Transformation approaches for securing data

So far, you've learned that data privacy is about preserving information any time a transaction occurs. You can use transformation techniques to protect sensitive information in a way that allows you to perform analysis safely. It's important for organizations to have the trust of their users when working with any of their data. As an analyst, it's your responsibility to use data honestly, transparently, and fairly. Each organization sets their own standards and policies based on the nature of, and the intended use of their data. In this reading, you'll learn more about data transformation approaches to ensure the security and privacy of your data.

Overview of PII

As a reminder, PII is personally identifiable information. PII data is unique to an individual, and is used to distinguish one person from another. Some common examples of PII that you may be familiar with include users' email address, mailing address, phone number, precise location, and full name. As an analyst, it is important to maintain the confidentiality of PII to ensure user data is protected against malicious intent. For example, a criminal with access to PII can commit a variety of theft and fraud by opening credit cards, bank accounts, applying for loans, and implementing email phishing attacks.

Even if the organization you work for collects personal information from users, it doesn't mean you will work directly with PII. If you do work with PII, you'll be granted access to the bare minimum of data needed to do your job. This level of access is known as "need to know" access. As a reminder, accessing data out of curiosity, or for a non-valid business reason, is likely a violation of organization policies.

Approaches to secure PII

The primary focus of securing PII is to transform the original data in a way that protects the identity of individuals, while maintaining a degree of usefulness for analysis processes. As an analyst, you can use the anonymization, masking, or pseudonymization method, or a combination of methods to ensure the privacy and utility of user data. Let's explore these in more detail.

Anonymization

Anonymization involves transforming data in a way that it can't be traced back to its original form. As an analyst, you can remove, encode, or replace PII to create a dataset that protects user privacy. Encoding entails transitioning data into a binary (1s and 0s) format.

One way to anonymize data is to replace all usernames in a dataset with generic placeholders, like “user 1,” “user 2,” and so on, until all names are replaced. The placeholders don’t reveal any personal information, but you can still use the data to perform your analysis.

In its anonymized form, data cannot be identified by its original format. When analysts combine additional information like other datasets, it’s possible to determine the original data. This makes anonymization alone one of the lesser secure options for data.

Masking

Data masking is the process of hiding original data with modified content, like characters or symbols, while maintaining a similar structure to the original data. For example, you might choose to mask social security numbers with Xs or asterisks (*) like XXX-XX-1234 or ***-***-1234. The masked social security numbers have the same structure as the original, except only the last four digits are seen.

Masking is an effective technique to protect sensitive information. So, it’s important to think about what degree of masking will protect your data while also maintaining usability for analysis.

Pseudonymization

Pseudonymization involves replacing private identifiers with fake identifiers, or pseudonyms, to reduce the linkability of the data to the original user. Pseudonymization offers more security than masking, but it’s less secure than anonymization.

Key takeaways

Understanding and implementing data security approaches like anonymization, masking, and pseudonymization is important for protecting PII in cloud environments. Each method offers varying levels of security and usability. As an analyst, it’s important to thoughtfully consider each method before choosing the right approach for your specific use case.

Resources for more information

Learn more about how to identify PII, the dangers of unprotected data, and steps you can take to protect your PII with these resources:

- An overview of PII and how to safely use and protect PII: How to safeguard personally identifiable information
- Read more about how identity theft happens, what criminals can do with your PII, and steps you can take to protect your data: FBI Tech Tuesday: Protecting Against Personally Identifiable Information (PII) Theft