

Data is all around us

INTRODUCTION TO DATA LITERACY



Jess Ahmet

Content Developer, DataCamp

	A	B	C	D	E	F	G	H
1	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category
2	768805383	Existing Customer	45 M		3 High School	Married	\$60K - \$80K	
3	818770008	Existing Customer	49 F		5 Graduate	Single	Less than \$40K	
4	713982108	Existing Customer	51 M		3 Graduate	Married	\$80K - \$120K	
5	769911858	Existing Customer	40 F		4 High School	Unknown	Less than \$40K	
6	709106358	Existing Customer	40 M		3 Uneducated	Married	\$60K - \$80K	
7	713061558	Existing Customer	44 M		2 Graduate	Married	\$40K - \$60K	
8	810347208	Existing Customer	51 M		4 Unknown	Married	\$120K +	
9	818906208	Existing Customer	32 M		0 High School	Unknown	\$60K - \$80K	
10	710930508	Existing Customer	37 M		3 Uneducated	Single	\$60K - \$80K	
11	719661558	Existing Customer	48 M		2 Graduate	Single	\$80K - \$120K	
12	708790833	Existing Customer	42 M		5 Uneducated	Unknown	\$120K +	
13	710821833	Existing Customer	65 M		1 Unknown	Married	\$40K - \$60K	
14	710599683	Existing Customer	56 M		1 College	Single	\$80K - \$120K	
15	816082233	Existing Customer	35 M		3 Graduate	Unknown	\$60K - \$80K	
16	712396908	Existing Customer	57 F		2 Graduate	Married	Less than \$40K	
17	714885258	Existing Customer	44 M		4 Unknown	Unknown	\$80K - \$120K	
18	709967358	Existing Customer	48 M		4 Post-Graduate	Single	\$80K - \$120K	
19	753327333	Existing Customer	41 M		3 Unknown	Married	\$80K - \$120K	
20	806160108	Existing Customer	61 M		1 High School	Married	\$40K - \$60K	
21	709327383	Existing Customer	45 F		2 Graduate	Married	Unknown	
22	806165208	Existing Customer	47 M		1 Doctorate	Divorced	\$60K - \$80K	
23	708508758	Attrited Customer	62 F		0 Graduate	Married	Less than \$40K	
24	784725333	Existing Customer	41 M		3 High School	Married	\$40K - \$60K	
25	811604133	Existing Customer	47 F		4 Unknown	Single	Less than \$40K	
26	789124683	Existing Customer	54 M		2 Unknown	Married	\$80K - \$120K	
27	771071958	Existing Customer	41 F		3 Graduate	Single	Less than \$40K	
28	720466383	Existing Customer	59 M		1 High School	Unknown	\$40K - \$60K	
29	804424383	Existing Customer	63 M		1 Unknown	Married	\$60K - \$80K	
30	718813833	Existing Customer	44 F		3 Uneducated	Single	Unknown	
31	806624208	Existing Customer	47 M		4 High School	Married	\$40K - \$60K	
32	778348233	Existing Customer	53 M		3 Unknown	Married	\$80K - \$120K	
33	712991808	Existing Customer	53 M		2 Uneducated	Married	\$60K - \$80K	
34	709029408	Existing Customer	41 M		4 Graduate	Married	\$60K - \$80K	

Data comes in all shapes and sizes



DataCamp ✅ @DataCamp · 29 jul.
Happy Friday DataCampers! 😊

What course will you be studying over the weekend?

The left page of the book contains the following text:

it goes with ridiculous hairy ears and noses, he says he snorted a rat, and I hate cats. like we were tank tops, because they also should wear the white hats and who should do it? I don't like tank tops. I think it's a shame that we're not tank tops. The fact that it's filled with Canadians, or hockey fans, and hockeys in dental braces, it gets that does not make it a good place to live. like they're wearing white shirts or socks with us, because they look like they're attending a funeral or something.

The right page of the book contains the following text:

JEROME HAD RETURNED HOME from the Pissalong coast that he would and Jérôme waiting for him. He did not know what he was doing, but he was still there. He sat down from his labors and scanned the horizon, no longer looking at the poor, ugly and beautiful, smart and droll, old help and self-acceptance, victims at Republicans and Democrats, Chevy and Ford, and so on. He had been a good boy, but now he was old and educated and uneducated, to name a few, who understood the concept of reification, that everything has a price, like everyone makes a living.

The bottom line is that we are all self-righteous people to secretly believe that we are better than others. That no one is inherently righteous and that comes from Jesus as a gift. An embrace this human condition, it's not a bad day to have our hearts in the place of us, so I was prepared against most of us.

He had done the same when he returned home after collecting the last of the day.

When harvest time ended and Jerome had stored away the last of his crops, he and his family traveled to Jerome's camp for the first time since he had come into Gathland, Tennessee, via the Santa Monica Parkway, for a day trip. "That's it's real pretty here," he said.

He voice filled my heart with everything I know to be good, and I knew that I should just hop in my Jeep, buckle up and go.

I should have asked him again to sit down in a chair, that, umm, one day, I think—ever day the thought of driving won't leave me alone. One day I won't care about the scars on my body. One day I won't be beautiful as Neville in some simple playing selection from Vivaldi. Wherever I go, I think. See you.

Right now a truck in my Jeep down curvy roads into Bryan City for a stop at Ingle's or The Center is all that I can handle.

A cardinal flies by, and his bright color makes me see only one thing—blood. Blood, fresh and dried, all over the seats of Liam's Mustang. I try to wash away the memory of that rat.

The right page continues with more text from the story.

Salvador was of variety—nor was he interested in finding me.

"You were a big help today," he said in Salvador passed by, the PVC pipe bobbing like jacaranda in his arms.

"Thank you, Father," came the accented reply accompanied by the characteristic smile and nod. The man's voice was light and scratchy, like that of a well-worn-up crooner who had traveled far and wide.

"We may partake in some festivities later, if you could care to join us. Just a little song and dance, nothing too elaborate."

"No, thank you," Salvador said.

"Okay, if you change your mind, you know where we'll be."

Another agreeable nod, and then he moved on.

The last of the day's light died away, and torches were lit to aid in the final bits of work and cleanup. Breimair kept watch, ever alert, over the flames, a momentary glimmer through the surface. When he dissolved successfully into the darkness, no one else took particular notice; they had seen it a thousand times. He would return tomorrow, after all.

HAMMOND SHUDDERED AWAKE. The pillow had been over his head bashed everything off the nightstand before tumbling away. His eyes shifted about like those of a frightened child. There was sunlight slanting through the blinds, catching patterns on the furniture and the carpet. But he didn't notice the light, the sun, the morning or anything else. He was out of breath, his heart hammering.

Then it started coming back—the phone call, the church, the coffee, the sickening emptiness that followed the loss of the most important person in his life. He had thought he had a handle on it, a healthy, rational response to it. It had enough.

I had to live through it once, he thought himself. And I have to endure repeat showings! Then he remembered where he was and why he was here.

In the hallway, he turned on the cold water and splashed his face. He appraised himself in the mirror—pale-gray hair-moon had formed under his eyes, and he'd look like a corpse. He remembered Neals standing offer to fill a prescription for sleeping meds—drugs that would put him into such a deep sleep it would take all the time in the world to wake him up.

and aunt? Then I had to meet her. I was still in the house, because I had heard about Snow Flower's home and family for so long, while I knew almost nothing about mine. I was so angry, so angry. I was filled with anticipation—at last I'd be going to Snow Flower's house!—the seemed vague about the details. "Snow Flower's husband and his wife will bring me to me," Snow Flower said.

"Do you think my mother-in-law will join us for dinner?" I asked. "I would be pleased and because she would see me with my loving."

"I am not a too busy. She has many duties, just as you will one day."

"But I'll get to meet your mother, older sister, and ..."

"I had expected to meet Mama and Ann would be part of Snow Flower's rituals. She seemed so much a member of our family. I thought she would want them there."

"Auntie Wang?" I asked. "The marchioness probably make several appearances."

THE GINGERLY 117

Los of people play the piano. It couldn't be that last."

"I don't know..." Her father obviously didn't believe a word of it.

"For just a moment, Jamie had a sort of all-over-again feeling, just like the moment of meeting, which had replied for the bakery job. He liked the same kind of women taking over as he watched his hand go up when he was asking it to play a little piano.

"They looked at him as if they'd never seen him before in their lives.

"Well, they both did," Anne looked at her dad for a little help.

"Well, I didn't forget. Who knew if she'd allow him to leave?"

"I'll be along, though, who would give him a better chance to have an eye on the elusive Anne Stewart. That in itself was worth it, even if it was just for a few weeks. Who knew if he could lead or besides, wasn't this something a normal person would do? A normal person with a normal life!"

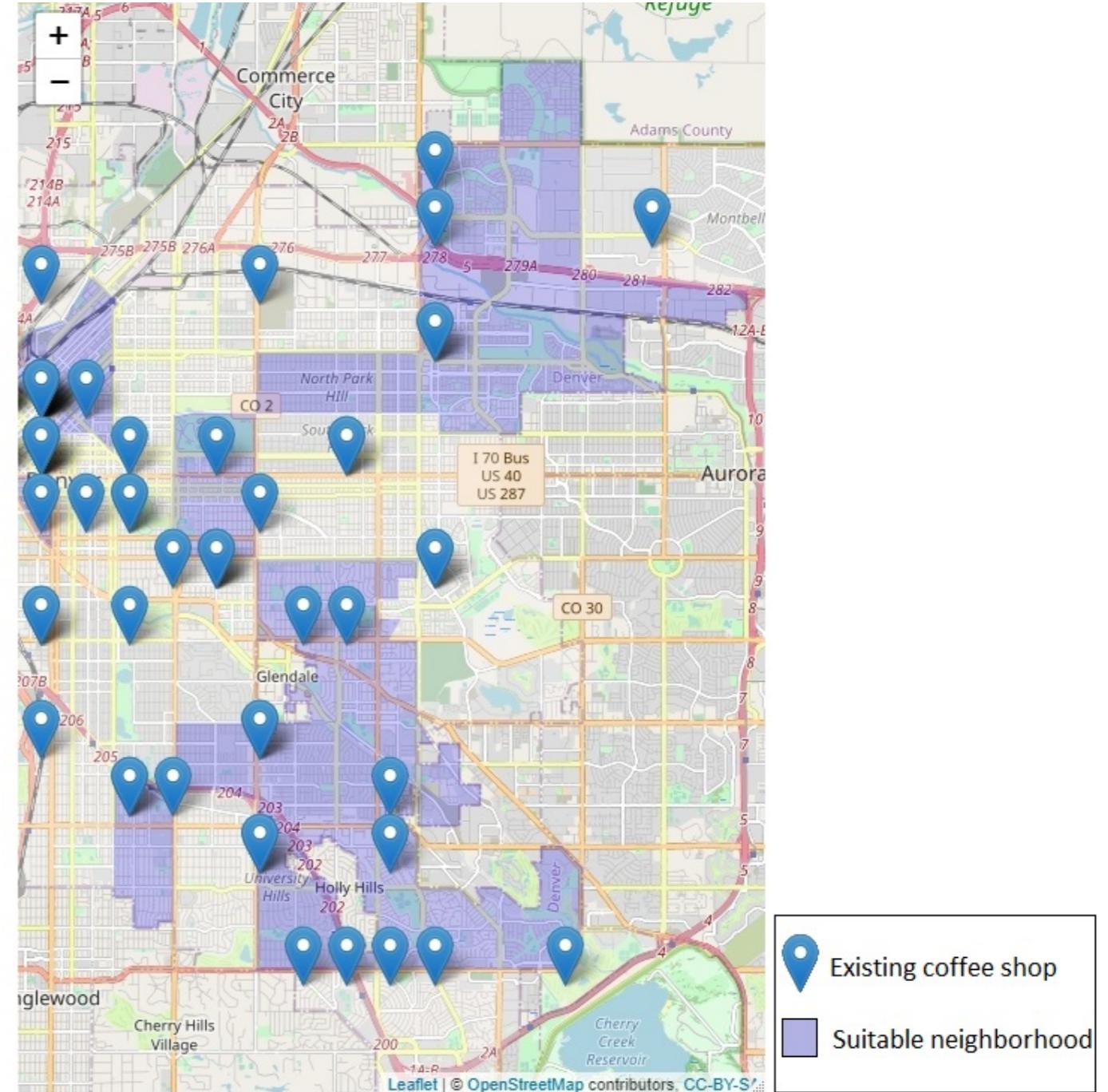
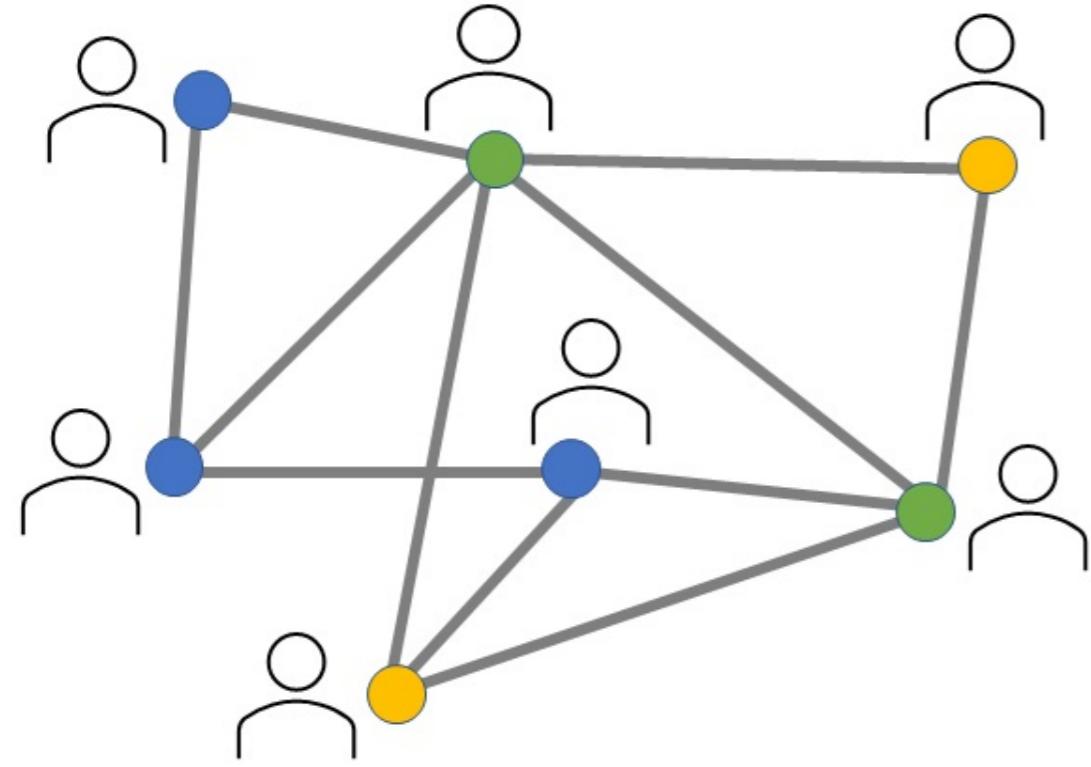
"Looks like you've got yourself a volunteer," Andy said. "Bettie that's alone, right? I'll give Lance a call if you want. Kind of clear the way for you, what do you think?"

"I don't think he does that, Dad." Anne held up her hand.

"I don't want you to get the wrong idea," Jamie thought he'd give them a way out, just in case. "I wouldn't land any Nashville gigs, I didn't

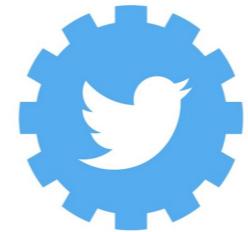
16

Data comes in all shapes and sizes



Data sources: open data

- Public databases and records
 - EU: data.europa.eu
 - US: data.gov
- Public APIs
 - Twitter
 - Google Maps
 - IMDB
 - Wikipedia
 - ...



Data sources: internal data

- Customer data
- Transaction data
- Website data
- Survey data



Let's practice!

INTRODUCTION TO DATA LITERACY

Common data types

INTRODUCTION TO DATA LITERACY



Jess Ahmet

Content Developer, DataCamp

Importance of the data type

- Different types of data required to get a complete picture
- Has an effect on:
 - How to collect the data
 - How to store the data
 - How to analyze the data



Structured vs. unstructured data

Structured

- Data in tabular form
- Easy to search and organize
- Requires less preprocessing
- Stored in relational databases

Unstructured

- Data without a pre-defined structure
- Difficult to search and organize
- Requires more preprocessing
- Stored in document databases

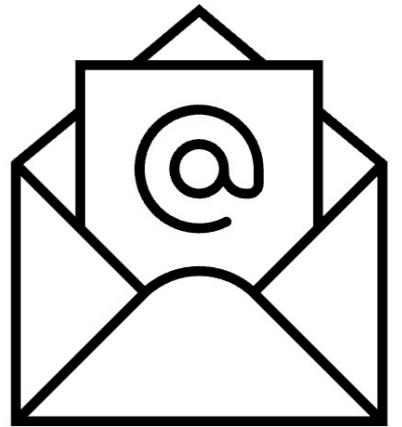
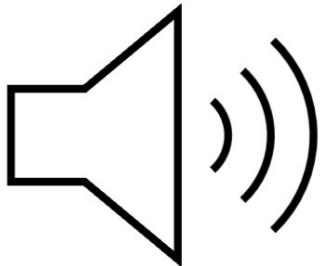
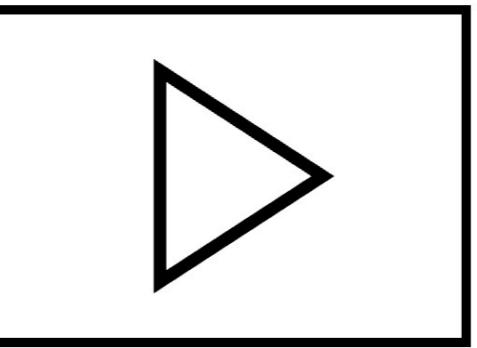
Examples of structured data

- Spreadsheets
- Data tables

A	B	C	D	E	F	G	H	
1	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category
2	768805383	Existing Customer	45 M		3	High School	Married	\$60K - \$80K
3	818770008	Existing Customer	49 F		5	Graduate	Single	Less than \$40K
4	713982108	Existing Customer	51 M		3	Graduate	Married	\$80K - \$120K
5	769911858	Existing Customer	40 F		4	High School	Unknown	Less than \$40K
6	709106358	Existing Customer	40 M		3	Uneducated	Married	\$60K - \$80K
7	713061558	Existing Customer	44 M		2	Graduate	Married	\$40K - \$60K
8	810347208	Existing Customer	51 M		4	Unknown	Married	\$120K +
9	818906208	Existing Customer	32 M		0	High School	Unknown	\$60K - \$80K
10	710930508	Existing Customer	37 M		3	Uneducated	Single	\$60K - \$80K
11	719661558	Existing Customer	48 M		2	Graduate	Single	\$80K - \$120K
12	708790833	Existing Customer	42 M		5	Uneducated	Unknown	\$120K +
13	710821833	Existing Customer	65 M		1	Unknown	Married	\$40K - \$60K
14	710599683	Existing Customer	56 M		1	College	Single	\$80K - \$120K
15	816082233	Existing Customer	35 M		3	Graduate	Unknown	\$60K - \$80K
16	712396908	Existing Customer	57 F		2	Graduate	Married	Less than \$40K
17	714885258	Existing Customer	44 M		4	Unknown	Unknown	\$80K - \$120K
18	709967358	Existing Customer	48 M		4	Post-Graduate	Single	\$80K - \$120K
19	753327333	Existing Customer	41 M		3	Unknown	Married	\$80K - \$120K
20	806160108	Existing Customer	61 M		1	High School	Married	\$40K - \$60K
21	709327383	Existing Customer	45 F		2	Graduate	Married	Unknown
22	806165208	Existing Customer	47 M		1	Doctorate	Divorced	\$60K - \$80K
23	708508758	Attrited Customer	62 F		0	Graduate	Married	Less than \$40K
24	784725333	Existing Customer	41 M		3	High School	Married	\$40K - \$60K
25	811604133	Existing Customer	47 F		4	Unknown	Single	Less than \$40K
26	789124683	Existing Customer	54 M		2	Unknown	Married	\$80K - \$120K
27	771071958	Existing Customer	41 F		3	Graduate	Single	Less than \$40K
28	720466383	Existing Customer	59 M		1	High School	Unknown	\$40K - \$60K
29	804424383	Existing Customer	63 M		1	Unknown	Married	\$60K - \$80K
30	718813833	Existing Customer	44 F		3	Uneducated	Single	Unknown
31	806624208	Existing Customer	47 M		4	High School	Married	\$40K - \$60K
32	778348233	Existing Customer	53 M		3	Unknown	Married	\$80K - \$120K
33	712991808	Existing Customer	53 M		2	Uneducated	Married	\$60K - \$80K
34	709029408	Existing Customer	41 M		4	Graduate	Married	\$60K - \$80K

Examples of unstructured data

- Images
- Videos
- Sound
- Text



Quantitative vs. qualitative data

Quantitative

- Describes something with numbers
- Can be measured or counted
- Wider range of statistics and analysis methods
- Example: distance between home and work

Qualitative

- Describes something with categories
- Can be observed
- More restricted range of statistics and analysis methods
- Example: color of your eyes

Examples of quantitative data

- Count of ice cream scoops
- Minutes before the ice cream starts to melt
- Length of the cone



Examples of qualitative data

- Flavor of the ice cream
- Color of the ice cream
- Cone or cup?



Let's practice!

INTRODUCTION TO DATA LITERACY

Managing data

INTRODUCTION TO DATA LITERACY



Jess Ahmet

Content Developer, DataCamp

Databases: basic concepts

- Database management system (DBMS): software for managing databases
- Different types of databases:
 - Relational vs. document database
 - Data warehouse vs. data lake
- Designing and optimizing database systems is typically the responsibility of a data engineer

ORACLE



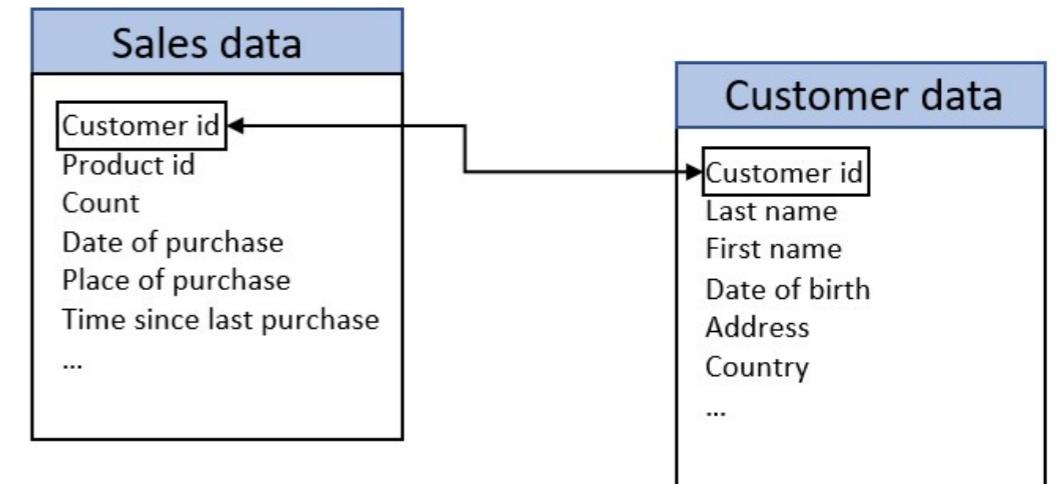
Microsoft®
SQL Server®



PostgreSQL



Relational database



Data storage in the cloud

- Data is stored on remote servers and accessed over the Internet
- Service provided by a specialized third party
- Cost-effective, but security dependent on third party



Google Cloud



Automation through data pipelines

- Pipelines move data from one database to another
- Automated collection and storage via the ETL process:
 - Extract
 - Transform
 - Load
- Ensures the availability of up-to-date and accurate data



Getting data from databases

- Querying: access and retrieve data
- Query with Structured Query Language (SQL)
- Further analysis with programming languages like R or Python



Dashboards

- Provides information at a glance, such as Key Performance Indicators (KPIs) and other metrics
- Receives data from a linked database
- Data is typically presented in a very visual
- Multipurpose tool: also used for exploratory analysis of the data and communicating
- Typically created and managed by a data analyst



Let's practice!

INTRODUCTION TO DATA LITERACY

Common data problems

INTRODUCTION TO DATA LITERACY



Jess Ahmet

Content Developer, DataCamp

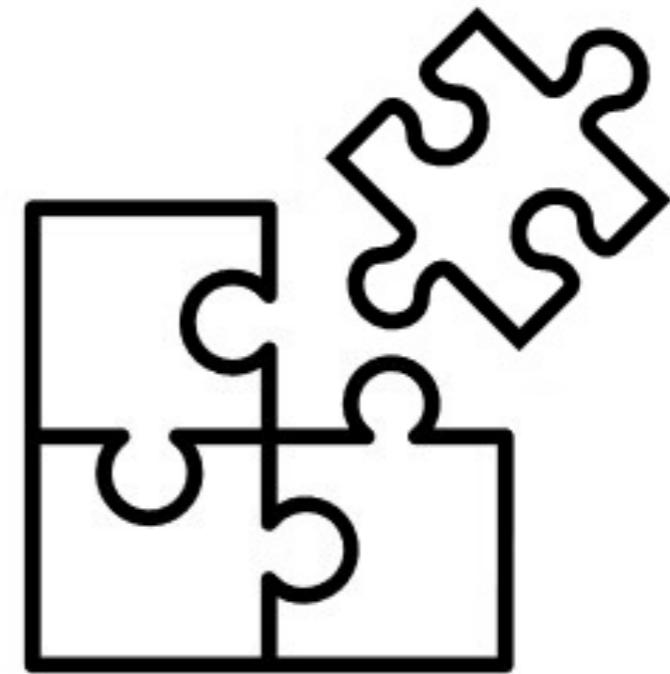
Dirty data

- Dirty data is:
 - Incorrect
 - Incomplete
 - Inconsistent
- Caused by human error, technical issues, or issues with the data collection process
- Garbage in, garbage out principle: dirty data can lead to wrong conclusions



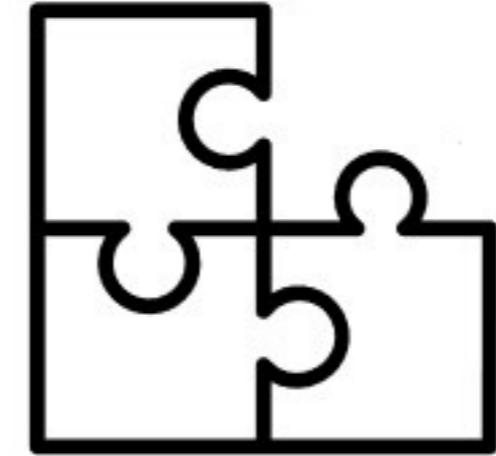
Data errors

- Data is incorrect or inconsistent
- Typically caused by human or technical error in recording the value or the format
- Techniques to counter:
 - If original value or valid format is known: correct data
 - If unknown: drop data



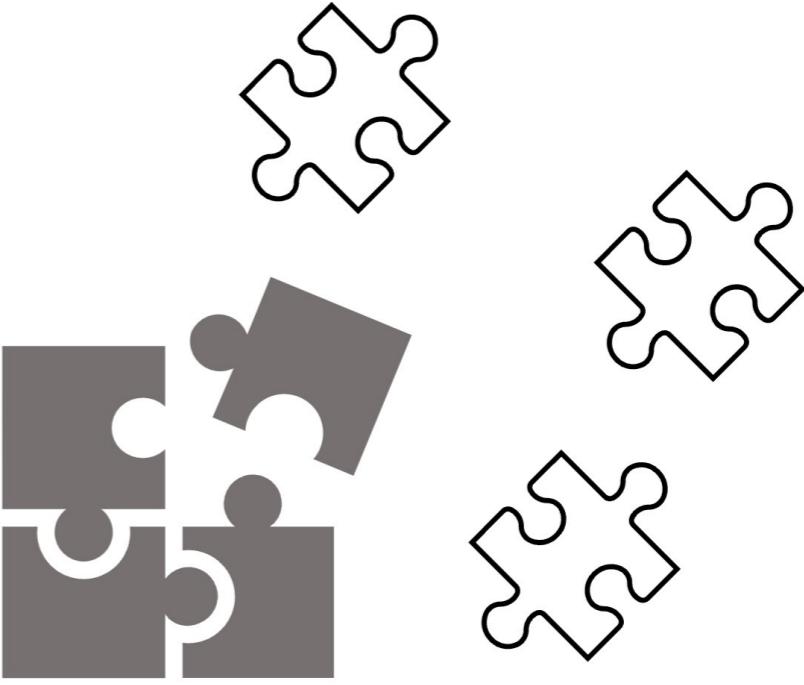
Missing data

- Data is incomplete
- Problematic if:
 - Many data points are missing
 - There are underlying patterns in the missing data
- Techniques to counter:
 - Dropping data
 - Imputation



Data bias

- Societal bias can be reflected in data as data bias
- Leads to unrepresentative data and therefore results
- Hard to detect and to resolve
- Techniques to counter:
 - Sound data collection process
 - Awareness in conclusions
 - Explainable AI models



Data cleaning

- Set of techniques to counter data problems
- Important preparation step for any data analysis
- But not all data problems are (completely) solvable
- It is always possible to do some kind of analysis



Let's practice!

INTRODUCTION TO DATA LITERACY