

Data basics

INTRODUCTION TO DATA



Maarten Van den Broeck

Senior Content Developer at DataCamp

Data is everywhere

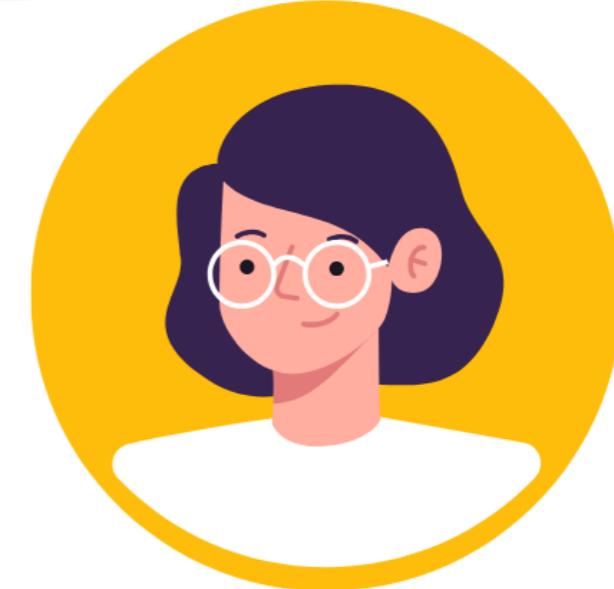


Data is everywhere

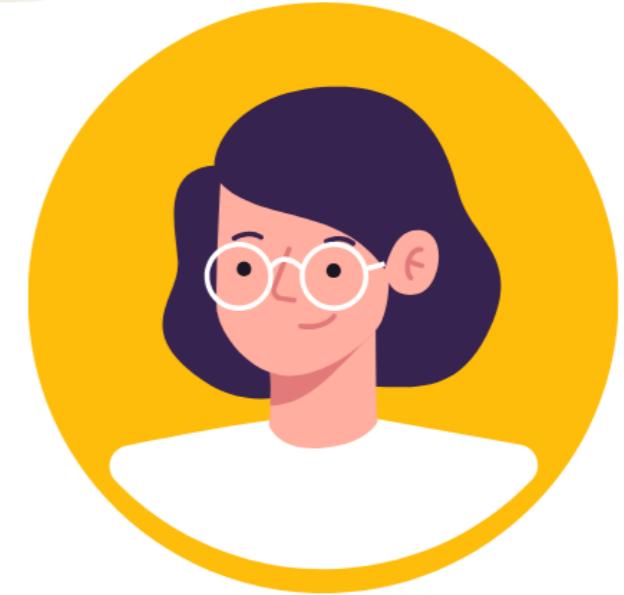
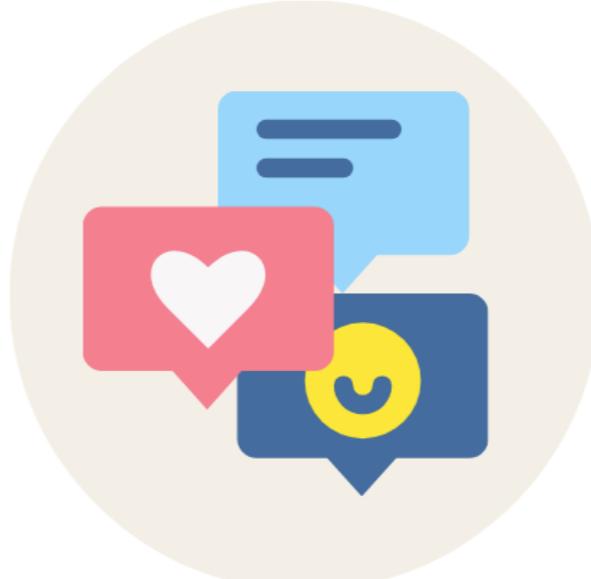
Name

Age

Hobbies



Data is everywhere



Data is everywhere



Data is everywhere



Data is everywhere



What is data?

- Derived from **datum**: *given, fact*



What is data?

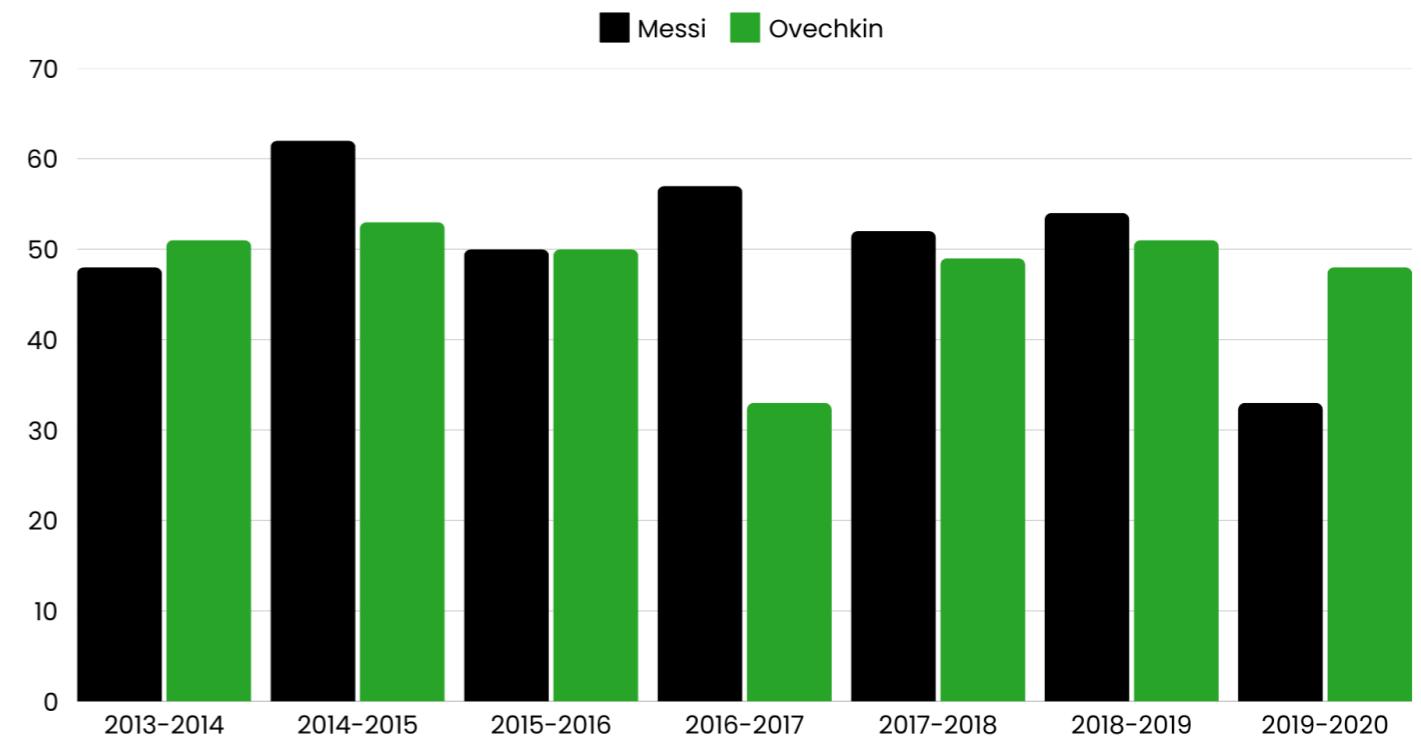
- Derived from **datum**: *given, fact*
- Valuable resource in this digital era¹



¹ The Economist, May 6th 2017: The world most valuable resource is no longer oil, but data

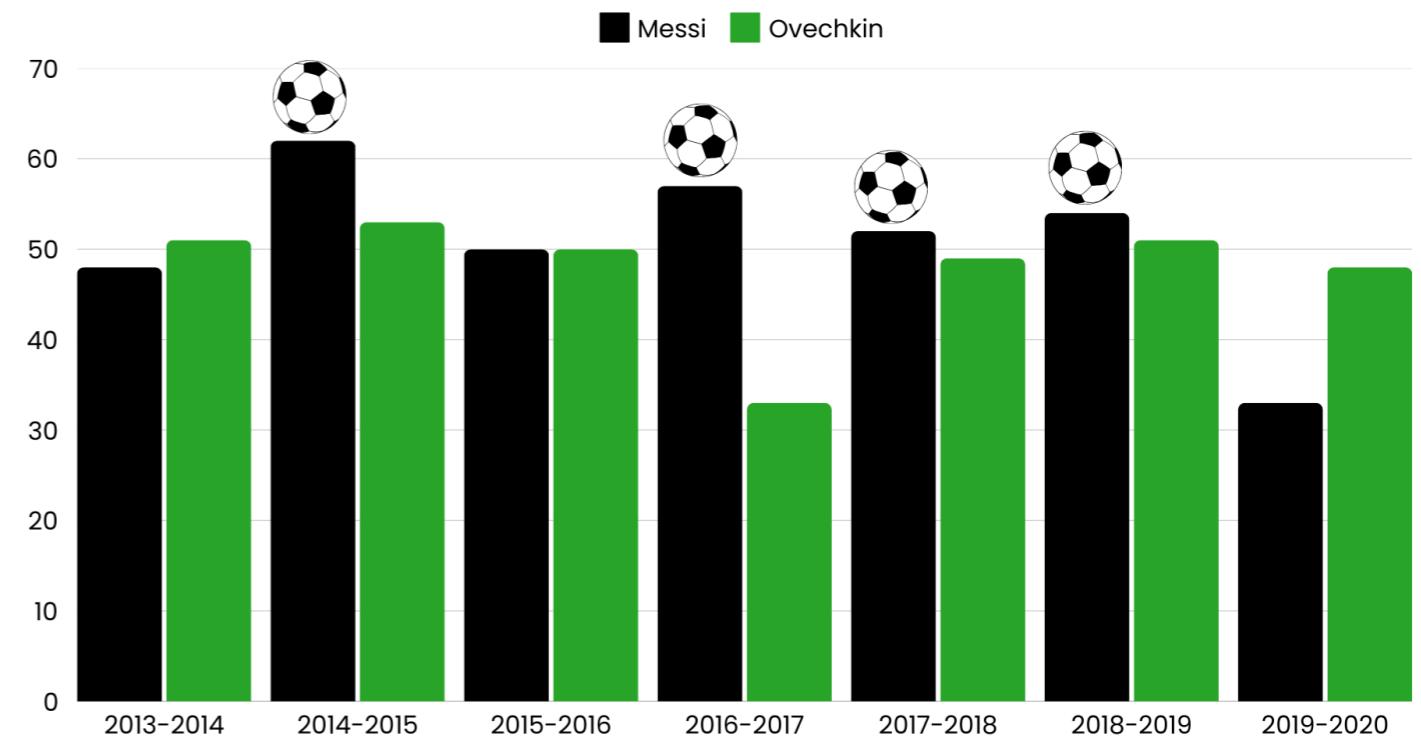
Data context

- Who is a great player?
 - Lionel Messi
 - Alexander Ovechkin



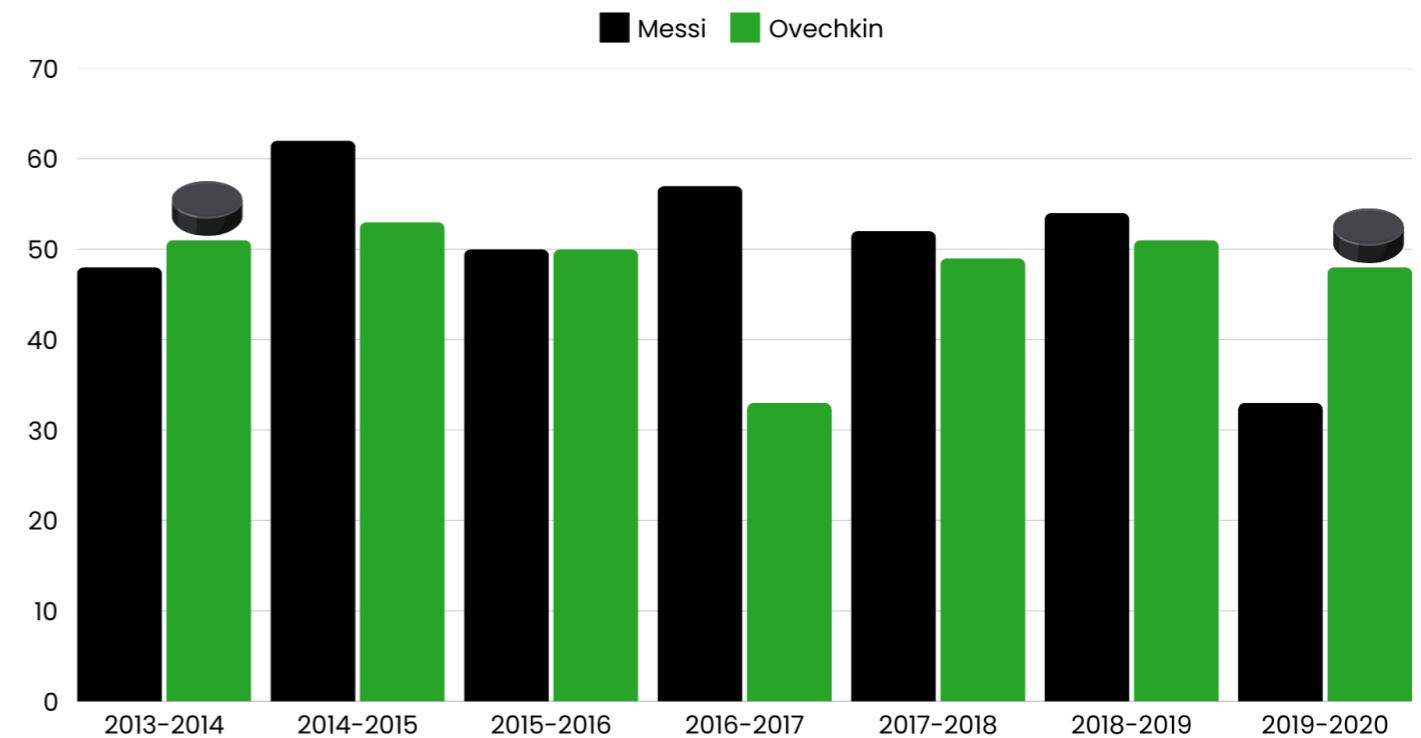
Data context

- Who is a great player?
 - Lionel Messi
 - Alexander Ovechkin



Data context

- Who is a great player?
 - Lionel Messi
 - Alexander Ovechkin

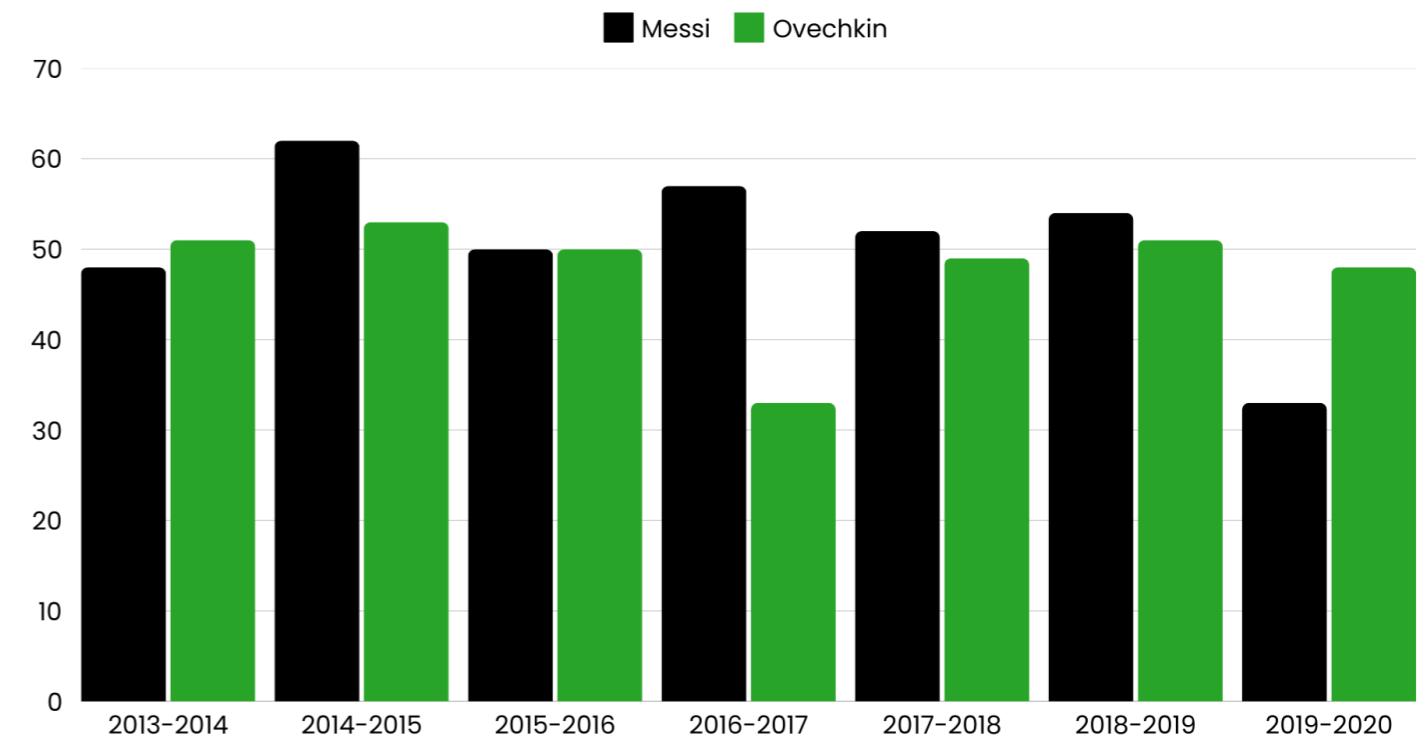


Data context

Information that provides meaning to data

- When the data was collected
- Where the data was collected
- ...

These characteristics of the data are called
the **metadata**



Types of data

Unstructured:

- Football match video
- Without labels or order

Structured

Structured:

- Table listing goals, times, players
- Organized and easier to analyze

Unstructured

Quantitative

Qualitative

Structured

- Common in spreadsheets
- Easy to filter and analyze

Examples:

- Sales records
- Employee attendance
- Weather data

Sales records

| ID | Product | Sales |
|----|---------|-------|
| 1 | T-shirt | 15 |
| 2 | Jeans | 2 |
| 3 | Shoes | 3 |
| 4 | Jacket | 1 |
| 5 | Hat | 5 |

Unstructured

- Harder to analyze
- Needs processing

Examples:

- Videos
- Interviews
- Pictures



Quantitative

- Also called numerical data
- Ideal for calculations and visualizations

Examples:

- Points scored
- Height
- Temperature

Qualitative

- Also called categorical data
- Useful for spotting patterns

Examples:

- Favorite sports
- Customer feedback

Let's recap

- **Structured:** organized and easy to analyze
- **Unstructured:** complex but insightful
- **Quantitative:** numerical and ideal for calculations
- **Qualitative:** describes categories and reveals trends

Structured

Unstructured

Quantitative

Qualitative

Let's practice!

INTRODUCTION TO DATA

The curious case of data growth

INTRODUCTION TO DATA

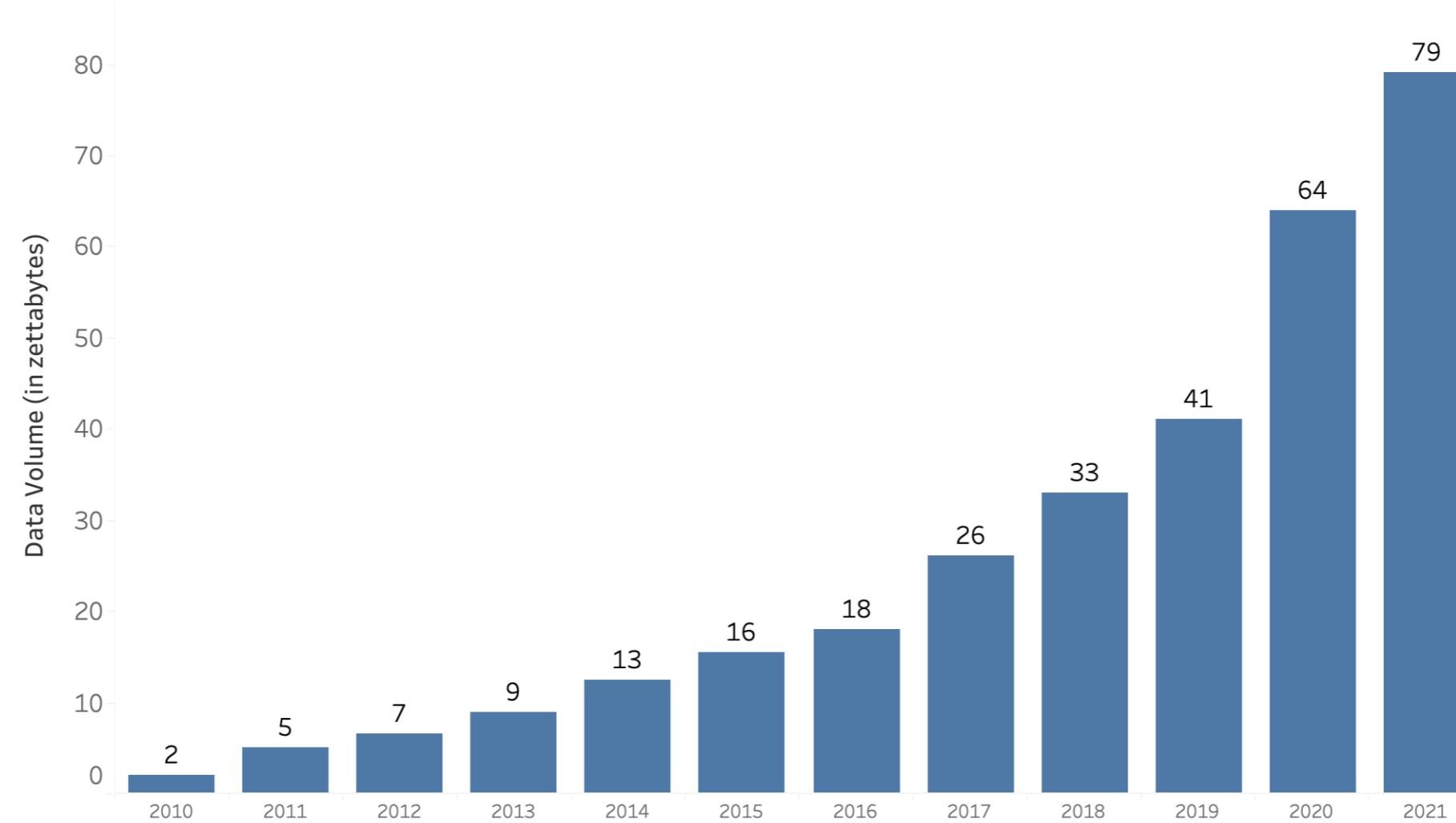


Maarten Van den Broeck

Senior Content Developer at DataCamp

The volume of data has grown exponentially

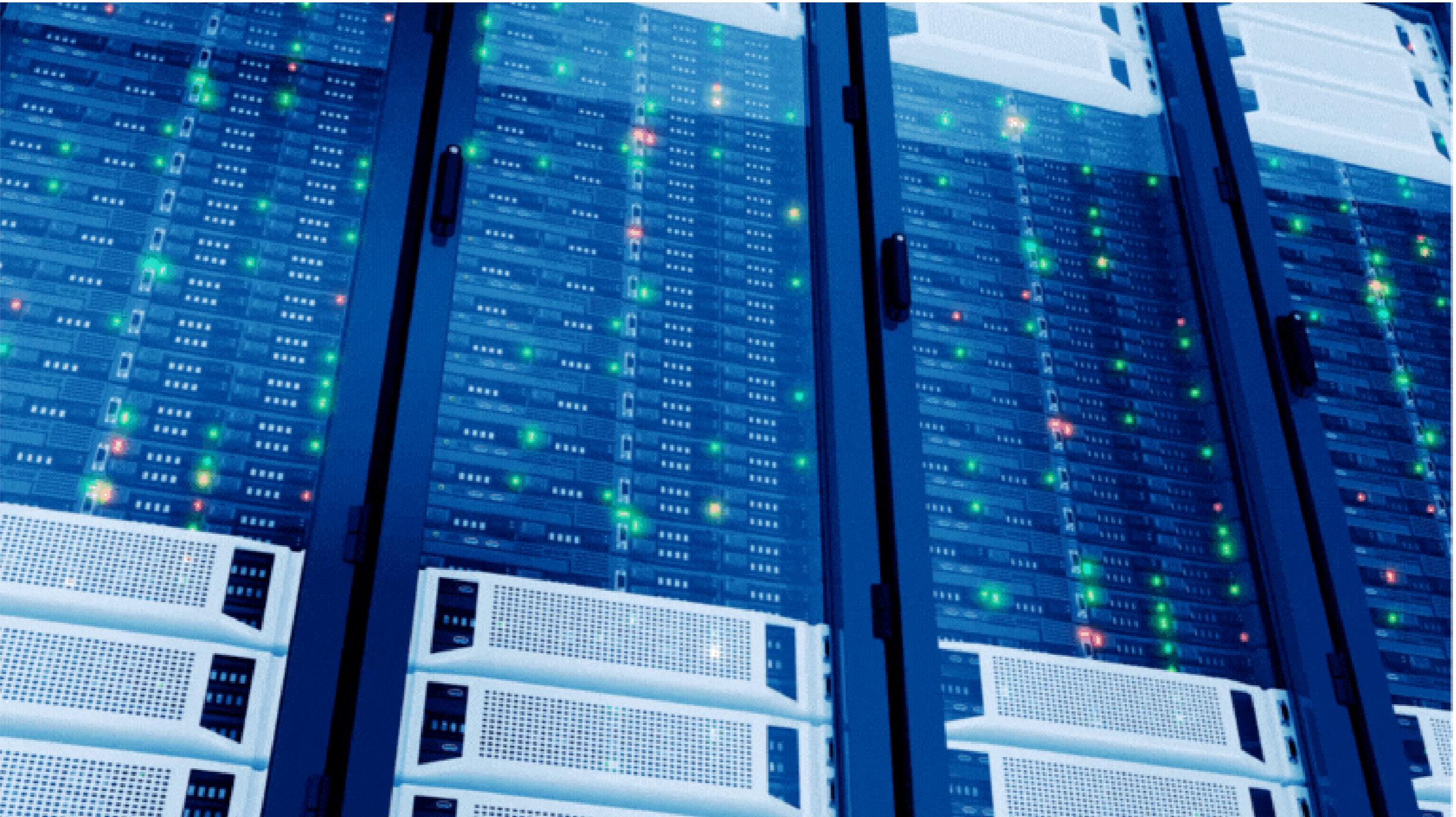
Volume of data/information created, captured, copied, and consumed worldwide by year



1 zettabyte = a one followed by 21 zero's in bytes = 1 billion terrabyte

¹ Source: Statista

Data storage is changing



Data storage is changing

Historical data storage

- Genetic information in DNA
- Cave and wall paintings



Data storage is changing

Historical data storage

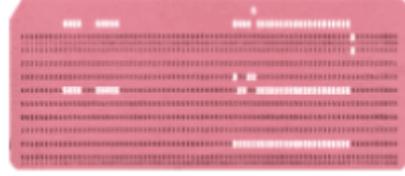
- Genetic information in DNA
- Cave and wall paintings
- Scrolls and books of papyrus/parchment



Data storage is changing

Historical data storage

- Genetic information in DNA
- Cave and wall paintings
- Scrolls and books of papyrus/parchment



19th and 20th century

- Punch cards

Data storage is changing

Historical data storage

- Genetic information in DNA
- Cave and wall paintings
- Scrolls and books of papyrus/parchment

19th and 20th century

- Punch cards
- Magnetic tape, floppy disks



Data storage is changing

Historical data storage

- Genetic information in DNA
- Cave and wall paintings
- Scrolls and books of papyrus/parchment

19th and 20th century

- Punch cards
- Magnetic tape, floppy disks

20th and 21st century

- More data on smaller media



Data storage is changing

Historical data storage

- Genetic information in DNA
- Cave and wall paintings
- Scrolls and books of papyrus/parchment



19th and 20th century

- Punch cards
- Magnetic tape, floppy disks

20th and 21st century

- More data on smaller media
- CDs and hard/solid state drives (local)

Data storage is changing

Historical data storage

- Genetic information in DNA
- Cave and wall paintings
- Scrolls and books of papyrus/parchment



19th and 20th century

- Punch cards
- Magnetic tape, floppy disks



20th and 21st century

- More data on smaller media
- CDs and hard/solid state drives (local)
- Data centers (cloud)



Data - where does it come from?

Ice cream shop #1 in New York



Ice cream shop #2 in New York



- Sells vanilla, chocolate, and strawberry
- Has a rough idea of sale transactions

- Sells 20+ ice cream flavors
- Also sells coffees and milkshakes
- Tracks all sales

Capturing data

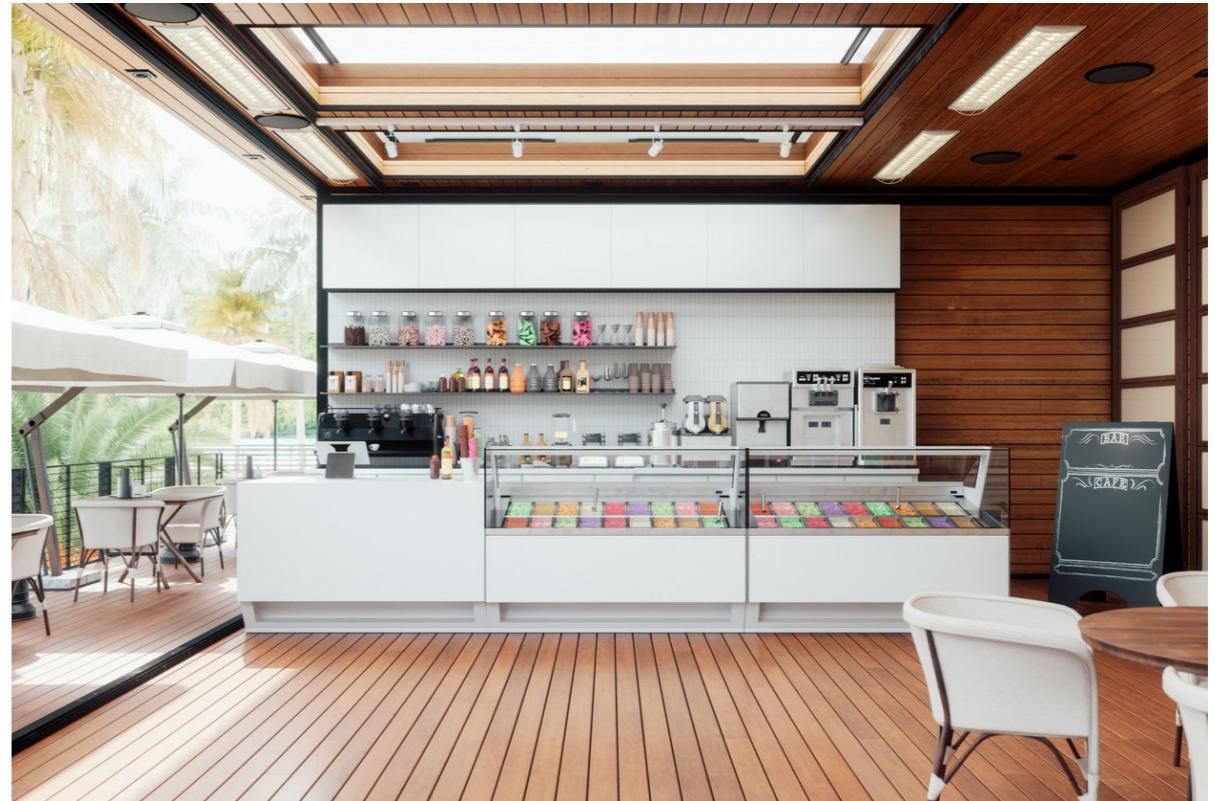
Data captured

- Sales per product type and ice cream flavor
- Stock per product type and flavor
- Weather data

Optimizations

- Avoid popular flavors being out of stock
- Replace poor selling flavors with new ones
- Predict sale spikes due to high temperature
- Optimize prices

Ice cream shop #2 in NY, USA

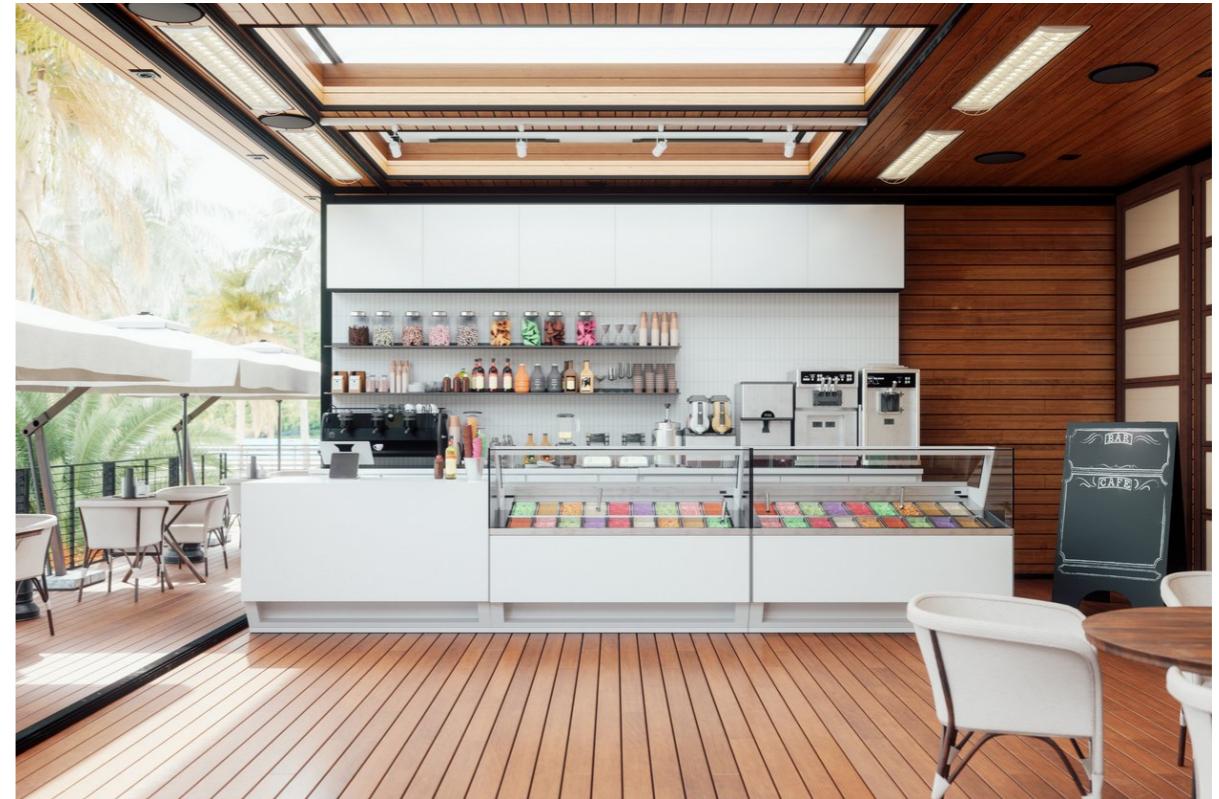


Which ice cream shop would fare better?

Ice cream shop #1 in New York



Ice cream shop #2 in New York



- Uses gut feeling to make decisions
- Randomly switches ice cream flavors

- Uses data to make decisions
- Searching for the best flavors

Companies are more complex than ice cream shops

3D Manufacturing companies

- Beam heat
- Layer thickness
- Structural stability



Financial institutions

- Mortgage applications
- Fraud detection



Let's practice!

INTRODUCTION TO DATA

The value of data

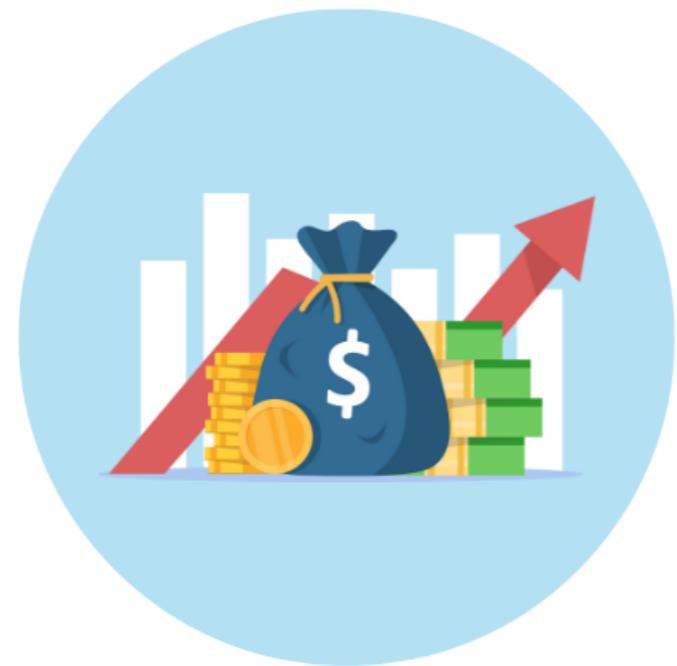
INTRODUCTION TO DATA



Maarten Van den Broeck

Senior Content Developer at DataCamp

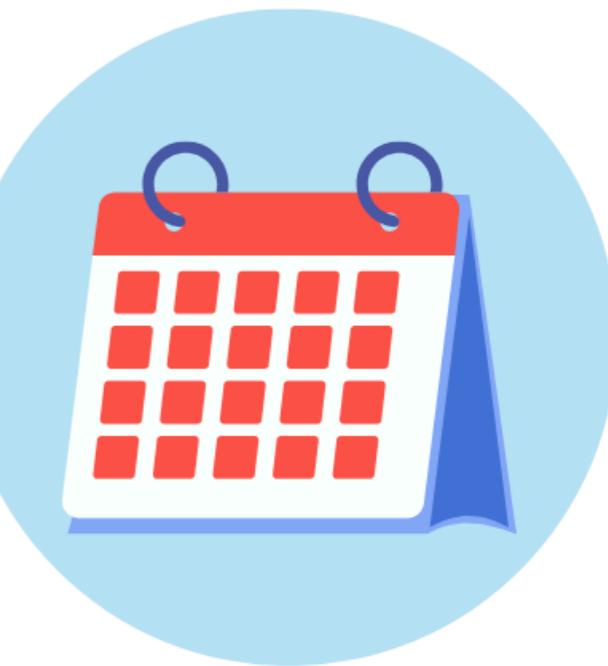
Individuals and organizations use data



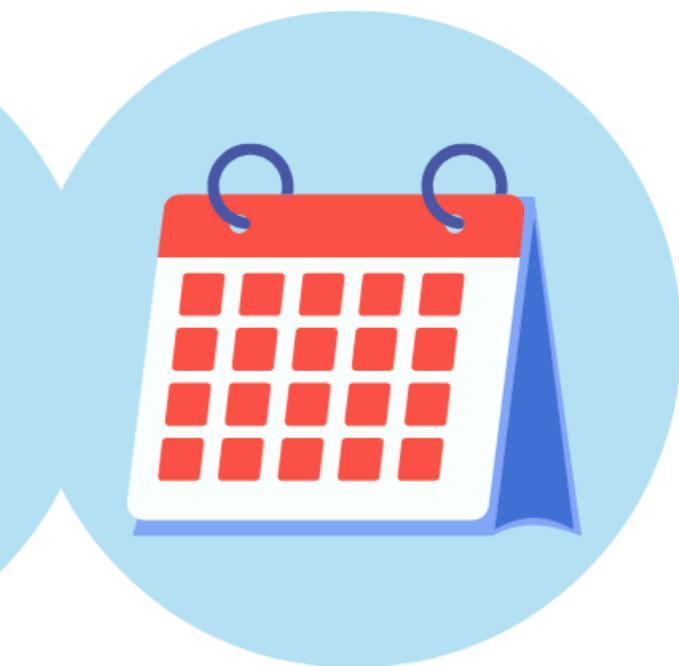
Individuals and organizations use data



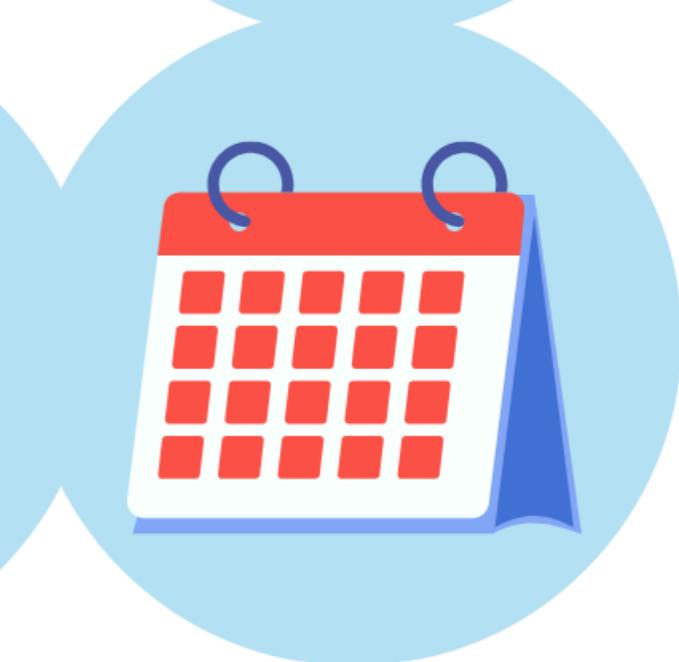
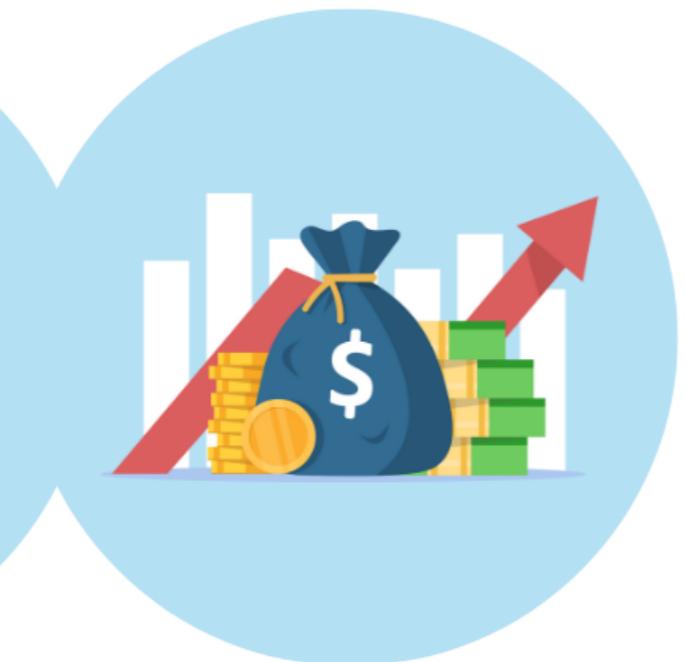
Individuals and organizations use data



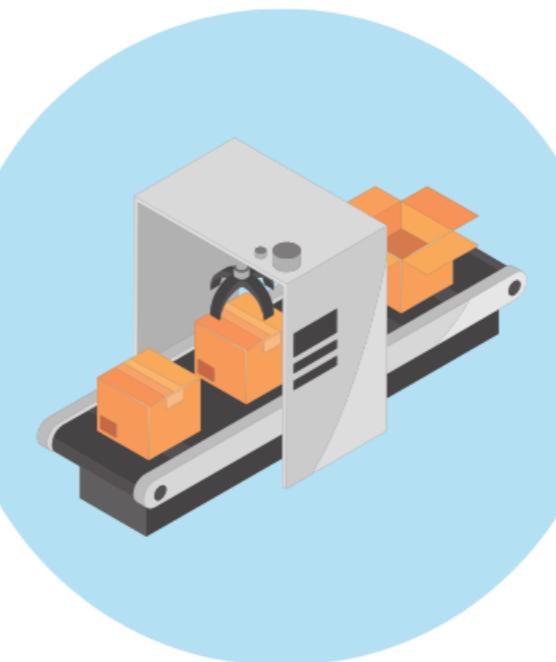
Individuals and organizations use data



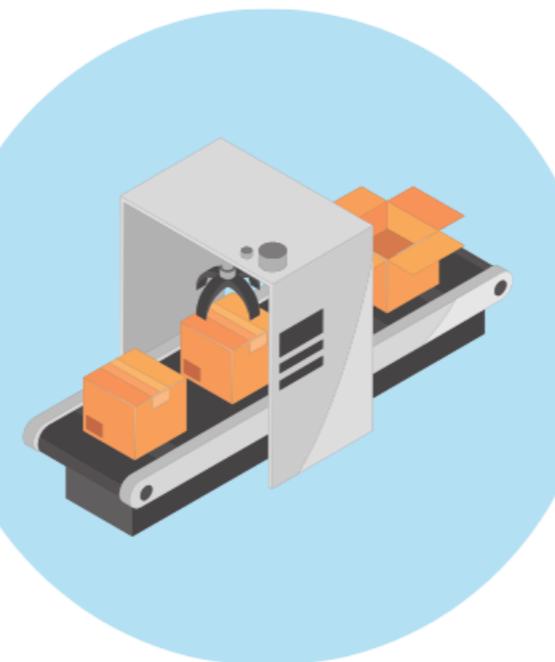
Individuals and organizations use data



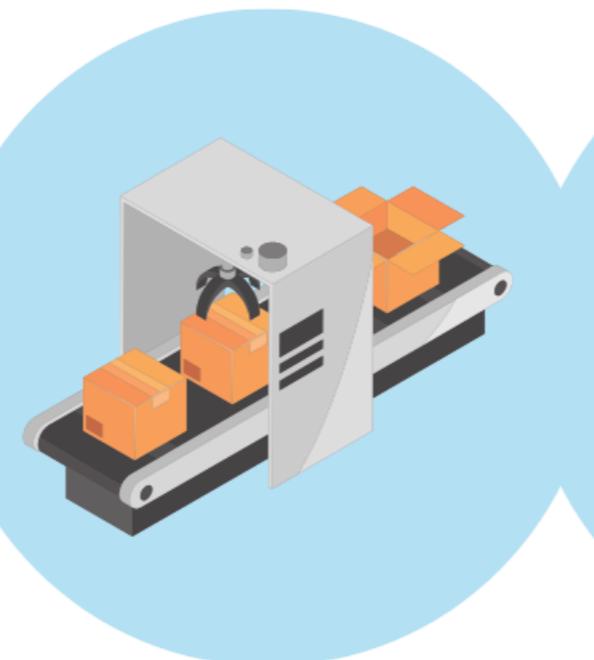
Individuals and organizations use data



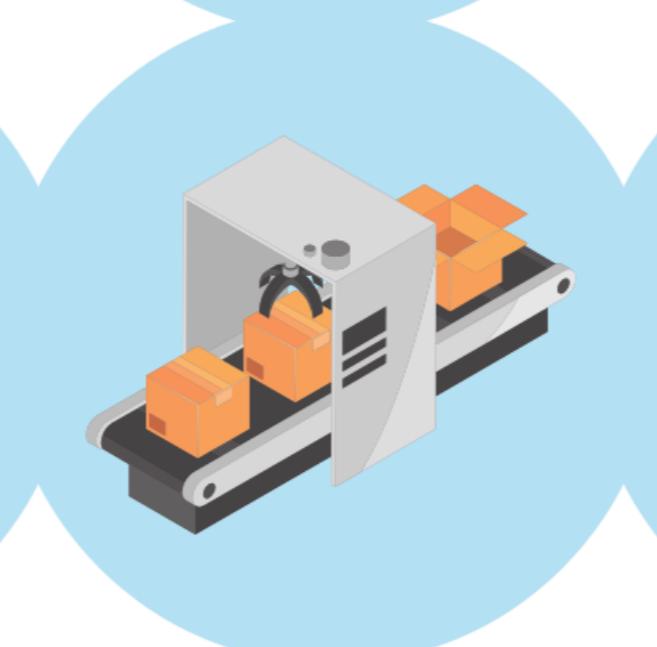
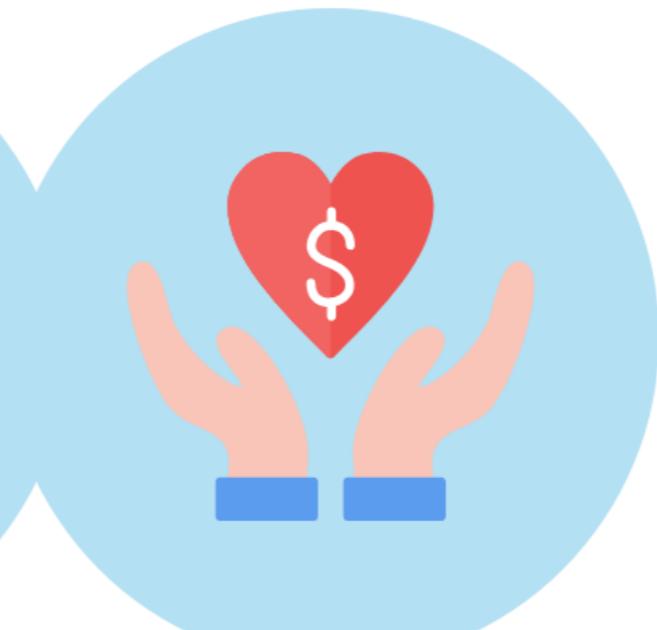
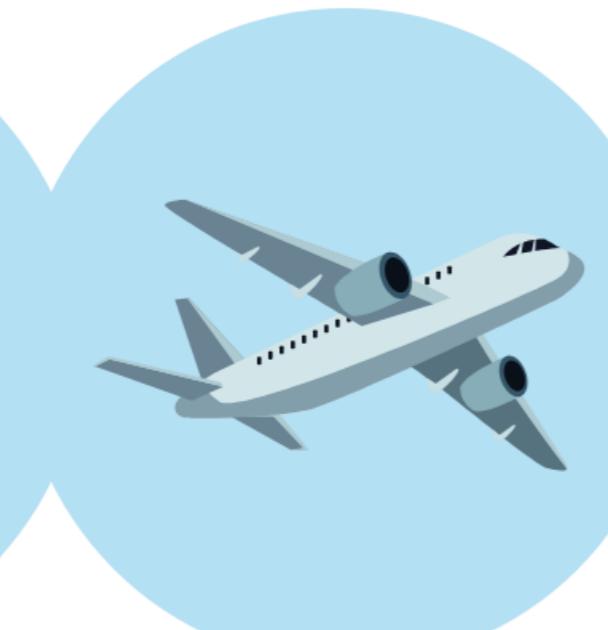
Individuals and organizations use data



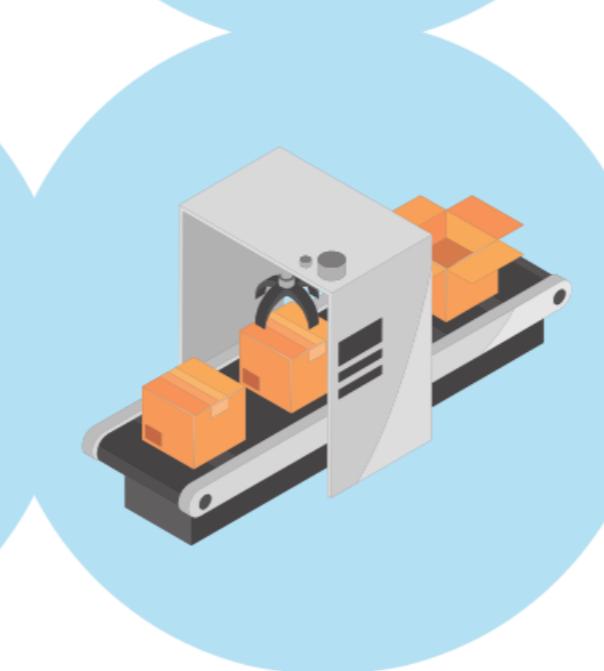
Individuals and organizations use data



Individuals and organizations use data



Individuals and organizations use data



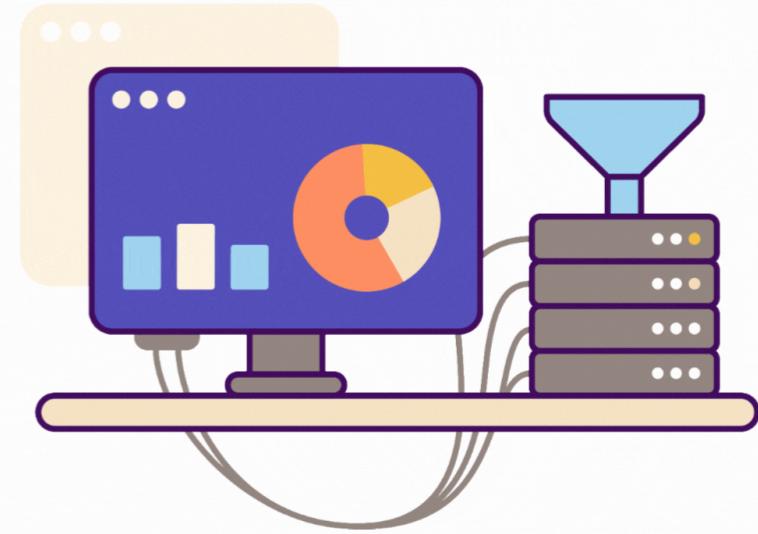
Data in organizations

Goal: support business objectives

- Profitability
- Social good
- Research
- Customer satisfaction & employee happiness

How: by improving decision making

- Measure return on investment (ROI)
- Optimize processes and find new opportunities



Data in healthcare

Example: wearable devices

- Monitor personal data

Goal:

- Detect and prevent health problems
- Turning patient care into precision medicine
- Advancing healthcare research worldwide



Data in supply chain

Example: monitoring various metrics

- Average inventory
- Inventory turnover ratio

Goal:

- Make sense of the massive amount of generated data
- Demand forecasting



The sequence of processes involved in the production and distribution of a product¹.

¹ Oxford Languages

Data in education

Example: DataCamp courses

- User feedback
- Incorrectly submitted answers
- Learner drop-off

Goal:

- Improve course design
- Identify struggle areas
- Personalize learning experience



Data is a competitive advantage

Why is that?

- Unique to an organization
- Unavailable for competitors
- Generate insights and actions based on that data



Let's practice!

INTRODUCTION TO DATA