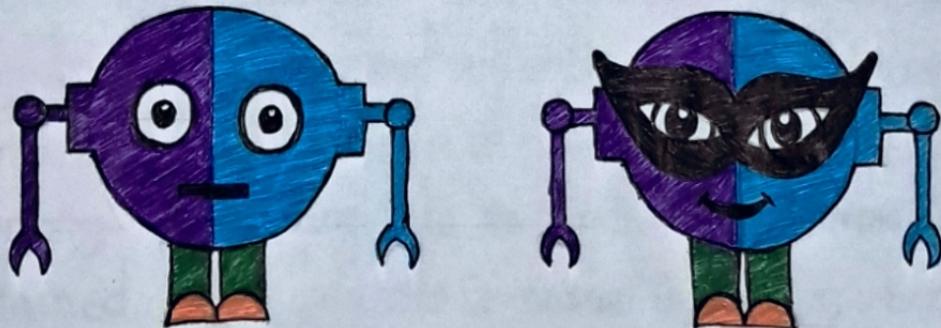
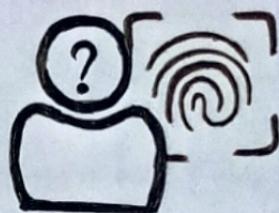


Data Anonymization

Data anonymization is the process of protecting people's private or sensitive data by eliminating PII (Personally Identifiable Information). Typically, data anonymization involves blanking, hashing, or masking personal information, often by using fixed-length codes to represent data columns, or hiding data with altered values.

Data Anonymization



Sensitive data → Anonymized data

Personally identifiable information (PII) is any information connected to a specific individual that can be used to uncover that individual's identity, such as their social security number, full name, or email address.



Personally Identifiable Information

Data professionals can take additional measures to protect users and their data. Data aggregation, for example, is the process of collecting and combining details from a significant number of users in terms of totals or summary. Aggregating data ensures that information contained within datasets is shown in groups; when coupled with other anonymization techniques, data professionals can ensure compliance with data privacy and anonymization standards.

Data anonymization is used in just about every industry. As a data analytics professional, you probably won't

personally be performing anonymization, but it's useful to understand what kinds of data are often anonymized before you start working with it. This data might include:

- Telephone numbers.
- Names.
- License plates and license numbers.
- Social security numbers.
- IP addresses.
- Medical records.
- Email addresses.
- Photographs.
- Account numbers.

Imagine a world where we all had access to each other's addresses, account numbers, and other identifiable information. That would invade a lot of people's privacy and make the world less safe. Data anonymization is one of the ways we can help keep data private and secure!