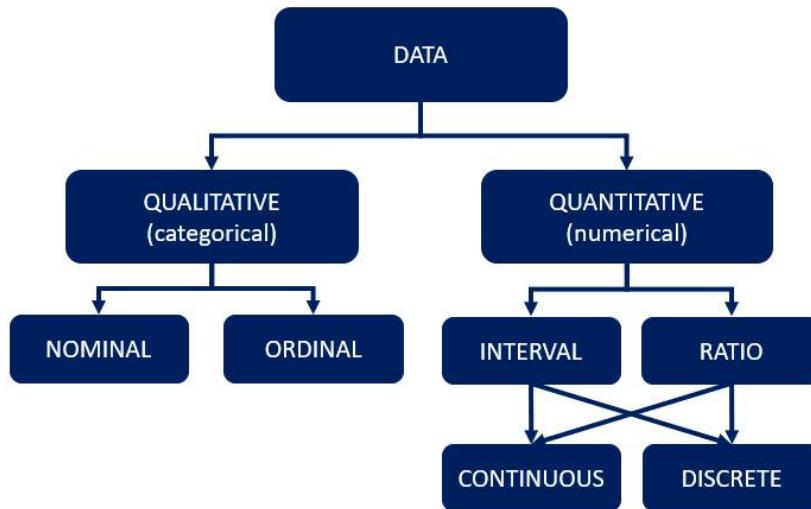


Figure 2.9, The different forms of data types

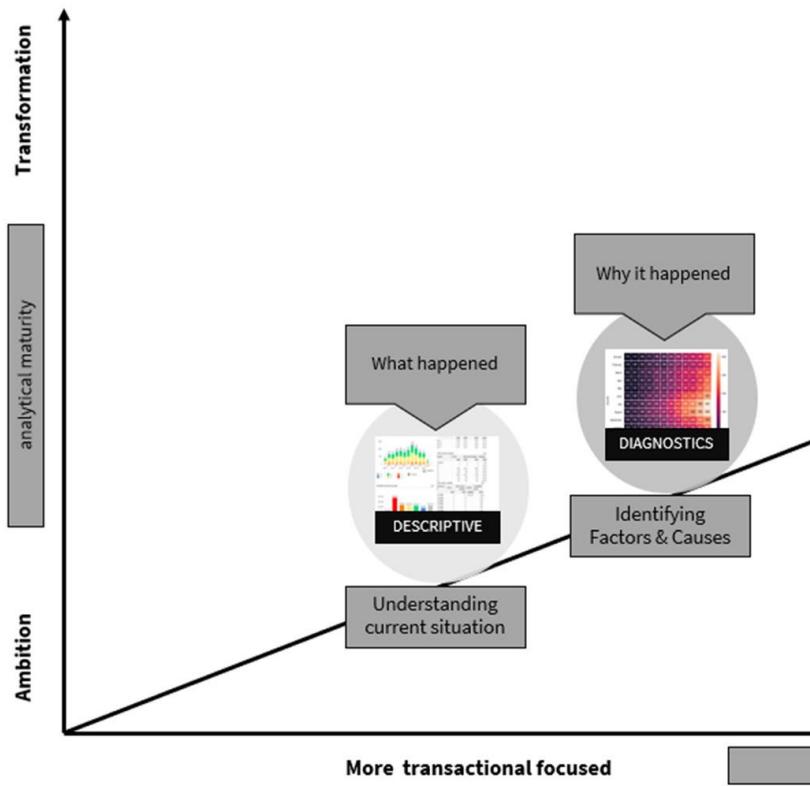


Figure 2.10, Exploratory analysis: the second step of your data journey, the phase of diagnostic analysis

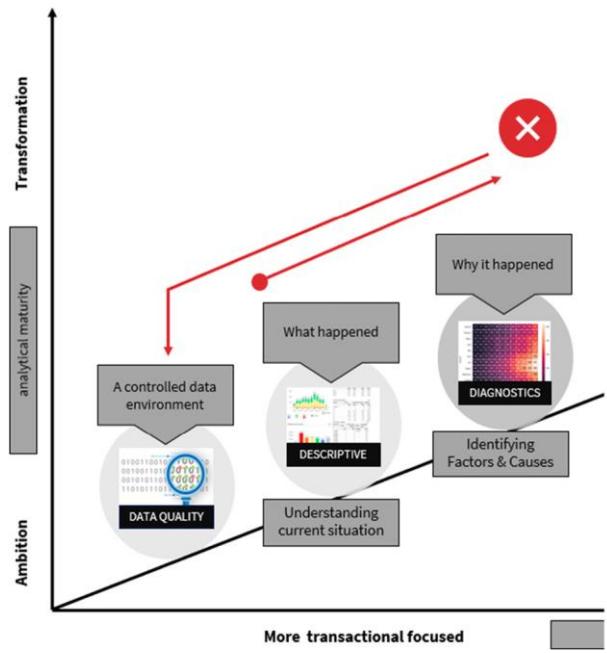


Figure 2.11, The third step of your data journey: the discovery of data quality issues

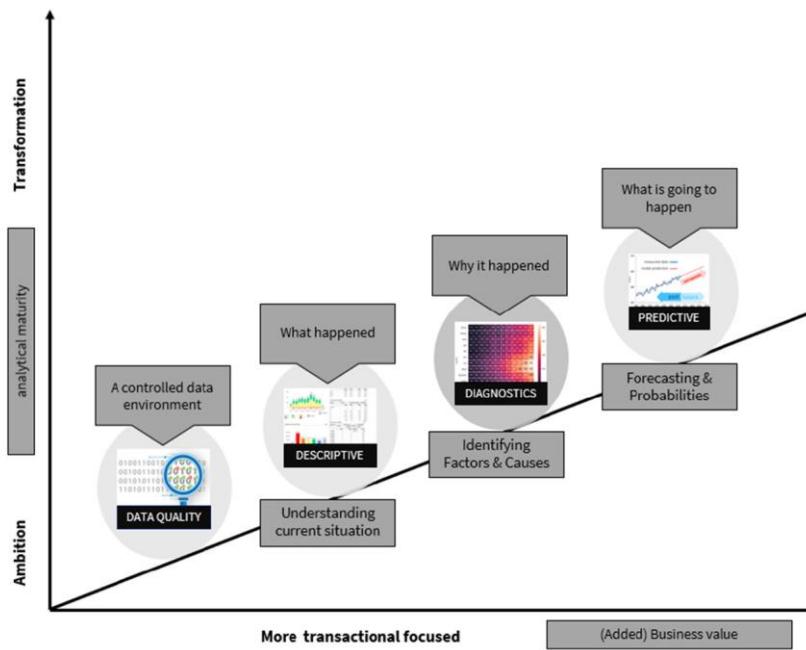


Figure 2.12, The fourth step of your data journey: the phase of predictive analytics

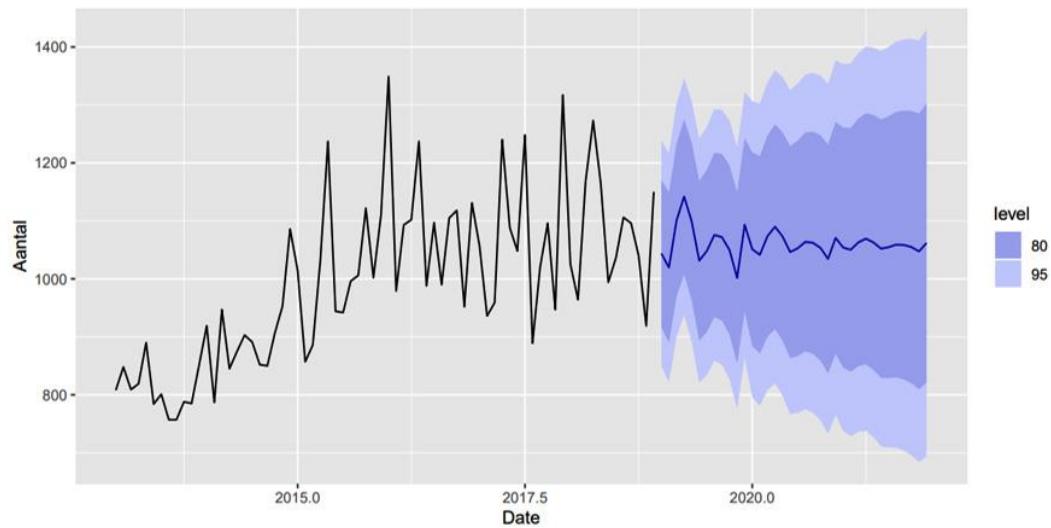


Figure 2.13, An example of a forecast with the number of possible requests with the expected numbers (note: a prediction is always wrong)

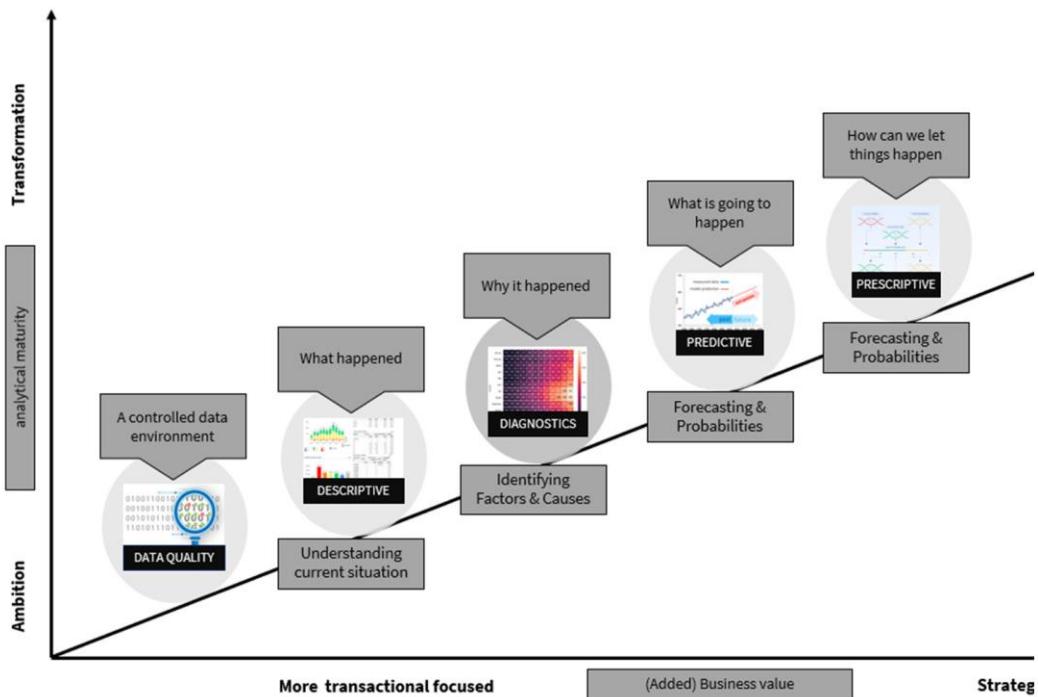


Figure 2.14, The fifth step of your data journey: the phase of diagnostic analysis

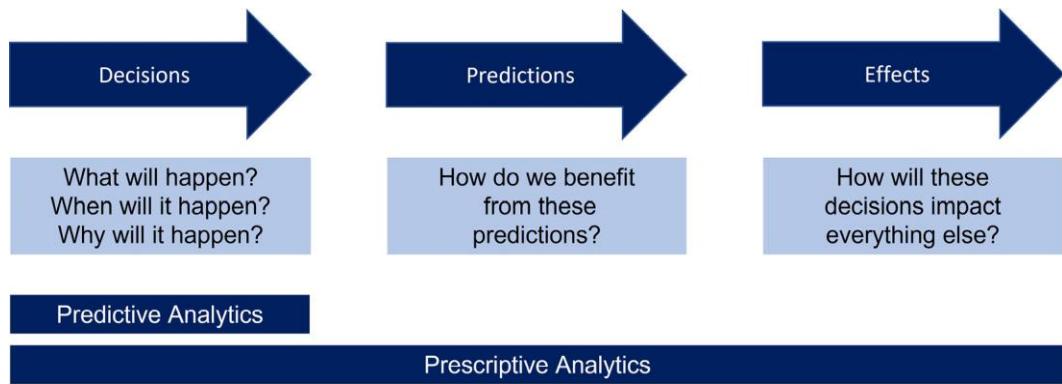


Figure 2.15, Describing the prescriptive model and things to think of

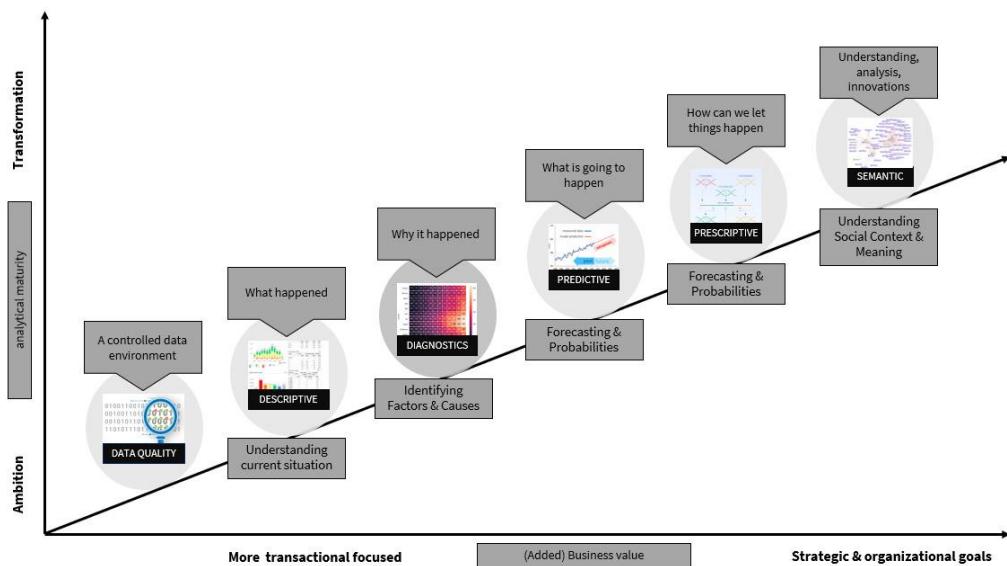


Figure 2.16, The last step of your data journey: the phase of semantic or artificial intelligence

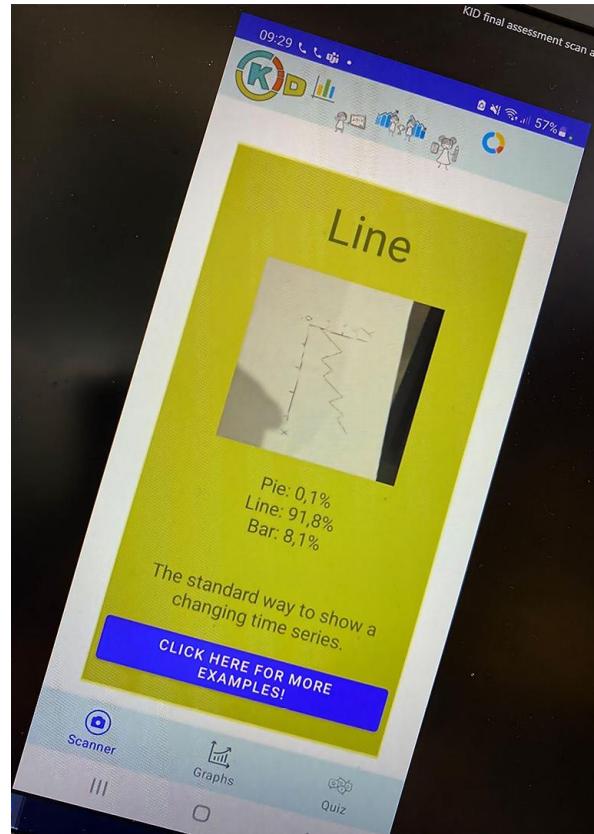


Figure 2.17, Scan app: Kids in Data

Call to Balloon	The total time from the emergency call to the performed treatment, the angioplasty
Response time	The time from the emergency call to the arrival of the ambulance
Call to Door	The time from the emergency call to transfer the hospital personnel
Door to Balloon	The time from transfer by the hospital personnel to angioplasty
ECG to Balloon	The elapsed time between the ECG (by ambulance personnel) to angioplasty

Figure 2.18, The used definitions for the performance indicators that we used in this project

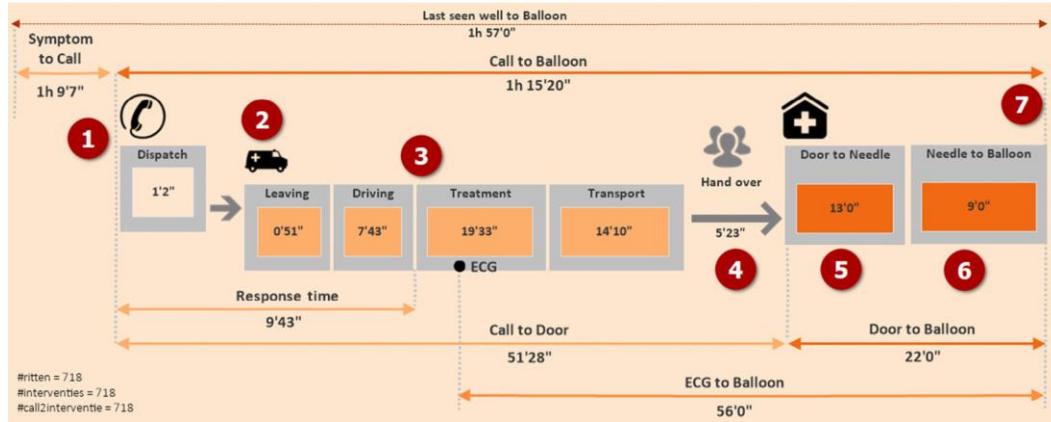


Figure 2.19, The visualized process schema of Call to Balloon

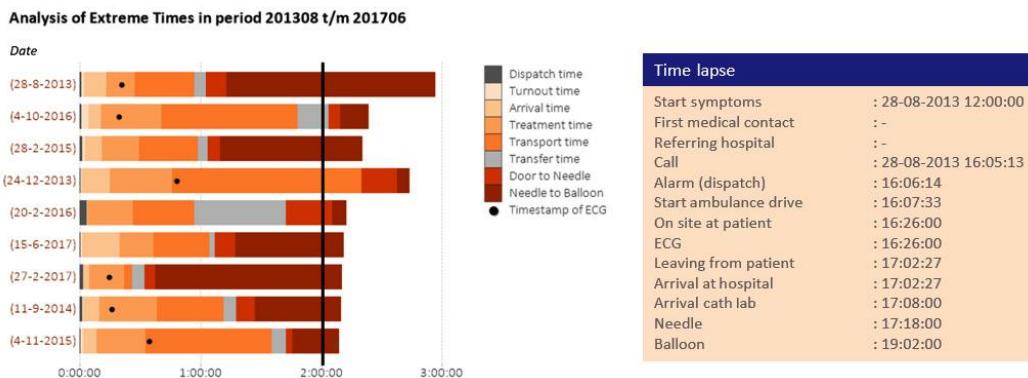


Figure 2.20, The visualization that the safety region used to analyze the process steps

Chapter 3

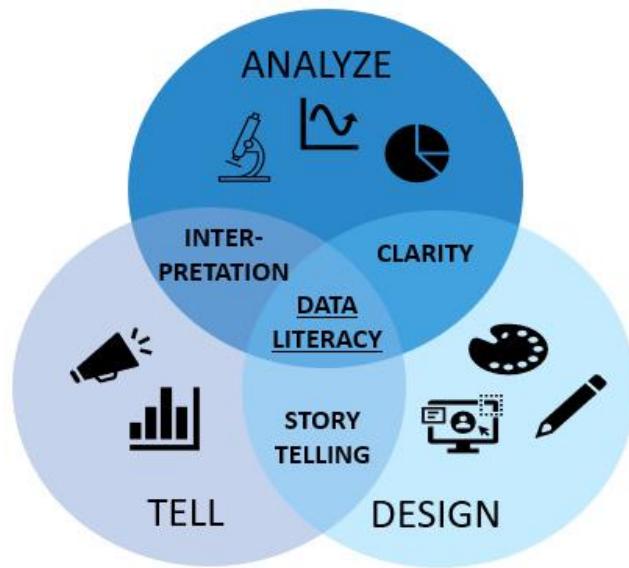


Figure 3.1 – Defining data literacy skills

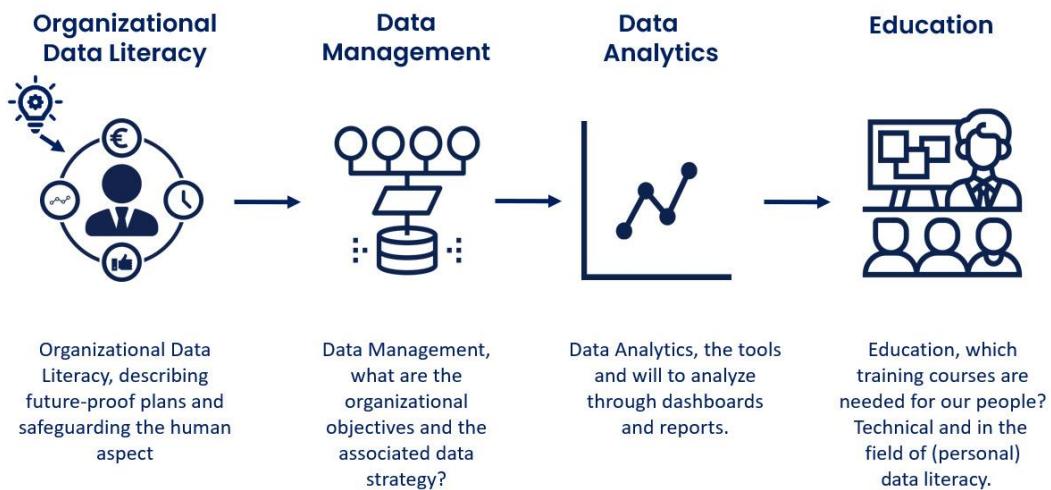


Figure 3.2 – Describing the four fundamental pillars



Figure 3.3 – The first pillar: organizational data literacy



Figure 3.4, The second pillar: data management



Figure 3.5, The third pillar: data analytics

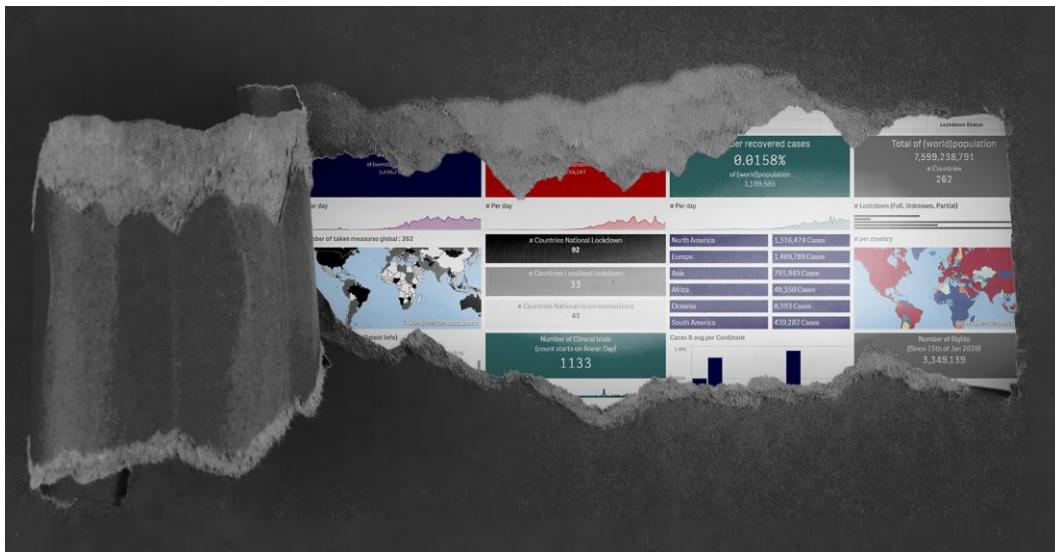


Figure 3.6, Dare to decide

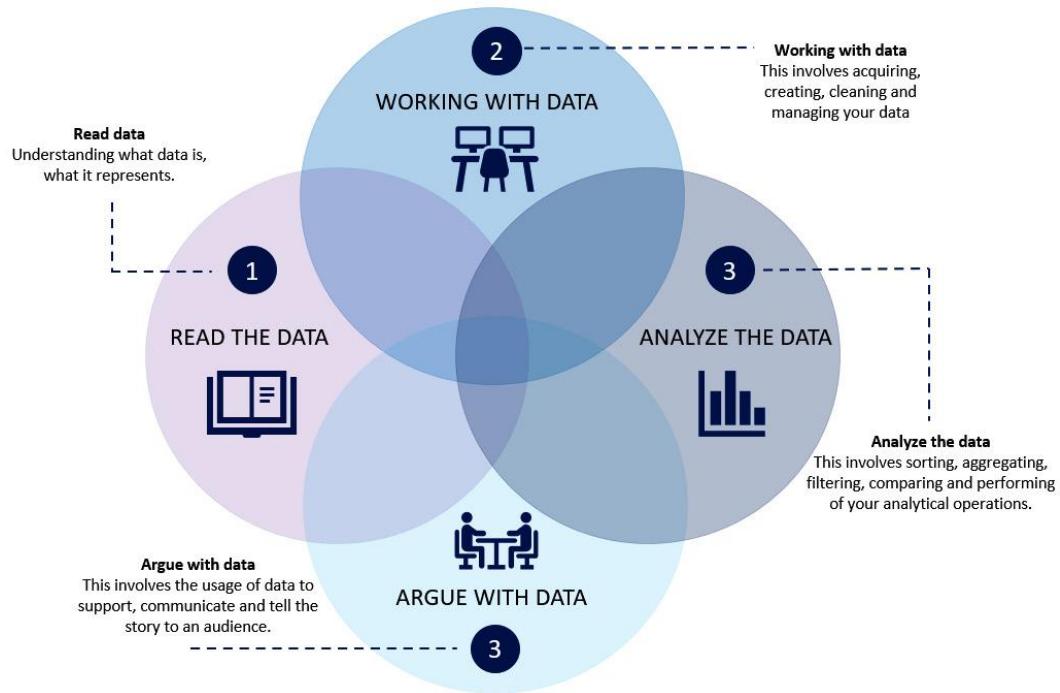


Figure 3.7, The four primary elements of data literacy

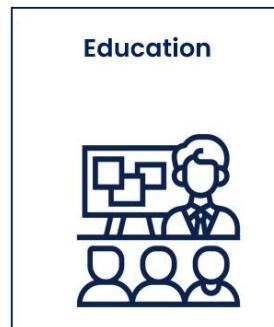


Figure 3.8, The fourth pillar: education

Non-Functional Requirements	
Number of expected users	<10 10-50 50-100 100-250 >250
Expected frequency of use	ad-hoc daily weekly monthly quarterly yearly
Expected mode of use	Standard reporting Search interactively
Up-to-datedness of the data (Refresh rate)	ad-hoc daily weekly monthly quarterly yearly
Application availability	
Addition and or explanation	

Figure 3.9, NFR project intake

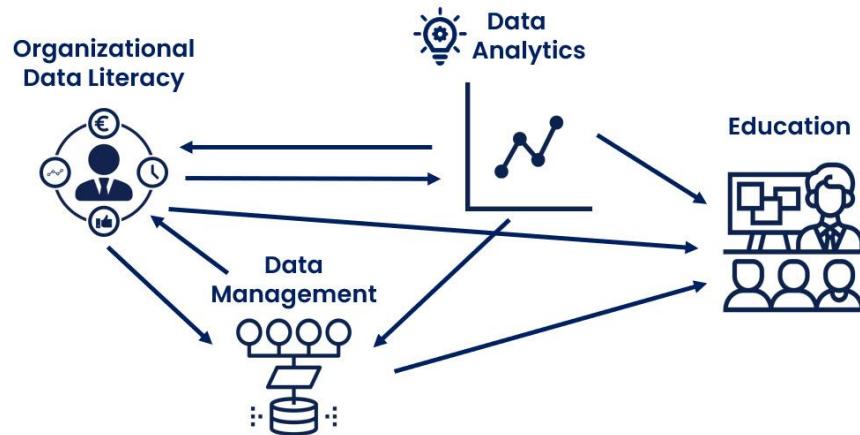


Figure 3.10, There are various ways to use the pillars

Further Reading

- For more information about first pillar, go to the presentation by Mike Capone, CEO of Qlik, October 2018 at <https://bit.ly/3pdNMqs>.
- For more information about second pillar, go to the *TechTarget*, October 2019 at <https://www.techtarget.com/searchdatamanagement/definition/data-management>.

- For more information about the third pillar, visit:
<https://towardsdatascience.com/why-visual-literacy-is-essential-to-good-data-visualization-5b9dfb5aa61>.
- The full story can be found at <https://www.linkedin.com/pulse/torn-between-two-angelika-kidas/>.
- The complete report on Gartner Magic Quadrant can be found on Gartner's website (<https://www.gartner.com/doc/reprints?id=1-292LEME3&ct=220209&st=sb>). Also, reports from Forrester (www.forrester.com) and BARC (<https://barc-research.com/research/b-trend-monitor/>) can help to make the choice that best suits your organization.
- There are reports available from Gartner that state that half of IT spending will be shifted to the cloud by 2025
(<https://www.gartner.com/en/newsroom/press-releases/2022-02-09-gartner-says-more-than-half-of-enterprise-it-spending#:~:text=Accelerating%20Shift%20to%20the%20Cloud,%20according%20to%20Gartner%2C%20Inc>).

Chapter 4

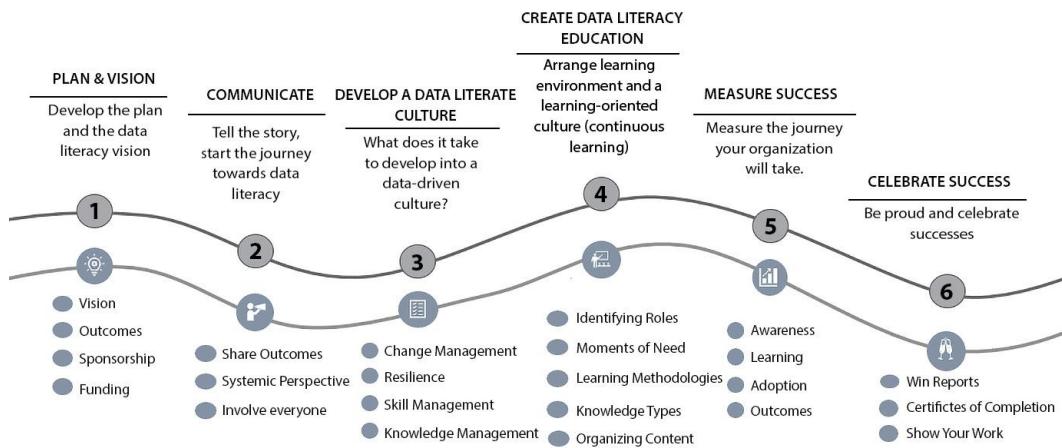


Figure 4.1 – Implementing organizational data literacy

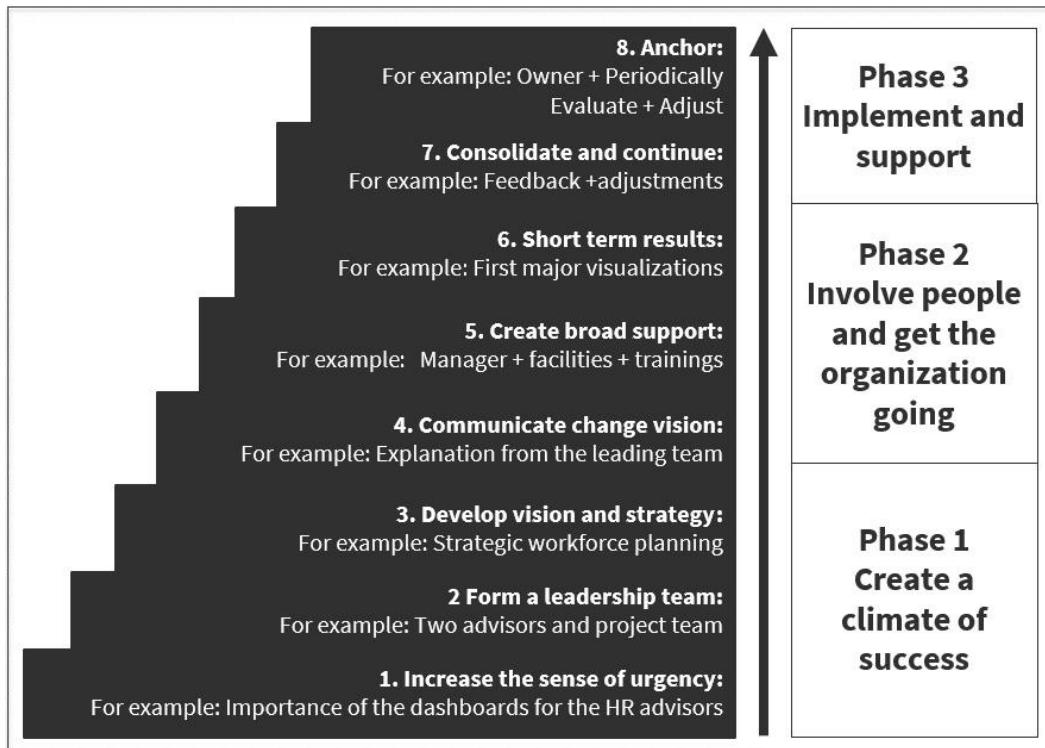


Figure 4.2 – Change management step



Flexibility & Adaptability	Changing business model. New work environment. New work tasks.
Work-life boundaries	New work times
Connect	Leaned on work groups, family, friends, and social groups
Hope	The pandemic will end eventually, and we will get through this

Figure 4.3 – Resilience during COVID-19



Figure 4.4 – DIKW pyramid

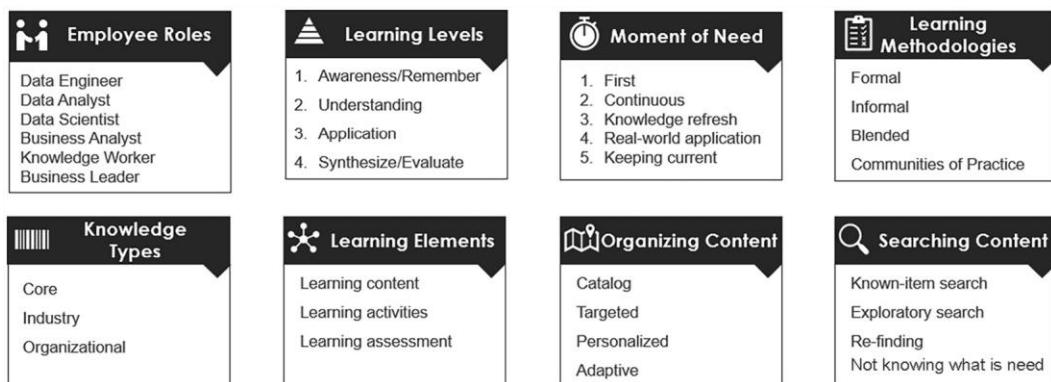


Figure 4.5 – Data literacy educational program dimensions



Figure 4.6 – The five moments of need of an employee

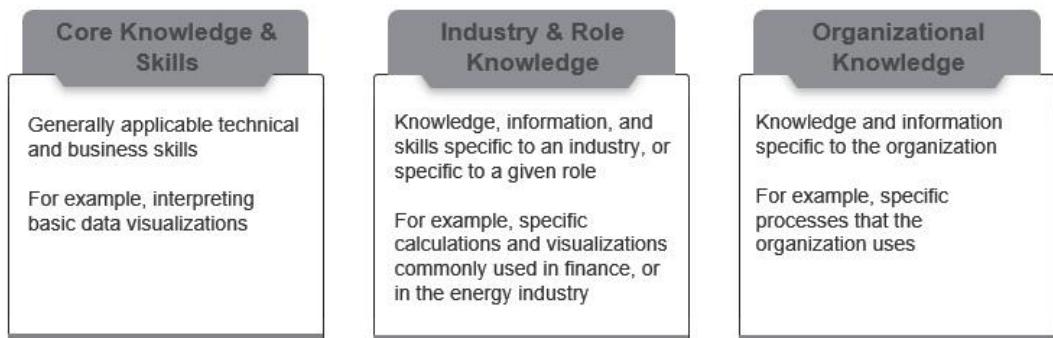


Figure 4.7 – Three types of content to include in your data literacy education

Push and Pull	As Close to Point of Need as Possible	Context-Specific	
Catalog	Targeted	Personalized	Adaptive
Everything listed	Role-based	Rules-based	Data Driven
Searchable	Product-based	Decision Tree	Improves overtime
	Experience-based		
	Geography-based		

Figure 4.8 – Ways to organize data literacy educational content

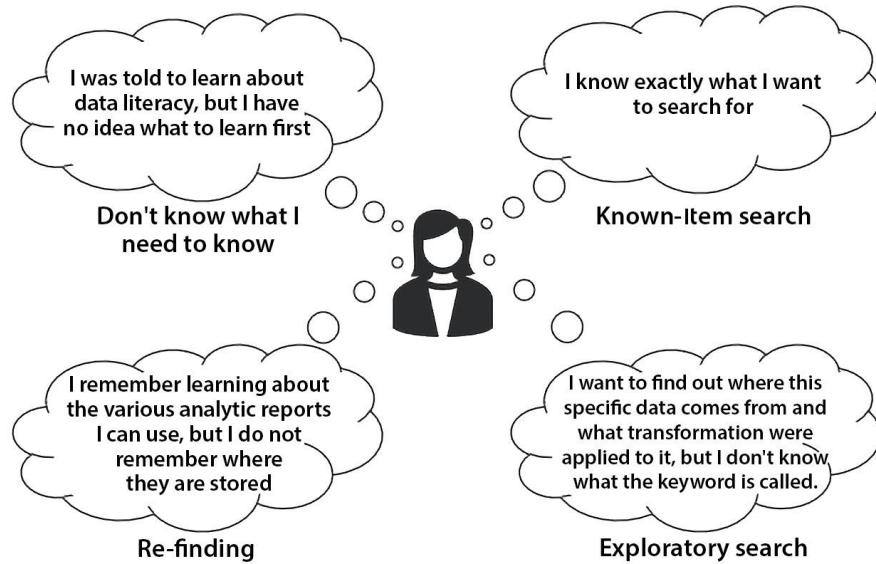


Figure 4.9 – How employees search for data literacy education

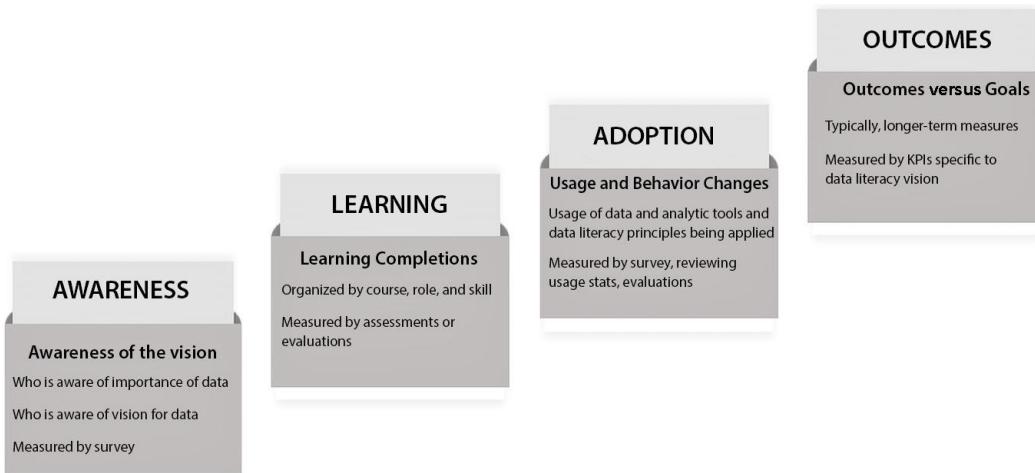


Figure 4.10 – Four levels of measurement of data literacy initiatives

Further reading

- Please go through the following resource for further reading:
- *Kotter, J.P. (2012), Leading Change, Harvard Business Review Press.*

Chapter 5

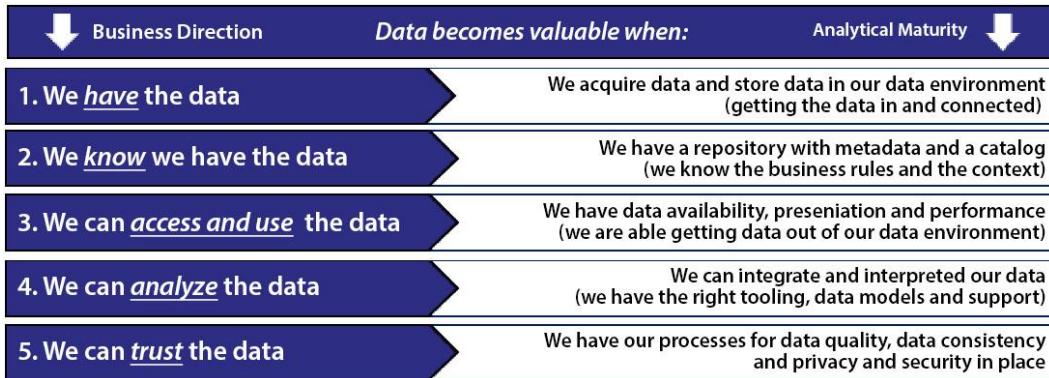


Figure 5.1 – Chart showing when data becomes valuable



Figure 5.2, Example of a data quality report



Figure 5.3, The five-step approach of data management

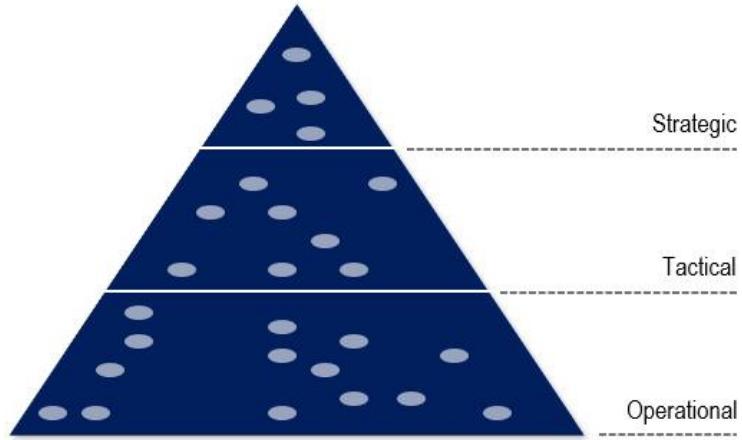


Figure 5.4, The information pyramid



Figure 5.5, The data quality process

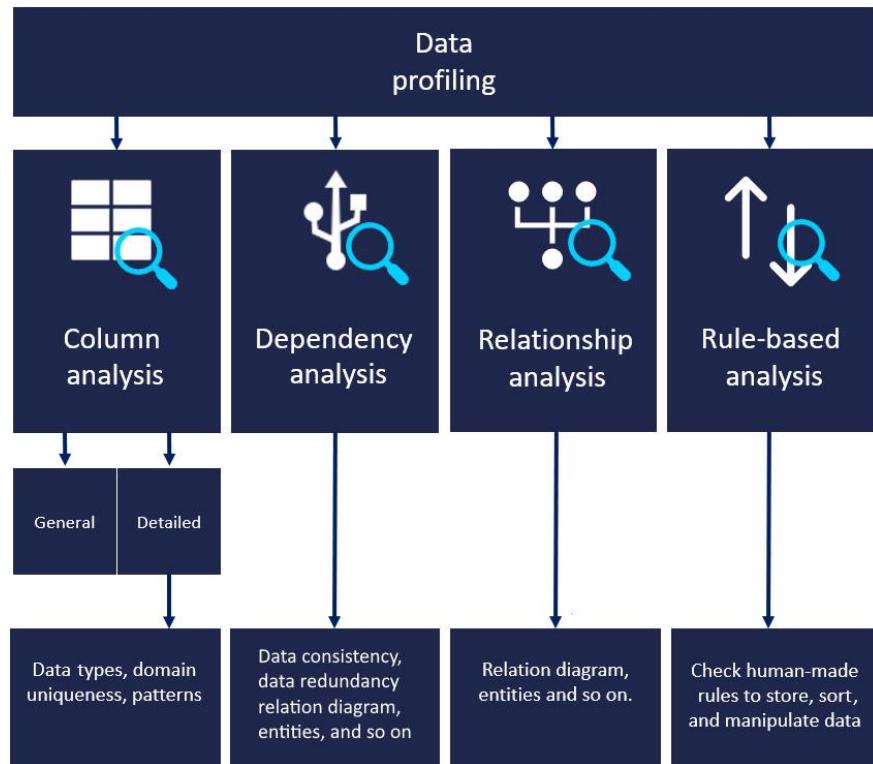


Figure 5.6, Data profiling

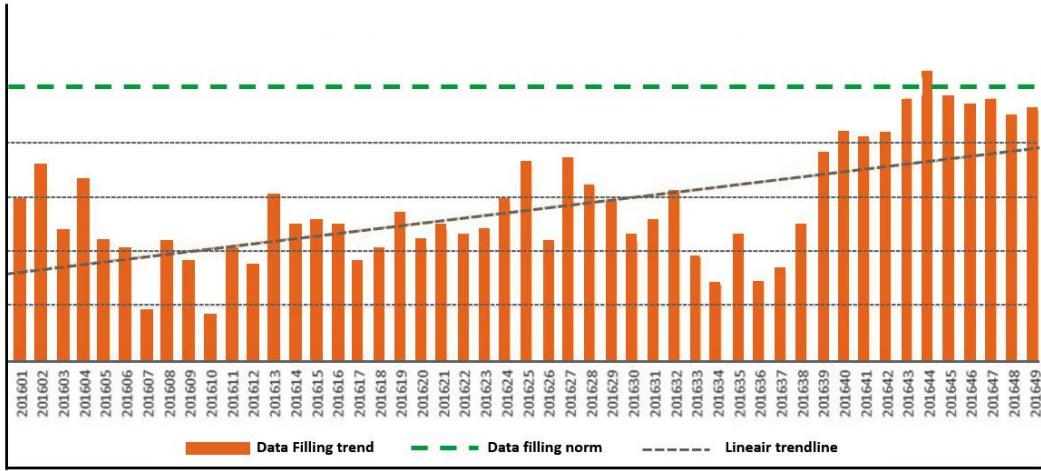


Figure 5.7, Data storage monitoring

Vizlib Writeback Table - Assignment of Studies							
Edit	NCT ID	Estimated Enrollment	Approver	Milestone 1 Risk	Milestone 2 Risk	Comment	Anticipated Completion
<input checked="" type="checkbox"/>	NCT04644328	20,000,000	Carole Lombard	● Low	○ Medium	On track to complete by year end	12/31/2020
<input checked="" type="checkbox"/>	NCT04303507	40,000	Barbara Stanwyck	[Pick Milestone 1 Risk]	[Pick Milestone 2 Risk]	Ready	5/31/2020
<input checked="" type="checkbox"/>	NCI046118U2	30,000	Bette Davis	● Low	○ Medium	Reduced patient population	11/30/2020
<input checked="" type="checkbox"/>	NCT04641481	25,800	Audrey Hepburn	● Low	⚠ High	Supply delays from COVID	12/24/2020
<input checked="" type="checkbox"/>	NCT04371419	15,475	Bette Davis	○ Medium	[Pick Milestone 2 Risk]	looked at new enrollments	10/13/2020
<input checked="" type="checkbox"/>	NCT04416308	13,000	Audrey Hepburn	● Low	● Low	Updated the enrollment	12/9/2020
<input checked="" type="checkbox"/>	NCT04400838	12,390	Audrey Hepburn	● Low	○ Medium	May have to pause study based ...	10/23/2020
<input checked="" type="checkbox"/>	NCT04536051	10,300	Cary Grant	○ Medium	⚠ High	Reduced study from population	12/7/2020

Figure 5.8, Using a write-back functionality

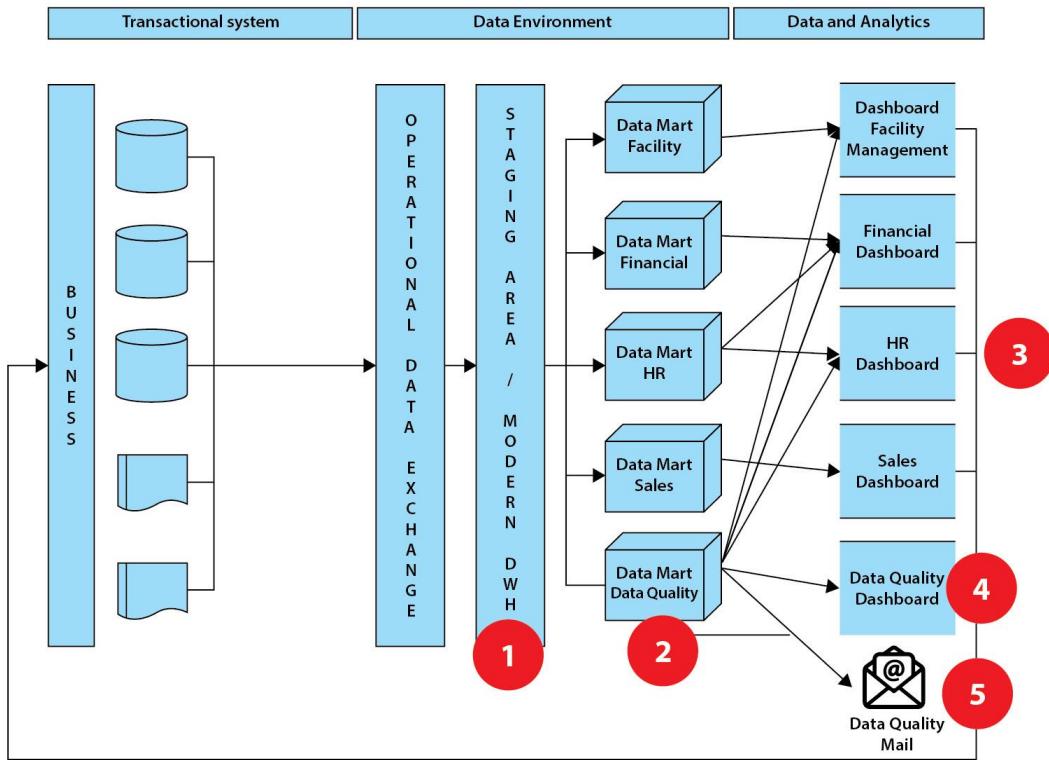


Figure 5.9, Standard data warehouse setup with quality checks



Figure 5.10, Data quality dashboard

Chapter 6

Strategic goal: live long enough to see my great grandkids

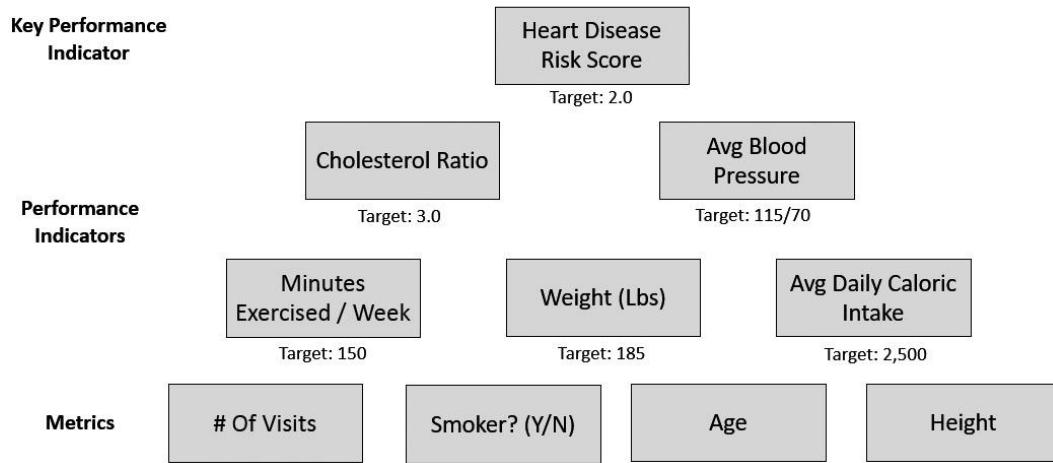


Figure 6.1 – Example performance indicators for personal health

Strategic goal:
Increase annual revenue by 25% compared to last year

DEPARTMENT	GOAL	PERFORMANCE INDICATORS
Sales	Close 15 deals each month	Conversion rate Target: 65%
		Avg length of sales cycle Target: 30 days
Marketing	Add 25 new qualified opportunities each month	# event attendees Target: 120
		# of unique website visitors monthly Target: 450
Customer Success	Customer retention at 95% each month	Customer health score Target: 85%

Figure 6.2 – Example performance indicators for an organization

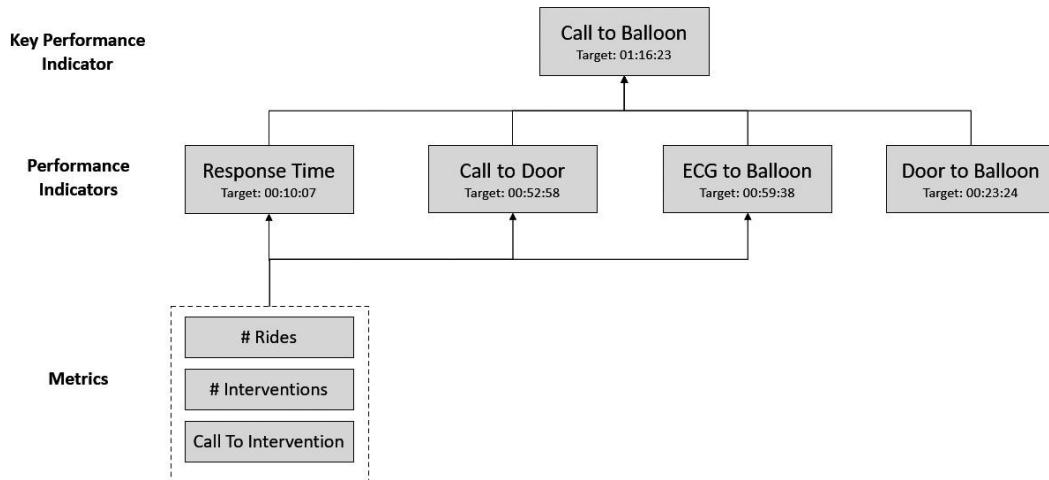


Figure 6.3 – Example performance indicators for a safety region



Figure 6.4 – Example 1 of KPI visualizations

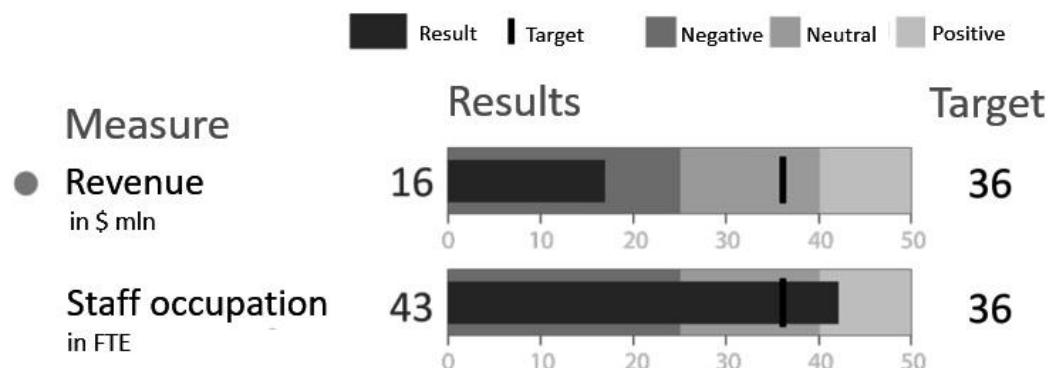


Figure 6.5 – Example 2 of KPI visualizations

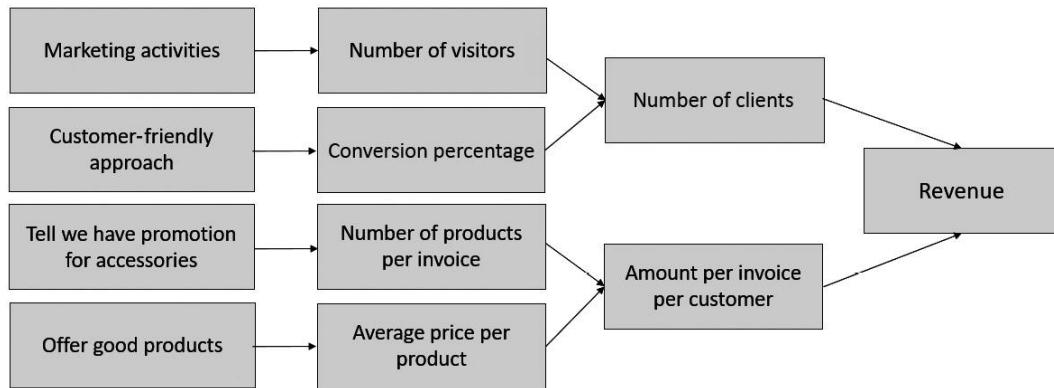


Figure 6.6 – Example leading and lagging indicators

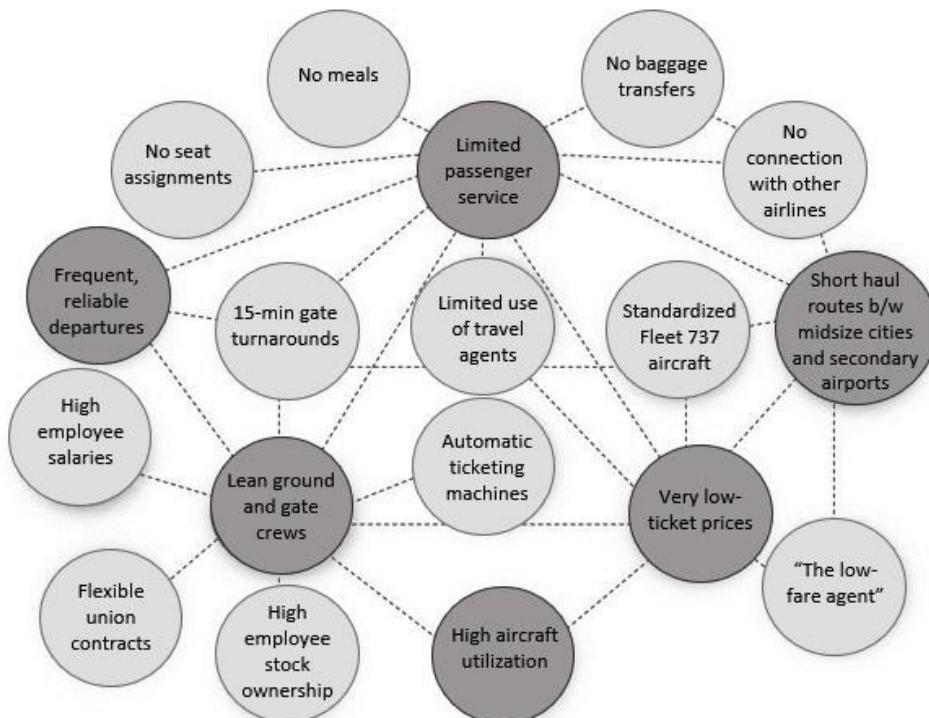


Figure 6.7 – Southwest Airlines activity systems map (Source: Michael E. Porter, "What is Strategy," *Harvard Business Review*, Nov— Dec, 1996, 61-78.)

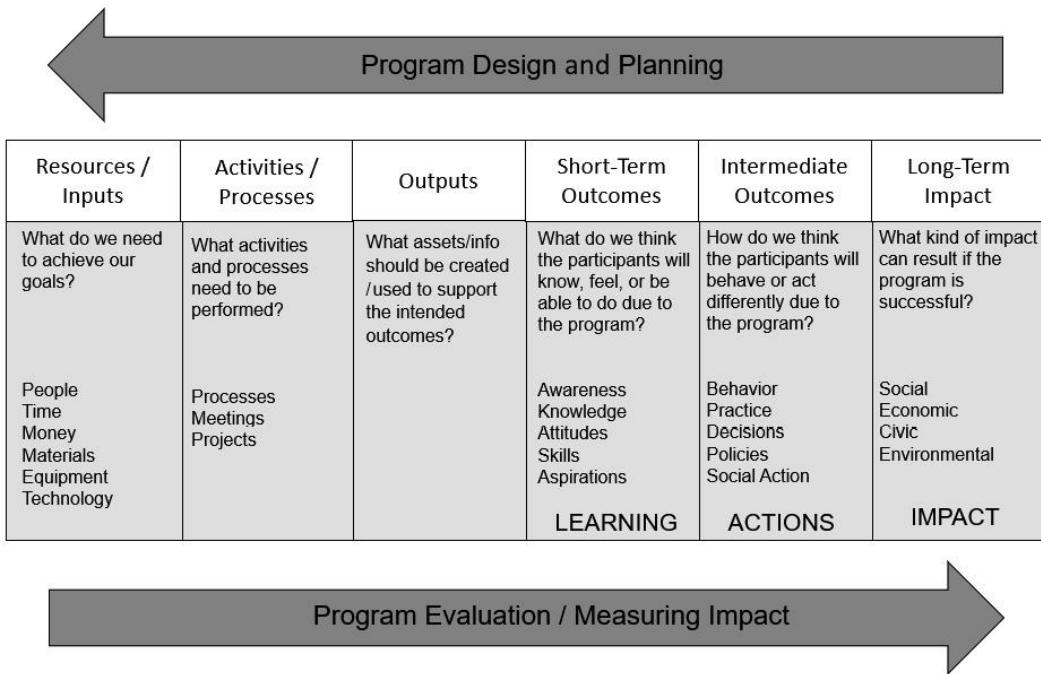


Figure 6.8 – Logic model template

	Outcomes	Starting at the end, the outcomes for the new coffee retailer are an avid customer fan base, repeat customers, and increases in sales revenue.
	Output	Next, the coffee retailer needs to determine what will help drive those outcomes. In their case, they see the number of coffees brewed and the number of coffees sold as measurable outputs.
	Processes	The coffee retailer would like customers to be able to come whenever they want and receive the same, consistent cup of coffee. They would also want to keep the process as efficient as possible to keep costs down.
	Inputs	The resources for the coffee retailer include things such as the coffee beans, water, workers, and the time invested. The activities include the entire coffee-making process.

Figure 6.9 – Logic model for a coffee retailer

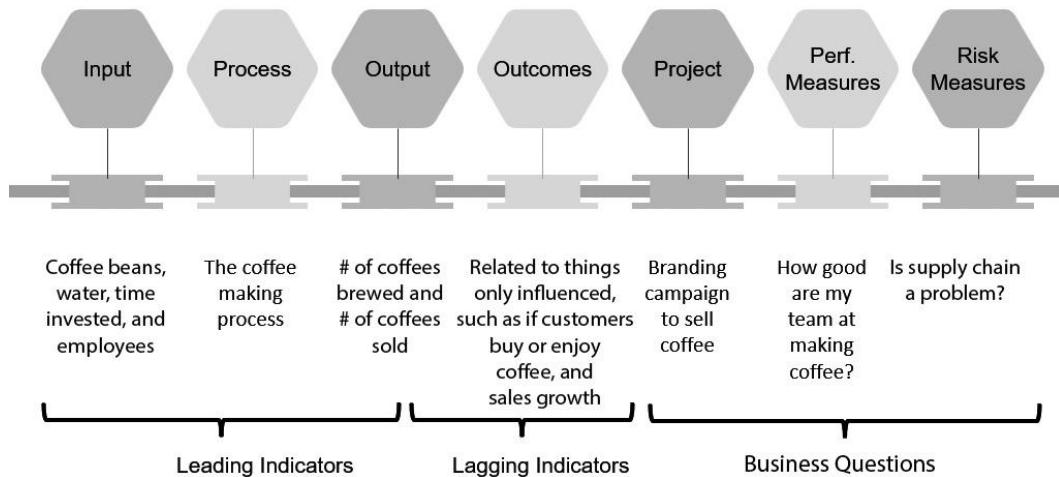


Figure 6.10 – Example logic model for a coffee retailer

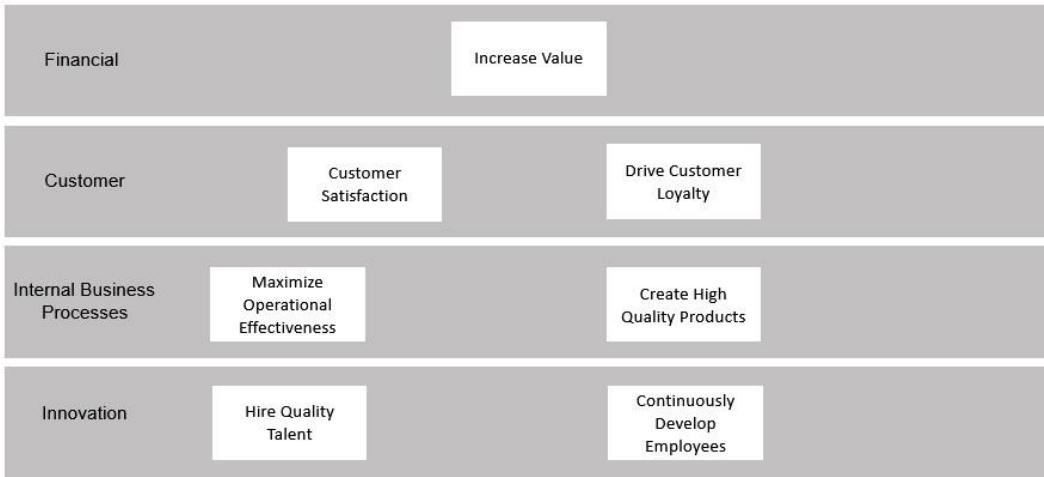


Figure 6.11 – Sample balanced scorecard

Further Reading

- “Overcoming the ‘Cobra Effect’ in Your Business,” Continuous Business Planning, accessed June 21, 2022:
- <https://www.continuousbusinessplanning.com/blog/02082019123204-overcoming-the-cobra-effect-in-your-business/>
- Porter, M. From Competitive Advantage to Corporate Strategy. Harvard Business Review. 1987.
- E. Mazareanu, “Enplaned Passengers on Domestic Flights – Airlines in U.S. 2011–2019,” Statista.com, April 1, 2020.
- W.K. Kellogg Foundation. *W.K. Kellogg Foundation Logic Model Development Guide*. Battle Creek: W.K. Kellogg Foundation. 1998.

Chapter 7

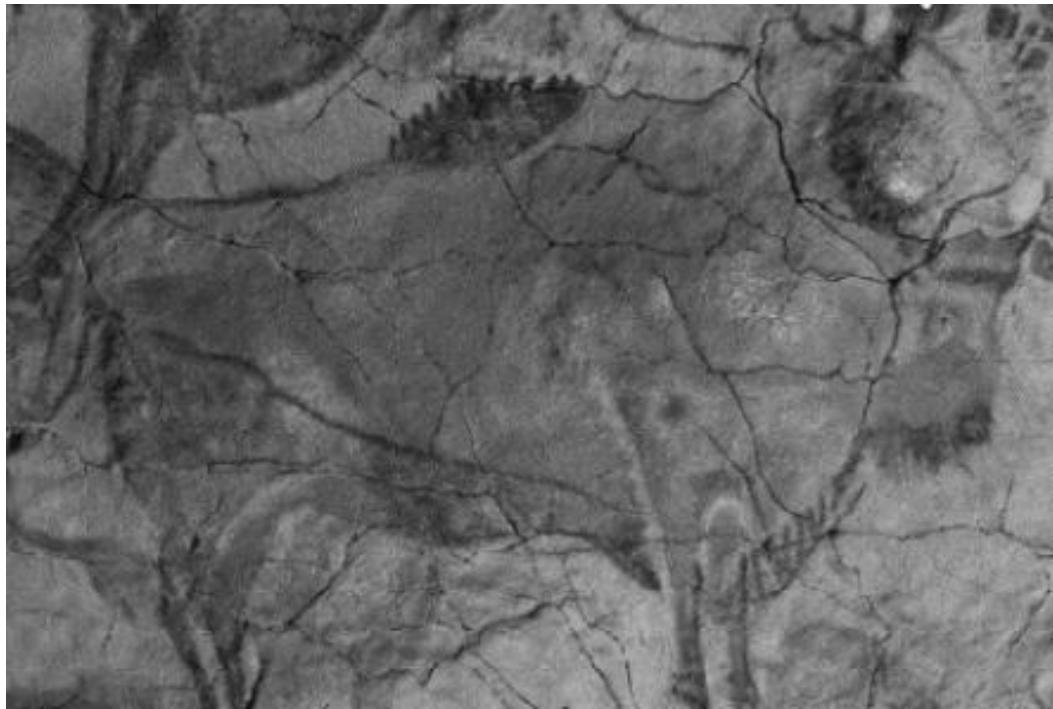


Figure 7.1 – A rock painting in Altamira, Spain. The source of the image can be found at <https://www.spain.info/en/places-of-interest/caves-altamira/>.

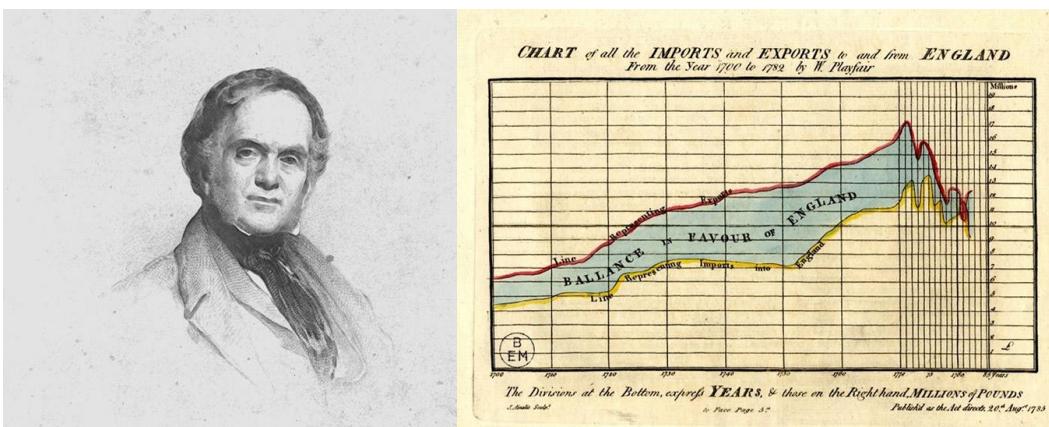


Figure 7.2 – Playfair's Atlas. The source of the image can be found at
<https://www.christies.com/en/lot/lot-5388575>.

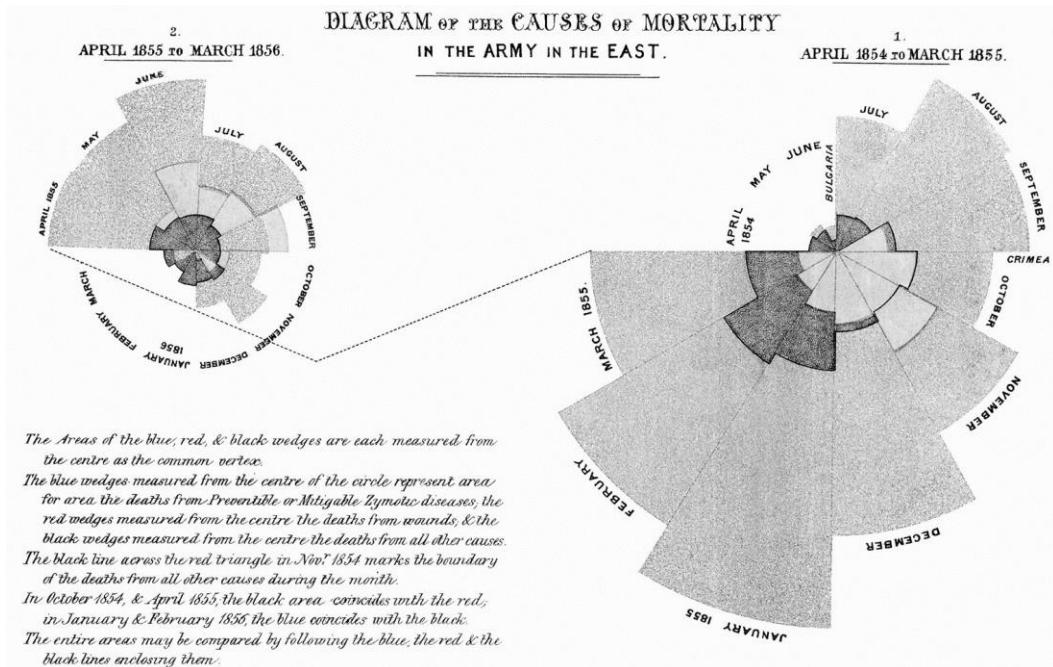


Figure 7.3 – Nightingale's Rose Diagram. The source of the image can be found at
<https://www.historyofinformation.com/image.php?id=851>.



Figure 7.4 – A representation of a bad graph and the correct version

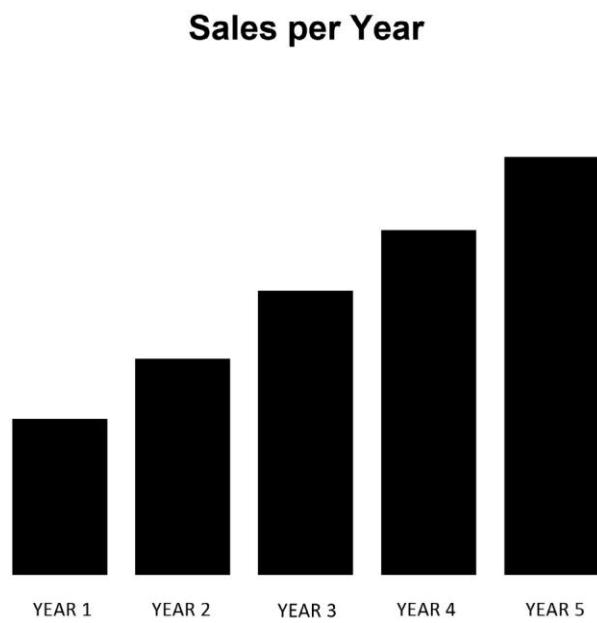


Figure 7.5 – Bad sales graph – 1

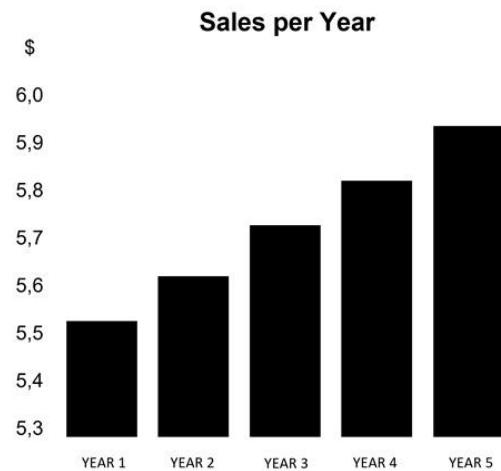


Figure 7.6 – Bad sales graph – 2

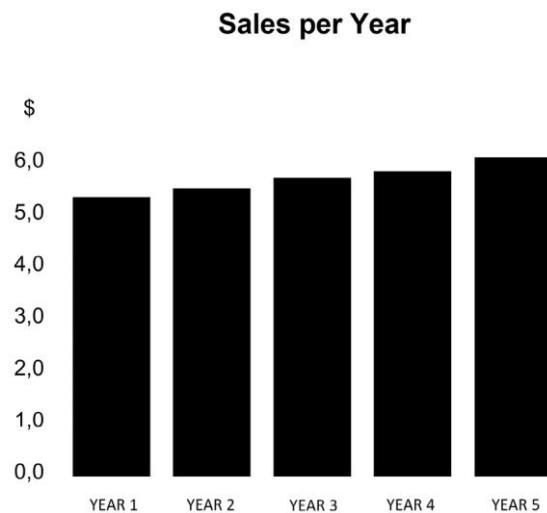


Figure 7.7 – Sales graph – final

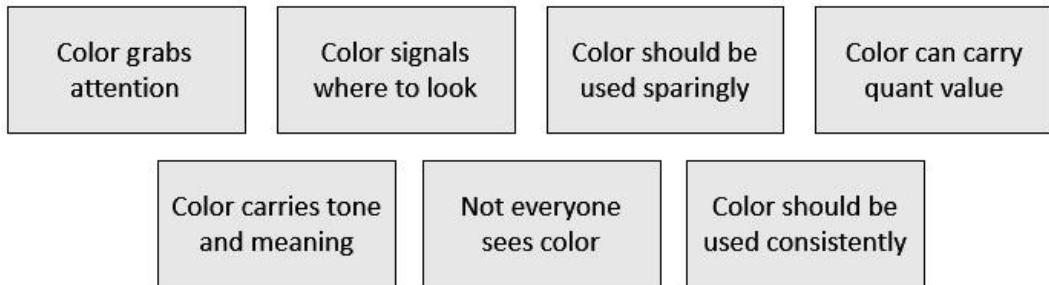


Figure 7.8 – Color does matter

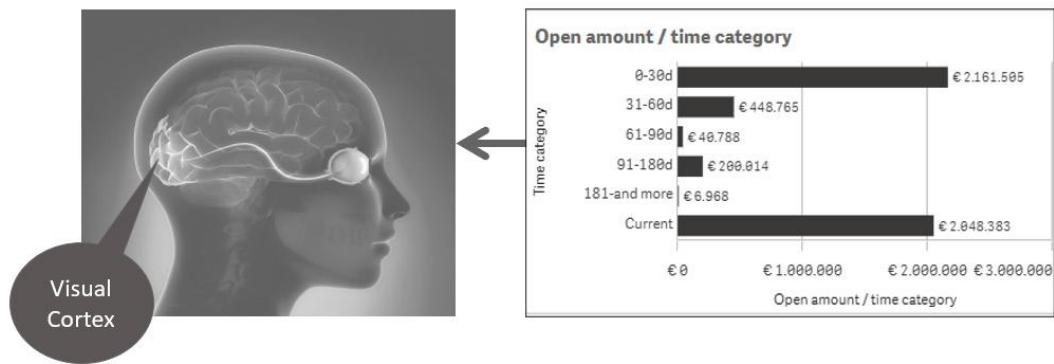


Figure 7.9 – Eye-brain connection



Figure 7.10 – Convex versus concave

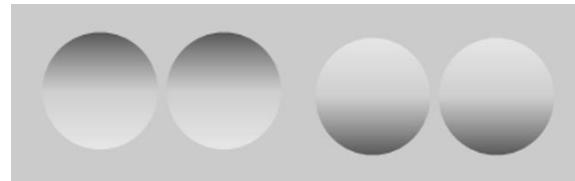


Figure 7.11 – Concave versus convex

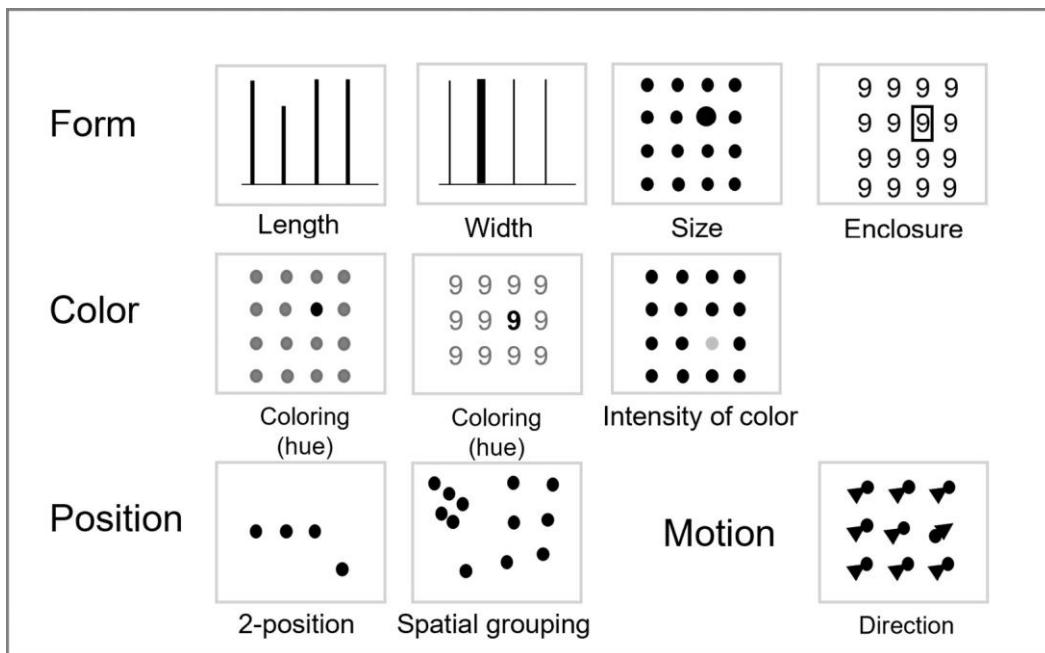


Figure 7.12 – Pre-attentive processing

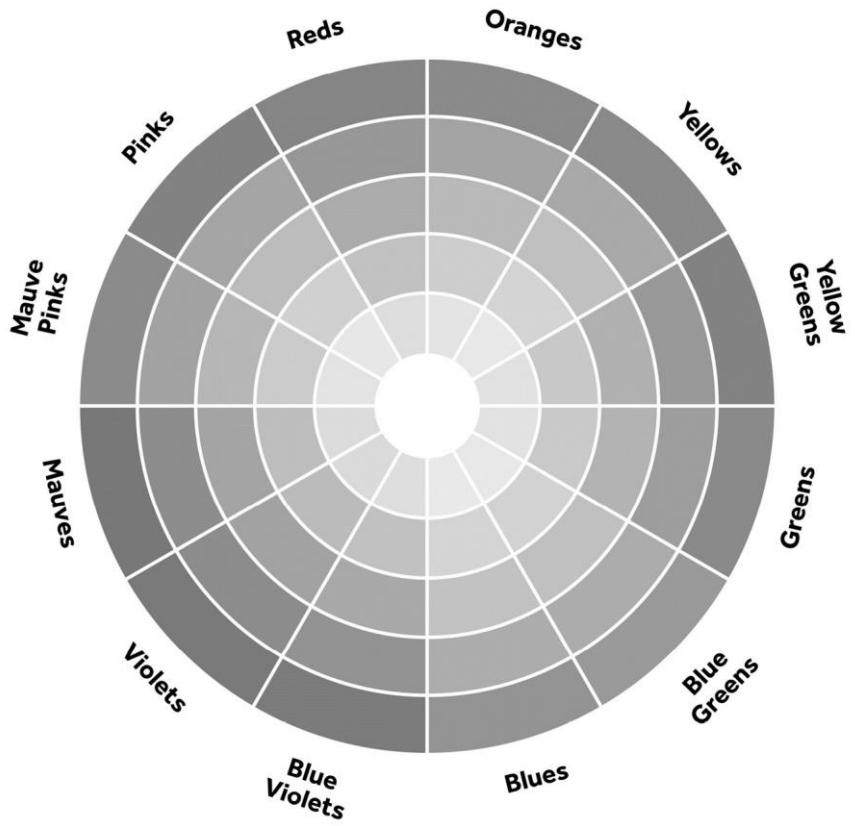


Figure 7.13 – Color wheel

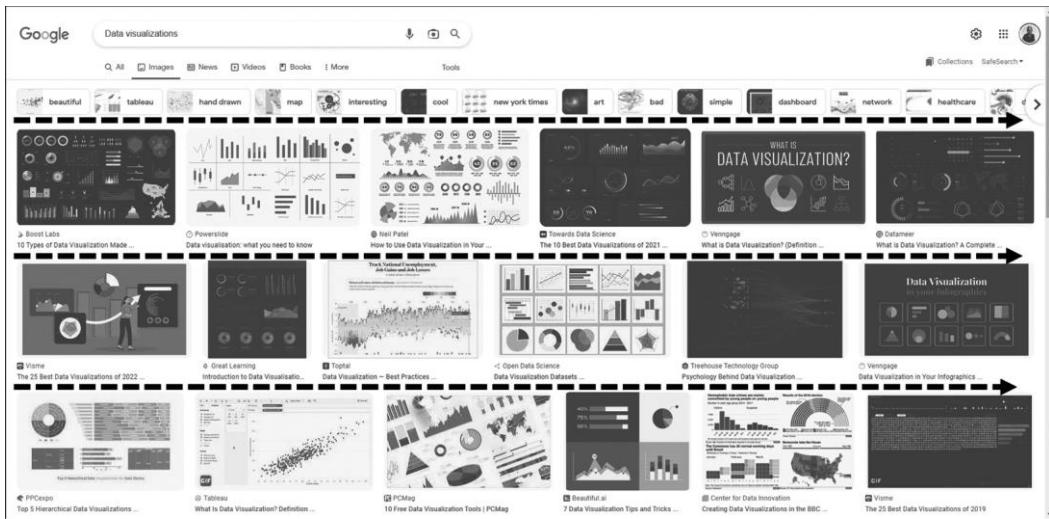


Figure 7.14 – When we read, we read from top left to right bottom

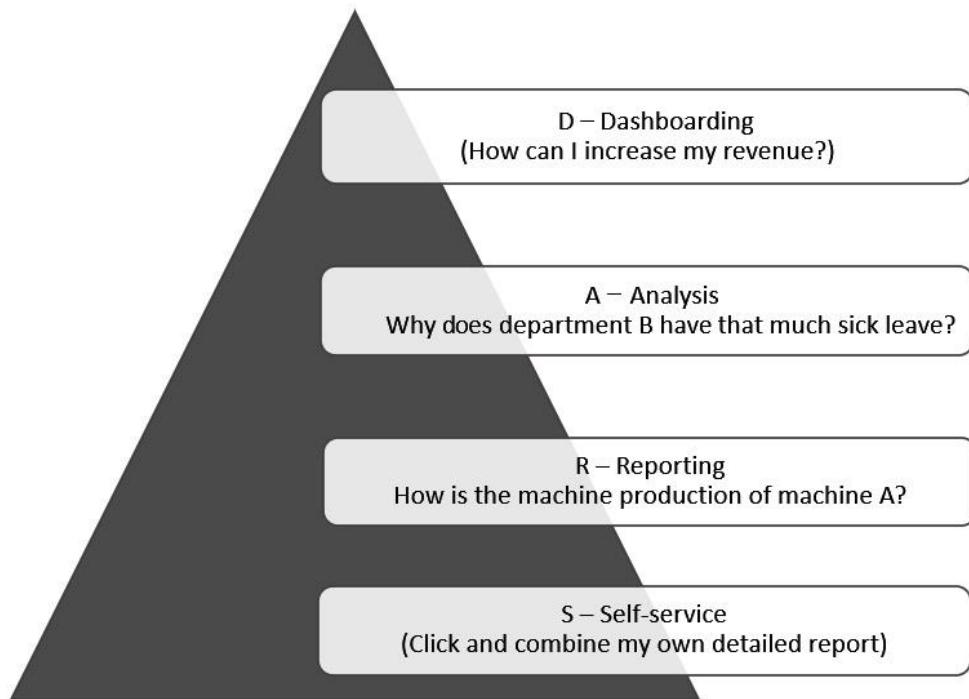


Figure 7.15 – DAR(S) principle

Who is my public	Farmers	Tourists	Explorers	Miners
Managing board	X			
Middle management	X			
Team manager		X		
Team lead				
Controller			X	
Process advisor			X	
Business Analysts			X	
Analysts			X	
Data Analysts			X	
Data Scientists				X

Figure 7.16 – Inmon groups to identify the level of interactivity and detail

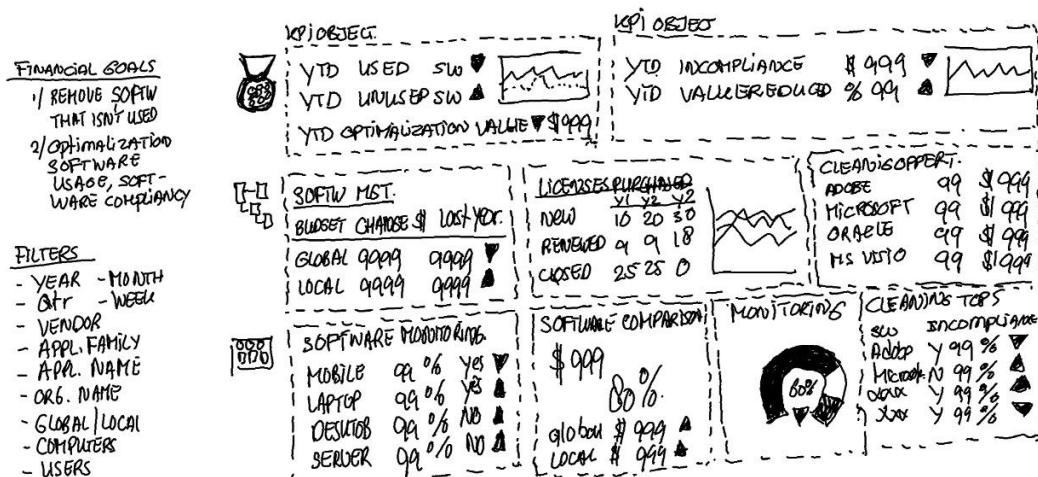


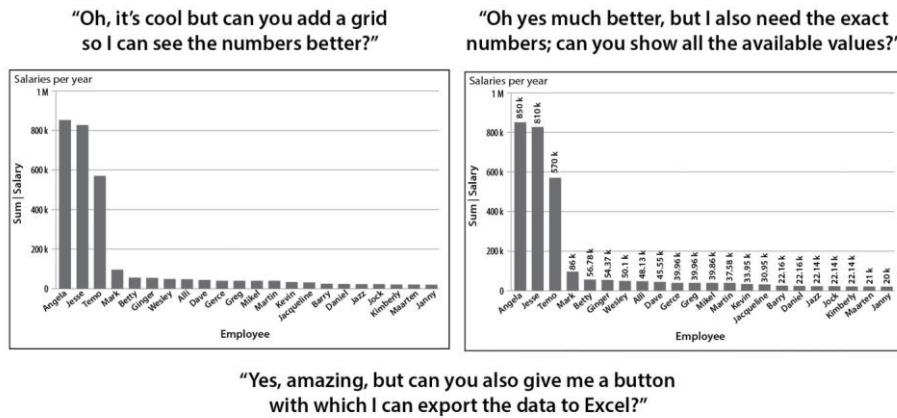
Figure 7.17 – Chatting and sketching



Figure 7.18 – Final result after a few chatting and sketching sessions



Figure 7.19 – The three phases of dashboard design



"Yes, amazing, but can you also give me a button with which I can export the data to Excel?"



Figure 7.20 – This is how it usually goes

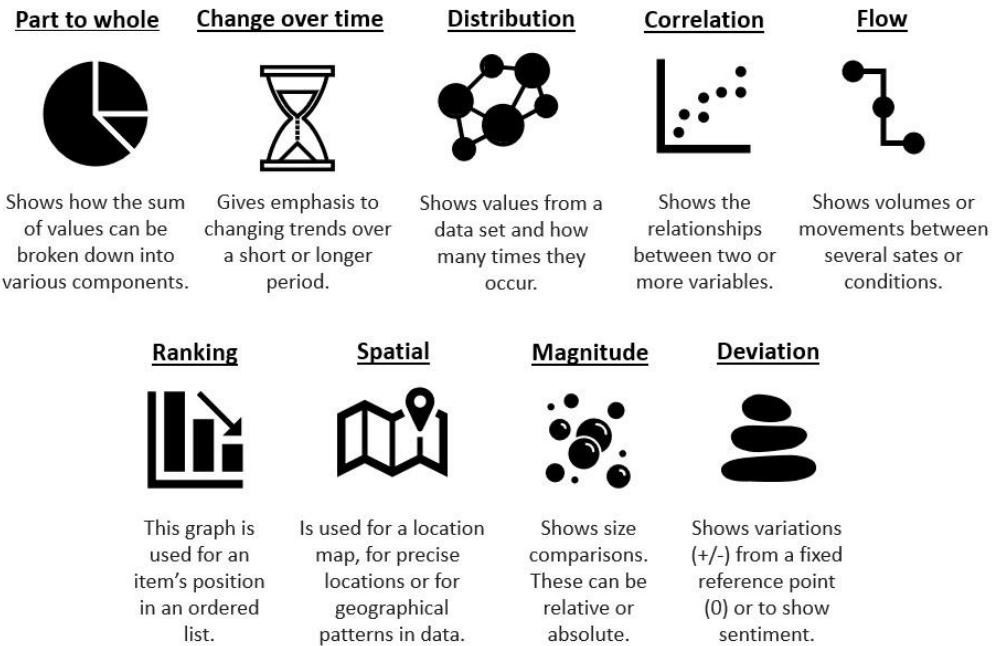


Figure 7.21 – The data visualization functional groups



Figure 7.22 – Two types of bar chart

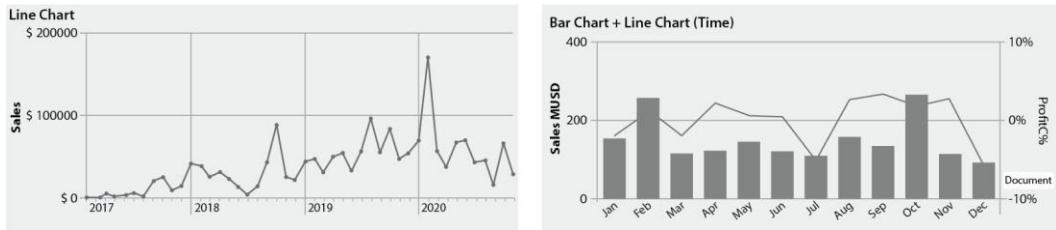


Figure 7.23 – Two types of line chart

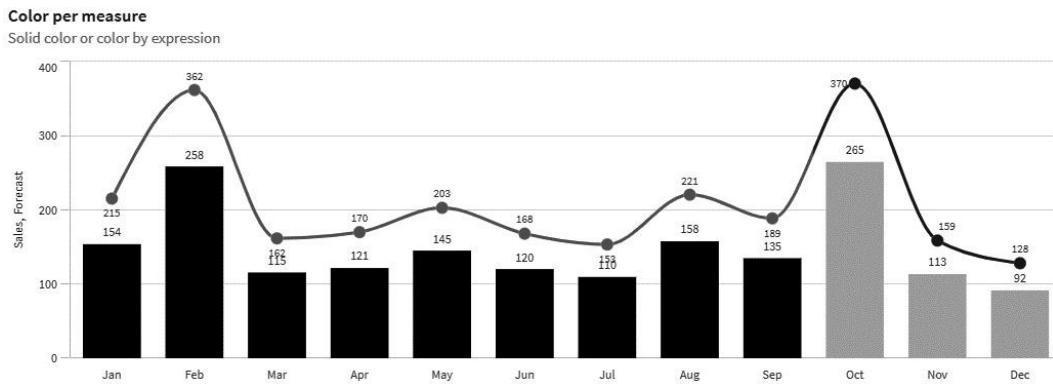


Figure 7.24 – Combination chart – colors per expression

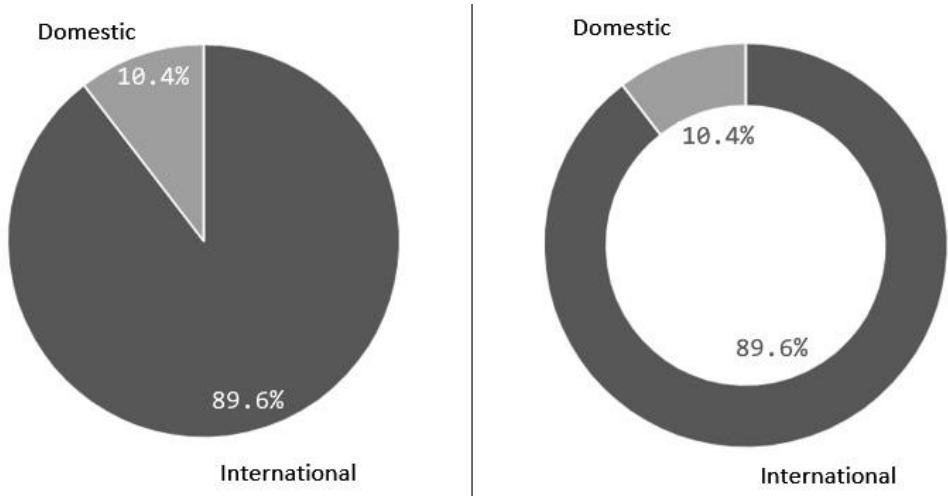


Figure 7.25 – Two examples of pie charts

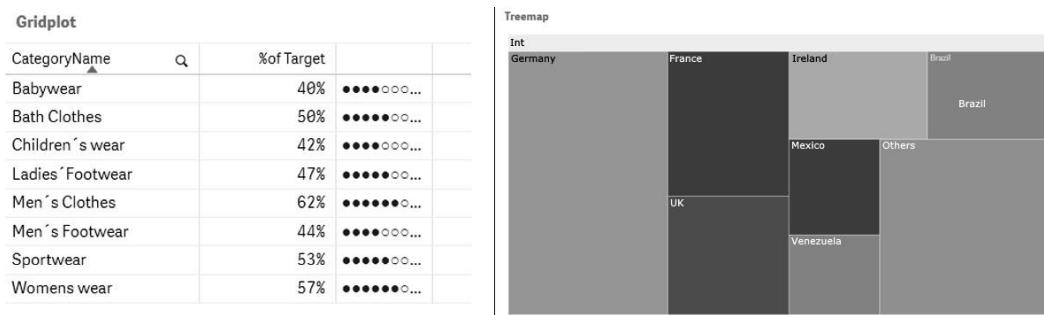


Figure 7.26 – An example of a grid plot and a tree map

	Q Q	Day Q	Sun	Mon	Tue	Wed	Thu	Fri	Sat	
			Q1	74	117	107	58	55	67	58
			Q2	54	63	56	37	33	55	70
			Q3	64	66	42	43	53	95	47
			Q4	53	39	116	50	74	80	59

Figure 7.27 – A calendar heatmap



Figure 7.28 – The first bed cleaning visualization

Avg Turnaround Time mins
72.06

Requests
81,785

Housekeepers
286

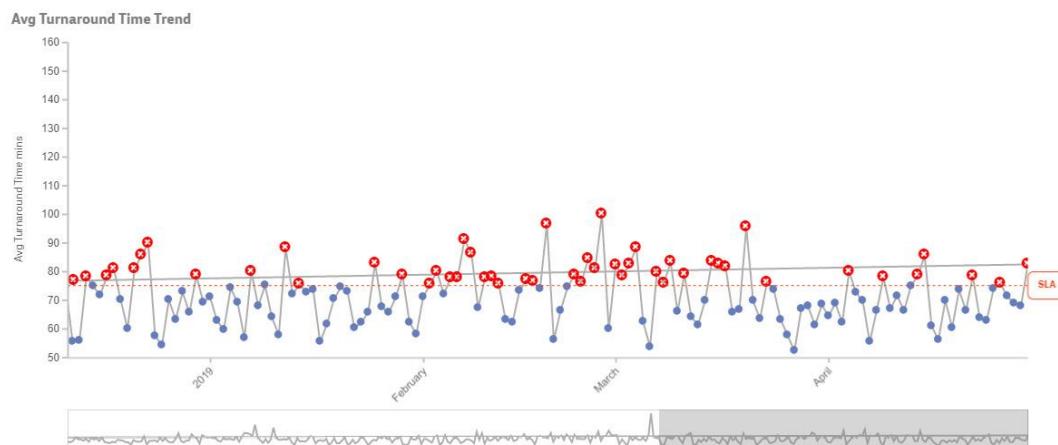
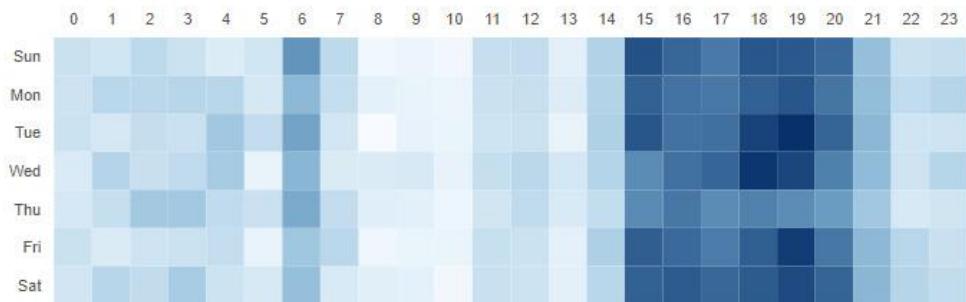
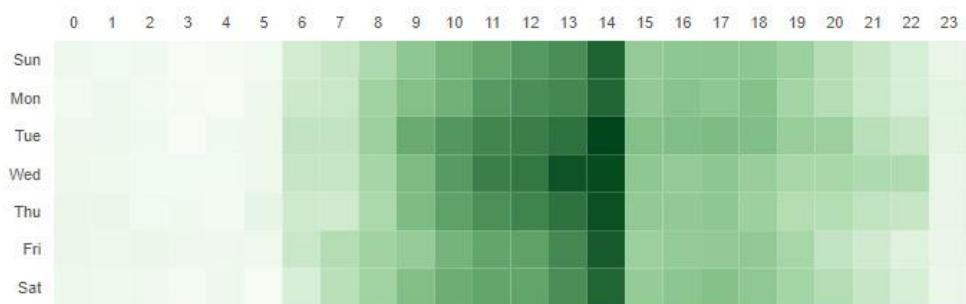


Figure 7.29 – Improving and adding a target line

Average Turnaround Time



Estimated Staffing Levels



Requests



Figure 7.30 – The usage of powerful heatmaps

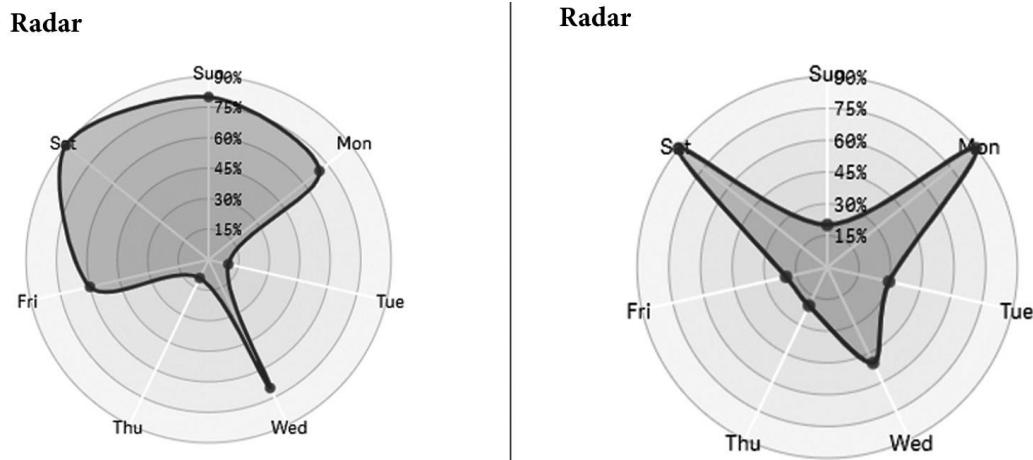


Figure 7.31 – Examples of radar charts

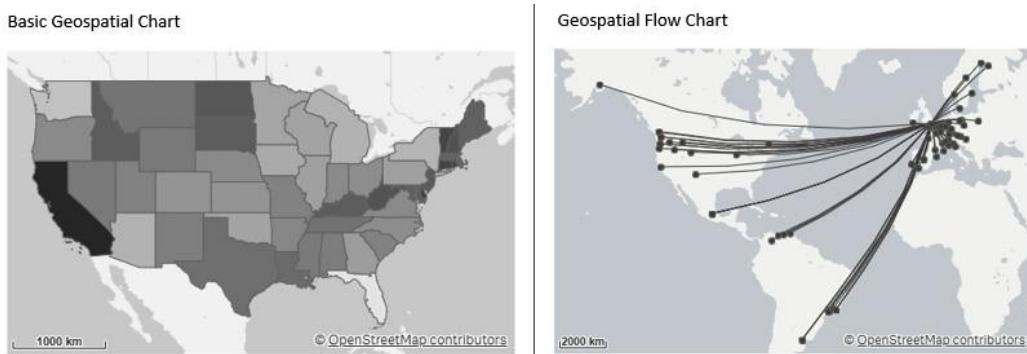


Figure 7.32 – Examples of geospatial charts

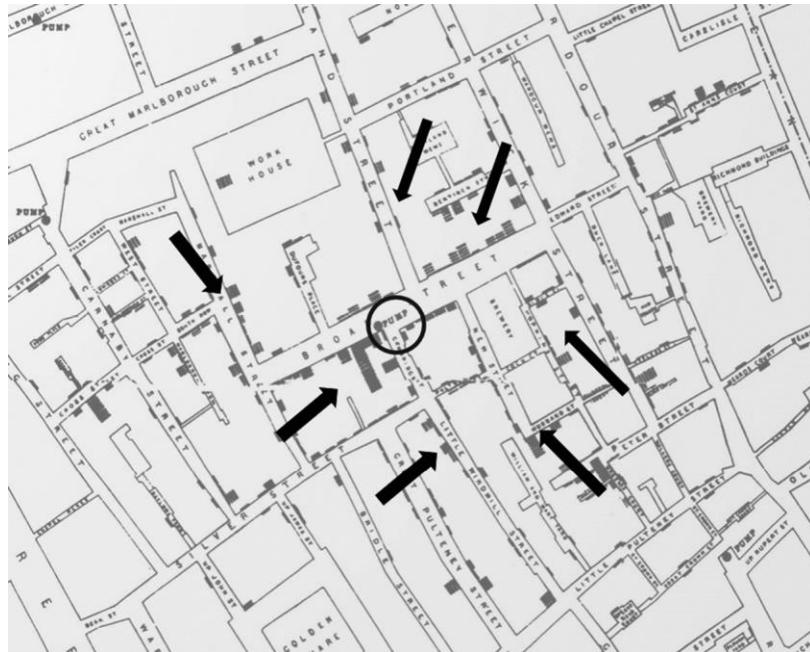


Figure 7.33 – The story of Dr. Snow

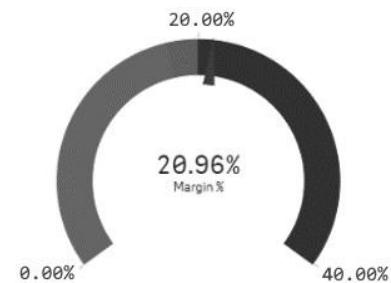


Figure 7.34 – Two types of KPIs

Year to Date Division Scorecard (thousands)									
Region	YTD Sales Trend	YTD Sales 2015 / Sales 2014	YTD Sales 2014	%	Fresh Food Sales 2015	YTD Fresh Food Sales 2014	%	Customer Satis 2015	YTD
Total		\$99,171	\$105,812	-6.3% ▼	\$83,991	\$80,947	3.8%	89.40%	
USA		\$22,156	\$21,785	1.7%	\$18,937	\$18,103	4.6%	89.86%	
Nordic		\$21,299	\$23,885	-10.8% ▼	\$17,922	\$17,119	4.7%	89.16%	
Japan		\$17,490	\$20,512	-14.7% ▼	\$18,124	\$17,619	2.9%	88.82%	
UK		\$17,217	\$18,154	-5.2% ▼	\$13,326	\$12,730	4.7%	89.52%	
Germany		\$10,712	\$10,721	-0.1% ▼	\$7,786	\$7,705	1.0%	90.08%	
Spain		\$10,297	\$10,756	-4.3% ▼	\$7,897	\$7,671	2.9%	89.13%	

Figure 7.35 – Table with mini-charts

Sales 2020

Grey scale represents bad, average & good.

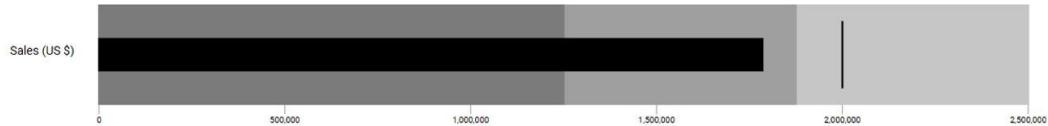
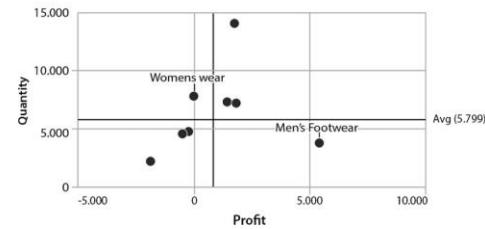


Figure 7.36 – An example of a bullet chart

Sale Performance 2020

The sales performance per product group



Sale Performance 2020

The sales performance per country

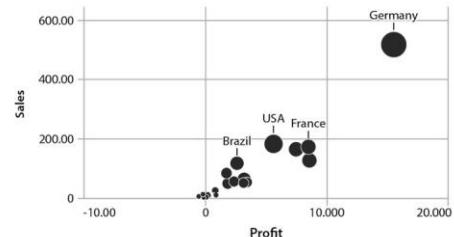


Figure 7.37 – Example of a scatter chart (left) and a bubble chart (right)

The Old Faithful Geyser – Yellowstone National Park

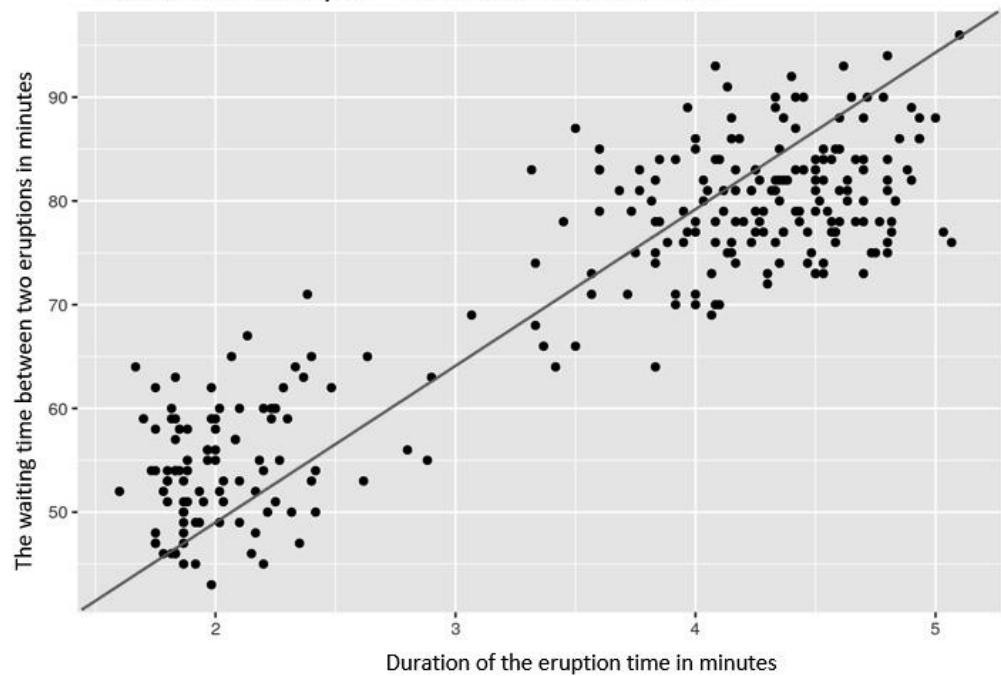


Figure 7.38 – A scatter plot of the Old Faithful geyser

Market share

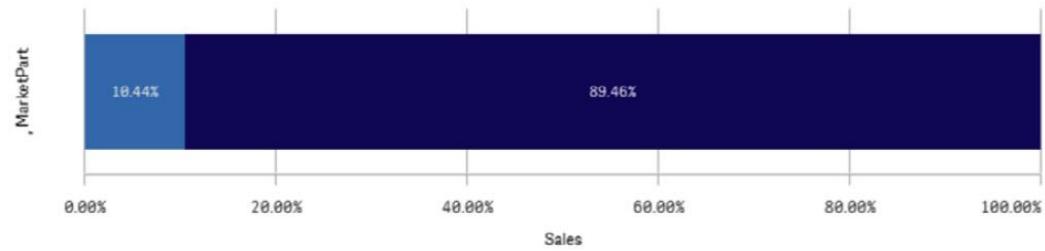


Figure 7.39 – An example of a stacked bar chart

Analysis of Extreme Times in period 201308 t/m 201706

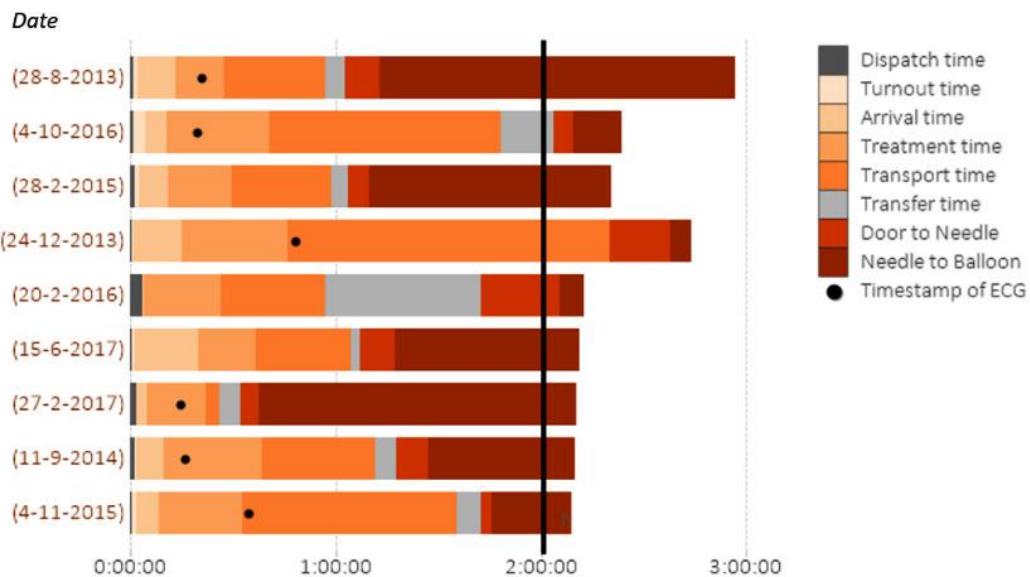


Figure 7.40 – Call to Balloon process breakdown with a stacked bar chart

Market share International versus Domestic

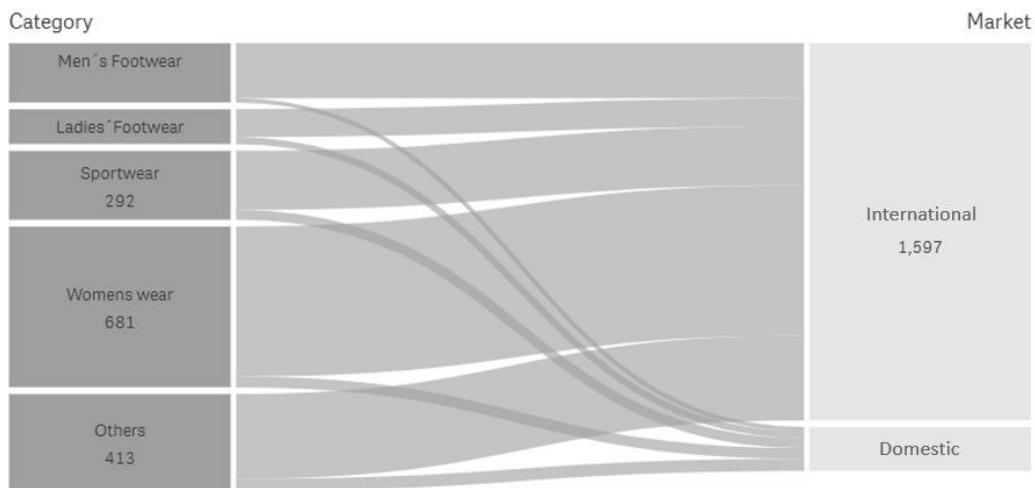


Figure 7.41 – The flow of consumer goods and the market share (domestic versus international)

Financial result 2021

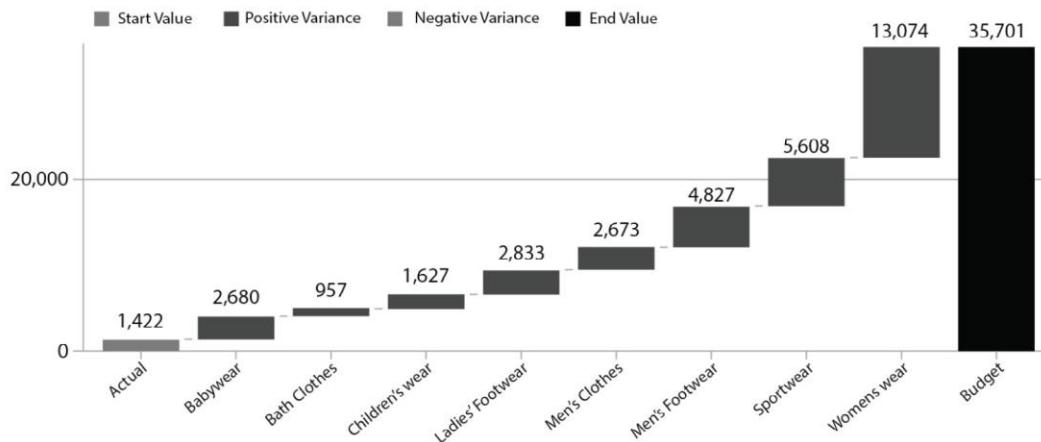


Figure 7.42 – An example of a waterfall chart

What are the number of Clinical Trials over time?

In the graph the number of COVID-19 cases are shown over time. Also the predictions (forecast) is given.

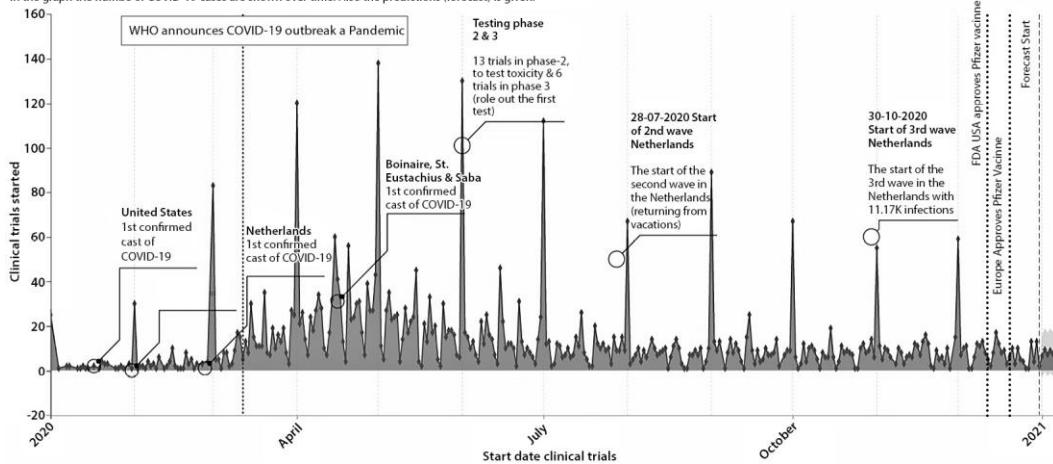


Figure 7.43 – Adding context to your charts

What is the Revenue per Main Category for the year 2020?

This graph shows the sales over the year 2020 for our productgroups in the USA market

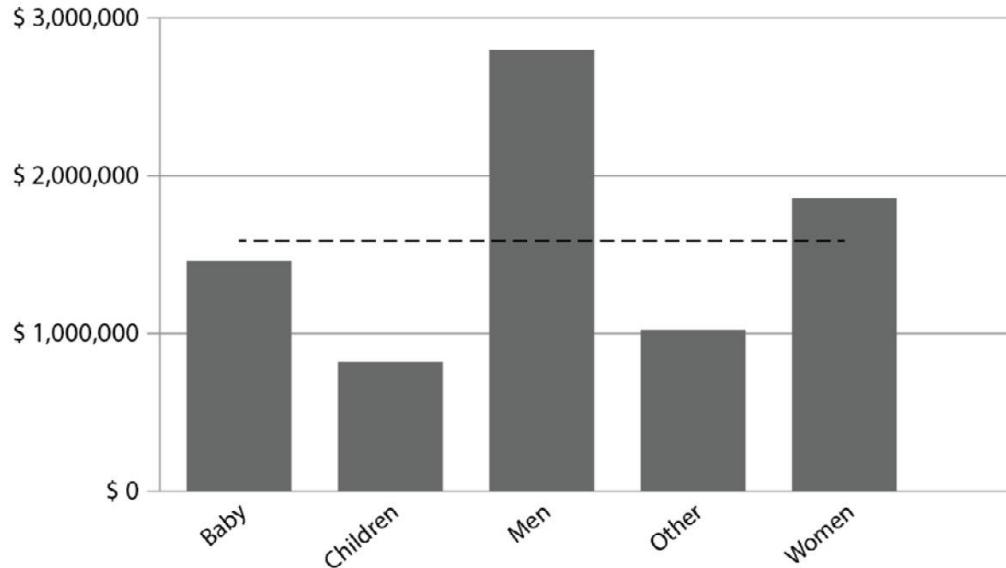


Figure 7.44 – An example of using a title, subtitle, and footer to give context

Further Reading

- Stephen Few inspired visual perception example in one of his presentations on his website, Perceptual Edge:
<http://www.perceptualedge.com/files/VisualAnalysisCourse.pdf>.
- You can find color wheel at www.colormatters.com. The following are some other fantastic websites to visit:
 - <https://color.adobe.com>
 - <https://coolcolors.co/>
 - <https://palettengenerator.com/>
- The *Financial Times* has built an amazing **visual vocabulary**, which can be found at <https://ft-interactive.github.io/visual-vocabulary/>.

- <https://infogram.com/examples/charts/bar-chart> provides some fascinating examples of when to utilize a bar chart.
- You can find the complete story at <https://infowetrust.com/project/breviary>.

Chapter 8

Personal questions:	Work-related questions:	Government/municipality questions:
<ul style="list-style-type: none">• Can I visit my elderly parents?• How can I protect myself and my loved-ones• Is it safe to travel by public transport?• Should I go to the office or do my shopping at the mall?• When will I be able to travel again abroad• Will I lose my job when my office must close?	<ul style="list-style-type: none">• How can I keep my employees safe?• How can I arrange a safe working environment for my staff?• Is it safe for my staff to travel by public transport?• Can we safely have clients at our office (or in our shops)?• Will my business survive when we must close?• How can we ask for support to our government?	<ul style="list-style-type: none">• How do we prevent people to get infected• How do we arrange the wellbeing of our inhabitants?• Who are our vulnerable elderly and how can we protect them?• What do we need to have testing facilities?• Should we close shops, schools and other facilities• What should we arrange to be able to vaccine our community?

Figure 8.1 – COVID-19-related questions

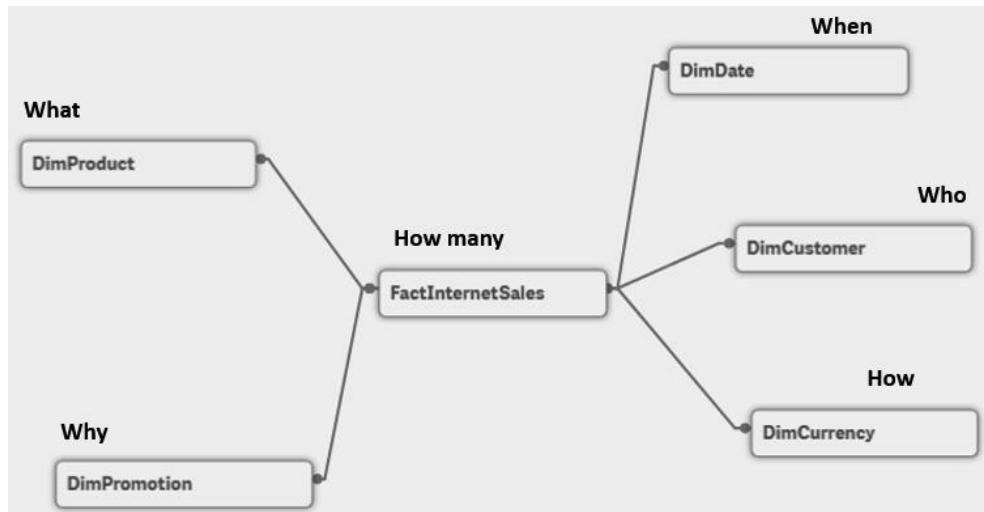


Figure 8.2 – Typical questions

Tuesday
March, 2nd
2021

Amsterdam News

Latest news & updates

x
x
x

Peter Woodward

Staggering increase in infections continues, the number of new corona patients in hospitals fluctuates!

The dramatic increase in infections continues, and the number of new corona patients in hospitals fluctuates, the number of new corona cases continues to rise dramatically now that the Omicron variant is most common.

Between Tuesday morning and Wednesday morning, 32,149 positive tests were reported. That is the highest number ever for a Wednesday. The number of new admissions of corona patients to the nursing wards has fluctuated in recent days. The experts say that most likely one of the reasons the numbers are rising this much is related to the increased testing capacity and the number of tests taken.



In the past seven days, the National Institute for Public Health was informed that the corona virus has been diagnosed 209,078 times. That equates to an average of 29,868 positive tests per day, the highest level ever. They even found in some areas a dramatic rise of 45%. This average is rising for the fifteenth day in a row during the COVID pandemic.

Figure 8.3 – News article and removing emotion

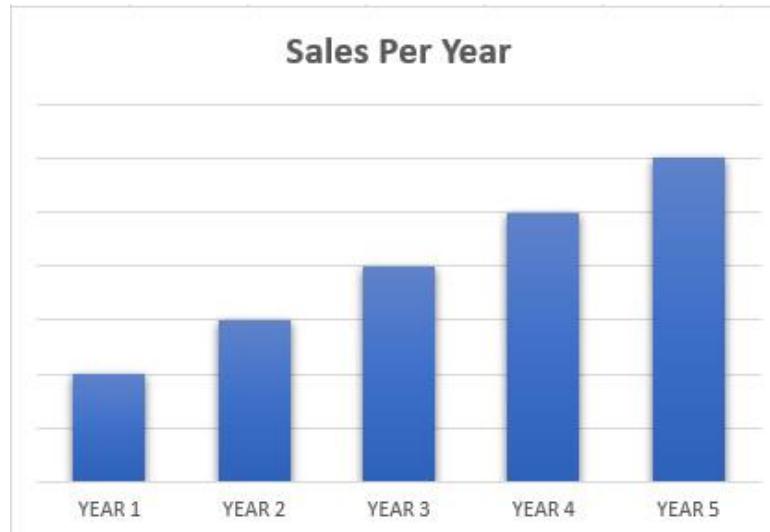


Figure 8.4 – Visualization without a y-axis scale

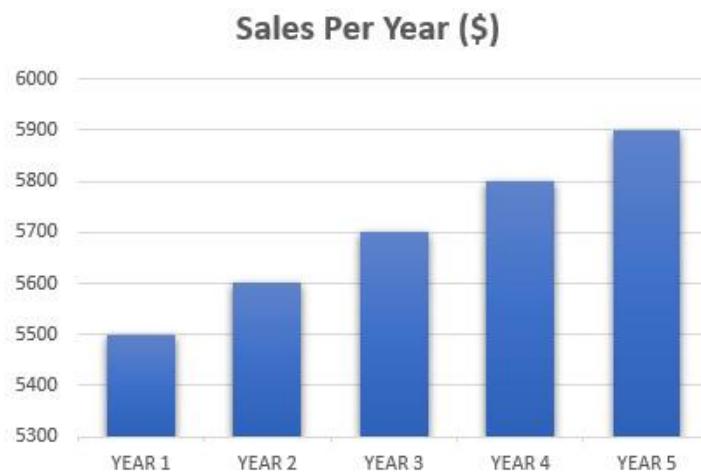


Figure 8.5 – Visualization with a scale not starting at 0

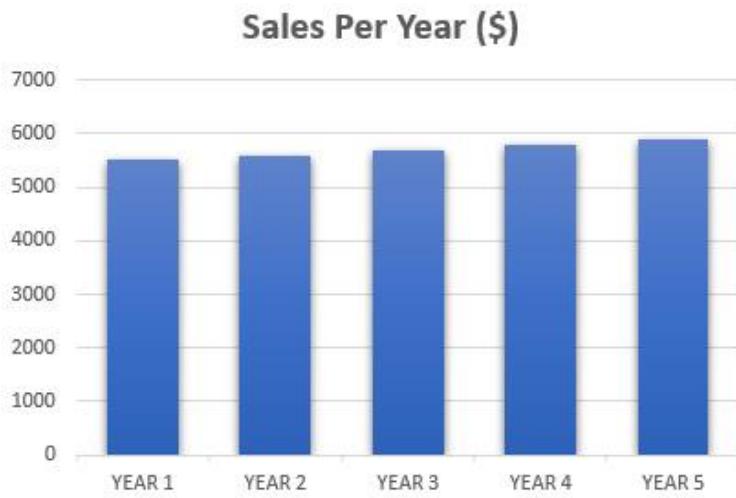


Figure 8.6 – Visualization with proper scaling

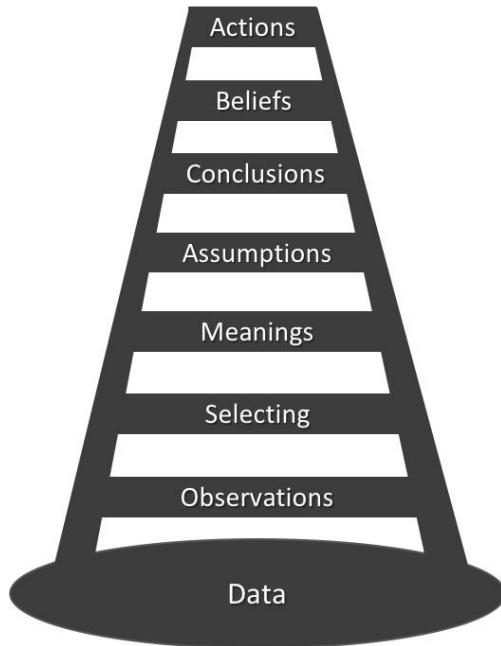


Figure 8.7 – Ladder of Inference

Further Reading

- *Wahl, D. (2016). Designing Regenerative Cultures. Triarchy Press.*
- *What's Your Problem?: To Solve Your Toughest Problems, Change the Problems You Solve. Thomas Wedell-Wedellsborg, Harvard Business Review Press, 2020.*
- *Shron, M. (2014). Thinking with Data. O'Reilly Media.*
- *Wald, Abraham (1943). A Method of Estimating Plane Vulnerability Based on Damage of Survivors. Statistical Research Group.*
- *Hanegan, K. (2020). Turning Data into Wisdom: How We Can Collaborate with Data to Change Ourselves, Our Organizations, and Even the World. Kevin Hanegan.*
- *Ross, Howard J. Everyday Bias: Identifying and Navigating Unconscious Judgments in Our Daily Lives. 2014.*
- *Senge, Peter M. The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization. Crown Publishing, 1994.*
- Maya Angelou interview. *O, the Oprah Magazine*. April 2011.
- *Scott, E. P. The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies. Princeton, NJ: Princeton University Press, 2008.*
- *Sloane, Paul. Lateral Thinking Puzzles, 1991.*

Chapter 9

Further Reading

- <https://www.vpnmentor.com/blog/data-privacy-security-stats/>
- <https://www.pieterjanssen.eu/2020/06/15/transitieproces-van-de-stasi-documenten-in-volle-gang/>
- <https://www.internetsociety.org/policybriefs/responsible-data-handling/>
- <https://www.vpnmentor.com/blog/data-privacy-security-stats/>
- <https://autoriteitpersoonsgegevens.nl/nl/nieuws/ap-beboet-transavia-om-slechte-beveiliging-persoonsgegevens>
- <https://www.vpnmentor.com/blog/data-privacy-security-stats/>
- <https://amp.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>
- <https://www.amnesty.nl/actueel/toeslagenaffaire-is-mensenrechtenschending-zegt-amnesty-international>
- <https://autoriteitpersoonsgegevens.nl/nl/nieuws/werkwijze-belastingdienst-strijd-met-de-wet-en-discriminerend>
- <https://nos.nl/artikel/2364658-nooit-meer-een-toeslagenaffaire-wat-moet-er-nu-gebeuren>

Figures

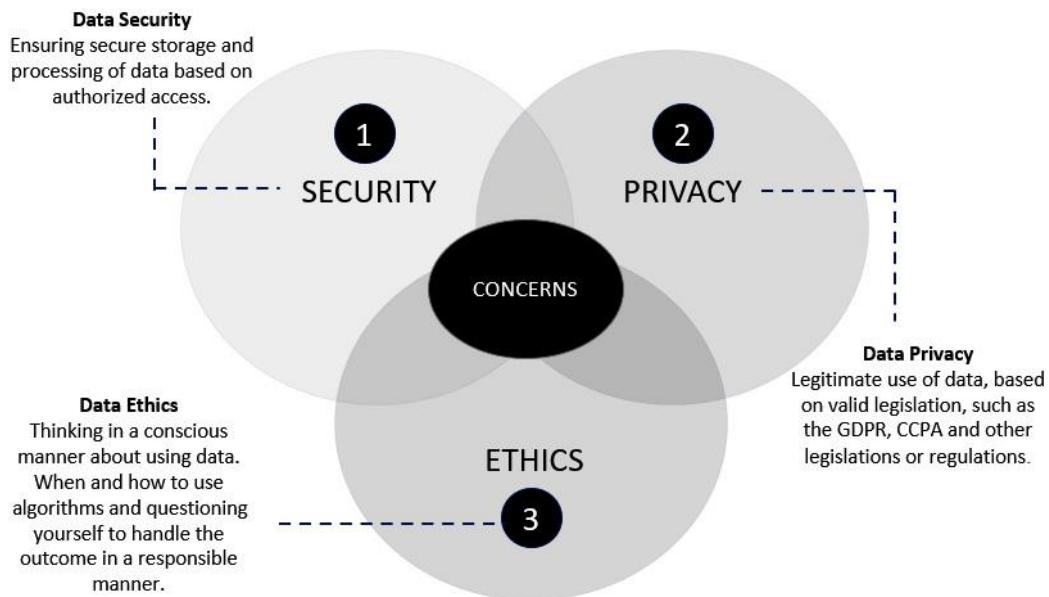


Figure 9.1 – Three important data concerns

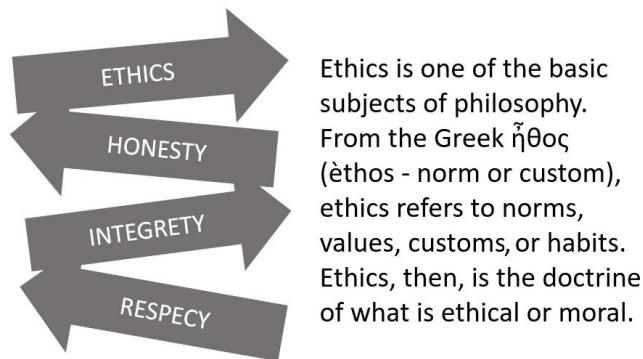


Figure 9.2 – What is ethics?

Checklists/Examples

Intake form example

An example of such an intake form with included checklists is available in *Appendix A*:

Data security check A data security incident is a serious security breach. When protected, sensitive, or confidential data is stolen, used, viewed, stored, or transmitted by someone who is not authorized, a breach occurs. Other security concerns include data leaks, information leaks (who can see what), sensitive information leaks, and data spillage. As a result, when it comes to data and information usage, we must identify some specific elements.	
Application/model/app Dashboards and reports are critical for any organization, team, or department. The work cannot be completed without this information.	<i>Essential Important Desirable Unclassified</i>
Confidentiality The dashboards and reports are private and only available to a select group of managers and employees.	<i>Essential Important Desirable Unclassified</i>
Access control What type of access control is required when confidentiality is required? What are the rules that we must follow? (For example, the manager of team A can see information about team A and so on.)	<i>Add information:</i>
Integrity Incorrect information will not be tolerated (accountable environment). The organization suffers significant damage if the	<i>Essential Important </i>

information is incorrect, incomplete, or late. Inaccurate data, such as financial transactions, undermines trust in the organization.	<i>Desirable / Unclassified</i>
Availability Dashboards and reports are essential. The organization will suffer severe consequences if the information is not available.	<i>Essential / Important / Desirable / Unclassified</i>
Additional information and/or explanation	

A checklist example can be found in the data privacy check table (the complete intake form and checklist are available in *Appendix A*):

Data privacy check When it comes to using personal data, we should take precautions or be aware from the start of our project that we must adhere to privacy regulations. We need to identify and categorize several categories. For example, in the Netherlands, we handle the following criteria: You may not just process personal data unless you have a legal basis. You will need the consent of the person in question. The processing serves a general, legitimate interest. Note here that simply moving, copying, or linking data is already a process of digesting or assimilating data.	
Name, Address, and City data (Name, Address, Postal Code, and City of Residence) Yes No	

Identification data (Passport, Driver's license, or Social Security number)	Yes No
Application details (Application letters, Resume)	Yes No
Contact details (E-mail address, Phone or Fax number)	Yes No
Salary details (Salary details, Social Payments, Income Taxes, Expense reimbursement)	Yes No
Social Media accounts (LinkedIn, Twitter, WhatsApp, Facebook)	Yes No
Image and sound recordings (Video, Photos, Passport photographs)	Yes No
Payment details (Bank name, Account Number, Name of Account holder)	Yes No
Addition and or explanation When a checkbox is answered with "yes," it is advised that you arrange a legislation and regulations check with your Privacy Officer.	

Legislation – Regulations check	
Privacy Impact Analysis (PIA) needed?	Essential Important Desirable Unclassified
Register of Data needed?	Essential Important Desirable Unclassified

Addition and/or explanation

Data ethics checklist

Data ethics checklist

Data generation, use, collection, analysis, and dissemination are all aspects of working with data. We can do this with both unstructured and structured data. When we do so, there is a chance that the decisions we make will have an impact on individuals and the world. As a result, we must be open about how we use data in our projects.

1	<p>Do we have to take care of laws and regulations in this data and analytics project?</p> <p>The first critical step is to determine whether any legalizations or regulations are applicable to the project.</p>	<p>Yes No Write extra/additional information here:</p>
2	<p>Is the data that we want to use available in an ethical manner and is that data suitable for usage?</p> <p>We must be mindful of who owns the data and ensure that it is used in the manner intended by the owner!</p>	<p>Yes No Write extra/additional information here:</p>
3	<p>Is it possible to identify and check bias in the data that we have collected or used for our models?</p> <p>People can be biased by their origin, and the same is true for data and the application of algorithms and business rules. The data we collect and store is not as objective as we believe! When using algorithms and business rules, we should be aware that the data we use to train the models</p>	<p>Yes No Write extra/additional information here:</p>

	<p>can have an impact on people and possible human bias magnifies, which results in undesirable outcomes.</p> <p>To conclude, we must be able to identify, test, verify, and discuss the results.</p>	
4	<p>Can we identify and demonstrate bias in our created model or in the used data?</p> <p>When we use data and apply various models, we may have used data that is biased. When we use that specific dataset and apply the learning models, the model produces biased results; for example, a bias based on gender, age, equality, or racial elements.</p> <p>We must be aware that we must consider documenting, discussing, and evaluating our data usage choices. The message here is to avoid doing things simply because you can!</p>	<p>Yes No</p> <p>Write extra/additional information here:</p>
5	<p>Can the legal rights of individuals be impinged by the use of data?</p> <p>When individuals' legal rights are at stake, the organization must have permission and the right to use data for specific purposes. For example, suppose an organization provides data to its direct partners, but privacy conditions other than the internal data usage may be addressed here. In the event of an incident, for example, they know address details and more detailed information about people, but certain data is not shared to protect those people's privacy (such as names, address details, and other things by which someone can be individually identified).</p> <p>It could also happen internally, for example, through logging information that is known at the employee level but is shared with users at the department or concern level.</p> <p>We should be aware of this in order to protect the privacy of each individual employee.</p>	<p>Yes No</p> <p>Write extra/additional information here:</p>

6	<p>Are we able to understand that the data that we want to use is suitable for the purposes of our project?</p> <p>When we begin an analytics or data science project, we must understand and ensure that the data we intend to use is appropriate for the purpose of our projects. Following that, we should be able to verify and validate the data for our project.</p> <p>For example, when records or values are missing, the outcome of our algorithms and business rules can have a significant impact on the results, potentially producing a biased result.</p>	<p>Yes No Write extra/additional information here:</p>
7	<p>Do we have a multi-disciplinary team present to discuss the apparent dilemmas regarding the possible usage of algorithms and the possible outcomes?</p> <p>When it comes to assessing and discussing our own work, we need to discuss the dilemmas and outcomes that can occur with a multi-disciplinary team.</p>	<p>Yes No Write extra/additional information here:</p>
8	<p>Explainable AI by Design. Are we able to define the role of the algorithm used and what processes are being followed (procedural transparency)?</p> <p>It is critical for data engineers and data scientists who train models to understand the model's behavior in order to detect errors or weaknesses. This is why we must correctly describe the used algorithms or business rules.</p> <p>When data scientists and data engineers train a model, it is critical to understand the model's behavior. They must identify any flaws or errors.</p>	<p>Yes No Write extra/additional information here:</p>
9	<p>Explainable AI. Are we able to explain the algorithm or business rules to the guardians, stakeholders, and others with whom they are concerned?</p>	<p>Yes No</p>

	Explainable AI is defined as the ability of a person to comprehend the reason for a decision. The decision is influenced by algorithms and business rules. To be able to understand the model's decision, we must be able to explain the decision. We can do so by design, but we can also do it post hoc by using an algorithm to understand the black-box model.	Write extra/additional information here:
	Addition and/or explanation	

Chapter 10

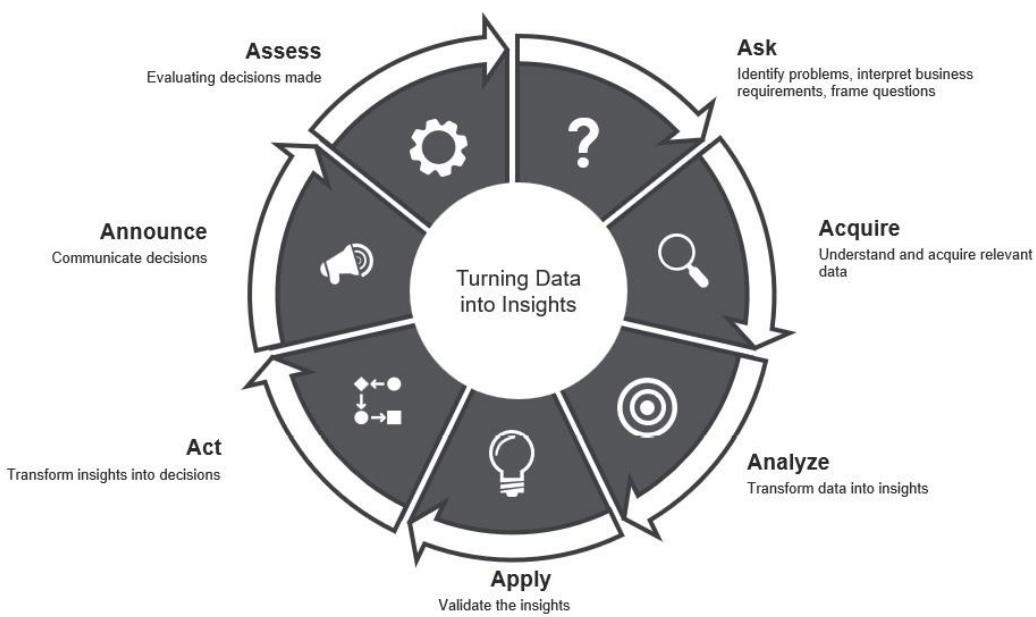
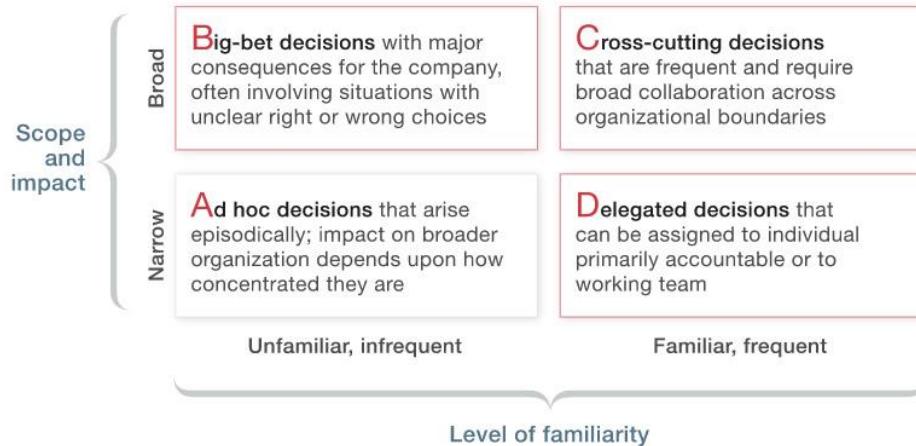


Figure 10.1 – Data-informed decision-making process

Strategic	Tactical	Operational
Increase organization's revenue by 15%	Selecting vendors as part of the organization's supply chain	Employee's schedule
Change organization's business model	Number of goods stored in inventory	Return policy

Figure 10.2 – Decision importance levels

The ABCDs of categorizing decisions.



McKinsey&Company

Figure 10.3 – McKinsey & Company's decision categories



Figure 10.4 – Identifying the right questions to ask

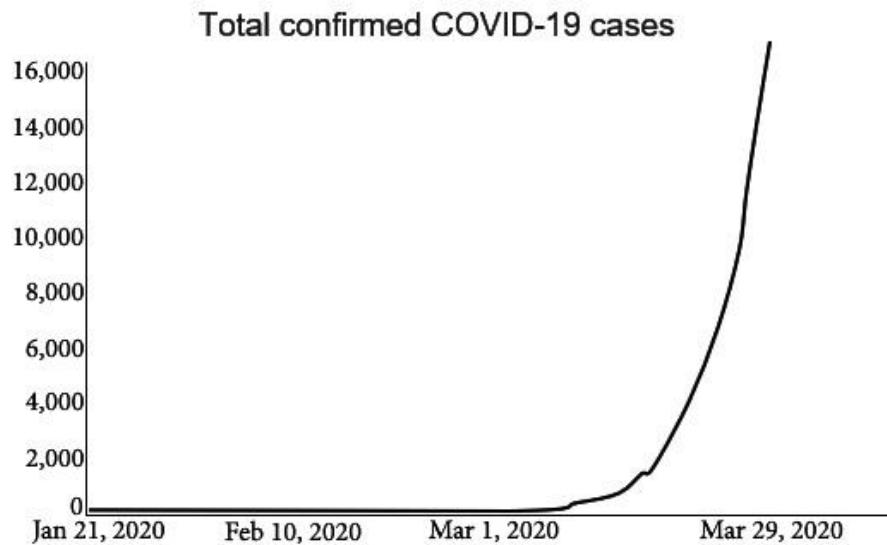


Figure 10.5 – Confirmed COVID-19 cases

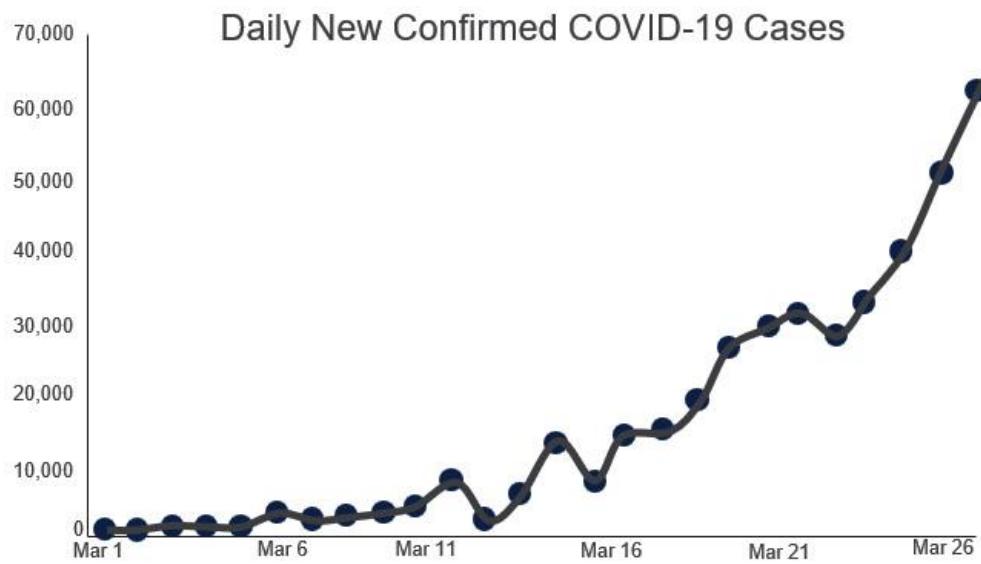


Figure 10.6 – Daily new confirmed cases

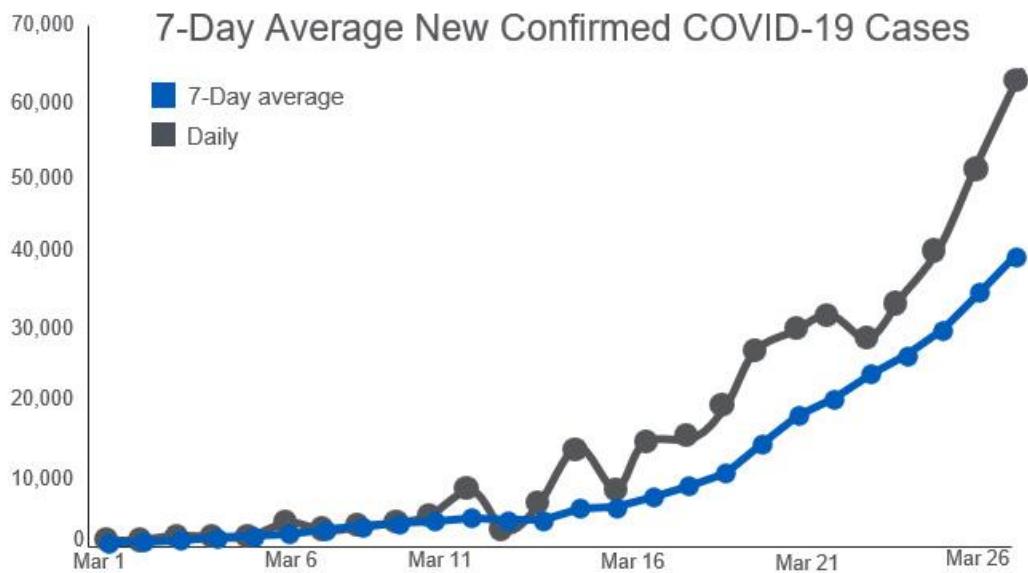


Figure 10.7 – 7-day average new confirmed COVID-19 cases

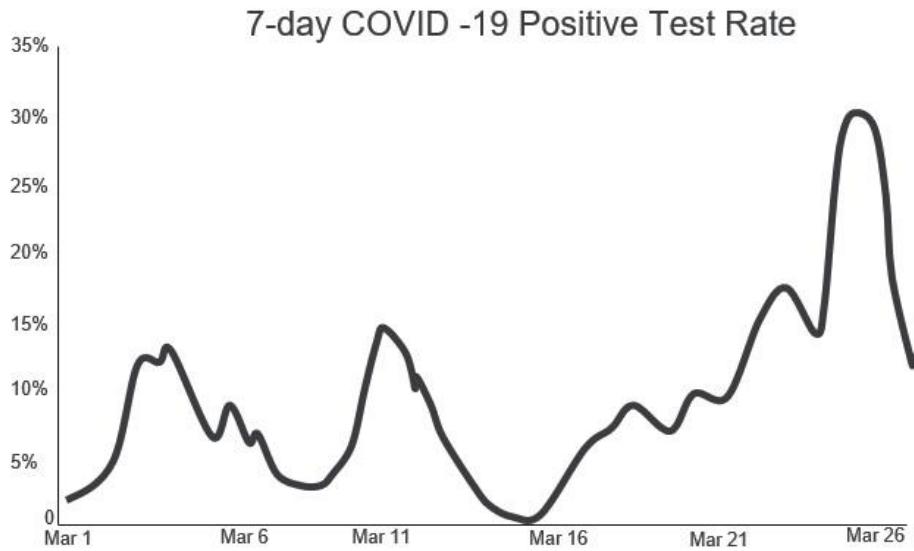


Figure 10.8 – 7-day COVID-19 positive test rate

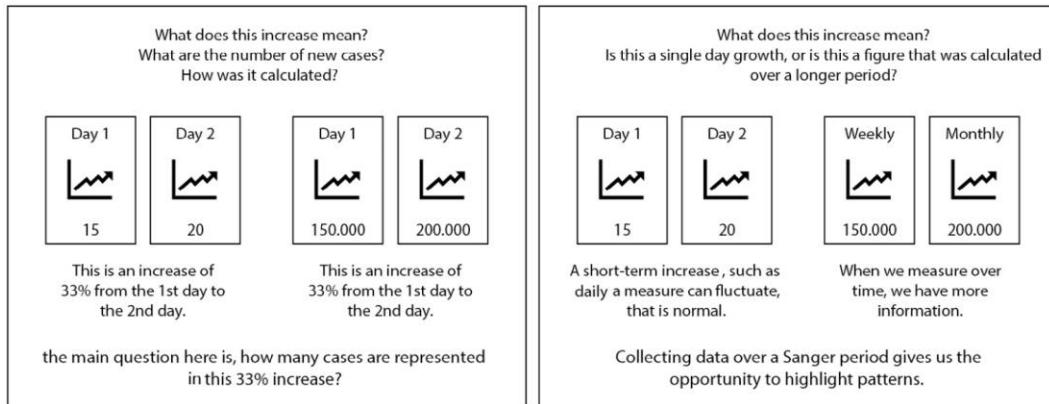


Figure 10.9 – Interpreting the data behind the visualizations



Figure 10.10 – Questions about the rate of transmission

Indicator	Personal Risk	Prevalence	Activity-specific Transmission	Health System Support
Values	Low, Medium, High	Low, Medium, High, Unknown	Low, Medium, High, Unknown	Low, Medium, High, Unknown
Required Data	Age Gender # Co-morbidities Number of people in immediate household	Positivity rate # of tests	Size of activity Indoor / Outdoor Preventative measures (Mask mandate, Social distancing)	% Beds not Occupied Quality of Care (Meds, Oxygen, Staff)

Figure 10.11 – Indicators for assessing a family vacation during COVID

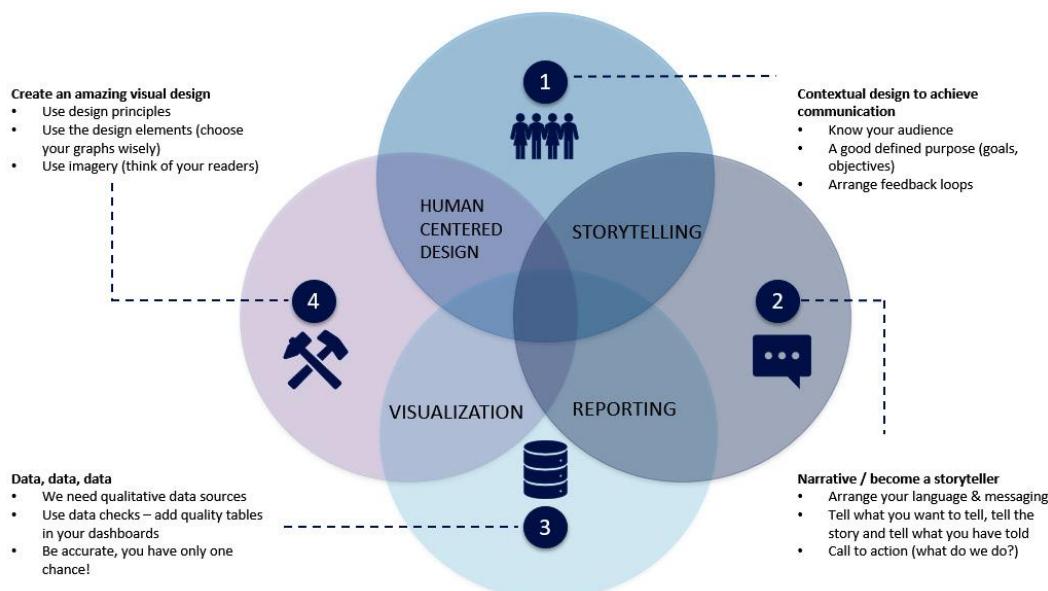


Figure 10.12 – Approach to analytic storytelling

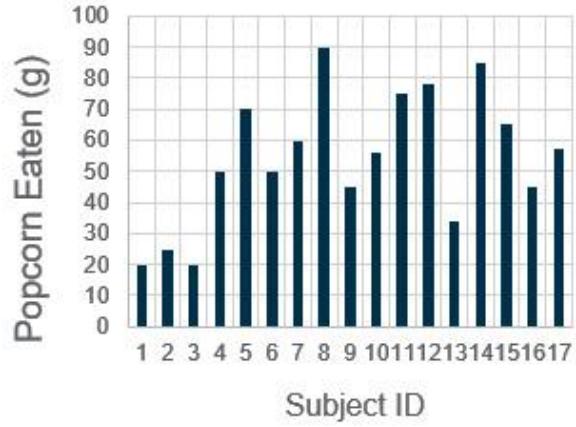


Figure 10.13 – Popcorn eaten by subject

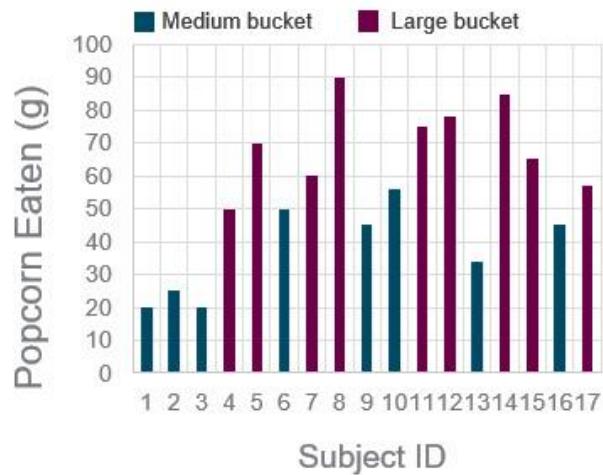


Figure 10.14 – Popcorn eaten by subject and by bucket size



People with larger buckets ate 53 percent more popcorn than people with the medium size

Figure 10.15 – Does bucket size matter?

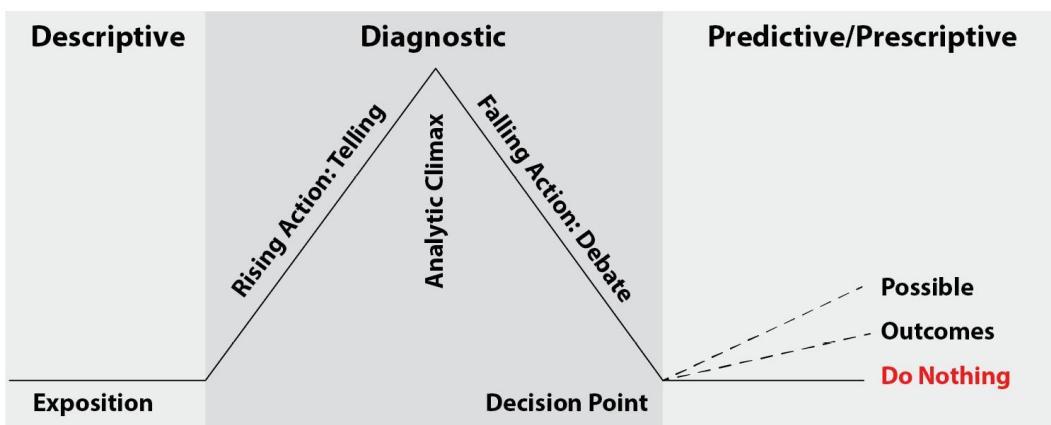


Figure 10.16 – Gartner’s approach to storytelling

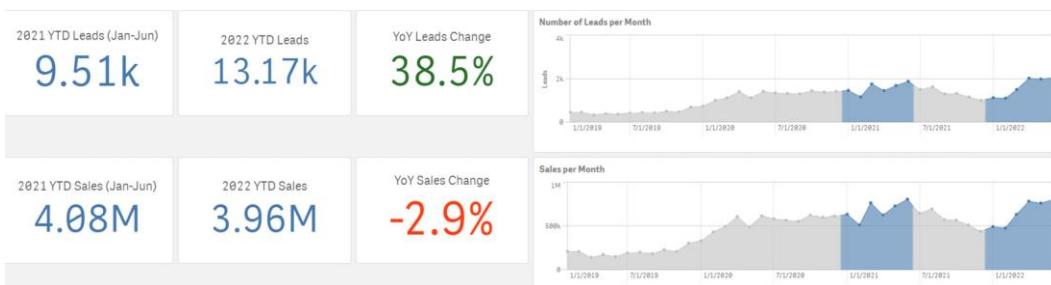


Figure 10.17 – Coffee equipment leads and sales



Figure 10.18 – Coffee equipment sales discount

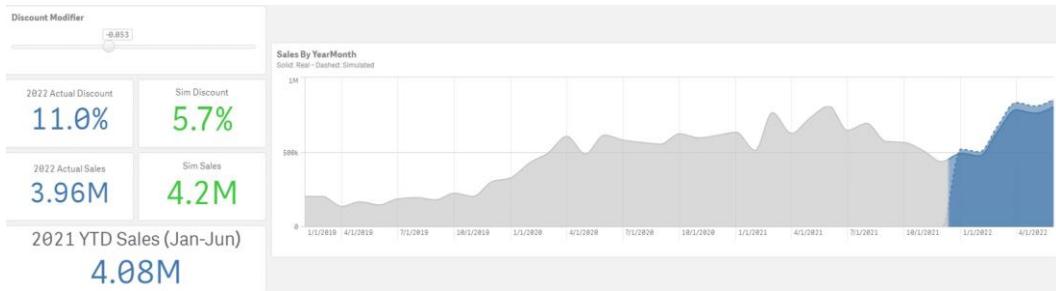


Figure 10.19 – Coffee equipment simulated sales

Further reading

- Snowden, D. and Boone, M., 2007. *A Leader's Framework for Decision Making*. [online] Harvard Business Review. Available at: <https://hbr.org/2007/11/a-leader-s-framework-for-decision-making>
- Heath, C., & Heath, D. (2009). Made to stick: *Why some ideas take hold and others come unstuck*. Random House Books

Chapter 11

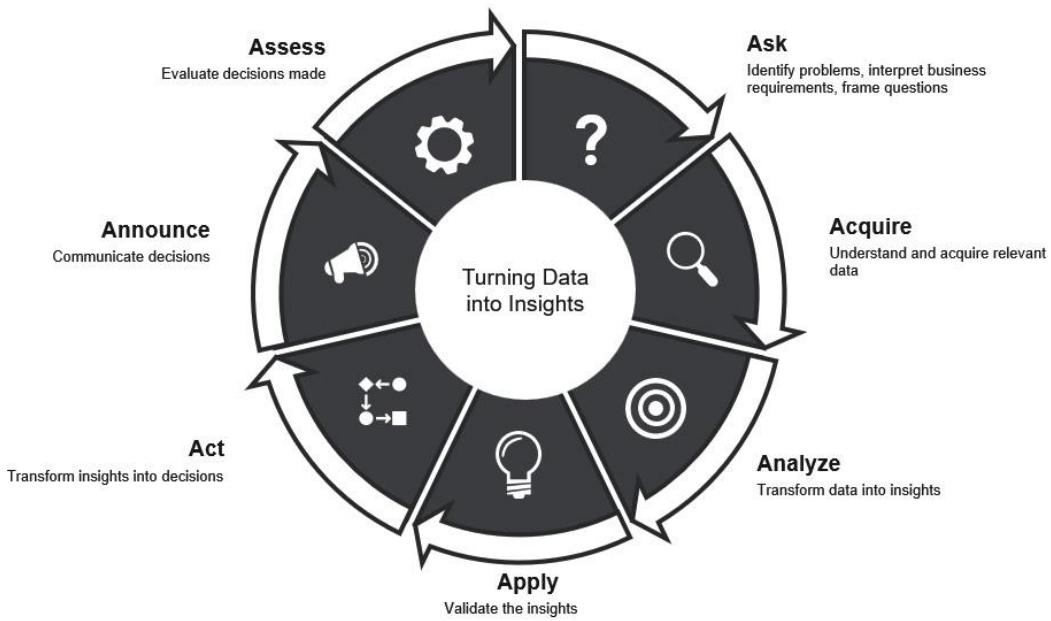


Figure 11.1 – Phases for turning data into insights

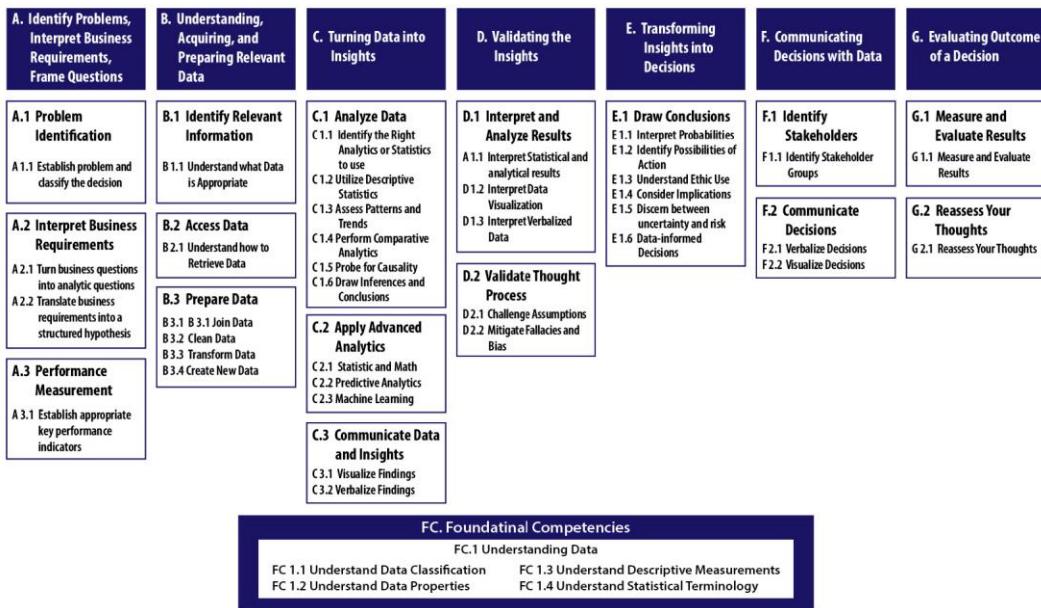


Figure 11.2 – Data literacy competency framework

	Understanding, Acquiring, and Preparing Relevant Data	Transforming Data into Insights	Transforming Insights into Decisions	Communicating Decisions with Data
Data Extraction				
Data Prep				
Data Analytics				
Data Visualization				
AI/ML				

Data Science				
Data Communication				
Interpreting Statistical and Analytical Results				
UX Design				
Understanding and Using Data Effectively				

Figure 11.3 – Data literacy technical skills



Figure 11.4 – Data literacy soft skills

	Identifying Problems and Interpreting Requirements	Understanding, Acquiring, and Preparing Relevant Data	Transforming Data into Insights	Transforming Insights into Decisions	Communicating Decisions with Data	Evaluating Outcomes of Decisions
Questioning						
Observation						
Analysis						
Numeracy						
Interpretation						
Reasoning						
Inference						
Evaluation						
Explanation						
Reflection						

Figure 11.5 – Critical thinking components

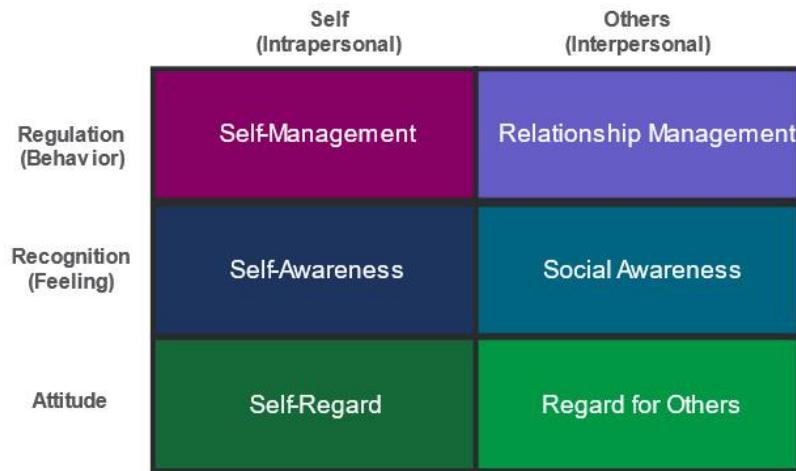


Figure 11.6 – Emotional intelligence components

Self-Awareness	Self-Management	Self-Regard	Social Awareness	Relationship Management	Regard for Others
Openness	Impulse Control	Self-Confidence	Empathy	Accountability	Flexibility
Self-Knowledge	Resilience	Self-Respect	Awareness	Conflict Management	Interdependence
Integrity	Stress Tolerance	Optimism	Anticipation	Influence	Trustworthiness
Introspection	Authenticity	Motivation	Mindfulness	Emotional Stability	Collaboration

Figure 11.7 – Emotional intelligence traits and skills

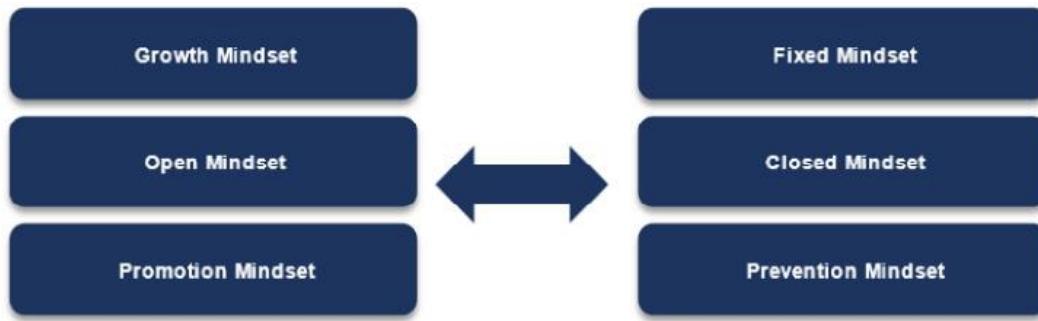


Figure 11.8 – Data literacy mindsets

Further Reading

- Armand Ruiz, *The 80/20 Data Science Dilemma*, (InfoWorld, September 26, 2017).
- W. Edwards Deming, *Out of the Crisis* (Cambridge, Massachusetts: The MIT Press, 2018).
- Scott E. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton, NJ: Princeton University Press, 2008).

Chapter 12 – Assessing Your Data Literacy Maturity

Assessing individual data literacy

As per our belief, data literacy is not one-size-fits-all. Not only is it specific to the role that you play with data, but there are also different levels. Therefore, a single data literacy assessment that is used by everyone, regardless of their level or how they interact with data, is not practical or useful. Assessing someone's skills around preparing data is not relevant to a person whose role with data involves only reviewing insights generated by others.

However, it is very useful for people new to data literacy to have a simple way to assess their foundational data literacy skills and mindsets, regardless of what role they play with data. In this chapter, we will provide you with such an assessment. In the end, we will provide you with a link to a more robust assessment you can complete online if you want to go deeper.

This 22-question self-assessment, shown in *Figure 12.1*, can be completed in about 15 minutes. For each statement listed, you should score it between 1 and 5 using the following descriptions:

1. Strongly disagree with the statement (1 point)
2. Slightly disagree with the statement (2 points)
3. Neither agree nor disagree with the statement (3 points)
4. Slightly agree with the statement (4 points)
5. Strongly agree with the statement (5 points)

Understand Data

This section assesses how well you understand raw data, including its type, classification, and any properties, as well as how well you understand the simple

statistical terminology of various descriptive statistics, as discussed in *Chapter 2, Unfolding Your Data Journey*.

- | | |
|---|--|
| 1 | Can identify whether data is qualitative or quantitative. |
| 2 | Can identify what level of measurement (nominal, ordinal, interval, ratio) a given piece of data has. |
| 3 | Can identify whether data is discrete or continuous. |
| 4 | Can understand simple statistical terminology and interpret the results of simple descriptive stats such as distributions, frequencies, ranks, percentiles, means, medians, modes, ranges, standard deviation, and interquartile ranges. |
| 5 | Can apply means, medians, and modes and how to choose the right one based on the data and the question. |
| 6 | Can identify data security, data privacy, and data ethics concerns related to the data being used. |

Interpret Business Requirements

This section assesses how well you can understand the question being asked, identify relevant data, and how you can create proper **key performance indicators (KPIs)**, as discussed in *Chapter 6, Aligning with Organizational Goals, and Chapter 10, Turning Insights into Decisions*.

- | | |
|---|---|
| 7 | Can transform a business problem into a quantifiable form as an analytics question. |
| 8 | Can create KPIs that adhere to the characteristics of a KPI. |
| 9 | Can identify whether data is relevant or not relevant to the decision to be made. |

Create and Interpret Data Visualizations

This section assesses how well you can both create and interpret data visualizations. Given this is a basic foundational data literacy assessment, the focus is only on basic visualizations, as discussed in *Chapter 8, Questioning the Data*.

- | | |
|----|--|
| 10 | Can select and create a basic visualization. |
| 11 | Can apply appropriate design components of visualizations, including the use of color, annotations, and preattentive attributes. |
| 12 | Can understand what information a basic visualization is meant to convey. |
| 13 | Can identify and describe simple patterns and trends in a visualization. |

Apply the Human Element of Data Literacy

<p>This section assesses how well you understand cognitive bias and its impact on your perspective of data and information, as well as how well you can identify and apply techniques to mitigate the impact of bias, as discussed in <i>Chapter 7</i>.</p>	
14	Can describe cognitive biases and how they can impact interpreting data and insights.
15	Can apply techniques to help mitigate cognitive bias.
Communicate with Data	
<p>This section assesses how well you can communicate insights and decisions using data, as discussed in <i>Chapter 10</i>.</p>	
16	Can determine the appropriate method and mechanism for communicating the decision.
Mindsets	
<p>This section assesses whether you have the right mindset to work with data.</p>	
17	Willingness to ask additional and clarifying questions.
18	Openness to new insights, even if they contradict your previous convictions.
19	Willingness to question your contextual knowledge and its influence on the interpretation of the data and insights.

20	Openness to the fact that your interpretation of the data and insights may be influenced by cognitive bias.
21	Willingness to question the significance of the results for the data or insights presented.
22	Objectivity in representing data, including no conscious manipulation to misrepresent the information to consumers.
TOTAL (Add up all the scores from above)	

Figure 12.1 – Foundational data literacy assessment

You can use the table shown in *Figure 12.2* to assess your foundational data literacy based on your total score:

Total Score	Foundational Data Literacy
0-44	You need to work on multiple skills and mindsets to increase your foundational data literacy.
45-69	You have some foundational data literacy skills, but still have a way to go to become foundationally data literate.

70-94	You have a majority of the foundational data literacy skills, but there is still room for growth in a few areas.
95-110	You have optimized all the foundational data literate skills. Congratulations! You can now look at the next level and grow from foundational skills to more intermediate and mastery levels.

Figure 12.2 – Foundational data literacy assessment scoring

If you are looking at areas to improve your data literacy, start by focusing on the statements and categories where you answered below a 4 or 5. Ideally, you should strive to have a score of either 4 or 5 for all of these 25 statements.

If you are interested in either a more comprehensive individual assessment that covers all roles, or if you are looking at going beyond foundational data literacy to more intermediate and advanced levels, then this is available at
<http://www.kevinhaneGAN.com/dataLiteracyAssessment>.

This assessment leverages the following seven phases of how data is interacted with, from raw data to the decision being made with the data:

- Identifying problems, interpreting business requirements, and framing questions
- Understanding, acquiring, and preparing the relevant data
- Turning data into insights
- Validating the insights
- Transforming insights into decisions
- Communicating decisions
- Evaluating the outcomes of decisions

In this comprehensive assessment, you must select which parts of the data journey relate to your role. Then, you will be asked to complete the assessment for those related competencies out of the total 16 competencies and 37 sub-competencies. You will get results that show you your data literacy level for those competencies, with recommendations on areas you can work on.

Assessing organizational data literacy

There are 2 options for anyone looking to perform an organizational data literacy assessment. We have a simple 16-question assessment that you can use to get an understanding of where your organization is in terms data literacy, and we have a more robust assessment organized into 5 dimensions and 25 subdimensions. This chapter will introduce you to both assessments.

Basic organizational data literacy assessment

This 16-question assessment, shown in *Figure 12.3*, can be completed in about 10 to 15 minutes. For each statement listed, you should score it between 1 and 5 using the following descriptions:

6. Strongly disagree with the statement (1 point)
7. Slightly disagree with the statement (2 points)
8. Neither agree nor disagree with the statement (3 points)
9. Slightly agree with the statement (4 points)
10. Strongly agree with the statement (5 points)

Organizational Culture and Strategy

This section assesses how well-suited the organizational strategy is to data literacy, including how well it supports a culture of data literacy, as discussed in *Chapter 4, Implementing Organizational Data Literacy* and *Chapter 6, Aligning with Organizational Goals*.

1	An executive sponsor exists within the organization for data literacy.
2	The majority of the leaders within the organization promote data literacy in their processes, including the ability for employees to openly question and challenge data.
3	Your organization has defined and communicated its data and analytics vision and ambitions.
4	Leverages KPIs for strategic and departmental goals.

Organizational Processes

This section assesses organizational processes that relate to how they use and support data, as discussed in Chapters 4, 7, 9, and 10.

- | | |
|---|---|
| 5 | Decisions that are communicated include data, visualizations, and insights. |
| 6 | Decision-making processes that use both data and a human element are present within the organization. |
| 7 | The organization has processes and resources in place to support data security and data privacy. |
| 8 | The organization has processes and resources in place to monitor and improve data quality. |
| 9 | The organization has processes and education in place to foster the ethical handling of data. |

Data and Analytics

This section assesses how the organization uses and provides access to data and analytics, as discussed in Chapters 2, 4, and 5.

- | | |
|----|---|
| 10 | Data is accessible to the majority of the organization to use in a self-service manner. |
| 11 | The organization has an established data strategy, including data management and data governance. |
| 12 | The organization leverages more than just descriptive analytics (including diagnostic, predictive, and prescriptive). |
| 13 | Data and analytics are embedded into the daily operations of the organization. |

Tools and Technology

This section assesses the organization's portfolio of tools and technology to support data literacy.

- | | |
|----|---|
| 14 | The organization invests in tools and technology that support the data journey and makes it easier for decision-makers to leverage data. |
| 15 | The organization invests in technical integrations that bring the data and analytics to the right users within their work environment with minimal manual intervention (this includes mashups and embedded dashboards and analytics, alerts and notifications, and tools to integrate the data with the analytics). |

Knowledge and Learning Programs

<p>This session assesses the organization's approach to how it educates, upskills, and supports its employees on their data literacy journey, as discussed in <i>Chapter 4, Implementing Organizational Data Literacy</i>.</p>	
16	The organization provides support in the form of training, coaching, or community discussions so that employees can upskill on data literacy, as well as the time for employees to take advantage of the support.
TOTAL (Add up all the scores from above)	

Figure 12.3 – Basic organizational data literacy assessment

You can use the table shown in *Figure 12.4* to assess your organization's data literacy based on your total score:

Total Score	Organizational Data Literacy
16-32	Immature. Your organization does not have a lot of focus on data literacy.
33-44	Developing. Your organization has some areas that support organizational data literacy but it still has a long way to go to maximize the value data literacy can provide to an organization.
45-69	Managed. Your organization sees the value in data literacy and has a focus on some key areas of organizational data literacy, but there are a few additional focus areas that can make a difference.
70-80	Mature. Your organization ranks as a fully mature data-literate organization that has a focus on just about all of the key areas of organizational data literacy.

Figure 12.4 – Basic organizational data literacy assessment scoring

If you are looking at areas to improve your organizational data literacy, start by focusing on the categories or statements where you answered below a 4 or 5. Ideally, you should strive to have a score of either 4 or 5 in all of these 10 statements.

If you are looking for a more robust organizational data literacy assessment that can highlight more specific strengths and weaknesses, along with recommendations, you can use the assessment covered in the next section.

Robust organizational data literacy maturity assessment

Our more robust organizational data literacy assessment will assess your organization's maturity with data literacy initiatives in 5 dimensions and 25 total subdimensions.

The list of subdimensions can be seen in *Figure 12.5*:

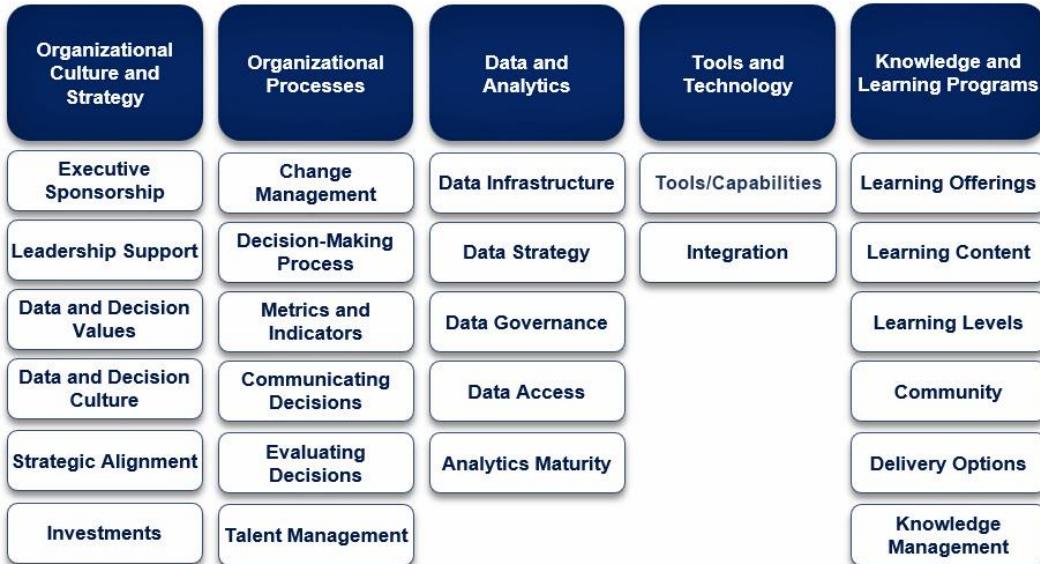


Figure 12.5 – Organizational data literacy dimensions and subdimensions

Ideally, the assessment would be filled out by someone in the organization with a good overview of the organization's culture and strategy, processes, data and analytics, metrics, skills, technology, data-driven culture, leadership, strategy, execution (of processes), analytical integration, analytical empowerment of employees, and education programs.

It is highly likely, especially in large organizations, that a single person may not know enough of all the dimensions to complete the assessment themselves. In these situations, it is OK to split the assessment up so that multiple resources are involved in filling it out. For example, maybe the chief data officer fills out the sections on data and analytics, maybe the chief information officer fills out the sections on tools and technology, and maybe a leader in education fills out the sections on knowledge and learning programs. Another way to split up the assessment would be to have a single individual or team responsible for the entire assessment, but rather than answering the questions themselves, they conduct interviews and perform audits and gather information to properly complete the assessment.

Let's move on to the actual assessment, which is organized by dimensions. The assessment is provided in this chapter but can also be taken online at <http://www.kevinhaneigan.com/maturityAssessment>.

Organizational culture and strategy

There are six subdimensions within the organizational culture and strategy dimension:

Executive Sponsorship

An executive sponsor is critical to an organization's data and literacy maturity. Three of their main tasks are active and visible participation, communicating support and the importance of data literacy, and building a coalition. Select the number that best represents the status of the executive sponsorship of data literacy in your organization:

- No executive sponsor exists. (1 point)
- The executive sponsor is only evident through funding authorization and resource allocation. (2 points)
- There is an executive sponsor who provides active and visible participation. (3 points)
- There is an executive sponsor who provides active and visible participation, builds a coalition, and communicates support to promote data literacy. (4 points)

Score	
--------------	--

Leadership Support

Beyond the executive sponsor, there needs to be buy-in and support from other leaders, including actively using data to make decisions. Select the number that best represents the status of the leadership support of data literacy in your organization:

- Managers have some awareness of the uses of data, but they don't see the value of it, and they don't believe data literacy is important for everyone. (1 point)

- Many managers know data and data literacy is important, so they are interested in it, and they even use data from time to time, but they are not convinced and do not practice the principles required (that is, they do not allocate time for their team to upskill on data literacy). (2 points)
- Most managers are engaged with data and ask the right questions about the data to harness the value in it. (3 points)
- Most managers understand how data can improve their organization and actively use it to drive questions and influence their decisions. Leaders communicate using data and let the data inform and drive their own decisions. Leaders hold teams accountable for decisions they make if they are using just opinions and not data. (4 points)

Score	
--------------	--

Organizational Values

Organizational values guide your business decisions. Three organizational values specific to data literacy organizations are as follows:

- Data-informed decision-making is valuable.
- Democratization of data across the organization is important.
- Sharing knowledge about data is recommended and seen as valuable.

Select the number that best represents the status of the organizational values of data literacy in your organization:

- None of the characteristics listed exist. (1 point)
- One of the characteristics listed exists. (2 points)
- Two of the characteristics listed exist. (3 points)
- All three of the characteristics listed exist. (4 points)

Score	
--------------	--

Data and Decision Culture

Culture is the way values are implemented in businesses. The following characteristics apply to the data and decision culture for a data-literate organization:

A test and learning environment.

- Hierarchies do not matter when it comes to decision-making.
- Leaders should not have the ability to impose an instinct-based veto (not using data).
- Data is openly questioned and challenged across the organization.
- Users can ask questions and get answers without having to ask a data expert or similar.

Select the number that best represents the status of the data and decision culture at your organization.

- None of the characteristics listed exist. (1 point)
- One of the characteristics listed exists. (2 points)
- Two of the characteristics listed exist. (3 points)
- All of the characteristics listed exist. (4 points)

Score	
--------------	--

Strategic Alignment

The following characteristics are true for data-literate organizations in terms of the strategic alignment of data literacy:

- A vision exists for where your organization wants to end up on its data literacy journey.
- Specific data literacy objectives have been created.
- A link has been made between data and staying ahead of the competition.
- Data is seen as an organizational asset.

- Shows a clear ROI on data investments.

Select the number that best represents the status of the data and decision culture in your organization:

- None of the characteristics listed exist. (1 point)
- One of the characteristics listed exists. (2 points)
- Two of the characteristics listed exist. (3 points)
- All of the characteristics listed exist. (4 points)

Score	
--------------	--

Investment

Data-literate organizations don't just talk about the importance of data and analytics and data literacy – they action it with investments to support those initiatives. Select the number that best represents the investments for data literacy in your organization:

- Most of the investment in data and analytics is budgeted on a project-by-project basis. (1 point)
- Most of the investment in data and analytics is budgeted within each department. (2 points)
- Most of the investments in data and analytics come from a centralized budget in IT or similar. (3 points)
- Most of the investments in data and analytics come from a centralized budget specific to data and analytics. (4 points)

Score	
--------------	--

Assessment score for organizational culture and strategy

Add up the total number of points for your responses across all six subdimensions and calculate the average score for this dimension, as shown in *Figure 12.6*. If you decided to

get multiple people to complete the same sections of the assessment, calculate the average score:

Dimension: Organizational Culture and Strategy		
1	Executive Sponsorship	1-4
2	Leadership Support	1-4
3	Data and Decision Values	1-4
4	Data and Decision Culture	1-4
5	Strategic Alignment	1-4
6	Investments	1-4
Total Dimension Score (Add all scores and enter total)		
Average total (Divide the "Total Dimension Score" by 6)		

Figure 12.6 – Organizational culture and strategy scoring

Organizational processes

There are six subdimensions within the organizational processes dimension:

Change Management	
As we mentioned earlier in this book, the biggest obstacles and adoption barriers that organizations face when they try to become data-literate are managerial and cultural. Both of these require strong change management. Select the number that best represents the change management in your organization:	
<ul style="list-style-type: none">• Little or no change management practice exists. (1 point)• Some elements of change management are applied to some projects. (2 points)• Comprehensive change management exists, but only for some projects. (3 points)• Change management is seen as an organizational competency with not only an organization-wide standard but also a dedicated team and is used at all levels of the organization. (4 points)	
Score	

Decision-Making Process

An organization is only truly data literate if it can use data to influence and inform decisions. This requires a decision-making process. Select the number that best represents the decision-making process in your organization:

- Very little structure, if any at all, in a decision-making process. (1 point)
- Some structure in the decision-making process but mostly in terms of terminology, not the actual process. (2 points)
- There's a structured decision-making process in some departments but not all. (3 points)
- There's a structured decision-making process that's used everywhere in the organization and tied to KPIs, performance data, and resource allocation. (4 points)

Score	
--------------	--

Metrics and Indicators

Data-literate organizations leverage data everywhere, including running the business based on metrics and KPIs. Select the number that best represents the status of the metrics and indicators in your organization:

- Has no established KPIs. (1 point)
- Has KPIs but only for strategic organizational performance. (2 points)
- Has KPIs for organizational and strategic performance, as well as departmental, but the departmental ones are not horizontally aligned. (3 points)
- Has KPIs for organizational and strategic performance, as well as departmental, and a deliberate practice is in place to trickle down and across all KPIs for vertical and horizontal alignment. (4 points)

Score	
--------------	--

Communicating Decisions

Data-literate organizations are highly skilled in communicating data and insights.

Select the number that best represents how your organization communicates with data:

- No organizational processes exist for communicating decisions with data and insights at any level. (1 point)
- Decisions on select projects are communicated using data and insights but only in an ad hoc manner. No formal process exists. (2 points)
- A process exists for communicating decisions using data and insights, but it is only used on select projects or within a subset of departments. (3 points)
- A process exists and is used everywhere across the organization for communicating decisions using data and insights. (4 points)

Score	
--------------	--

Evaluating Decisions

Data-literate organizations have processes to evaluate and learn from decisions after they are made. The following characteristics apply to how data-literate organizations evaluate decisions:

- The process is evaluated, not just the decision.
- Post-mortem on major decisions.
- Information is collected and analyzed related to the decision after it is made.
- A process exists for evaluating decisions after they are made.

Select the number that best represents the status of evaluating decisions in your organization:

- None or one of the characteristics listed exists. (1 point)
- Two of the characteristics listed exist. (2 points)
- Three of the characteristics listed exist. (3 points)

<ul style="list-style-type: none"> • All four of the characteristics listed exist. (4 points) 	
Score	

Talent Management

Data-literate organizations properly manage their organization's talent as it relates to data literacy. Less focus is put on the specific job and role, and more focus is put on the relevant competencies and skills. The following characteristics apply to how data-literate organizations think about talent management:

- The Criteria for assessing potential employees includes looking for specific data literacy competencies and skills.
- The organization has a list of competencies and/or skills that are required to meet organization goals.
- The organization has a method for analyzing employees against required competencies and/or skills.
- The organization has an approach for closing the competency and skills gaps via learning opportunities.

Select the number that best represents the status of evaluating decisions in your organization:

- None or one of the characteristics listed exists. (1 point)
- Two of the characteristics listed exist. (2 points)
- Three of the characteristics listed exist. (3 points)
- All four of the characteristics listed exist. (4 points)

Score	
--------------	--

Assessment score for organizational processes

Add up the total number of points for your responses across all six subdimensions and calculate the average score for this dimension, as shown in *Figure 12.7*:

Dimension: Organizational Processes
--

1	Change Management	1-4
2	Decision-Making Process	1-4
3	Metrics and Indicators	1-4
4	Communicating Decisions	1-4
5	Evaluating Decisions	1-4
6	Talent Management	1-4
Total Dimension Score (Add all scores and enter total)		
Average total (Divide the “Total Dimension Score” by 6)		

Figure 12.7 – Organizational process scoring

Data and analytics

There are five subdimensions within the data and analytics dimension:

Data Infrastructure

Data-literate organizations require a proper data infrastructure to be able to process, analyze, and act on data. Select the number that best represents the data infrastructure in your organization:

- Data is collected on an ad hoc basis with disparate tools. (1 point)
- There are a few people who understand how data is accessed in the organization. (2 points)
- Data is accessible to the majority of the organization, but it is not readily discoverable, compliant, understood, or actionable throughout the organization. (3 points)
- There's trusted centralized storage with quality data that is discoverable, trusted, understood by the business, and accessible. (4 points)

Score

Data Strategy

Data-literate organizations require a solid data strategy to define how they manage data as a strategic asset. Select the number that best represents the data strategy in your organization:

- No strategy in place. (1 point)
- Departments have their own data strategies. (2 points)
- Organization-wide data strategy. (3 points)
- An organization-wide data strategy with its own organization, leader (CDO), and budget. (4 points)

Score	
--------------	--

Data Governance

Data-literate organizations require an effective data governance strategy to ensure the data is usable and protected so that it is trusted. Select the number that best represents the data governance in your organization:

- There are no data quality rules and processes and no documented strategy for data governance. There is no centralized approach to data, leading to redundant data. (1 point)
- Various departments have implemented rules and processes for data quality and governance, but nothing exists across the entire organization. (2 points)
- A centralized approach to data quality and governance exists, but either not all the departments utilize it or the centralized approach is only used for specific projects. (3 points)
- A centralized approach to data quality and governance exists, and the entire organization has embraced it, providing users with a trusted single source of truth when it comes to data and insights. (4 points)

Score	
--------------	--

Data Access/Data Democratization

The following characteristics apply to how employees within data-literate organizations access data:

- Data is accessible across the organization to use with self-service analytics tools.
- Internal and external integration.
- Defined schemas of inputs.
- Identifiable catalog of datasets.
- Derivability of data.

Select the number that best represents the status of data access in your organization:

- None or one of the characteristics listed exists. (1 point)
- Two of the characteristics listed exist. (2 points)
- Three or four of the characteristics listed exist. (3 points)
- All five of the characteristics listed exist. (4 points)

Score	
--------------	--

Analytics Maturity

Data-literate organizations leverage the full range of analytic capabilities to help come up with insights. These range from more simple descriptive analytics to more advanced predictive and prescriptive analytics. Select the number that best represents the analytics maturity in your organization:

- Basic analytics is used in some parts of the organization, but not everywhere. The analytics are typically in the form of descriptive analytics via ad hoc reporting on historical data. (1 point)
- Analytics includes diagnostic analytics as well. (2 points)
- Analytics includes predictive analytics as well. (3 points)
- Analytics includes prescriptive analytics as well. (4 points)

Score	
--------------	--

Assessment score for data and analytics

Add up the total number of points for your responses across all five subdimensions and calculate the average score for this dimension, as shown in *Figure 12.8*:

Dimension: Data and Analytics		
1	Data Infrastructure	1-4
2	Data Strategy	1-4
3	Data Governance	1-4
4	Data Access	1-4
5	Analytics Maturity	1-4
Total Dimension Score (Add all scores and enter total)		
Average total (Divide the “Total Dimension Score” by 5)		

Figure 12.8 – Data and analytics scoring

Tools and technologies

There are two subdimensions within the tools and technologies dimension:

Tools

The following tools and capabilities should exist within a data-literate organization:

Data Storage Tools

- Data Integration Tools
- Master Data Management Tools
- Data Mining Tools
- Data Catalog
- Data Visualization Tools
- Analytics Tools
- Advanced Analytic Tools
- Artificial Intelligence/Machine Learning
- Decision Intelligence Tools

Select the number that best represents the status of tools and capabilities in your organization:

- Up to 3 of the tools and capabilities listed exist. (1 point)
- 4 or 5 of the tools and capabilities listed exist. (2 points)
- 6 or 8 of the tools and capabilities listed exist. (3 points)
- 9 or 10 of the tools and capabilities listed exist. (4 points)

Score	
-------	--

Integrations

Having the right tools and technology is important, but data-literate organizations require that those tools and technologies be integrated into the workflow of the users. This helps empower employees by bringing the insights to them in an environment they are used to, rather than them having to switch between multiple applications.

Select the number that best represents the maturity of technology integrations in your organization:

- Data and insights are provided manually in the form of reports. Very little integration exists. (1 point)
- Some levels of automated reports and dashboards exist within the organization, but they are not embedded; rather, they are emailed. (2 points)
- A few key analytics are embedded directly into workflows and applications. The rest are handled via manual or automated reports and dashboards. (3 points)
- The majority of analytic capabilities are embedded directly into workflows and applications that the users use. (4 points)

Score	
-------	--

Assessment score for tools and technologies

Add up the total number of points for your responses across the two subdimensions and calculate the average score for this dimension, as shown in *Figure 12.9*:

Dimension: Tools and Technologies		
1	Tools and Capabilities	1-4
2	Integration	1-4
Total Dimension Score (Add all scores and enter total)		
Average total (Divide the “Total Dimension Score” by 2)		

Figure 12.9 – Tools and technologies scoring

Knowledge and learning programs

There are six subdimensions within the knowledge and learning programs dimension:

Learning Offerings	
Data-literate organizations provide various learning options for their employees to upskill their data literacy. Select the number that best represents the maturity and diversity of data literacy learning offerings in your organization:	
<ul style="list-style-type: none">• No offerings are provided or recommended. (1 point)• Third-party offerings are recommended but they must be completed on your own time. (2 points)• Offerings are built into the company’s L&D programs and certain employees are allowed to take them during work hours (but not all employees). (3 points)• Organization-wide data upskilling programs are open and available to all employees to take during work hours. (4 points)	

Score	
Content Included The following types of learning content should exist within a data-literate organization:	

- Learning about data and analytics tools and technology.
- Learning about industry and business knowledge.
- Learning about organizational knowledge.
- Learning about soft skills.

Select the number that best represents the status of data literacy learning content in your organization:

- Data and analytics tools and technology training only. (1 point)
- Data and analytics tools and technology training plus either industry, business, organizational knowledge, or soft skill training. (2 points)
- Data and analytics tools and technology training plus two of the following: industry, business, organizational knowledge, or soft skill training. (3 points)
- All four types of learning content exist. (4 points)

Score	
--------------	--

Levels and Assessments

A data-literate organization has learning offerings that contain various learning levels to allow individuals to learn what they need at the time (that is, awareness, understanding, application, and expertise). Select the number that best represents the levels of data literacy learning offered in your organization:

- No learning levels. Just one level of training. (1 point)
- Two learning levels (for example, basic and advanced) exist but there's no assessment to validate you learned what you needed at the current level. (2 points)
- Three or more learning levels exist but there's no assessment to validate you learned what you needed at the current level. (3 points)
- Three or more learning levels exist with an assessment to validate what you learned at each level. (4 points)

Score	
-------	--

Community

Data-literate organizations provide mechanisms for employees to get informal learning and ask questions to peers in their moment of need. This can include forums, a coach or mentor, or a community of practice, as examples. Select the number that best represents the support for data literacy communities in your organization:

- Forums exist but no one is accountable for monitoring them. (1 point)
- Forums exist and there are employees dedicated to monitoring them. (2 points)
- A coaching or mentoring program for data literacy exists. (3 points)
- Communities of practice or similar exist. (4 points)

Score	
-------	--

Learning Delivery Options

Data-literate organizations should provide multiple delivery options for employees to consume data-literate training. Those delivery options can include the following:

- Instructor-led
- Self-paced
- Workshops
- Webinars
- Blended learning

Select the number that best represents the status of learning delivery options in your organization:

- None exist. (1 point)
- One or two exist. (2 points)
- Three exist. (3 points)

<ul style="list-style-type: none"> • Four or five exist. (4 points) 	
Score	

Knowledge Management Practices

Data-literate organizations provide a mechanism to capture the insights, knowledge, and wisdom of employees in an organizational knowledge base. Select the number that best represents the maturity of knowledge management in your organization:

- No mechanism exists. (1 point)
- A mechanism exists, but only in certain departments. (2 points)
- A mechanism exists for the entire organization, but adoption is low. (3 points)
- A mechanism exists for the entire organization, and the entire organization uses it. (4 points)

Score	
--------------	--

Assessment score for knowledge and learning programs

Add up the total number of points for your responses across all six subdimensions and calculate the average score for this dimension, as shown in *Figure 12.10*:

Dimension: Knowledge and Learning Programs		
1	Learning Offerings	1-4
2	Learning Content	1-4
3	Learning Levels	1-4
4	Community	1-4
5	Delivery Options	1-4
6	Knowledge Management	1-4
Total Dimension Score (Add all scores and enter total)		
Average total (Divide the “Total Dimension Score” by 6)		

Figure 12.10 – Knowledge and learning programs scoring

Analyzing the maturity score

Once the assessment is completed, the organization can analyze the results. The goal of this step is to identify the most important areas to start working on improving.

In the online assessment, each dimension's score is calculated on a range from 1 to 10. With a consistent range for each dimension, we can plot them all together in a radar plot, as shown in *Figure 12.11*, to see where your organization's data literacy strengths and weaknesses are:

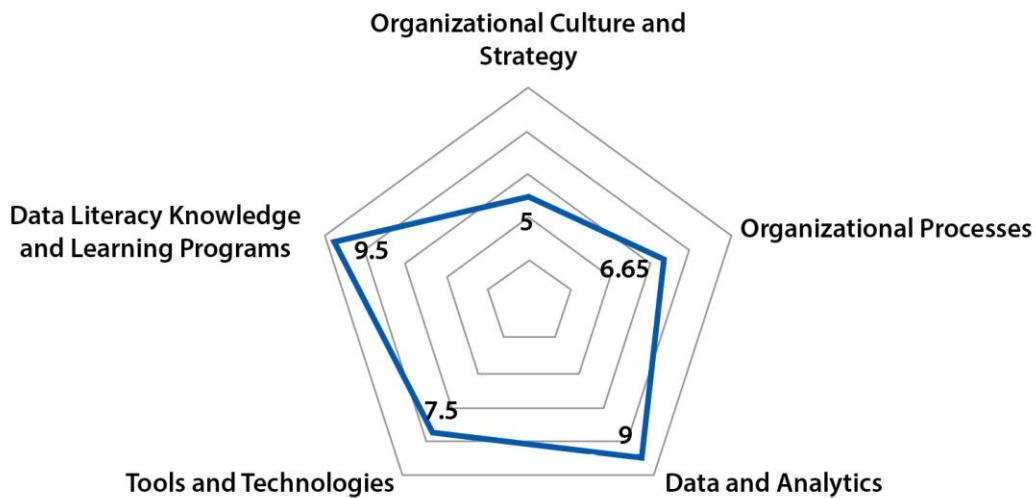


Figure 12.11 – Organizational data literacy assessment results

If you are not using the online assessment and are scoring this manually, just use the average scores from each dimension. As an example, *Figure 12.12* shows the scores from a completed manual assessment. Here, you can see the average scores from each dimension listed:

Average of the Organizational Culture and Strategy Dimension	2.0
Average of the Organizational Process Dimension	2.66
Average of the Data and Analytics Dimension	3.6
Average of the Tools and Technologies Dimension	3.0
Average of the Knowledge and Learning Programs Dimension	3.8
Overall Organizational Data Literacy Maturity Score (Take the average of all the dimension averages)	3.01

Figure 12.12 – Sample organizational data literacy assessment scoring

The table displayed in *Figure 12.13* shows the maturity level for each dimension calculated from the average score. The same table can be used to show the overall maturity level of the organization if you use the overall average score across all the dimensions:

Dimension Average Score	Maturity Level
1.00 – 1.59	Ad Hoc
1.60 – 2.19	Developing
2.20 – 2.79	Managed
2.80 – 3.39	Advanced
3.40 – 4.00	Optimized

Figure 12.13 – Organizational data literacy maturity levels

There is no common answer to the question of *what area we should focus on improving first*. All of the dimensions that have been assessed are important for organizations. However, we believe it is near impossible to mature in your organizational data literacy without the right strategy, support, and culture, so it makes sense to look for opportunities to increase your score in that dimension. Beyond that, the online assessment will give you more specific feedback and recommendations of where to start regarding your areas of improvement based on your scores.

However, if you are scoring the assessment manually and not using the online assessment, you should first focus on which dimensions are lower than the others.

Figure 12.14 shows the results of the sample assessment:

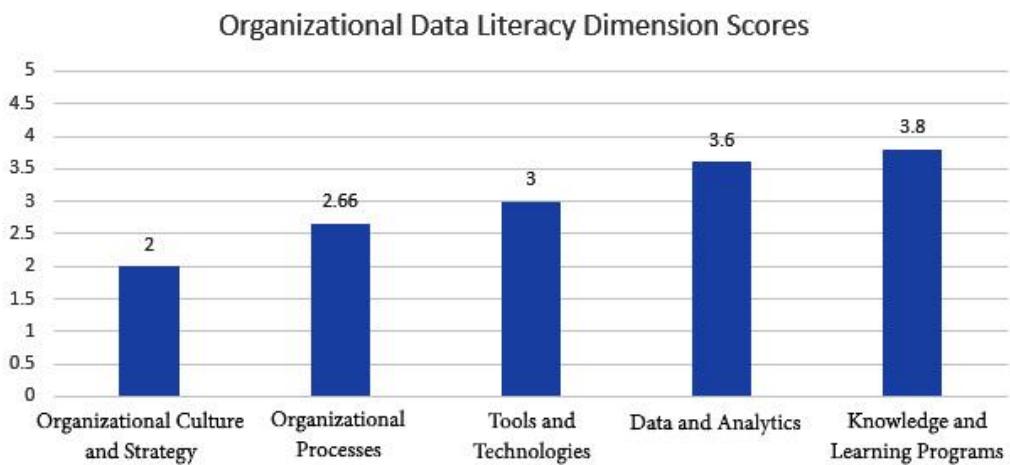


Figure 12.14 – Organizational Data Literacy dimension scores

The results highlight that **Organizational Culture and Strategy** is the lowest scoring dimension with a score of **2.0** on the higher end of the results compared to **Knowledge and Learning Programs**, which has a score of **3.8**. In this case, the recommendation would be to drill down into the **Organizational Culture and Strategy** dimension and start to make improvements there first.

To drill down into a dimension, you should look at each subdimension's score compared to the average across all the scores. For example, if you have identified you need to improve on your **Organizational Culture and Strategy** dimension, you can use the chart shown in *Figure 12.15* to visualize your scores for each of its six subdimensions and identify which you should prioritize first:

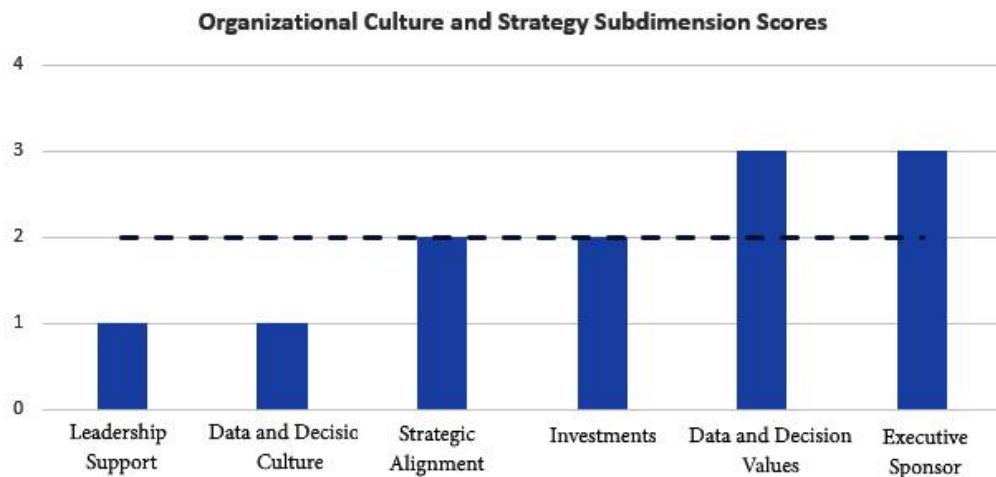


Figure 12.15 – Organizational Culture and Strategy Subdimension Scores

Addressing the subdimensions with the lowest scores would make the most sense as a place to start. Those subdimensions most likely need immediate attention. Once those subdimensions have been selected, the organization should meet and kick off a plan to address the weaknesses. In the example shown in *Figure 12.15*, the lowest subdimensions are **Leadership Support** and **Data and Decision Culture**.

Summary

In this chapter, we learned how to assess both the data literacy maturity of an organization as well as an individual's data literacy competency level. For both, we provided a very simple assessment that should take no longer than 15 minutes, along with more robust assessments that will give you more specifics on your gaps and where to focus next. Ideally, you will take the assessments initially so that you have a baseline and then retake the assessment at future intervals – for example, every 6 months or every year – to see how well you are closing the gaps.

In the next and final chapter, we will focus on how you should approach data and analytics projects within your organization, as well as what risks to look out for.

14 – Appendix A – Templates

We have used a variety of templates throughout this book.

The following templates are the ones that are all available. Please feel free to use these ideas as best practices, as *best practices are good ideas*:

- Project intake form
- Layout for a business case
- Layout for a business case scenario description
- A business case financial analysis
- Layout for a risk assessment
- Layout for a summary business case (for the conclusion and advice)
- Layout information and measure plan
- Layout for a KPI description
- Table with the Inmon groups and a description of their roles

Project intake form

Part I – To be completed by the client

Project

Name:

Adjustment for existing application(s) Yes, name:

No

Client details

Organizational unit client

Cost center number client

Client (budget manager)

Client role

Delegated client

Role delegated client

Contact person client

Role contact person

Relationship with other project(s) Yes, name:

No

Project description

Briefly describe what the client wants to see realized.

Description of business unit

Briefly describe the function of the business unit for which the application is intended.

Non-functional requirements

Number of expected users	<10 10-50 50-100 100-250 >250
Expected frequency of use	ad-hoc daily weekly monthly quarterly yearly
expected mode of use	Standard reporting Search interactively
Up-to-datedness of the data (refresh rate)	ad-hoc daily weekly monthly quarterly yearly

Addition and or explanation

Data security check

A data security incident is a serious security breach. When protected, sensitive, or confidential data is stolen, used, viewed, stored, or transmitted by someone who is not authorized, a breach occurs. Other security concerns include data leaks, information leaks (who can see what), sensitive information leaks, and data spillage. As a result, when it comes to data and information usage, we must identify some specific elements.

Application/model/app

Dashboards and reports are critical for any organization, team, or department. The work cannot be completed without this information.

Essential |
Important |
Desirable |
Unclassified

Confidentiality

Essential |
Important |

The dashboards and reports are private and only available to a select group of managers and employees.	Desirable Unclassified
Access Control	Add information:
What type of access control is required when confidentiality is required? What are the rules that we must follow? (Manager of team A – can see information about team A, and so on).	
Integrity	Essential Important Desirable Unclassified
Incorrect information will not be tolerated (accountable environment). The organization suffers significant damage if the information is incorrect, incomplete, or late. Inaccurate data, such as inaccurate financial transactions, undermines trust in the organization.	
Availability	Essential Important Desirable Unclassified
Dashboards and reports are essential. The organization will suffer severe consequences if the information is not available.	

Addition and or explanation

Data privacy check

Some countries have special legislation and regulations when it comes to data usage and data privacy. For example, in the Netherlands, there is a regulation called **General Data Protection Regulations (GDPR)**. When it comes to regulations in the US it is a patchwork of rules. There is actually no specific privacy law except in relation to children but in the State of California, they have implemented data privacy regulations.

When it comes to using personal data we should take care and be aware upfront that we need to consider the privacy regulations for our project. There are several categories that we have to identify and classify.

Name, address, and city data

Yes

(Name, address, postal code, city of residence)	<input type="checkbox"/> No
Identification data	<input type="checkbox"/> Yes
(Passport, driver's license, or social security number)	<input type="checkbox"/> No
Application details	<input type="checkbox"/> Yes
(Application letters, resume)	<input type="checkbox"/> No
Contact details	<input type="checkbox"/> Yes
(E-mail address, phone or fax number)	<input type="checkbox"/> No
Salary details	<input type="checkbox"/> Yes
(Salary details, social payments, income taxes, expense reimbursement)	<input type="checkbox"/> No
Social media accounts	<input type="checkbox"/> Yes
(LinkedIn, Twitter, WhatsApp, Facebook)	<input type="checkbox"/> No
Image and sound recordings	<input type="checkbox"/> Yes
(Video, photos, passport photographs)	<input type="checkbox"/> No
Payment details	<input type="checkbox"/> Yes
(Bank name, account number, name of account holder)	<input type="checkbox"/> No

Addition and or explanation

When a checkbox is answered with “yes” it is advised to arrange a legislation and regulations check with your Privacy Officer.

Legislation – regulations check

Is privacy impact analysis needed?	Essential Important Desirable Unclassified
Is a register of data needed?	Essential Important Desirable Unclassified

Addition and/or explanation

Data ethics checklist

Data generation, use, collection, analysis, and dissemination are all aspects of data. We can do this with both unstructured and structured data. When we do so, there is a chance that the decisions we make will have an impact on individuals and the world. As a result, we must be open about how we use data in our projects.

- 1 **Do we have to take care of laws and regulations in this data and analytics project?** Yes
 No

The first critical step is to determine whether any legalizations or regulations are applicable to the project.

Write any additional information here:

- 2 **Is the data that we want to use available in an ethical manner and is this data suitable for usage?** Yes
 No

We must be mindful of who owns the data and ensure that it is used in the manner intended by the owner!

Write any additional information here:

- 3 **Is it possible to identify and check bias in the data that we have collected or used for our models?** Yes
 No

People can be biased by their origin, and the same is true for data and the application of algorithms and business rules. The data we collect and store is not as objective as we believe! When using algorithms and business rules, we should be aware that the data we use to train the models can have an impact on people and possible human bias can be magnified, which results in undesirable outcomes.

Write any additional information here:

To conclude, we must be able to identify, test, verify, and discuss the results.

- 4 Can we identify and demonstrate bias in our created model or in the data used?**
- When we use data and apply various models, we may have used data that is biased. When we use that specific data set and apply the learning models, the model produces biased results. For example, a bias based on gender, age, equality, or racial elements.
- We need to be aware that we must consider documenting, discussing, and evaluating our data usage choices. The message here is to avoid doing things simply because you can!
- 5 Can the legal rights of individuals be impinged by the use of data?**
- When an individual's legal rights are at stake, the organization must have permission and the right to use data for specific purposes. As an example, suppose an organization provides data to its direct partners, but different privacy conditions than the internal data usage may be addressed here. In the event of an incident, for example, they know the addresses and more detailed information from people, but certain data is not shared to protect those people's privacy (such as names, address details, and other things on which someone can be individually identified).
- It could also happen internally, for example, by logging information that is known at the employee level but is shared with users at the department or concern level. We should be aware of this in order to protect the privacy of each individual employee.
- 6 Are we able to understand that the data we want to use is suitable for the purposes of our project?**

Yes

No

Write any additional information here:

Yes

No

Write any additional information here:

Yes

No

	<p>When we begin an analytics or data science project, we must understand and ensure that the data we intend to use is appropriate for the purpose of our projects. Following that, we should be able to verify and validate the data for our project.</p> <p>For example, when records or values are missing, the outcome of our algorithms and business rules can have a significant impact on the results, potentially producing a biased result.</p>	<p>Write any additional information here:</p>
7	<p>Do we have a multi-disciplinary team present to discuss the present dilemmas, and explore the possible usage of algorithms and the possible outcomes?</p> <p>When it comes to assessing and discussing our own work, we need to focus on the dilemma and the outcomes that can occur with a multi-disciplinary team.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <p>Write any additional information here:</p>
8	<p>Explainable AI by design? Are we able to define the role of an algorithm used and what processes are being followed (procedural transparency)?</p> <p>It is critical for data engineers and data scientists who train models to understand the model's behavior in order to detect errors or weaknesses. This is why we must correctly describe the used algorithms or business rules.</p> <p>When data scientists and data engineers train a model, it is critical to understand the model's behavior. They must be able to identify any flaws or errors.</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <p>Write any additional information here:</p>
9	<p>Explainable AI. Are we able to explain the algorithm or business rules to the guardians, stakeholders, and others whom it concerns?</p> <p>Explainable AI is defined as the ability of a person to comprehend the reason for a decision. The decision is influenced by algorithms and business rules. To be able to</p>	<input type="checkbox"/> Yes <input type="checkbox"/> No <p>Write any additional</p>

understand the model's decision, we must be able to explain the decision. We can do so by design, but we can also do it post-hoc by using an algorithm to understand the black-box model.

information here:

Addition and/or explanation

Risks

Are organizational changes to be expected that could affect the progress of the project?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Has a project manager/leader/coordinator been appointed on the client side?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Is the contact person present full-time? Or is there a full-time backup?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Can a response time to questions of a maximum of 3 days be guaranteed?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Does the contact person have knowledge of the source systems and the source data?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Is all data available in the source systems?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> No, but the development/adaptation of the source systems is planned and will be released on the following date:
Are there interfaces with the source systems?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> No, we have to arrange this

Risks

Are there dependencies with external suppliers?

Yes

No

Are there dependencies with external suppliers?

Yes

No

Source system data (per source system)

Source system name

Source system type

- Database
- Interface, an automated connection
- File, manual (e.g. MS Excel, CSV, and so on)
- Cloud
- Other

Owner source system (department)

Internal/external management organization

- Internal – name:
- External – name:

Contact person

Is the source system data model available?

Yes

No

Are field definitions available?

Yes

No

Particularities

Size of data delivery (number of files)

Number of expected rows

Additional information

Source system data (per source system)

Source system name

- Source system type
- Database
 - Interface, an automated connection
 - File, manual (e.g. MS Excel, CSV, and so on)
 - Cloud
 - Other

Owner source system (department)

- Internal/external management organization
- Internal – name:
 - External – name:

Contact person

- Is the source system data model available?
- Yes
 - No
- Are field definitions available?
- Yes
 - No

Particularities

Size of data delivery (number of files)

Number of expected rows

Additional information

Part II – to be completed by the contractor/internal supplier, and so on.

Other project data

Project number

Date of intake

The appointed project manager or project leader

Date intake forwarded to the contact person

Name of the person who did the intake

Relationship with other projects? Yes, name of project:

No

The first functional requests that are known

Indicate in the table below as completely and concretely as possible which functional requirements you set for the application. The MoSCoW method is used to set priorities:

Must have this – this requirement must be delivered in the final result

Should have this if at all possible – this requirement is highly desirable

Could have this if it does not affect anything else – this requirement should only be addressed if there is enough time

Would have this – this requirement will not be discussed now, but could be interesting in the future and is noted in the backlog

Complexity:

Low/middle/high/very high

The standard level of change complexity is middle

Description of the functional requirements	MoSCoW	Complexity	Intended for the type of use?
--	--------	------------	-------------------------------

The first functional requests that are known

Indicate in the table below as completely and concretely as possible which functional requirements you set for the application. The MoSCoW method is used to set priorities:

Must have this – this requirement must be delivered in the final result

Should have this if at all possible – this requirement is highly desirable

Could have this if it does not affect anything else – this requirement should only be addressed if there is enough time

Would have this – this requirement will not be discussed now, but could be interesting in the future and is noted in the backlog

Complexity:

Low/middle/high/very high

The standard level of change complexity is middle

Description of the functional requirements	MoSCoW	Complexity	Intended for the type of use?
--	--------	------------	-------------------------------

Other elements that were agreed upon

Other elements that were agreed upon

STARR TEMPLATE

Name project :

Name project member :

Date :

S Situation

Describe the circumstances under which the your project took place (reason, parties involved, where, etc.)

T Task

Describe the exact assignment you worked on (and who was your project leader, client), or that you assigned to yourself. Indicate what your role/position was in this project

A Activities

Describe the approach you used (activities; tools; preparation)

R Results

How was the result assessed, by whom and on what basis?

R Reflect

Why do you consider this your best practice, how do you look back to the work that you have done? Relate to the complexity of the situation and/or development in feedback you've received.

Layout for a business case

11. *Reasons for your project* – A brief introduction, a description of the problem, and the project's main goal (why).
12. *The options or scenarios* – The following elements create a scenario: an introduction, pros and cons, and strategic contribution. *Figure 13.10* shows an example of how to write your scenarios or options.
13. *Qualitative benefits* – Try to emphasize the advantages that your project will bring to your company (without the financial calculations for now).
14. *Risks* – Which high-impact risks did you identify? There are several risks that we should investigate, control, and describe.
15. *Financial analysis* – There are a few things to consider when conducting a thorough financial analysis of your project. It not only makes all project costs visible, but it also explains why you need to include them in your budget. The same principle applies to the future value that your business case can provide. Always plan three years ahead of time, and consider the costs of maintaining dashboards and reports. The majority of the benefits will be quantitative (in eurodollars), but some will be qualitative.
16. *Conclusion and advice* – It's time to wrap up the business case and recommend an option to your board of directors and management team.

Layout for a business case scenario description

Name

Option 1

Description

Pros

Cons

Strategic contribution

A business case financial analysis

In the next example, we will guide you through a financial analysis for a fictive small project. In *Chapter 13, Managing Data and Analytics Projects*, the business case is described in more detail.

Describe and calculate the one-time project costs:

QUANTITATIVE ANALYSIS	YEAR 1	YEAR 2	YEAR 3	TOTAL
-----------------------	--------	--------	--------	-------

One-time project costs

	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
TOTAL ONE-OFF COSTS	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00

Describe and calculate the annual recurring costs:

RECURRING COSTS	YEAR 1	YEAR 2	YEAR 3	TOTAL
	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
TOTAL RECURRING COSTS	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00

Describe and calculate the quantitative benefits:

QUANTITATIVE BENEFITS	YEAR 1	YEAR 2	YEAR 3	TOTAL
-----------------------	--------	--------	--------	-------

Value

	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
TOTAL BENEFITS (VALUE)	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00

When one-time costs, recurring cost, and the possible value is calculated it's time to perform a quantitative analysis. With this analysis, you will be able to determine if the project could be of value to your organization.

QUANTITATIVE ANALYSIS	YEAR 1	YEAR 2	YEAR 3	TOTAL
BENEFITS				
Total of benefits	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
TOTAL BENEFITS	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
COSTS				
ONE-TIME PROJECT COSTS	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
RECURRING COSTS	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
TOTAL COSTS	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
NET BENEFITS (+) / COSTS (-)	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00

Layout for a risk assessment

Which high-impact risks did you identify? Rate their chance of occurring from 1-5, along with the potential impact also from 1-5, and multiply to get the total score.

Nr.	Risk	Chance of occurrence	Impact	Score
1	Extracting historical data can cause performance issues.	5	5	25
2	Inconsistent data from several sources that do not align. Therefore, the insights will not be trusted.	4	4	16
3				
4				

Layout for a summary business case

The following is an example of using a table with the business case elements to make a thorough choice based on an array of advice.

	Option 1	Option 2	Option 3	Option 4
Description	No changes	Having a new tool	Change existing visualizations	Complete new infrastructure and cockpit
Benefits	None	High	Medium	High/medium
Cons	High	Low	Medium	Very high
Strategic contribution	None	High	Medium	Medium
Actors	None	4	5	8
Project duration	None	4 sprints of 2 weeks	2 sprints of 2 weeks	6 sprints of 2 weeks
Costs	None	Medium (write the value of your financial analysis here)	Low (write the value of your financial analysis here)	Very high (write the value of your financial analysis here)
Maintenance	None	Medium (write the value of your financial analysis here)	Low (write the value of your financial analysis here)	High (write the value of your financial analysis here)

Layout information and measure plan

Elements that should be written down in an information measure plan are as follows:

17. An introduction:

18. A short description of your information measure plan

19. A problem analysis

20. The objectives of your project

21. The scope (what are you going to do and what not)

22. The information necessity for management:

23. Describe the KPIs and reports that are needed

24. Who your users are (have a look at *Figure 7.16*)

25. Describe the desired level of granularity (the level of detail)

26. Describe the nonfunctional requirements (see the intake form that we added here in the template chapter)

27. User stories/use cases/requirements:

28. User story 1

29. User story 2

30. And so on

31. Data sources:

32. Desired sources

33. Data quality check

34. Describe the missing data

35. Describe the metadata

36. The solution:

37. Describe the dashboard and reports (follow the DARs principle – see *Chapter 7*)

38. Describe and sketch the data flow
39. Design the dashboard (using the chatting and sketching technique that is described in *Chapter 7*)
40. Appendix : A
41. Design decisions
42. Remarks

Layout for a KPI description

Name of KPI

Definition of KPI

Owner of KPI

What is the purpose of usage?

Frequency

Department

Type of graph

Additional information

Unit of measure

Data source

Required tables

Reporting period

Data owner

Data steward

Date of KPI approval

Norm

KPI

Table with the Inmon groups and a description of their roles

Who is my public?	Farmers	Tourists	Explorers	Miners
Managing board	X			
Middle management	X			
Team manager		X		
Team lead				
Controller			X	
Process advisor			X	
Business analysts			X	
Analysts			X	
Data analysts			X	
Data scientists				X

A short description of the Inmon classifications used in the table: **Farmers** have defined, predictable requirements. **Tourists** are practically equivalent to farmers, but they must utilize filters to look at the data differently and understand the findings. **Explorers** seek to examine existing indicators from several perspectives (dimensions) and interact thoroughly with dashboards and reports for data-informed decision-making. **Miners** are more of a scientific field; they are our data scientists, and they want a lot of freedom to investigate anomalies in the data (looking for the golden egg).

15 – Appendix B – References

Throughout this book, we have cited or referred to additional materials to help you on your data literacy journey. This appendix serves as a single collection of all these resources. They are categorized as follows:

- *Inspirational books* – Books that we believe are valuable to read to help you increase your data literacy journey.
- *Online articles and blogs* – This is a collection of useful articles and blog posts on topics that are also covered in this book. They can be used to expand your knowledge.
- *Online tools* – This is a collection of helpful online tools that you can use in your various data and analytics projects.
- *Online sites* – This is a collection of websites where we believe you will find useful content related to data literacy.

This list is also available at www.kevinhanegan.com/dataliteracyinpracticereferences. This is especially helpful for people who have the paperback version of this book, so you can click on any of the URLs directly.

Inspirational books

- *Cruise. Control (Dutch)* by Merlijn Gillissen and Charles van der Ploeg
- *De Intelligente, datagedreven organisatie (Dutch)* by Daan van Beek
- *De Big Data Revolution (Dutch)* by Viktor Mayer-Schonberger and Kenneth Cukier (also available in English)
- *Data Strategy* by Bernard Marr
- *Datacratisch Werken (Dutch)* by Daan van Beek
- *Data Quality. The Field Guide* by Thomas C Redman Ph. D
- *Designing Regenerative Cultures* by Daniel Wahl
- *Everyday Bias: Identifying and Navigating Unconscious Judgments in Our Daily Lives* by Howard Ross

- *Lateral Thinking Puzzles* by Paul Sloane
- *Leading Change* by John P Kotter
- *Now You See It* by Stephen Few
- *Out of the Crisis* by Edward Deming
- *Show Me the Numbers* by Stephen Few
- *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* by Scott Page
- *The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization* by Peter Senge
- *Thinking with Data* by Max Shron
- *Turning Data into Wisdom* by Kevin Hanegan

Online articles and blogs

- *6 Reasons Why BI and Analytics Projects Fail and How to Avoid It* by Mark Tossel: <https://www.salesforceben.com/6-reasons-why-bi-and-analytics-projects-fail-and-how-to-avoid-it>
- *A Complete Guide to Bar Charts* by Mike Yi: <https://chartio.com/learn/charts/bar-chart-complete-guide/>
- *An Appeal to the Eye. William Playfair promotes his charts* by RJ Andrews: <https://infowetrust.com/project/breviary>
- *BitMetric's Visualization of Gartner's Business Analytics Magic Quadrant*: <https://gap.bitmetric.nl/extensions/magicquadrant/index.html> by Bitmetric
- *Business Intelligence and Analytics Image*: https://www.researchgate.net/figure/Business-intelligence-and-analytics-Source-Davenport-and-Harris-2007_fiq1_311962711

- *Chapter 5 (The Big Bang: William Playfair, the Father of Modern Graphics) from A History of Data Visualization & Graphic Communication:*
<https://friendly.github.io/HistDataVis/ch05-playfair.html>
- *Concave vs Convex:* <https://www.grammarly.com/blog/concave-vs-convex/>
- *Data on Old Faithful Geyser:*
<https://www.nps.gov/yell/planyourvisit/exploreoldfaithful.htm>
- *Data Privacy and Security Statistics:* <https://www.vpnmentor.com/blog/data-privacy-security-stats>
- *Data Tables Design Basics* by Taras Bakusevych: <https://taras-bakusevych.medium.com/data-tables-design-3c705b106a64>
- *Data Visualization Overview* by Kate Brush and Ed Burns:
<https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization>
- *Diagram of the Causes of Mortality in the Army in the East:*
<https://www.historyofinformation.com/image.php?id=851>
- *How to Increase the Success Rate of Data Science Projects:*
<https://www.techtarget.com/searchbusinessanalytics/feature/How-to-increase-the-success-rate-of-data-science-projects>
- *Important Examples of Prehistoric Rock Art:* <https://www.spain.info/en/places-of-interest/caves-altamira>
- *Moneyball (the movie):* [https://en.wikipedia.org/wiki/Moneyball_\(film\)](https://en.wikipedia.org/wiki/Moneyball_(film))
- *Netherlands Data Protection Authority Publishes 2021 Data Breach Report:*
<https://www.dataguidance.com/news/netherlands-dpa-publishes-2021-data-breach-report>
- *Our Top Data and Analytics Predictions for 2019 by Andrew White:*
https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/

- *Policy Brief: Principles for Responsible Data Handling:*
<https://www.internetsociety.org/policy/briefs/responsible-data-handling>
- *QlikView Case Study: Uitvoeringsinstituut Werknemersverzekeringen (UWV) Competence Center builds more than 50 applications in just 2 years and provides 1,650 people with reporting & analysis capabilities:* <https://www.qlik.com/us-/media/files/customer-success/global-us/uwv-success-story-en.pdf> and <http://docplayer.net/30148188-Discover-uwv-qlikview-competence-center-gcc.html>
- *The 7 Most Pressing Ethical Issues in Artificial Intelligence:*
<https://kambria.io/blog/the-7-most-pressing-ethical-issues-in-artificial-intelligence>
- *The Lost Art of Questioning* by Kevin Hanegan: <https://www.qlik.com/blog/the-lost-art-of-questioning>
- *Torn... Between Two...* by Angelika Klidas:
<https://www.linkedin.com/pulse/torn-between-two-angelika-klidas>
- *Visual Battle: Table vs Graph* by Cole Nussbaumer:
<https://www.storytellingwithdata.com/blog/2011/11/visual-battle-table-vs-graph>
- *VizLib Library: Supporting data literacy for all:* <https://home.vizlib.com/vizlib-library-supporting-data-literacy-for-all/>
- *What to Consider When Creating Tables: Dos & Don'ts of Table Design* by Lisa Charlotte Muth: <https://blog.datawrapper.de/guide-what-to-consider-when-creating-tables>
- *Why Do So Many BI Initiatives Fail?* <https://www.silvon.com/blog/bi-initiatives-fail> by Pat Passet
- *Why Visual Literacy is Essential to Good Data Visualization* by Ben Dexter Cooley: <https://towardsdatascience.com/why-visual-literacy-is-essential-to-good-data-visualization-5b9dffb5aa6f>

Dutch articles and blogs

The following list includes articles and blogs written in Dutch:

- *13 Vision Statement Examples* by Jasper Bronkhorst:
<https://www.geldreview.nl/leiderschap/visie-statement-voorbeelden>
- *Call to Balloon – Insight into the Chain of Acute Myocardial Infarction Care (Dutch)* by Guus Schrijvers: <http://guusschrijvers.nl/wp-content/uploads/2016/10/Martin-Smeekes-Meten-en-weten-in-de-spoedzorgketen.pdf>
- *How You Become the Most Inspiring Organization in your Industry (Dutch)*:
<https://www.identitive.nl/hoe-ook-meest-inspirerende-organisatie-branche-worden>
- *Increase in infections continues, number of new corona patients in hospitals fluctuate (Dutch)*: <https://www.telegraaf.nl/nieuws/2012726519/stijging-besmettingen-zet-door-aantal-nieuwe-coronapatiënten-in-ziekenhuizen-schommelt>
- *Predictive Analytics: Five Inspiring Examples (Dutch)*:
<https://biplatform.nl/517083/predictive-analytics-vijf-inspirerende-voorbeelden.html>
- *Small Online Discovery Tour in Stasi Museum*:
<https://www.pieterianssen.eu/2020/04/06/kleine-online-ontdekkingstocht-in-stasi-museum>
- *Transition Process of the Stasi Documents in Full Swing*:
<https://www.pieterianssen.eu/2020/06/15/transitieproces-van-de-stasi-documenten-in-volle-gang>
- <https://mijngemeente.vrnhn.nl/verhalen/call-balloon-tijdwinst-bij-patiënten-met-een-hartinfarct>
- <https://www.rd.nl/artikel/479438-in-2020-produceert-de-wereld-40-zettabytes-aan-digitale-data> by Reformatorisch dagblad

Online tools

The following list includes the online tools you can use:

- *Adobe Color* – Online color wheel and color theme generator: color.adobe.com
- *Colors* – Online color palette generator: colors.co
- *Financial Times Chart Doctor* – A repository for images related to the Financial Times Visual Vocabulary: <https://github.com/Financial-Times/chart-doctor/tree/main/visual-vocabulary>
- *Palette Generator* – Create color palettes based off uploaded images palettengenerator.com
- *Sketch App Sources* – Design files and other UI/UX resources and templates: www.sketchappsources.com
- *Statista* – Site that shows statistics around the volume of data and information created and consumed: www.statista.com/statistics/871513/worldwide-data-created

Online sites

You can find the following sites and information online:

- *Alberto Cairo's website* – Content related to information design and data visualization: www.albertocairo.com
- *Dataversity* – Data Education for Business and IT Professionals: www.dataversity.net
- *Freecode Camp* – Educational site for learning web development: www.freecodecamp.org
- *IBM* – Helps organizations become data-driven businesses: www.ibm.com
- *IDC* – A global market intelligence firm: www.idc.com
- *Perceptual Edge Data Visualization Consultancy*: www.perceptualedge.com

- *Qlik* – Data and Analytics Software Company: wwwqlik.com
- *SAS Analytics Academic Programs*
https://www.sas.com/nl_nl/learn/academic-programs.html
- *The Data Literacy Project* – A community dedicated to creating a data-literate world, with blogs, assessments, courses, and more: thedataliteracyproject.org
- *The Power of AI and IoT* – AI and IoT Knowledge Base (Dutch): thepowerofai.nl

