Critical data security and privacy principles

You have learned how data analytics can be used for good causes, like assisting nonprofit organizations. Also, you learned that data professionals need to protect privacy within data and remain aware of other considerations, like data bias and making assumptions about data.

As a data analytics professional, you have a responsibility to handle data ethically. Data ethics refers to well-founded standards of right and wrong that dictate how data is collected, shared, and used. Throughout your career you will work with a lot of data. This sometimes includes PII, or **personally identifiable information,** which can be used by itself or with other data to track down a person's identity. One element of treating data ethically is ensuring that the privacy and security of that data is maintained throughout its lifetime. In this reading, you will learn more about the importance of data privacy and some strategies for protecting the privacy of data subjects.

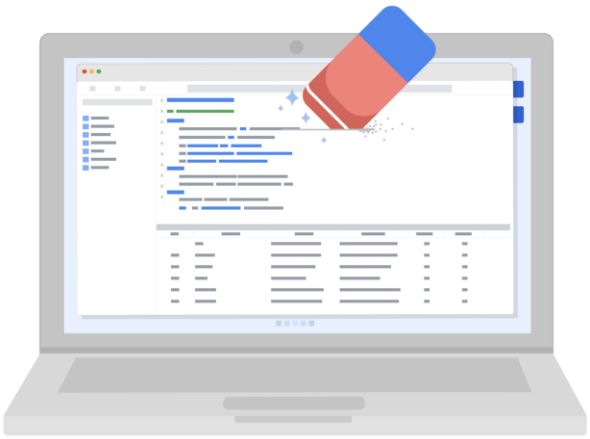
Privacy matters

Data privacy means preserving a data subject's information and activity any time a data transaction occurs. This is also called information privacy or data protection. Data privacy is concerned with the access, use, and collection of personal data. For the people whose data is being collected, this means they have the right to:

- Protection from unauthorized access to their private data
- Freedom from inappropriate use of their data
- The right to inspect, update, or correct their data
- Ability to give consent to data collection
- Legal right to access the data

In order to maintain these rights, businesses and organizations have to put privacy measures in place to protect individuals' data. This is also a matter of trust. The public's ability to trust companies with personal data is important. It's what makes people want to use a company's product, share their information, and more.

Protecting privacy with data anonymization



Organizations use a lot of different measures to protect the privacy of their data subjects, like incorporating access permissions to ensure that only the people who are supposed to access that information can do so. Another key strategy to maintaining privacy is data anonymization.

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

Key takeaways

For any professional working with data about actual people, it's important to consider the safety and privacy of those individuals. That's why understanding the importance of data privacy and how data that contains PII can be made secure for analysis is so important. We have a responsibility to protect people's data and the personal information that data might contain.