INCUBETA

Data Discrepancies

Between Reporting Surfaces in GA4

A guide to better understand and optimize GA4 reporting.

Introduction

Anyone researching or working with GA4 will be well aware of the significant disparities between Universal Analytics (UA) and GA4. *However*, there is currently limited documentation on the variances within GA4 itself, and bridging the gap is no easy feat. This is where Incubeta can step in – with our latest whitepaper that'll help you – an end user – comprehend the data treatment procedures across different Reporting Surfaces in GA4. A resource that helps you truly understand the intricacies of GA4 data.

For ease of use, we have divided this paper into two main sections: Reporting Surfaces and Data Processing.

Reporting Surfaces

This section will provide a detailed explanation of which data processing can be applied to the given surface by GA4, with a focus specifically on those that can cause discrepancies between surfaces. As such, things like 'segments' will not be discussed, as this is a treatment a user applies to their data, rather than data processing being done by GA4 directly.

Data Processing

In the following section, we will outline the data processing done by GA4 between different reporting surfaces. Much like in the first section, the focus is on the data processing by GA4 itself.

Note:

This guide is true at the time of writing – June 2023 – and we will endeavor to update it at unscheduled intervals. If you would like a specific section of this report updated, or have any questions about GA4, please email analytics@incubeta.com and our team will be happy to help.

01

Contents

Reporting Surfaces

Within the GA4 platform, Reporting Surfaces refer to the different areas or interfaces in which data is presented and analyzed. These surfaces provide distinct views and capabilities for interpreting and visualizing data.
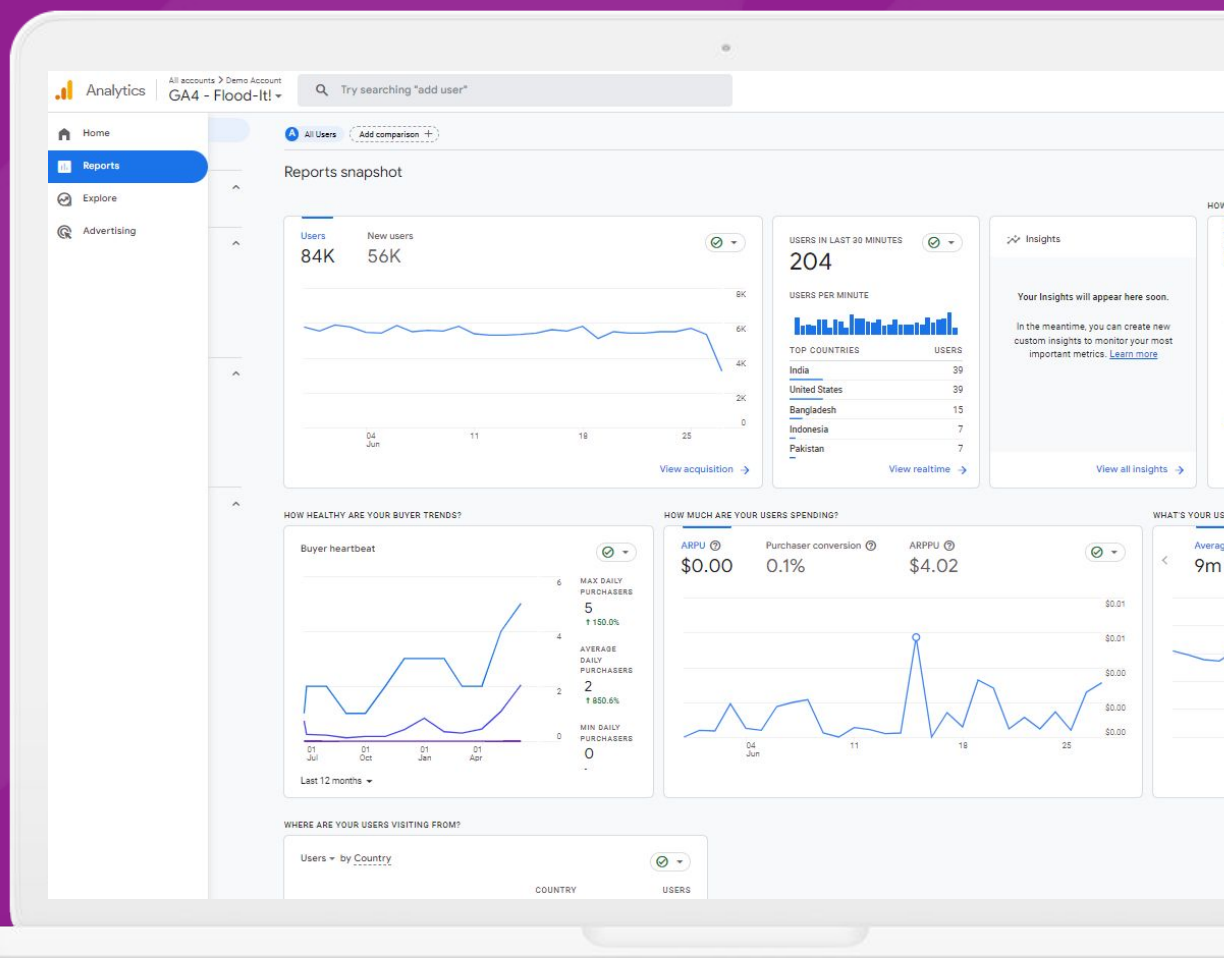
Underlying data & data processing

Google defines Reporting Surfaces as (1) Standard Reports, (2) API, (3) Explorations and (4) BigQuery.

These four Reporting Surfaces present data in distinct manners, primarily because of variations in data aggregation and processing methods. For example, Standard Reports and the API utilize aggregated data, whereas Explorations and BigQuery make use of event and user-level data, also known as raw data. This discrepancy in data sources contributes to the differences observed in how information is displayed across these surfaces.

The table below is an overview of reporting surfaces and the data processing each surface is subject to.

By Google	Standard Reports	Explore	Data API	BigQuery
Data Scope	Aggregated	Raw & user-level data	Aggregated	Raw & user-level data
Thresholding Applied to prevent anyone viewing a report or exploration from inferring the identity of individual users based on demographics, interests, or other signals present in the data.	Possible	Possible	Possible	Not Applied
High Cardinality	Can occur + show as the "(other)" row after filtering or applying a secondary dimension	No	Can occur + show as the "(other)" row after filtering or applying a secondary dimension	No
Sampling Used when the number of events returned by an exploration exceeds the limit for your property type.	No	Yes	Data API methods for Standard Reports will never show sampled data	No
Cross-Platform Reporting with Google Signals Google Signals are session data from sites and apps that Google associates with users who have signed in to their Google accounts, and who have turned on Ads Personalization.	Can be applied	Can be applied	Can be applied	No
Conversion Modeling Allows for accurate conversion attribution without identifying users.	Included	Included	Included	Not included
Behavioral Modeling Behavioral modeling for consent mode uses machine learning to model the behavior of users who decline analytics cookies based on the behavior of similar users who accept analytics cookies.	Included in Reporting Module + Real time. Not in Audiences, Segments with a sequence, Retention reports, Predictive Metrics	Partially-included, only in free-form tables	Can be applied if set up with a client-side tag either gtag.js or client-side GTM	Not included

Standard Reports

Standard Reports, accessible within the navigation panel on the left-hand side, are **derived from aggregated data** and generally do not encounter sampling, with the exception of funnel reports, which occasionally may be subjected to it.

The primary purpose of Standard Reports is to provide top-line data quickly. This Reporting Surface is not intended for analyzing detailed, granular data. Consequently, while sampling is typically absent in Standard Reports, Cardinality (the "other" row) may still be a factor. In the event that Cardinality becomes an issue and more detailed data is required, it is advisable to utilize Explore to eliminate Cardinality. GA4 360 customers can also take advantage of 'expanded datasets' to permanently open the "other" row entries in up to 100 reports. Another option to remove Cardinality for GA4 360 users is to request unsampled reports. Cardinality is explained in detail in the Google support article **HERE**.
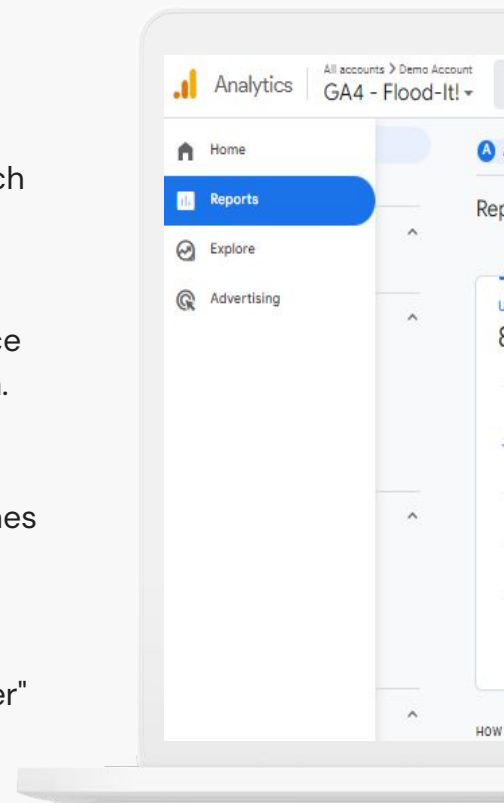
Note:

If granular data is required, use Explorations Reporting Surface, which is not subject to Cardinality. Standard Reports are subject to Thresholding if Google Signals is applied, reports are based on DDA and Conversion Modeling is included if criteria is met. Behavioral modeling will be applied if Consent Mode is installed and enabled.

Standard Reports are subject to Thresholding if Google Signals is applied. Thresholding is there to protect user's privacy – the data thresholds are predefined within the system and cannot be modified or adjusted by the user. You can increase the date range to include more data, which can help to not trigger the warning.

Standard Reports Cont...

The default attribution model in Standard Reports is Data Driven Attribution (DDA) model. You can select a different attribution model and conversion window at the property level to apply to all reports that use event-scoped traffic dimensions (for example, Source, Medium, Campaign, and Default Channel Group).

Standard Reports are subject to Conversion Modeling. The overall quantity of conversions remains unaffected, but rather the way they are assigned to a channel changes. Disabling Conversion Modeling can only be achieved by selecting the 'Observed' reporting identity in property settings. Currently there is no warning indicating the inclusion of modelled data in a report

Behavioral Modeling will be included in the Data API if Consent Mode is enabled, the quantity of conversions meets Google's thresholds, and the reporting identity is set to 'Blended'.

Data API

GA4 APIs open the door to programmatic management of your account(s). There are four different GA4 APIs: Data API, Admin API, User Deletion API and Measurement Protocol. In this section we focus on the Data API, as this is the surface subject to data processing.

The Data API offers the optimal solution for automated account management and data extraction. For instance, it proves invaluable in monitoring Transaction Revenue and other currency metrics during periods of significant currency exchange rate fluctuations. It is important to note that the Data API adheres to the same limitations as Standard Reports

Note:

Many users might not realize that they use the Data API. If you use a Google Analytics 4 connector in tools like Looker Studio or Tableau, the data will be imported via the Data API. However, if you are unsure, please check how the data is imported into your tool, as it can also come via a BigQuery connector- or in another way. This is important because if you are retrieving the data from GA4 via the data API, the data will be refreshed when you retrieve it. However, if you query it from a database, the data will not be re-processed. This can lead to discrepancies in what you see in a dashboard and the GA4 interface.

The data available in the Data API mirrors that of Standard Reports, and is subject to the same limitations. Cardinality is possible, however there will be no (other) rows for funnel method. The (other) row can be returned for core reporting and real time. Equally with sampling, it may apply to funnel method. No sampling will be applied for core reporting and real time. GA4 360 customers can also take advantage of 'expanded datasets' to permanently open the "other" row entries in up to 100 reports. This is particularly useful when connecting to tools like Looker Studio, and retrieving this information on regular basis. Another option to remove Cardinality for GA4 360 users is to request unsampled reports.

Data API Cont...

Data extracted via Data API is subject to Thresholding if Google Signals is applied. Thresholding is there to protect user's privacy – the data thresholds are predefined within the system and cannot be modified or adjusted by the user. You can increase the date range to include more data, which can help to not trigger the warning.

As with the previous cases, Data API is based on Data Driven Attribution (DDA) model. You can select a different attribution model and conversion window at the property level to apply to all reports that use event–scoped traffic dimensions (for example, Source, Medium, Campaign, and Default Channel Group).

Data API, like Standard Reports, is subject to Conversion Modeling. Overall quantity of conversions remains unaffected, but rather the way they are assigned to a channel. Disabling Conversion Modeling can only be achieved by selecting the 'Observed' reporting identity in property settings. Currently there is no warning indicating the inclusion of modeled data in a report.

Behavioral Modeling will be included in the Data API if Consent Mode is enabled, quantity of conversions meets Google's thresholds and reporting identity is set to 'Blended'. For more information on API Dimensions and Metrics, click **HERE**.

Explore

Explore, also called 'Explorations', is a set of powerful techniques that go beyond typical reports. It helps users uncover valuable insights and understand customer behavior in a more in-depth way. Explore is accessible within the navigation panel on the left-hand side.
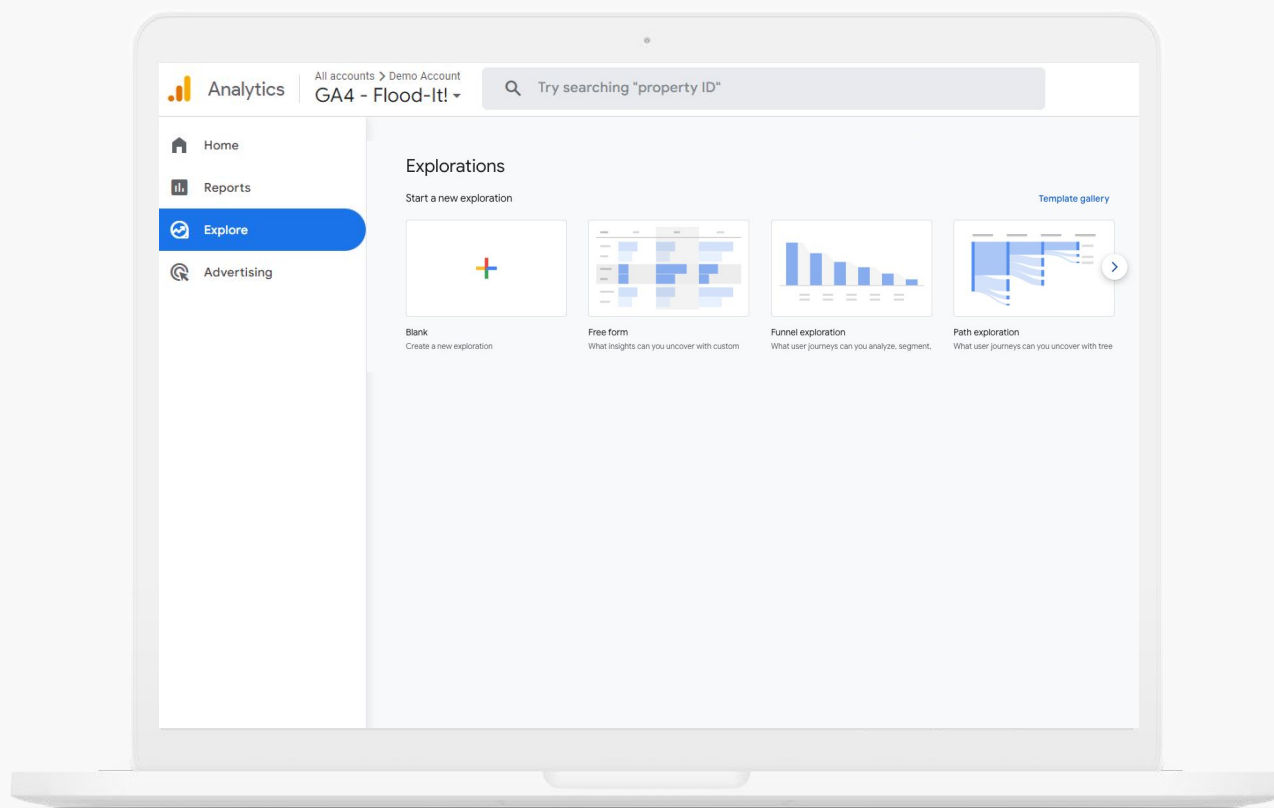
Explore reports are based on **event and user-level data** (raw data) and as such the queries might return somewhat different results than when comparing to the Standard Reports or API Reporting Surfaces. Explorations will not experience Cardinality (<u>see page 15</u>). They can, however, be subject to Sampling if the number of events returned by an exploration exceeds the limit for your property type.

Explore empowers users to find answers to their highly specific, on-the-fly, and advanced queries by harnessing a range of techniques like pathing, cohorts, and more. It is accessible for both Standard and 360 properties. 360 properties have the added advantage of adjusting sampling sizes to deliver quicker results.

Explore Cont...

Explorations, like Standard Reports and the Data API, are subject to Thresholding () if Google Signals is applied. Thresholding is there to protect user's privacy – the data thresholds are predefined within the system and cannot be modified or adjusted by the user. You can increase the date range to include more data, which can reduce the likelihood of thresholding.

Explore reports, like the previous two Surfaces, are based on Data Driven Attribution (DDA) models. You can select a different attribution model and conversion window at the property level to apply to all reports that use event-scoped traffic dimensions.

Explorations are also subject to Conversion Modeling. The overall quantity of conversions remains unaffected, but rather the way they are assigned to a channel. Disabling Conversion Modeling can only be achieved by selecting the 'Observed' reporting identity in property settings. Currently there is no warning indicating the inclusion of modeled data in a report.

Behavioral Modeling will be included in Explorations if Consent Mode is enabled, quantity of conversions meets Google's thresholds and reporting identity is set to 'Blended'. Behavioral Modeling will impact metrics like User, Sessions and Conversion counts.

If you do not want to include modeled data in the reports, choose the 'Observed' reporting identity in property settings.

At the point of writing this Whitepaper, no warning was present in reporting to flag whether modeled data is being presented.

10

BigQuery

BigQuery serves as the data warehouse that stores all the information underlying GA4. It houses the "raw data" that forms the foundation for GA4 reporting. Both GA4 360 and standard users have access to this functionality; however, to begin collecting data, the BigQuery export needs to be linked to a property.

Since BigQuery holds raw data, the numbers presented in this context can significantly diverge from what is available in the User Interface. There are no restrictions imposed on BigQuery in terms of Cardinality or Sampling. This means that neither limitations on the number of unique values nor data sampling will be encountered within BigQuery.

Please be aware that there is **no data related to Google Signals** in BigQuery. This may lead to the same user being counted multiple times across different devices, as Google Signals tracks users across various devices. As a result, when viewing the data in BigQuery, it is possible to observe a higher user count compared to reports based on Google Signals data.

It's important to note that the modeling features, such as DDA Model, Conversion Modeling, and Behavioral Modeling, are not available within BigQuery. Therefore, the traffic touchpoints in the data stored in this context will not be influenced by these modeling techniques.

When Consent Mode is enabled, the data stored in BigQuery includes cookieless pings collected by Google Analytics. Each session is assigned a unique user_pseudo_id. It's important to note that there may be variations between standard reports and granular data in BigQuery. For instance, the reports may show fewer active users compared to the data in BigQuery. This is because the modeling process aims to predict multiple sessions from an individual user who has chosen to decline cookies.

Note:

The data in BigQuery is likely to differ significantly from the reporting surfaces in the UI. This is typically because the modeling is not present in BigQuery. As a result, User, Session and Event counts are likely to be higher in BigQuery than in the GA4 UI.

Summary

Google defines Reporting Surfaces as Standard Reports, API, Explorations and BigQuery.

These four Reporting Surfaces display data in somewhat different ways, which is due to the differences in how the data is aggregated and processed. While Standard Reports and API use Aggregated data, Explorations and BigQuery use event and user-level data (raw data).

This means that data extracted from Standard Reports and API will be the same. The data source of Explore reports differs from Standard Reports and the API, and as such will deliver different results - however, the discrepancies are not expected to be significant.

Although BigQuery uses the same dataset as Explore, we are expecting potentially significant discrepancies between these two surfaces. This is because BigQuery does not include DDA, Google Signals, Conversion or Behavioral Modeling.

Note:

Each reporting surface in GA4 serves a distinct purpose and, as a result, handles data slightly differently. This can result in variations or discrepancies between the reporting surfaces. It's important to become familiar with the data processing procedures specific to each surface in order to gain a better understanding of these discrepancies and effectively explain them. By understanding the intricacies of data processing for each reporting surface, you can more accurately interpret and communicate any variations observed in the data.

In the following section, we will provide an overview of the data processing carried out by GA4 and delve into the specific ways it influences the surfaces to which it is applied. We will explore the details of this data processing and its impact on the various reporting interfaces.

Data Processing

Treatment of data varies across reporting surfaces, which can lead to discrepancies if two different reporting surfaces are compared. In this section we will go through the system data processing, but also highlight a couple of external factors which can lead to data discrepancies between surfaces.

GA4 Data Processing

Within GA4 reporting surfaces, there are two distinct data sources – aggregated and raw. However, beyond these sources, several additional data treatments come into play, exerting an influence on GA4 data. Some of these treatments have a direct impact on how data is presented in reports, such as Cardinality or Sampling, while others involve modeling user or session behavior, like Conversion and Behavioral Modeling.

In the upcoming section, we will delve into these data treatments, addressing whether the GA4 UI provides warnings when a treatment is applied and, where applicable, whether there are methods to circumvent these data treatments.

We discuss:

- ➔   Cardinality

- ➔   Sampling

- ➔   Thresholding when Google signals is enabled

- ➔   Attribution Modeling

- ➔   Conversion Modeling

- ➔   Behavioral Modeling

What is Cardinality?

Each report dimension (e.g. User source, User medium, User campaign, Gender, etc.) has a number of potential values that can be assigned to it. **The total number of unique values for a dimension is known as its cardinality.** Dimensions with a large number of possible values are known as high-cardinality dimensions. High-cardinality dimensions may affect quality and granularity shown in different parts of Analytics, depending on the system limits – this is 500 unique values in one day at the point of writing this Whitepaper. Examples of these dimensions could be page path, user id, etc.

The cardinality limit applies to all data collected in a property for the date range specified in the report, not just the data that's rendered in a given report. This is why you may have a report with a few rows but Cardinality would still be applied.

Note:

Data in both the Analytics reports and the Google Analytics Data API may be affected by the Analytics system limits, **resulting in data aggregating under the (other) dimension value**. This happens when the total values of all dimensions in a property exceeds 500 unique values per day; however, it must be noted that the metric totals for all the data retrieved will still be correct.

		↓ Views	Users	New users	Views per user	Average engagement time	Unique us...
	Page path + query string and screen class ▾ +						
	Totals	30,867,383 100% of total	3,038,420 100% of total	1,546,024 100% of total	10.16 Avg 0%	2m 47s Avg 0%	100...
1	(other)	9,687,419	2,307,830	573,197	4.20	1m 38s	
2	/	4,412,397	670,045	256,061	6.59	0m 41s	
3	StartupViewController	3,594,493	203,423	15,661	17.67	0m 17s	

Search...   Rows per page: 10 ▾   Go to: 1   ‹ 1-10 of 692 ›

* Most reports query smaller aggregate tables designed specifically for those reports. The row limit under these circumstances is 50K for GA4 standard and 1M for GA4 360. When you change the default query for a report, for example, by adding comparisons or a secondary dimension, the report may then query a larger aggregate table. The row limit under these circumstances is 2M.

15

What can we do about Cardinality?

Google has provided some new best practices to avoid Cardinality issues:

➔ Use default reports where possible. The default reports are supported by specific data sources, which will decrease the likelihood of data aggregating under the (other) dimension value.

➔ If your standard or 360 property shows the (other) row, **consider prioritizing the data that's most important to your business.** The cardinality limit applies to all data collected in the property for the date range specified in the report, not just the data rendering in one report. For example, *avoid collecting high-cardinality custom dimensions*, like a unique identifier for each user. Instead, use our User ID feature.

➔ If your standard properties hit the (other) row, consider upgrading to Analytics 360 to take advantage of automatic custom tables. For 360 properties, **automatic expanded data sets** are enabled automatically when you experience the (other) row so that data aggregates under (other) far less frequently.

➔ Google Analytics 360 users who see the (other) row in their reports can run the same queries in explorations. If an exploration begins to sample data, you can request unsampled exploration results.

16

What is Sampling?

In Google Analytics 4, the standard reports are always unsampled, except the funnel report which might experience sampling – this is because funnel reports load raw event and user-level data. The data API will be subject to the same limit as the Standard Reports.

However, when you build Exploration reports for analysis, you may experience sampling if your date range exceeded the last 60 days, or you are exceeding 10 million events (for free GA4) and the report you created is not a pre-existing standard report. Event limit for GA4 360 customers is up to 1 billion. In general, anything you do in Explore that replicates a standard report will be unsampled.

Note:

Sampling is indicated by a **red shield** in the top right hand corner of your report, but you need to expand to understand the detail.
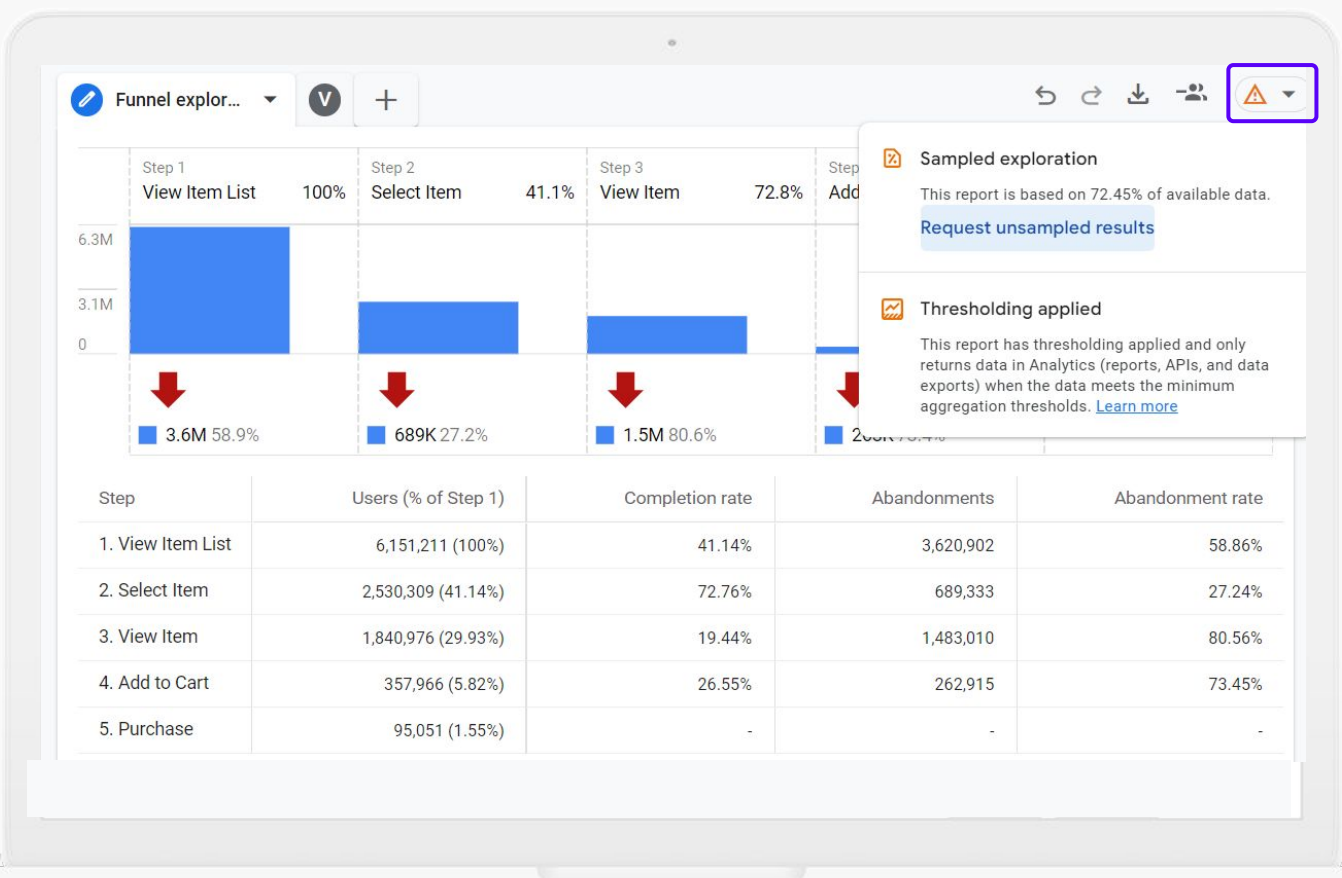
What can we do about Sampling?

Here are best practices to avoid Sampling issues:

➔ The quickest way to reduce sampling is to reduce the date range. This will automatically decrease the amount of data processed by the report, and decrease the likelihood of hitting sampling.

➔ If possible, reduce the quantity of metrics and dimensions your report is processing – this will reduce the quantity of data being calculated, and decrease the likelihood of sampling.

➔ Google Analytics 4 360 users can request unsampled results, therefore please consider upgrading if you are frequently experiencing these issues.

➔ Google Analytics 4 360 users can adjust the level of precision and speed: more detailed results will process up to 1B events to give the most precise results, while faster results will process 100M events – the latter is the default option.

➔ Use BigQuery – no sampling is present in BigQuery. However, please note that the data in BigQuery is unlikely to reproduce the other reporting surfaces because it does not contain data from Google Signals, DDA, Conversion Modeling or behavioral Modeling if you use Consent Mode.

➔ Please note: the use of filters will not reduce the overall data being processed, therefore this isn't a methodology which will reduce sampling.

Thresholding when Google signals is enabled

Thresholding occurs in GA4 when Google Signals is enabled, and applies to certain dimensions to protect users' privacy. Data thresholds are system defined, and cannot be adjusted. During our initial tests, we compared implementations with and without Google Signals enabled and found that the overall reported numbers were remarkably similar. Thresholding primarily resulted in the removal of rows with low volumes, rather than significantly affecting the overall metrics.
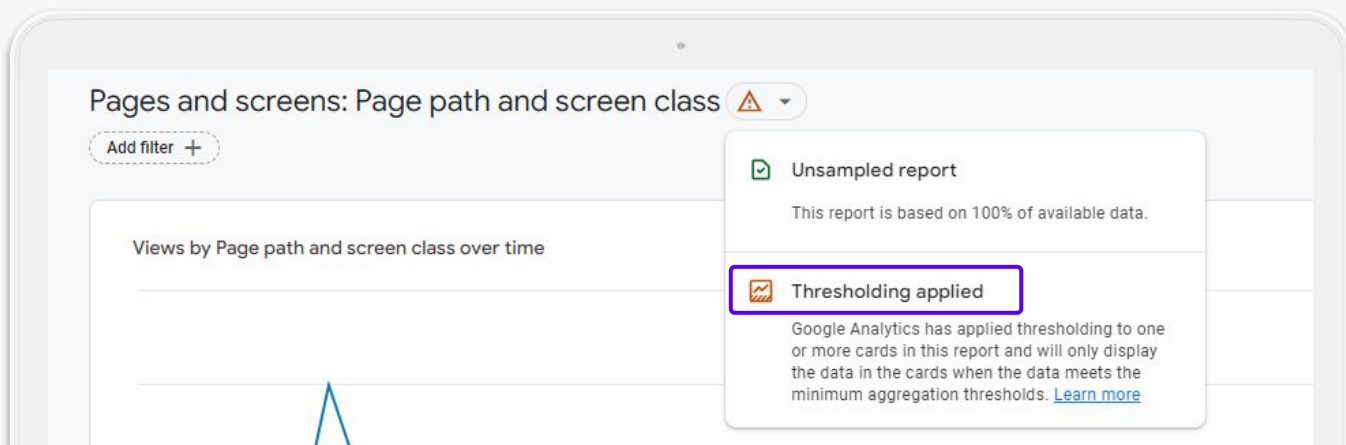
When viewing a report or an exploration within a limited date range and low user or event counts, data thresholds may be applied. By expanding the date range, you can potentially increase the number of users who triggered an event. This expanded range can allow you to access the previously thresholded data. This is more likely to be true for properties with high level of traffic - when tested, extending the date range on properties with 20,000 users per month didn't remove the thresholding on any date range between a week and three months.

Google Signals

You can't talk about Thresholding without talking about Google Signals. Google Signals comprises session data from websites and applications that Google links to users who have logged into their Google accounts and have enabled Ads Personalization. This association of data with these authenticated users facilitates cross-device reporting, cross-device remarketing, and the ability to export conversions across devices to Google Ads.

19

Thresholding Cont...

Google Signals can be disabled in GA4 properties, however, it is not recommended. Google Analytics for Audiences in Google Ads can't be used if Google Signals is disabled and any information relating to gender, age or other demographic data will no longer be available.

Turning off Google Signals is not retroactive, which means that the past data will still be affected by thresholding.

What can we do about Thresholding?

Here are best practices to avoid Thresholding:

➔ The quickest way to reduce Thresholding is to increase the date range. This will automatically increase the amount of users and events processed by the report, and therefore decrease the likelihood of hitting Thresholding. This might not work for properties with moderate level of traffic.

➔ Google Signals can be disabled, however it is not recommended - the drawbacks are described on the previous section.

➔ Another option is to switch the Reporting Identity to device-based - meaning that the overall user counts will be less accurate, however you will not experience Thresholding, as Google Signals is not included in this Reporting Identity. The benefit of this is that Reporting Identity is retroactive, therefore it will re-process data already collected, and remove the Thresholding. This can be changed as many times as desired without impact.

> **Note:**
>
> This methodology isn't without disadvantages. Testing on one property, the amount of overall 'views' for 'page path and screen class' increased by 23%, while 'event count' declined by 15%, and conversions increased by 165% in one month timeframe. Property with 20k users, 820k events 2.4k conversion per month, comparing 'pages and screens' standard report.

Reporting Identity

Determines which methods Analytics uses to associate events with users. Learn more

How would you like to identify your users?

⦿	Blended		ⓘ 2 inactive methods	⌄
	This identity evaluates: user ID, Google signals, device ID, modeled data.			
◯	Observed		ⓘ 1 inactive method	⌄
	This identity evaluates: user ID, Google signals, device ID.			

Save   Cancel   Show all

Note: the 'Device-based' reporting identity doesn't show up as an option when first navigating to the Reporting Identity section. To view the 'device-based' option, you need to click on 'Show All' in right-hand bottom corner, below the 'Blended' and 'Observed' options. This will show the 'device-based' option.

Attribution Modeling

Attribution involves the process of assigning credit for conversions to various advertisements, clicks, and factors that contribute to a user's journey in completing a conversion. When a new GA4 property is created, the default Attribution Model is Data Driven Attribution (DDA). You can change this model, as well as adjust the lookback window, in the 'Attribution Settings' section in property settings.

Typically Attribution will only impact comparison between BigQuery and other surfaces, as BigQuery does not have an attribution model included in the data. There might be small discrepancies between Explore reports and Standard Reports and API, however it is expected that these discrepancies would be insignificant and come from the difference in the data source (raw vs aggregated). Please note: attribution channel can be updated up to 7 days after the conversion is recorded. This is because Analytics is processing data and using it for model training. This can lead to data discrepancies within the same report at different times.

Note:

Starting in May 2023, the first click, linear, time decay, and position-based attribution models are being phased out.

While there is no warning that an attribution model is in use, it can be assumed that this is the case if conversion metrics are present and it's not stated otherwise. For example – Acquisition reports 'Overview' and 'User Acquisition' report 'first user default channel group', however it is clearly described in the interface.

The Conversions details reports and Explorations are among the reports impacted by changes to the attribution model. In Explorations, you can find a comprehensive list of dimensions that are compatible with attribution. It's important to note that user- and session-scoped traffic dimensions, such as Session source or First user medium, remain unaffected by modifications to the reporting attribution model.

Attribution Modeling Cont...

Depending on the attribution model chosen, certain metrics will experience changes when used with event-scoped traffic dimensions. These metrics include Conversions, Total revenue, Purchase revenue, and Total ad revenue. When transitioning to a non-last click attribution model, you may encounter decimals or fractional credit for the first time in these columns. This is because credit for a specific conversion is distributed among contributing ad interactions in accordance with the chosen attribution model.

Conversion Modeling

Conversion Modeling estimates online conversions that can't be observed directly, for example due to the browser restrictions like ITP/ETP/3P cookie deprecation, ATT regulations, x–device and consent. It uses Google's AI to quantify the impact of marketing efforts when a subset of conversions can't directly be tied to a traffic source. This feature affects data after July 2021 – data before that isn't impacted.

Conversion Modeling will, much like Attribution, impact Standard Reports, API and Explore but not BigQuery. This means that BigQuery will record the same amount of overall conversions, but they will not be matched to the same channels. Conversion Modeling allows to model the channel based on a sub–set of observable conversions. Google uses holdback validation methodology to ensure accuracy of Modeling.

> ### *Note:*
>
> Conversion Modeling does not change the overall amount of conversions recorded by Google Analytics. It will model only observable conversions – the amount in BigQuery and other surfaces will be the same (subject to the aggregated vs raw data source differences). The difference will be which channels the conversions are assigned to.

It is also important to mention if Conversion Modeling is in use, the numbers can change up to 7 days after a conversion has occurred. **For increased accuracy, Google recommends to select a date range beyond or prior to the previous week.**

At the point of writing this Whitepaper there was no warning in the reports whether the data presented is impacted by Conversion Modeling. Conversion Modeling will be applied automatically if the property meets the limits and Google deems the modeling accurate. If there isn't a discrepancy between BigQuery channels and what is seen in the UI, it is likely that Conversion Modeling hasn't been applied for that property. There is currently no way to disable this feature or distinguish modeled from observed data. Disabling Conversion Modeling can only be achieved by selecting the 'Observed' reporting identity in property settings.

24

Conversion Modeling Cont...

It is important to mention that Conversion Modeling will not model conversions based on unconsented pings – this is done by behavioral Modeling, and requires Consent Mode to be present. Conversion Modeling only works on consented users and does not change the overall volume of conversions observed by Google Analytics.

Behavioral Modeling

Behavioral Modeling is specifically used to model the data of users who decline consent, developed to help marketers tackle the loss of data when users decline analytics cookies. In order to apply Behavioral Modeling in your property, you must install the advanced implementation of Consent Mode and meet the modeling criteria. Like Conversion Modeling, Behavioral Modeling is available in Standard Reports, API and Explore, but not in BigQuery. This means that the quantity of users or sessions will be higher in BigQuery than in the other surfaces.

This is because BigQuery data contains cookieless pings collected by Google Analytics when consent mode is enabled and each session has a different user_pseudo_id. Modeling may lead to differences between standard reports and granular data in BigQuery, such as fewer active users in reports than BigQuery as modeling tries to predict multiple sessions from an individual user who declined cookies.

There is currently no warning in the reports to indicate that Behavioral Modeling is applied. To disable it, users can select 'observed' reporting identity, which will remove the unobserved data. This process is retroactive – it will affect historical data – and can be reverted to 'blended' (or device) at any point. The underlying data is not modifies by selecting a different reporting identity – only what is displayed changes.

The following criteria needs to be met for Behavioral Modeling to be present in GA4:

➜ **The property collects at least 1,000 events per day with analytics_storage='denied' for at least 7 days.**
➜ **The property has at least 1,000 daily users sending events with analytics_storage='granted' for at least 7 of the previous 28 days.**
➜ **It may take more than 7 days of meeting the data threshold within those 28 days to train the model successfully; however it's possible that even the additional data won't be sufficient for Analytics to train the model.**

Behavioral Modeling starts from the date a given property becomes eligible, and data will not be available should the property stop meeting eligibility criteria.

Useful Resources

Google has create a host of useful resources to help understand GA4 in all contexts. Here are the top assets used while writing this article:

VIDEO (90 mins) – 'Understanding your Google Analytics 4 data better' – GA4 walk through with product creators Googlers outlining the differences to UA, Reporting Surfaces, and other question. **WATCH NOW**

VIDEO (60 mins) – 'GA4 Launches & Reporting Surfaces' – Mikel and Romain discuss Reporting Surfaces and show a demo of the UI. **WATCH NOW**

PLAYBOOK – GA4 Reporting Playbook is a helpful guide to better understanding the reporting in GA4. **READ NOW**

ARTICLE – Reporting Surfaces Comparison – the absolute basic to understand different Reporting Surfaces and the differences in data treatment between them. **READ NOW**

RESOURCE HUB – A helpful resource for comparing metrics, mostly UA to GA4, and understanding the different metrics definitions between the products. **READ NOW**

GUIDE – How conversion modeling works – a simple guide to how conversions are modeled in GA4. **READ NOW**

ARTICLE – Behavioral Modeling for consent mode – great explanation of how Modeling is engaged when Consent Mode is enabled. **READ NOW**

ARTICLE – GA4 and GA4 360 limitations – a handy comparison of limits for enterprise and non-enterprise GA4 versions. **READ NOW**

BLOG – We can't fail to mention Simo Ahava's blog. The technical guru of Analytics, his articles are extremely detailed and high-quality guides to anything GA4 related. **READ NOW**

INCUBETA

Incubeta is a marketing partner built specifically to help businesses upgrade their growth. An international team of experts in marketing, technology, data and creative, Incubeta is a team of over 800 creators, thinkers, makers, and doers.

Present in 20 offices worldwide, Incubeta is a market leading specialist that uses the power of digital to unlock and amplify business growth potential through bespoke, localized digital solutions with global expertise. With 20+ years of experience in the digital space, Incubeta puts the ownership and control of the customer experience back in the hands of the advertiser.

As the largest and longest established Google Marketing Platform (GMP) sales partner in EMEA and APAC, Incubeta has proven that they are leading specialists by working with top-tier brands and helping advertisers tackle complex projects such as in–housing, personalization, data activation, and advanced digital media.

Author: Monika Mesnage, Analytics Manager

Get in touch: analytics@incubeta.com