Analytics Reports / Protecting user privacy in report data

Article

Protecting user privacy in report data

Understand measures that help protect user privacy.

Overview

The data reports in the Analytics Reports framework help you manage and run your apps. In some reports, Apple applies specific privacy measures to protect individual user identities. This article provides details about these privacy measures and lists the reports to which they apply.

Data in the framework’s reports is aggregated, and similar users are in groups, which means individual user data is not easily identifiable. To balance data privacy with usability, the framework applies two approaches as needed to specific reports:

Thresholding

Omits values with data from fewer than 5 users or 5 unique devices.

Noise

Adds a small random number to metrics.

These privacy measures maximize utility as the volume of data increases, while also helping protect users identities. The framework applies the same privacy measures across daily, weekly, and monthly reports for the same type of content, ensuring that utility increases as more data is collected over longer periods of time. To maintain stability, the threshold and random noise that the framework initially applies remains constant across multiple requests for the same report.

> **Note**
>
> For privacy information for reports not listed in this article, see the Privacy section in each report.

Understand standard and detailed report types

In some reports, data might be more sensitive than it is in others. To help protect user privacy, the framework provides two types of reports:

Standard

Does not include sensitive fields easily relatable to user data.

Detailed

Includes all fields and applies threshold and noise measures to help protect uniquely identifiable user data.

Detailed reports provide values with a small amount of noise and omit data from values that have fewer than 5 users or 5 unique devices. Except for the specific cases detailed below, metrics in the detailed reports are noised so that, approximately:

- 68.2% of values are expected to be within +/- 2 of the true values.
- 95% of values are expected be within +/- 4 of the true values.
- 99.7% of values are expected to be within +/- 6 of the true values.

For example, if a specific row in a report has a true count value of 500,000, then 99.7% of the time the noised count is somewhere between 499,994 and 500,006. In reports with multiple different metrics, these are noised independently as appropriate.

Here is an example that shows noised and not noised installation and deletion data, where the amount of noise applied to the Counts and Unique Devices is 2 and 1, respectively:

Noised state	Counts	Unique Devices
Not noised	5000	2000
Noised	5002	2001

The following report groups have both standard and detailed reports:

- App Clip Usage
- App Sessions
- App Store Discovery and Engagement
- App Store Downloads
- App Store Installations and Deletions
- App Store Pre-orders

- App Store Purchases
- App Store Web Preview

Understand App usage report privacy approach

Users who opt in to share app usage data with Apple and developers do so with the understanding that they share their data in a non-personally identifiable way. To help ensure user privacy, the framework only generates both standard and detailed app usage reports when events exist from at least five users for the respective report. If fewer than five users contribute to a usage metric for a specific day, week, or month, then the relevant report does not generate. Additionally, for detailed reports, individual rows only populate if they have a value from 5 or more devices. These reports use this privacy approach:

- App Clip Usage
- App Crashes
- App Sessions
- App Store Installations and Deletions

Understand App Store Purchases detailed report

Because the App Store Purchases report contains several metrics that are highly relatable, for example, paying users, purchases, sales in USD, and proceeds in USD, the detailed report calculates noise once and applies it to all metrics. This approach ensures that the report has internal consistency across metrics, even with noise applied:

Metrics

Paying users, purchases, sales in USD, and proceeds in USD

Noise

Adds a small random number to paying users and purchases metric values then uses this number to calculate the noised sales and proceeds.

Paying user data is noised so that, approximately:

- 68.2% of values are expected be within +/- 2 of the true values.
- 95% of values are expected be within +/- 4 of the true values.
- 99.7% of values are expected to be within +/- 6 of the true values.

For example, if a specific row in a report has 5,000 true paying users, then 99.7% of the time the report shows a value somewhere between 4,994 and 5,006. The framework then proportionally applies this noise to the purchases, sales in USD, and proceeds in USD metrics.

Here is an example showing true and noised data, where the amount of noise applied to paying users is -3. The noised price is 30 dollars less than the true as it is computed as the noise value (-3) * purchase price ($10):

Noised state	Paying users	Purchases	Sales in USD	Proceeds in USD
Not Noised	5,000 paying users	10,000 purchases	$100,000 sales in USD	$10 purchase price
Noised	4,997 paying users	9,997 purchases	$99,970 sales in USD	$10 purchase price

The following report has dollar-valued metrics:

- App Store Purchases

Understand session duration in specific detailed report types

In the App Sessions and App Clip Usage detailed reports, the total session duration metric can be more variable with user data when you compare it to other metrics such as unique devices. The framework calculates session duration noise accordingly while ensuring data usability. Other metrics in these reports are noised as previously described. The framework calculates a tailored level of noise for the session duration metric for each detailed report row. This approach reflects how different users use your app:

Metrics

Total session duration

Noise

Standard deviation calculated by dividing the average session duration for each detailed report row by 4

For each detailed report row, the total session duration is noised so that, approximately:

- 68.2% of the time the reported value is expected to be within +/- the corresponding deviation of the true value.
- 95% of time the reported value is expected to be within +/- twice that deviation of the true.
- 99.7% of time the reported value is expected to be within +/- three times that deviation of the true value.

Here’s an example noise deviation calculation: If a specific report row has 100,000 true total duration seconds and 5,000 sessions, then the average session duration is 20 seconds. The framework divides 20 seconds by 4 (noise divisor) to get 5 seconds, and uses this value as the standard deviation of the applied integer noise. In summary:

```
100,000 (duration in seconds) / 5000 (sessions) = 20 (average session duration in seconds)
```

```
20 (average session duration in seconds) / 4 (noise divisor) = 5 (seconds of noise deviation)
```

Here’s an example that shows noised and not noised sessions data using the calculated noise deviation. With this deviation, the amount of noise the framework applies to session duration is 10. Sessions is separately noised with 1:

Noised state	Sessions	Total session duration (seconds)
Not Noised	5,000	100,000
Noised	5,001	100,010

For the report row with 100,000 true total duration seconds, approximately:

- 68.2% of the time the reported value appears within +/- 5 seconds (between 99,995 and 100,005).
- 95% of the time the reported value is within +/- 10 seconds (between 99,990 and 100,010).
- 99.7% of the time the reported value is expected to be within +/- 15 seconds (between 99,985 and 100,015).

The following reports have a session duration metric:

- App Clip Usage
- App Crashes
- App Installs Performance
- App Sessions

See Also

Essentials

Data Completeness and Corrections

Understand how the Analytics Reports API provides complete data sets.