Defining Data: Takeaways

Concepts

DATA

- **Datum** is the singular form of **data**.
- One datum is a single fact that is recorded, measured, or collected through some type of observation, either by humans or machines.
- Data is a collection of recorded, measured, or collected **facts**.
- Personal opinions, beliefs, judgments, instincts, and viewpoints aren't data because they aren't recorded, measured, or collected facts.
- We can use data to reason, calculate, analyze, make decisions, predict, and plan for the future.
- The quantity, speed, and availability of data today has changed how businesses operate and how people live.

STRUCTURED DATA

- **Structured** data is tabular.
- Structured data has a predetermined format (i.e., tabular) and is represented by rows and columns.
- Tabular data is typically stored in spreadsheets or databases.

UNSTRUCTURED DATA

- Data is **unstructured** when it cannot be fit into a neat organization of rows and columns.
- Email messages, blog posts, images, audio clips, videos, satellite imagery, location data, sensor readings, and website logs are examples of unstructured data.
- It has recently been estimated that more than 80% of all new data produced today is unstructured.

SEMI-STRUCTURED DATA

- **Semi-structured data** doesn't have the structured format of tabular or spreadsheet data, but it contains some structure in the form of tags and metadata that help people group, describe, and analyze the data.
- **Metadata** provides basic information about the data, such as when and why it was created, who created the data, where the data was created, data file size, and other information.
- Unstructured data that is accompanied by metadata becomes semi-structured.

DATA TYPES

- There are different types of data.
- Knowing what type of data we have is key to processing and analyzing it correctly.

QUANTITATIVE AND QUALITATIVE DATA

- The two most basic data types are **quantitative** and **qualitative** data.
- Quantitative data is **numerical**.
    - We use numbers to represent quantitative data.
    - Quantitative data can be counted or measured.
    - An example of quantitative data is *speed* because we can measure it (e.g., kilometers per hour).
    - Quantitative data is also called **numerical** data.
- Qualitative data is **categorical**.
    - Qualitative data cannot be counted or measured. Instead, it classifies or categorizes different objects.
    - We use names or labels to represent qualitative data.
    - An example of qualitative data is *hair color* because we name the different categories of hair color (black, brown, blonde, gray, white, etc.).
    - Qualitative data is also called **categorical** data.

CATEGORICAL DATA: NOMINAL OR ORDINAL

- Categorical (qualitative) data can be **nominal** or **ordinal**.
- When data is nominal, the categories of the data can be named or listed, but there is no inherent order to them.
    - For example, *car manufacturer* is nominal. Car manufacturers can be named (e.g., Honda, Toyota, Nissan, Kia, etc.), but they don't have an inherent order.
- Ordinal data has categories, but they can be ordered — from smallest to largest, least to most, worst to best, or in other ordered ways.
    - For example, *T-shirt size* is ordinal. We use labels to describe it (e.g., small, medium, large), and these labels can be ordered from smallest to largest.

NUMERICAL DATA: DISCRETE OR CONTINUOUS

- Numerical (quantitative) data can be **discrete** or **continuous**.
- Discrete data is counted in whole numbers. Whole numbers do not allow any fractions or decimals.
    - For example, *number of siblings* is discrete because we have to use whole numbers to count our siblings (e.g., 1, 2, 3, and so on). It's impossible to have 2.3 siblings or $\frac{3}{2}$ siblings.
- Continuous data can take on any value between two whole numbers. In other words, it can have decimals and fractions.
    - For example, *height* and *length* are continuous, as they allow for and are often measured with decimals (e.g., 125.8 pounds) and fractions (e.g., $\frac{17}{2}$ inches)

Resources

- What Is Data?

- Difference Between Fact and Opinion
- Nominal, Ordinal, Interval, Ratio Scales with Examples
- WHY IS DATA IMPORTANT FOR YOUR BUSINESS?
- What is structured, semi structured and unstructured data
- Metadata