

Bernard Marr

Data Strategy

Course Notes

365  DataScience

Table of Contents

ABSTRACT	3
1. The 5 strategic data use case areas	4
1.1. Using data to improve decisions.....	4
1.1.1. Curated dashboards vs. self-service data exploration	5
1.1.2. Asking key business questions	7
1.2. Using data to understand your customers and markets.....	11
1.3. Using data to provide more intelligent services	12
1.4. Using data to make more intelligent products.....	14
1.5. Using data to improve your business processes.....	16
2. Monetizing your data.....	18
3. Defining your data use cases	18
4. Sourcing and collecting data	20
5. Metadata.....	24
6. Data governance	31
7. Turning data into insights	33
7.1. Text analytics.....	33
7.2. Sentiment analytics	34
7.3. Image analytics	34
7.4. Video analytics.....	34
7.5. Voice analytics	35
7.6. Data mining	35
7.7. Business experiments.....	36
7.8. Visual analytics.....	37
7.9. Correlation analysis	37
7.10. Regression analysis.....	38
7.11. Scenario analysis	38
7.12. Time series analysis	39
7.13. Monte Carlo simulation.....	40
7.14. Linear programming.....	41
7.15. Cohort analysis	41
7.16. Factor analysis.....	41
7.17. Neural network analysis	42

7.18. Deep learning	42
7.19. Reinforcement learning	42
8. Creating the technology and data infrastructure.....	43
8.1. How to collect data.....	43
8.2. How to store data.....	46
8.3. How to process data.....	46
8.4. Communicating data.....	47
9. Building the data competencies in your organization	49
9.1. Skills shortage.....	49
9.2. Building internal skills and competencies	50
9.3. Outsourcing your analysis	51
10. Executing and revisiting your strategy.....	53
10.1. Putting the data strategy into action	53
10.2. Why data strategies fail	54
10.3. Creating a data culture.....	55

ABSTRACT

In this course, you will learn the importance of using data to leverage its potential in business by approaching data in a strategic way, which will help you build a successful business. You will learn how to choose the right data to:

- Make informed better decisions.
- Offer smarter products & services.
- Improve business processes.
- Gain new revenue streams.

Keywords: data, data strategy, business decisions, business processes, data governance

1. The 5 strategic data use case areas

Although all data is important somehow that does not necessarily mean that we need all the data available to improve businesses. In order for data to be valuable to your business, it should help you:

- Address certain business needs.
- Solve your biggest challenges.
- Achieve your strategic goals

To be able to determine a good strategy, you first need to understand your business objectives.

5 strategic data use case areas

- 1- Using data to make better informed, fact-based decisions.
- 2- Helping you understand your markets and customers.
- 3- Using data to offer smarter services and intelligent products.
- 4- Using data to improve and automate your business processes.
- 5- Data monetization.

1.1. Using data to improve decisions

To improve decision-making in organizations, you need to define how you will make data available to employees and decision-makers. There are two ways to do this:

1- Self-service

In this case, you make data accessible to many people in the company. They can choose how to use it to solve their business problems.

2- Curated dashboards

You do not provide access to raw data, but instead work on determining the right data to be used and design sophisticated dashboards intended for decision makers.

In general, both approaches can be used simultaneously.

1.1.1. Curated dashboards vs. self-service data exploration

Good curated dashboard criteria:

1- Identify information needs

To identify exactly what information you need, you must first define the company's key business questions (KBQs). To do this, you need to:

- I. Consider company's strategy and goals.
- II. Define the questions related to each goal that need to be answered.

2- Identify the datasets or metrics to answer the questions

There are three mistakes to avoid when using data for the purposes of decision making:

- 1- Do not provide executives self-service tools hoping they will manage to answer questions on their own.
- 2- Do not include irrelevant data in curated dashboards

Common methods of capturing and recording data:

- Barcodes
- Payment gateways
- Online
- Apps
- Sensors

Challenges of self-service:

- 1- Building the right culture of delegated decision-making.

Employees should be empowered and feel they can contribute to the decision-making process. They need to believe their work is significant to the business's success.

- 2- Ensure improved data literacy

To be able to make decisions, employees need to have solid foundations in data-related skills. This requires training on how to look for the right data, analyze it, and interpret insights.

- 3- Provide the right tools for your employees

In addition to training, you need to provide the right tools and software to your employees, so that they can work with the data in the right way. And note that new tools require additional training.

4- Provide access to the right quality of data

In the field of data, quality is more important than quantity. To ensure you are providing the best data quality, you need to:

- Use a single source of truth
- Invest in master data management
- Conduct data audits

1.1.2. Asking key business questions

To make data-driven decisions, we must first define the questions that need to be answered.

What are key business questions?

- 1- KBQs capture business information needs.
- 2- KBQs are the biggest unanswered questions that managers want answers to.
- 3- KBQs provide guidance for collecting meaningful and relevant data.
- 4- KBQs put data into context facilitating communication and direct decision-making.

6 simple steps to develop KBQs

- 1- Start with strategy
 - a. Clarify strategic objectives.
 - b. Start developing KBQs.

Example

Objective: improve market share for product X

Possible KBQs:

- I. What is our market share?
- II. Who are our main competitors?
- III. Which of our products have an upward or downward trend in the market?
- IV. Will there be demand for service X in 5 years?
- V. How do we best price our products and services?
- VI. What marketing and sales channels are most effective?

Key KBQs:

- I. To what extent are we growing our relative market share for product X?
- II. What are the factors that make our customers buy from us VS our competitors?

2- Engage people in the KBQs design

- Don't design KBQs in the boardroom alone.
- Engage a wider group and ask them: "Which questions do they think are most important?"
- Collect feedback on the designed KBQs from subject-matter experts.

3- Try to phrase your KBQs as open questions

There are 2 types of questions:

Closed and open-end questions. They differ in character, usage and response.

- Closed-end questions

- Seek a short and specific response such as a single word or a short phrase.

E.g. Yes/No questions

- Easy to answer
- Seek simple facts such as What, When, Where?
- The person who asked the question is in control of the conversation

For example:

Are our customers satisfied?

- Open questions
- Seek an open-ended response.
- Invite the respondent to thoughts, explanations, and expressed opinions.
- They often start with what, why, how, and describe?
- The respondent is in control of the conversation.

Example:

What drives customer satisfaction?

4- Focus your KBQs on the present and the future

Ask:

To what extent are we increasing our market share?

Instead of:

Did we increase our market share?

5- Keep your KBQs short and clear

- Use only one question.
- Use clear language, no jargon or abbreviations.
- Stay away from management buzzwords.

Giving people access to data

Key takeaways:

- Instead of just giving people access to the data with a data analysis tool, you can help them by creating starting screens that focus on the frequently asked questions.
- When presenting the data, use suitable graphs to facilitate decision-making.
- Create an initial default visual that shows the data in the best way.
- Think about who should have access to the data in your organization.
- In general, give employees access to more data to democratize data-driven decision-making.

Curating the most important data insights

A well curated dashboard should:

- Include the most important insights.
- Start with a question.
- Have the most appropriate data visualizations, which show the data over time and provide comparisons with benchmarks or targets.
- Include annotations for each graph.
- Contain text-based descriptions because not everyone can read graphs or digest visualizations. The descriptions outline the key insights and interpretations people should be aware of.

1.2. Using data to understand your customers and markets

Data is now giving companies the ability to:

- Better understand their customers.
- Predict consumer trends in the market.
- Predict when customers are likely to buy and when they are likely to leave.

Example in using data in tuna fishing

In the past:

Fishing crews could only rely on:

- Intuition
- Past experiences
- Knowledge passed down through previous generations.

Result: Luck > Facts

Fishing boats had little information about:

- Weather conditions
- Number of other boats

In the present:

It is possible to track tuna via satellite.

- The industry is developing more advanced tracking technology that does not even need tracking sensors.
- Boats get constant updates about weather and wave conditions.

- Boats have data on how many other boats are already fishing.
- Crews can switch on sonar that measures the density of the fish under the boat so they can release the nets and hooks when there is maximum fish density

Example of using data in a butcher's shop

Problem:

- A supermarket opened next to the butcher and competition increased.
- The butcher did not know the shop footfall

Solution:

- I. Since literally everyone has a mobile phone, the butcher installed a \$100 device in the shop that could track and count how many people walked past his shop.
- II. He tried different marketing messages to see the real-time impact of each message.
- III. They pulled data from the Met office, which gives them weather predictions. This helps them decide when to produce sausages.

1.3. Using data to provide more intelligent services

AI is a driving factor for success in the service sector. It helps companies by:

- Enabling service-based businesses to improve their offering.
- Developing entirely new services and revenue streams.

3 ways businesses are delivering a better service through data and AI

1- Delivering a highly personalized offering

Data provides incredible opportunities to get to know your customers:

- Likes and dislikes
- Activity
- Engagement
- Factors encouraging engagement

Stitch Fix

- It is a subscription service that delivers handpicked clothing to your door.
- You can detail your size, style, preferences, and lifestyle.
- Using AI, the system pre-selects clothes that will fit and suit you.
- Then, a human personal stylist chooses the best option from that pre-selected list.

2- Giving customers more value

Successful companies in the future will be the ones able to develop a meaningful value-adding service relationship with their customers.

Vitality health

Unlike traditional health insurance providers, Vitality pays for wellness rather than paying out for sickness by using data and AI to track and reward customers healthy behavior.

For e.g. Members can earn points based on how many steps they took each day. These points can be redeemed against discounted services that support a healthier lifestyle.

3- Predicting customers' needs

1. The more data you have, the more accurate those predictions become, which allows you to provide services that perfectly anticipate your customers' needs
2. Amazon has patented what it called anticipatory shipping. Amazon needs to know how to predict, pack, and ship products before you place an order for them.

3. KONE Company provide predictive maintenance service after introducing sensors into its approximately one million lifts around the world.

Because these sensors monitor how the machinery is working, KONE can better:

- Predict and manage maintenance needs
- Provide a smoother maintenance service for customers

1.4. Using data to make more intelligent products

4 huge benefits of smart products

1. Making your customers' lives easier

Smart products in our home can gather information on what is going on around them and respond accordingly.

For e.g. a smart thermostat can heat your house to the perfect temperature without you having to program it.

Today's consumers expect smart solutions to a whole host of everyday tasks and activities including:

- Changing their baby's nappy
- Training for a marathon

2. Building better products

Using data to build a smart product is a fantastic way to build a more in-depth understanding of your customers. This knowledge can and should feed into your product design.

How does the process work?

- By building data processing capabilities into your products, you have the ability to collect masses of information on your customers' habits and preferences:
 - How do they use your product?
 - How often do they use it?
- All this data can be used to improve product design and develop new products that better meet your customers' needs.

3. Responding to customers' needs quicker

- The customer journey is sped up thanks to our permanent attachment to our phones. Google call these brief *I want to know/ I want to do/ I want to buy/ I want to go/ I want to learn* Micro-moments and they are becoming a vital part of marketing.

- As consumers, we increasingly expect brands to respond to us instantly and offer us exactly what we want.

4. Adding new revenue streams

Smart products often enable add-on, data-driven services.

For example, Apple has transitioned from a straight product manufacturer into a provider of music, TV, and fitness streaming services.

So there is a huge crossover between smart products and smart services. If you can make your product smarter, it could pave the way for a lucrative move into smart services.

1.5. Using data to improve your business processes

7 ways data is transforming everyday business processes for the better

1. Improving meetings

AI can help us cut down the tiresome admin involved before, during, and after meetings.

For example:

- Voice assistants such as Google Duplex can schedule appointments for you.
- Voicera's Eva Assistant can listen in on your meetings, capture key highlights and actions, create and share actionable notes afterwards.

2. Enhancing sales and marketing

Many of the PEC CRM solutions now incorporate data analytics, enabling sales teams to automatically create valuable insights.

- Salesforce's Einstein AI technology tool can predict which customers are more likely to generate more revenue and which are most likely to take their business elsewhere.

3. Assessing and improving customer service

Automation help companies automatically assess the quality of human customer service call.

- The Transcosmos AI solution can automatically assesses the quality of service given at speed and with human accuracy.

The AI can detect inappropriate and problematic customer service with more than twice the accuracy of a voice recognition system.

4. Improving product development processes

Generative design uses data to generate multiple designs from a single idea and avoids the expense of creating prototypes that do not deliver.

5. Automating content generation

Thanks to data and AI, machines are now capable of generating, engaging informative text to the extent that organizations like Forbes are now producing articles with the help of AI.

6. Enhancing the manufacturing process

The latest generations of robotic systems are capable of working alongside humans and interacting seamlessly and safely with the human workforce.

Thanks to data analysis and AI technologies like computer vision, collaborative robots (cobots) are aware of the humans around them and can react accordingly.

7. Refining recruitment

Unilever partnered with AI recruitment specialist *pymetrics* to create an online platform capable of conducting an initial assessment of candidates.

Thanks to this automated screening of candidates, around 70,000 person-hours of interviewing and assessing candidates have been cut per year.

2. Monetizing your data

There are two main approaches to data monetization:

- I. Create extra value for your organization using data.

Example: Facebook

They improve advertisers' experience by leveraging user data and providing targeted advertising services.

- II. Sell data to your customers or other interested third parties.

Example: Visa

They collect transaction data and then sell insights from that data to retailers.

3. Defining your data use cases

When organizations develop their data strategy, there are two fundamental steps:

- I. Identify the potential applications (or use cases) of data in your business.
 - a. Consider what your business is trying to achieve.
 - b. Consider the unique challenges your business is facing.

- c. Identify potential solutions through the use of data. Explore the many ways in which data could help your organization achieve its key strategic goals.

II. Begin to whittle use cases down to just a few top priorities.

How would the use of data help the business achieve its objectives, grow and prosper?

Examples:

- Solve a business-critical problem.
- Answer a specific business question.

Note:

- I. You need to rank your use cases order of their strategic importance to the business.
- II. You need to prioritize the use cases that represent the biggest opportunities for your business, or will help you solve your biggest business challenges.
- III. Identify one or two quick wins from your data use cases. That helps identify short-term, smaller data projects that are relatively quick and easy, and inexpensive to implement.
- IV. Identify and prioritize data use cases at least once a year.
- V. Look for common themes, overlaps, and complementarities across the data use cases in terms of:
 - Data requirements
 - Data governance

- Technology
- Skills and capacity
- Implementation and change management issues

4. Sourcing and collecting data

To collect the right data:

1. Automation is key.
2. There is not one type of data that is superior to other types.
3. Find the right type of data that fits your needs.
4. It is not easy to identify the type of data that will move the needle in your case.

Different types of data:

1. Structured data

Easiest to gather and analyze

2. Unstructured data

Can be more expensive to acquire or collect

In order to fully meet your strategic information needs, you need to be creative and combine different types of data.

Example:

Structured internal data (sales) + Structured external data (demographics)

Unstructured internal data (customer feedback) + Unstructured external data (social media analysis)

Structured, semi-structured, and unstructured data

Structured data	Semi-structured data	Unstructured data
<u>Features:</u> <ol style="list-style-type: none"> 1- Organized data. 2- Can be put in tabular form easily. 3- Easy to analyze 4- Can be stored in spreadsheets or databases 5- Easily discovered through search engines 6- A strict database format 	<u>Features:</u> <ol style="list-style-type: none"> 1- Have some elements of a structure. 2- It is a combination of structured and unstructured data. 3- Companies combine the analysis of structured data with the insights of unstructured parts of the data. 	<u>Features:</u> <ol style="list-style-type: none"> 1- Unorganized data. 2- Cannot be put in a tabular form easily. 3- Difficult to analyze in the past. 4- Represents 80-90 % of the world's data. 5- Can be stored in data lakes. 6- Not easily discovered through search engines. 7- Not easily categorized. 8- Rarely cover a single subject.
<u>Examples:</u> <ul style="list-style-type: none"> • Spreadsheets • Databases 	<u>Example:</u> A photograph taken on the phone (unstructured data) that is: <ul style="list-style-type: none"> • Time-stamped • Geo-stamped • Device-stamped And those are (structured data)	<u>Examples:</u> <ul style="list-style-type: none"> • Text files • Emails • Social media posts • Photos • Video footage • Sensor information

Example of advanced technology to analyze unstructured data:

Facebook is using deep learning technology in two tools:

- Deep face
To recognize faces
- Deep text
To recognize content, sentiment, emotions

Internal & external data

Internal data	External data
Owned or collected by the organization	All the information that exists outside of your organization
Can be structured and unstructured data	It is publicly available or privately owned by 3 rd parties.
Useful in the long run and collected for as long as necessary.	Dependent from 3 rd party specialists for access
Independent of a third party's willingness to share data with you.	Paid in most cases
Cheap or free to access	Less control of the way of collecting and the quality
Has a fixed cost or a decreasing cost over time	Less reliable for strategically important and business-critical insights
Very useful but limited	Unreliable in the long run (prevented access, raised cost)
Examples: <ul style="list-style-type: none"> • Sales data of your team (structured) • Video footage of your store (unstructured) 	Examples: <ul style="list-style-type: none"> • 3rd party agency spreadsheet (structured) • Twitter posts about your competitors (unstructured) • Social media posts

	<ul style="list-style-type: none">• Government census data
	Benefits: <ul style="list-style-type: none">• External data gives any business the capability to access and mine data for insights.• No storing, managing or securing data.

Notes:

- Do not rely entirely on internal data only because it creates biased views and would not give you a realistic picture of how your products are received.
- The sweet spot for best insights often comes from analyzing a mix of internal and external data.
- Most customer actions today leave a digital trace. This digital trace can and is being collected and analyzed.

Examples:

- Where did I access the data from?
- What sites did I visit?
- How did I interact with those sites?
- Which products did hold my attention the longest?

Computer algorithms can provide information in 2 main directions, content and tone.

Some of the more technically sophisticated retail stores are storing all the CCTV camera footage and analyzing it to better understand how their customers walk through the shops, where they pay attention, and for what duration.

The most tech-savvy are facial recognition algorithms to track individual shoppers' behaviors.

The biggest challenge when it comes to photo and video data is that they can create massive files, which can be tricky to store.

Sensor data:

- They are tiny devices that are attached to a product or device and transmit a particular type of data.
- Sensor data is an important source for very powerful insights to improve productivity and maintenance.

Examples of sensors in a mobile phone:

- GPS
- Accelerometer
- Gyroscope
- Proximity sensor
- Near field communication (NFC)

5. Metadata

What is Metadata?

- It is data that describes and summarizes information about other data.
- It helps IT systems uncover what users are looking for.
- One of the fastest-growing segments of enterprise data management.
- Drives digital asset management.

- Allows analysts to unlock meaning in big data.
- Allows data to be identified, discovered and associated across an enterprise.
- Without it, big data is unusable or unmanageable.

Streaming data

- Dynamic data that is generated continuously from a variety of sources.
- Each record needs to preserve its relations to the other data and sequence in time.

Examples:

- Social media feeds
- Sensors
- Cameras

Streaming analytics

- When analytics is brought to the data to generate insights while the data is in the "stream" instead of stored.
- Streaming analytics makes companies more agile and responsive to the realities in the environment where they operate.

Real-time and streaming data use cases

1. Maintenance: Identifying issues in real-time
Saves > time and money

2. Healthcare: live vital monitoring of patients

Examples> Glucometers, connected scales, heart rate monitor

3. Retail: live interaction with customers

Examples> location-based marketing, trend insights, operational efficiency improvements.

4. Social media: real-time monitoring of posts

Example> quickly taking action on offensive and fake news.

5. Finance: benefiting and securing finance

Examples> trading, monitoring transactions and identifying fraud

6. Energy and power: maintenance of equipment and understanding of consumer demand.

Example> optimizing fossil fuel

7. Personalization of products and services: better respond to consumer demand

Example> streaming services to recommend the next thing to watch

8. Transportation and supply-chain

Examples> internet of trains, safer autonomous vehicles

9. KPIs: proactive analysis

Example> making decisions based on real-time KPIs such as financial or operational performance data.

Streaming analytics platform components:

- Data collection
- Data processing
- Data management
- Data analytics
- Machine learning

Gathering internal data

Internal data can be collected from

- Systems
- Customers
- Employees
- Products

Examples of internal data that can be easily collected:

- Conversation data
- Photo and video data
- Transaction data (one of the most valuable sources of information)

It is useful in:

- Measuring sales
 - Monitoring stock levels
 - Predicting what you need to order
- Financial data

Accessing external data

Can be acquired for free from:

Government datasets (publicly available)

Social media platforms

- Text data
- Photo data
- Video data

Google trends

It helps you understand trends in your industry and anticipate changing customer needs.

Weather data

It can be used to forecast sales

Sources of external data

30 sources of useful external data

- 1- [World bank open data](#)
- 2- [International monetary fund](#)
- 3- [National center for education statistics](#)

- 4- [The UK data archive](#)
- 5- [Five thirty eight](#)
- 6- [FBI uniform crime reporting](#)
- 7- [Bureau of justice](#)
- 8- [Qlik data market](#)
- 9- [NASA Exoplanet Archive](#)
- 10- [UN comtrade database](#)
- 11- [Financial times market data](#)
- 12- [Google trends](#)
- 13- [Twitter](#)
- 14- [Google scholar](#)
- 15- [Instagram](#)
- 16- [Open Corporates](#)
- 17- [Glass door](#)
- 18- [IMDB datasets](#)
- 19- [Open library](#)
- 20- [Labeled faces in the wild](#)
- 21- [Microsoft Marco](#)
- 22- [Machine learning dataset repository](#)
- 23- [eBay market insights](#)
- 24- [National history museum data portal](#)
- 25- [CERN open data](#)

- 26- [One million audio cover images](#)
- 27- [Complete public Reddit comments corpus](#)
- 28- [Microsoft Azure data markets](#)
- 29- [Irish electric vehicle charge point status](#)
- 30- [London air](#)

What to do if the data you want does not exist?

Example: *Sprigg Company*

Problem:

Farmers in developing countries did not have access to the same data as their colleagues in developed nations.

Soil data. Soil sample analysis takes weeks. Need immediate insight on what crops to plant or how to treat an existing crop.

Solution:

Sprigg developed mobile test sensors with IoT devices for testing the soil remotely with immediate results. Then, the result is sent to the central repository for further analysis.

The firms which have:

- Vision and innovation culture

Can:

- Collect, analyze, and use never-before-collected data

Will have:

- Rights and competitive insights which benefit the business immensely

6. Data governance

- We need to be aware when working with data, especially personally identifiable information, requires some significant legal and regulatory compliance.
- A potential failure could mean serious consequences such as legal lawsuits, fines, and losing consumer trust.
- Regulations are being introduced to tighten up how companies collect, store and use data.
- When working in practice, specialist legal advice is highly recommended.
- Businesses are successful with insights obtained from third-party data.
- When building a business with key business processes relying on data, it is crucial that you are the owner of that data.
- Make sure you own the data for
 - ✓ Important business operations
 - ✓ Critical decision making
 - ✓ Monetization

Data ownership fundamental aspects

1. Own any data that is necessary to your business
2. Have the necessary rights and permissions to collect and use the data.

Ensuring the correct rights are in place

Best practice for verifying data rights

It is important to have metadata which can include

- When and where was the data collected?
- What permissions were given?

General Data Protection Regulation (GDPR)

- It is a new legislation by the EU to protect personal data.
- It defines standardized data protection laws.
- For companies failing to comply, a fine of up to 20 million euros is applied.

Remember:

- Private data has to be protected.
- Private data must be used for the purpose for which it was handed over.
- Communicate to users what data you are going to collect from them and how it will be used.

How to develop consumer trust?

Example: The Royal Bank of Scotland

- They developed a new customer data strategy called "Personology".
- Its aim is to give some real value back to customers and show them what they do with this data.
- They use big data tools to comb through all customers' transactions to gain valuable insights for their customers.

7. Turning data into insights

7.1. Text analytics

Text data

- It is unstructured data.
- In the past, text data did not fit with relational databases and spreadsheets.

Examples:

- Journals
 - Blogs
 - Emails
- Today, models are able to access text for commercially relevant patterns.

Ways to perform text analytics:

1. Text categorization
2. Text clustering
3. Concept extraction
4. Sentiment assessment
5. Document summarization

Example in real life: Private equity and venture capital industry

PE and VC funds are using text analytics to uncover good investment opportunities based on text data available on the internet.

7.2. Sentiment analytics

- The goal of sentiment analysis (opinion mining) is to gain an idea about the attitude towards a particular topic.
- It allows businesses to understand better how their stakeholders feel in a particular situation.
- Social media uses sentiment analysis to make refined targeting and improve advertisers' return on investment.

7.3. Image analytics

- It is the analysis of information contained in photographs, medical images, or graphics.
- A digital photo contains information about:
 - Location
 - Time
 - Identity
 - Action
- This data can give very valuable insights but only if it answers the right strategic questions.

7.4. Video analytics

- It is the analysis of information contained in video footage.
- A few years ago, video data was too expensive to store on a company's servers.

- The advent of cloud technologies democratized access to storage and has made video footage data easier to store.
- Video analytics allows the calculation of conversion rates for different items clients look at and considered purchasing.

Example: Retail shops

Retail shops can collect video footage of client behavior.

7.5. Voice analytics

- It obtains information from audio recordings of spoken content.
- By analyzing large volumes of audio data, we can categorize the topics discussed and the emotional state of the speakers.
- Voice analytics needs to be performed with great caution by authorities.
- By applying large-scale voice analytics, the firm will also benefit from:
 - Reduced management costs.
 - More efficient customer service

7.6. Data mining

Techniques and methods used to apply data mining include

- Artificial intelligence
- Statistics
- Database systems
- Machine learning

The three phases of data mining:

- I. Initial exploration
- II. Model building and validation
- III. Deployment

Note:

The interpretation and business logic of the results from data mining requires additional analysis, some context and consideration.

7.7. Business experiments

To avoid bad decisions, businesses apply techniques such as

- Experimental design

It consists in applying a change in one part of the organization and comparing the results against other parts of the organization where the change has not been applied yet.

- A/B testing

A particularly popular approach in digital marketing testing one version of a page/functionality against another. Through constant testing, marketers can:

- Refine different aspects of the customer journey.
- Optimize conversion rates.
- Achieve better return on investment.

7.8. Visual analytics

- Visualization is the most efficient way to communicate vast quantities of information.
- Using a visual helps our brain process and understand information more quickly.
- Visualization enables business analysts to:
 - Explore data without spending too much time on numbers.
 - Extract valuable insights.

7.9. Correlation analysis

- A statistical technique aimed at establishing whether a pair of variables are related.

For example:

The extent to which a price relates to a quantity sold for a particular product.

Whether job applicants' scores are related to future employee performance.

- Correlation establishes a statistical relationship but does not prove causation.
- Correlational coefficient is a key output in correlation analysis. It provides an indication of the strength of the relationship between variables.

Correlation analysis advantages

1. Low cost
2. Simple to understand
3. Gives you great insights

Correlation analysis disadvantages

1. Does not prove causation by itself.
2. Needs to start with a pre-defined set of variables.

7.10. Regression analysis

- Regression analysis is the best method to study whether a dependent variable is driven by a set of independent variables.
- It is highly beneficial to consider additional datasets as this might uncover a new perspective of the current situation.

Example:

Regression analysis can be used to understand the profitability of a certain market segment or a certain product.

7.11. Scenario analysis

- Known as horizon analysis or total return analysis, it is an analytical framework that consists of building a model that shows how a final output would look like in several different cases.
- It outlines future outcomes based on different factors and assumptions.
- It allows decision-makers to appreciate the impact of certain variables and to see what the expected final outcome is in a given state of the world.
- It can be used to assess the possible realizations of different strategic choices or to create a combination of scenarios.

Trajectory of revenue

It is one of the most important drivers that determine company value. It has 3 scenarios:

- Best case scenario
- Base case scenario
- Worst case scenario

Scenario planning is a five-stage process

- I. Define the problem
- II. Gather the data
- III. Separate certainties from uncertainties
- IV. Develop scenarios
- V. Use the outcome in your planning

7.12. Time series analysis

- It is a technique that allows us to assess changes over time or predict future outcomes based on what has happened in the past.
- It allows us to quantify the impact of management decisions on future outcomes.
- It relies on past data and the data of a certain variable in the past to predict the future.

Forms of time series analysis

- Naïve time series
- Autoregressive models
- Moving average models

Where to use time series analysis

- Statistics
- Pattern recognition
- Mathematics
- Finance
- Weather forecasting
- Earthquake prediction

7.13. Monte Carlo simulation

- It is a mathematical problem-solving and risk-assessment technique that approximates the probability or the risk of a certain outcome using computerized simulations of a random variable.

Advantages:

1. Fully explores the probability distribution of a certain variable.
2. Helps us manage risks.

7.14. Linear programming

- A math technique used to find an optimal solution to a problem that has several known constraints.
- Helps businesses solve problems such as maximization of a function.
- Helps companies allocate their resources in a manner that maximizes revenue or reduces cost.

7.15. Cohort analysis

- The process of analyzing a group's collective behaviour over time.
- The most frequently used in customer analytics.
- By analyzing data points that share certain common traits, analysts can gain a better idea of that group's behaviour.

7.16. Factor analysis

- A collection of techniques used for data reduction and structure detection when working with big datasets and a large number of variables.
- Helps us study the relationships between variables to identify the variables of common factors.

7.17. Neural network analysis

- Artificial neural networks are computer programs modelled on the human brain and the way humans learn.
- Artificial neurons are arranged in a series of layers:
 - Input layer (receives information)
 - Hidden layers (transform information)
 - Output layer (uses information)

7.18. Deep learning

- Deep learning is a subset of machine learning. Where ANNs, algorithms inspired by the human brain, learn from large amounts of data.
- The deep learning algorithm would perform a task repeatedly, each time tweaking it a little to improve the outcome.
- Deep learning allows machines to solve complex problems even when the dataset is very diverse, unstructured, or inter-connected.
- The more deep learning algorithms learn, the better they perform.

7.19. Reinforcement learning

- It is a form of machine learning where the optimal behavior or action is reinforced by a positive reward.
- Machines use reinforcement learning algorithms to determine the ideal behaviour based upon feedback from the environment.

- Depending on the complexity of the problem, reinforcement learning algorithms can keep adapting to the environment over time to maximize the reward.
- Reinforcement learning requires a lot of data.

8. Creating the technology and data infrastructure

8.1. How to collect data

Data source

- The way data arrives at your company
- Data can be structured, unstructured, internal or external
- Data can be collected in a variety of sophisticated ways. for e.g. sensors

Different data acquisition methods

- Customer records
- Sales information
- Bank statements
- Accounting entries
- Extractions from an ERP system
- Employee information
- Customer and employee satisfaction reports
- Social media channels
- Emails
- Video recordings of operations

- Data collected with sensors and other equipment
- E-commerce data

Popular data collection techniques

- Sensors
- Applications
- CCTV video
- Beacons
- Website cookies
- Social media images

Note:

The way you intend to use the data your company is collecting will also predefine whether it is a good idea to outsource or not.

Real-life example > *Transport for London*

Their data sources:

- Ticket systems
- Traffic signals
- Sensors attached to vehicles
- Customer feedback
- Social media channels

Database

- A structured repository that contains structured data in rows and columns.
- Example: spreadsheets

Data warehouse

- A large repository where data is stored in files that are well organized.
- Data is only loaded into the warehouse when use for the data has been identified.

Data mart

- A subset of a data warehouse.
- Contains data that is designed for a particular part of an organization or a particular purpose.

Data lake

- A way of storing large volumes of data in their natural un-curated state.
- A data lake holds data in an unstructured way.
- It accepts and retains all data from all data sources, supports all data types and schemas are applied only when the data is ready to be used.

<u>Point of comparison</u>	Data lake	Data warehouse
1. Data	<ul style="list-style-type: none"> • Retains all data 	<ul style="list-style-type: none"> • Contains only processed data
2. Agility	<ul style="list-style-type: none"> • More flexible • Can be configured and reconfigured as necessary 	<ul style="list-style-type: none"> • Structured but difficult to be changed
3. Users	<ul style="list-style-type: none"> • Data scientists 	<ul style="list-style-type: none"> • Business users and analysts
4. Security	<ul style="list-style-type: none"> • More vulnerable 	<ul style="list-style-type: none"> • More secure

8.2. How to store data

Companies can choose between two options to store data:

- I. Building in-house server infrastructure
- II. Relying on external storage systems (storing data in the cloud)

Cloud solutions are designed with the idea to democratize data storage, where you pay as much as you use.

Distributed storage

- Using cheap, off-the-shelf components to create high-capacity data storage, which is controlled by a program that keeps track of where everything is and finds it upon request.
- The 2 terms “distributed storage” and “cloud storage” are often used in the same context.

Note:

One of the strongest advantages of cloud computing is the usage of the power of lots of different machines to perform tasks.

8.3. How to process data

Open-source frameworks

- Widely available today for data analysts and data scientists.
- Helps avoid heavy investments in expensive bespoke data analysis infrastructure

Data set acquisition stages

- Cleaning or pre-processing the data.
- Building the analytics model.
- Extracting meaningful conclusions from the data analysis.

MapReduce tool

A method for analyzing data where you select the elements of the data that you want to analyze and put it into a format from which insights can be gleaned.

Big suppliers of analytics:

- ❖ IBM
- ❖ Oracle
- ❖ Google (Big Query)
- ❖ Cloudera
- ❖ Hadoop

8.4. Communicating data

Business dashboard

A simple visual display of the most important information that decision-makers need to help them achieve objectives.

Operational dashboard

- Allows monitoring of day-to-day processes and outputs of your business.
- Provides information that allows you to fix issues before they become problems.

Strategic dashboard

Looks to the future and seeks to identify obstacles and challenges that you may face on the way to achieving your strategic goals.

KPIs dashboards

- ❖ A great way to communicate insights from key performance indicators to the people that need them.
- ❖ Give decision-makers quick access to the critical indicators or instruments of the business.

5 tips on how to build an excellent dashboard:

1. Make sure your dashboard fits on a single screen or page.
2. Include the most critical and insightful KPIs necessary for achieving your operational and strategic objectives only.
3. Choose an appropriate and accessible way to display the dashboard.
4. Make the dashboard easy to look at, navigate and understand.
5. Focus on information delivery and understanding.

Reporting lessons from journalists

- 1- Create headlines (this is what the data is telling us).
- 2- Use clear data visualizations (turning data into something we can understand)
- 3- Bring a narrative (to capture the essence of the story)

Note:

- Software can bring key performance indicators to life through powerful communication and collaboration features.
- Most of the tools available today are web-based, so you can access the information from your internet browser anywhere at any time.
- Software can therefore make ongoing data crunching, reporting and communication much easier.

9. Building the data competencies in your organization

9.1. Skills shortage

- The supply of capable data scientists is unable to catch up with the demand.
- Small and mid-size businesses find it hard to build a competent team of professionals. They have limited resources and cannot attract people by relying on big brand names.
- The data scientist's role is poorly defined. There're a bunch of roles in the field that have different meanings in different companies.

A true data scientist is someone who has:

- Domain knowledge expertise in the field of data & programming
- Business skills
- Analytical skills

The number of data science students has increased which improved the supply of prepared data professionals for businesses.

Skills that every data-focused team needs to process to turn data into insights

- Business skills

Understand how the business creates value for its customers and which are the main strategic goals for the organization.

- Analytical skills

A data scientist should be able to investigate cause and effect, reason with an open mind, and spot data patterns.

- Computer science proficiency

Data collection, storage, analysis, and communication are done with computers.

- Solid statistics and mathematics understanding

A data scientist needs to be able to define relevant populations, sample sizes, acceptable confidence levels, and apply machine learning algorithms.

- Creativity

Since data science is an emerging field, there are no hard and fast rules about what a company should use big data for. So, having an open and creative mind goes a long way.

9.2. Building internal skills and competencies

Problem

It is very difficult to find candidates who have all five skills described previously.

Solution 1

Recruit different professionals who have some of these skills

Solution 2

Hire data professionals who do not have all 5 skills and train them internally with the hope that they will be able to learn fast enough.

Note:

The most valuable trait of an employee is the ability to desire and grow

Solution 3

Identify individuals within the organization who possess some of the five core skills.

This can be a preferable option compared to recruiting someone from the outside

9.3. Outsourcing your analysis

When building a data science team, you can hire in-house or external services.

When to outsource:

- If you are struggling to recruit well-prepared professionals at a reasonable price.
- When you do not have any subject matter experts in-house.

Large companies

- Work with large datasets
- Have significant know-how from past projects

Small companies

- Specialize in a particular industry
- Have a competitive edge against bigger service providers

To make a good decision when outsourcing:

- You should be as informed as possible.
- You should get reviews on the consultancy from other companies.
- Ask for previous case studies to see how the company added value to their clients.
- Make sure the external partner understands the industry.
- Make sure you define your deadlines.

Note:

Working with external experts is a good opportunity to train your team. As this will give you the possibility to rely on an in-house solution after a while.

A good leader in the intelligence revolution

- I. Is aware of the intelligence revolution and understands its importance.
- II. Approaches data analysis and AI strategically and not technically.
- III. Identifies how to best use AI and data analysis in their business.
- IV. Finds a strategic sponsor and strong influencer for AI and data analysis initiatives.

Chief intelligence officer:

- Communicates with the AI/data teams, the leadership and the business teams.

- Sets the vision for using AI
- Educates about the importance of AI
- Places the ethical frameworks for AI usage
- Build the right skills, capacity and technology infrastructure
- Oversees execution and delivery
- Manages stakeholders

10. Executing and revisiting your strategy

10.1. Putting the data strategy into action

Harmful attitudes that kill a data strategy

- “We are not a data company”

Today, every business is a data business. Every industry can learn more about its clients, revenue patterns, demand, and improvements.

- “Working with data and implementing a data strategy in practice is too expensive”

Today, even small businesses can access cloud services and open-source software that make it relatively inexpensive to work with data.

- “We already have more data than we need”

The trick is in having the right data and using it to obtain insights that will help you achieve your goals.

- “Everyone else is already ahead of us”

You do not know how far ahead your competitors are. They might only be in the exploratory stages of implementing their data strategy or they might be focusing on the wrong use cases.

- “Our customers are not asking for it”

Customers are definitely interested in products and services that are more intelligent, customized to their needs and cheaper.

10.2. Why data strategies fail

A strategy is:

- The plan by which a company achieves its goals.
- The principles of executing a data strategy are broadly the same as executing any other strategy.
- The data strategy is the framework that puts together the different building blocks of performing data analysis as an entity. It tells you how to get from point A to point B.

Potential reasons why data strategies fail:

- Lack of communication within the organization.
- Lack of communication between departments within the company.
- A potential remedy for this is the introduction of end-to-end process owners who are cross-functional leaders taking responsibility for interdivisional communication.
- Lack of specialized skills at the right time.

Prevention measures:

- Good communication
- High level of buy-in across the organization

10.3. Creating a data culture

How to establish a strong data culture?

- Involve some of the brightest employees during the process of forming a data strategy and implementing it.
- Negativity can be contagious. If people on your team become skeptical and pessimistic, try to understand what problems they have in their work and how data can be helpful to make their job easier.
- Build trust by being transparent about the data you are collecting and the reasons for it.

Two strands to consider when revisiting your data strategy:

- Do you have different business needs?
- Has technology been evolved?

How often should your company review its data strategy?

- It depends on your industry and the way you use data.
- The faster your industry changes, the bigger role data plays in your business, and the more often should you review your data strategy.

Three different ways data strategy can evolve:

- Some questions are one-offs, and once you have collected and analyzed the data and you act on the insights, no further ongoing data analysis is required.
- Other questions will be around ongoing issues that you want to continue to measure and monitor.
- Cases where your data for your initial questions, throws open new questions, or where they will push you in a new direction, with new data and analytics needs.

Notes

- Every few years, companies need to assess whether their current infrastructure is in line with industry standards and upcoming trends.
- Investments in new technology can be seen as a burdensome capital expenditure but it often allows firms to reduce operating costs.
- Existing infrastructure is not going to be able to cope with the amount of data that will need to be transmitted if the IoT continues to grow with this trajectory.

Key technological advances that shift the data strategy of organizations:

- Cloud computing
- Edge computing
- Blockchain technology
- Machine learning
- Internet of things
- Affective computing

- Virtual reality
- Cognitive computing
- Robotics

Using data for good

- Identify and diagnose diseases
- Allow us to develop vaccines in month
- Helps us address climate change
- Helps us eliminate hunger
- Identify and stop the spread of disinformation, fake news and online bullying
- Helps us make our world more equal and less discriminatory

Learn DATA SCIENCE anytime, anywhere, at your own pace.

If you found this resource useful, check out our **e-learning program**. We have everything you need to succeed in data science.

Learn the most sought-after data science skills from the **best experts in the field!**
Earn a **verifiable certificate** of achievement trusted by employers worldwide and future proof your car



Danielle Thé
Esade Ramon
Llull University



Bernard Marr
Cambridge University



Tina Huang
University
of Pennsylvania



Ken Jee
DePaul University



**Anastasia
Kuznetsova**
Université
Côte d'Azur

Comprehensive training, exams, certificates.

- ✓ 160+ hours of video
- ✓ 599+ Exercises
- ✓ Downloadables
- ✓ Exams & Certification
- ✓ Personalized support
- ✓ Resume Builder & Feedback
- ✓ Portfolio advice
- ✓ New content
- ✓ Career tracks

Join a global community of 1.8 M successful students with an annual subscription
at 60% OFF with coupon code **365RESOURCES**.

~~\$432~~ **\$172.80**/year



Start at 60% Off

VAT may be applied



Bernard Marr

Email: team@365datascience.com

365[✓]DataScience