Data anonymization

What is data anonymization?

You have been learning about the importance of privacy in data analytics. Now, it is time to talk about **data anonymization** and what types of data should be anonymized. **Personally identifiable information**, or **PII**, is information that can be used by itself or with other data to track down a person's identity.

Data anonymization is the process of protecting people's private or sensitive data by eliminating that kind of 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.

Your role in data anonymization

Organizations have a responsibility to protect their data and the personal information that data might contain. As a data analyst, you might be expected to understand what data needs to be anonymized, but you generally wouldn't be responsible for the data anonymization itself. A rare exception might be if you work with a copy of the data for testing or development purposes. In this case, you could be required to anonymize the data before you work with it.

What types of data should be anonymized?

Healthcare and financial data are two of the most sensitive types of data. These industries rely a lot on data anonymization techniques. After all, the stakes are very high. That's why data in these two industries usually goes through **de-identification**, which is **a process used to wipe data clean of all personally identifying information**.