Varify.io - Happy Horizon

BigQuery for Experiment Analysis

Let's meet

Your hosts

Steffen Schulz
Co-founder Varify.io

Bas Linders
Product Owner CRO | Data scientist

happy horizon

We’re Happy Horizon

The horizon. That’s your higher purpose.
The destination you’re always on your way to.
The goal that seems out of reach, but always gives you direction.
And determines your next step.

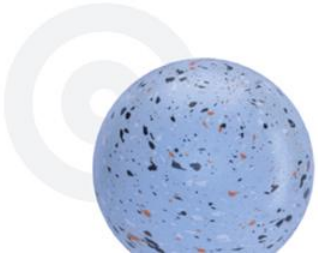
Our offices

Our ambition is to be a 'one-stop-shop' in the digital field. Spread throughout the Netherlands (and beyond). So that we are always near our clients and talents. Our group currently consists of 7 hubs:

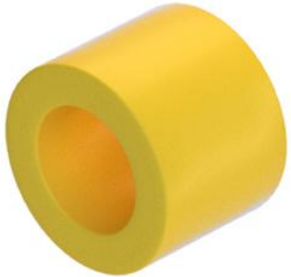
Eindhoven, Amsterdam, Tilburg, Apeldoorn, Arnhem, The Hague and Utrecht

We’re experts in

- Strategy & Innovation
- Data & Automation
- Branding, Content & PR
- Online Marketing
- Development & Technology



Our clients



MSD
NOVARTIS
abbvie
AstraZeneca
teva
Sanquin
SANOFI GENZYME
Medtronic
pidz.
UMC Utrecht
VIFOR PHARMA
WEBER
Incyte
Hygienius
U CENTER
invisalign
KYOWA KIRIN
LEO
SANDOZ
Galápagos
ORGANON
STICHTING KIJKSLUITER
St. Anna Zorggroep
SANANET
Seagen
eurofins
CSU
CONTINU PROFESSIONALS
NIERSTICHTING
Je nieren zijn je leven.
fatboy
SHIMANO
ESDEC
INNOVATIVE MOUNTING SYSTEMS
Rode Kruis
Fontys
HET VERGETEN Kind
Sligro
Sligro Food Group
PSV
Eneco
Luxaflex
sanidirect
pets place
expert
vidaXL
ABAB
ACCOUNTANTS EN ADVISEURS
Leger des Heils
WARNER MUSIC GROUP
KWF KANKER BESTRIJDING
Roche
woon EXPRESS
Hout handel Online.nl
GRANDO KEUKENS & BAD
MSD
KWANTUM
Daar woon je beter van!
PG PinkGellac
nederlandse brandwonden stichting
VISIT BRABANT
SISSY-BOY
Intergamma
LOAVIES
xenos

VARIFY.iO® - A/B Testing Platform

- **Unlimited Traffic**
- **No Cookies**
- **Single Source of Truth**
- **Big Query Integration**

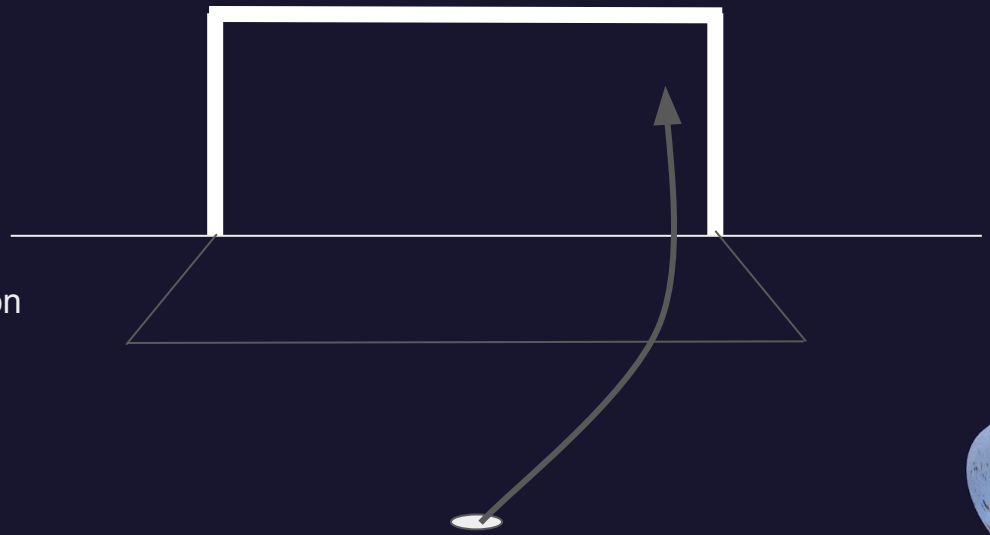
OMR REVIEWS
TOP RATED
CONVERSION OPTIMIZATION
Q1/25

MADE & HOSTED IN GERMANY

CARGLASS® DUDEN ELLE NIVEA FOCUS ESN® ELITE SPORTS NUTRITION

Our topics today

- Experimentation & data collection
- A single source of truth
- Differences between GA4 & BQ
- Connecting experiments & BQ
- Tips / How to (not) fail



“Left untouched, all websites will decline in performance over time.”

- Matt Scaysbrook

BigQuery for Experiment Analysis

Experimentation & Data Collection

happy horizon

CREATIVE DIGITAL AGENCY

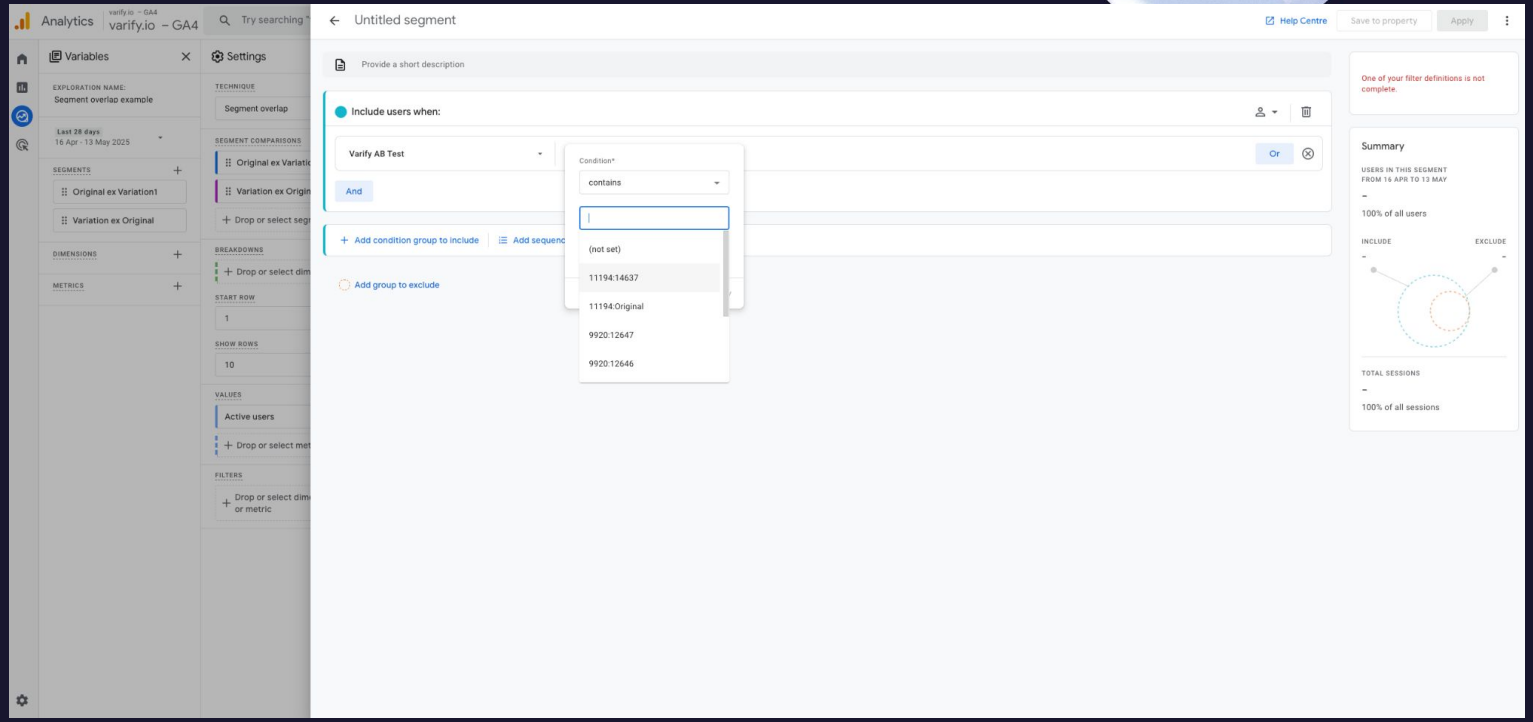
Why even bother with other sources?

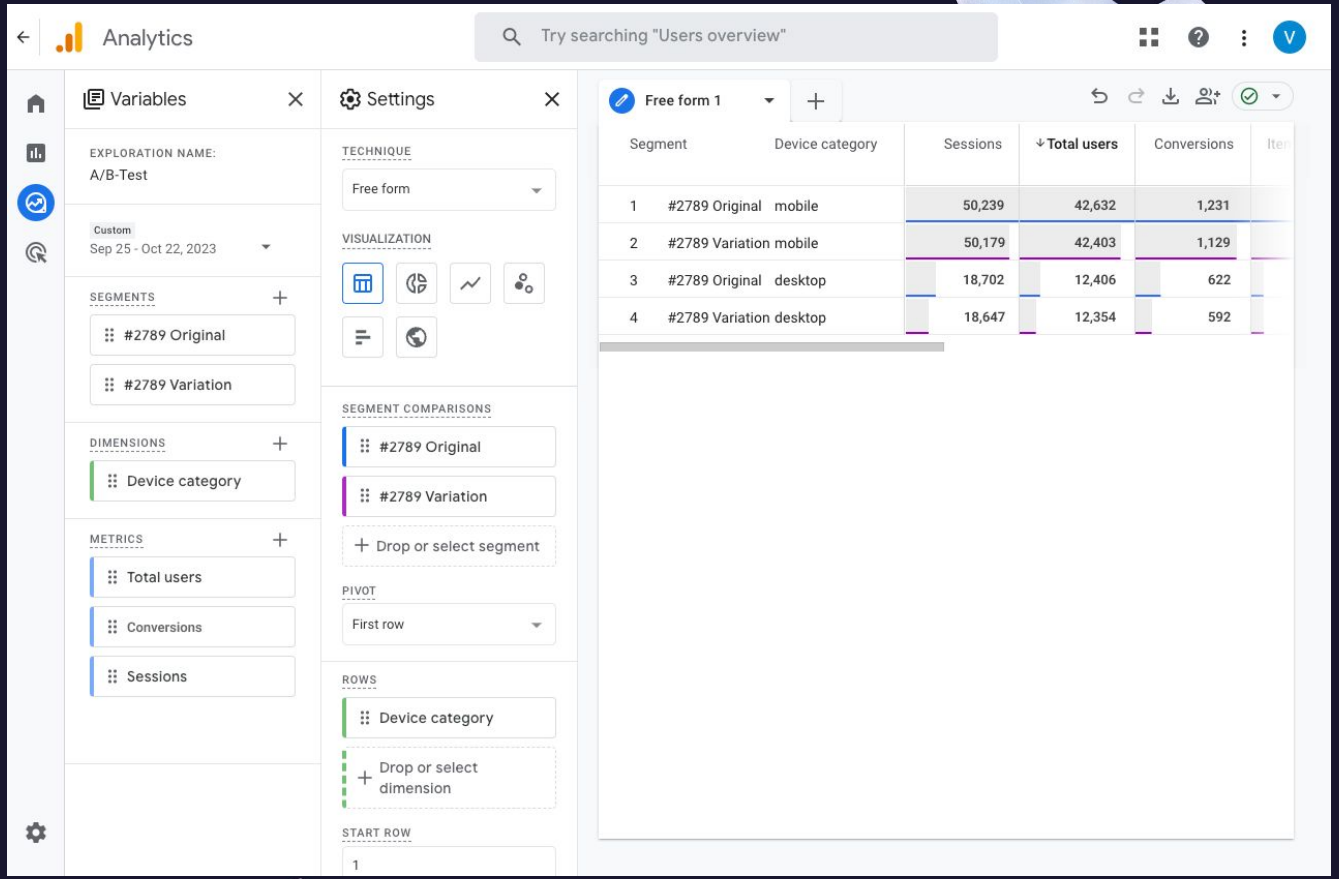
- **Single Source of Truth:** Conflicting data can arise between tools
- **Deeper Insights:** Some tools offer deeper segment analysis (traffic, source, medium)
- **Data Ownership:** Limited control over data interpretation in many tools
- **Goal & Event Setup:** Requires technical expertise, varies by tool
- **Black Box:** Sampling methods and data transparency differ across platforms

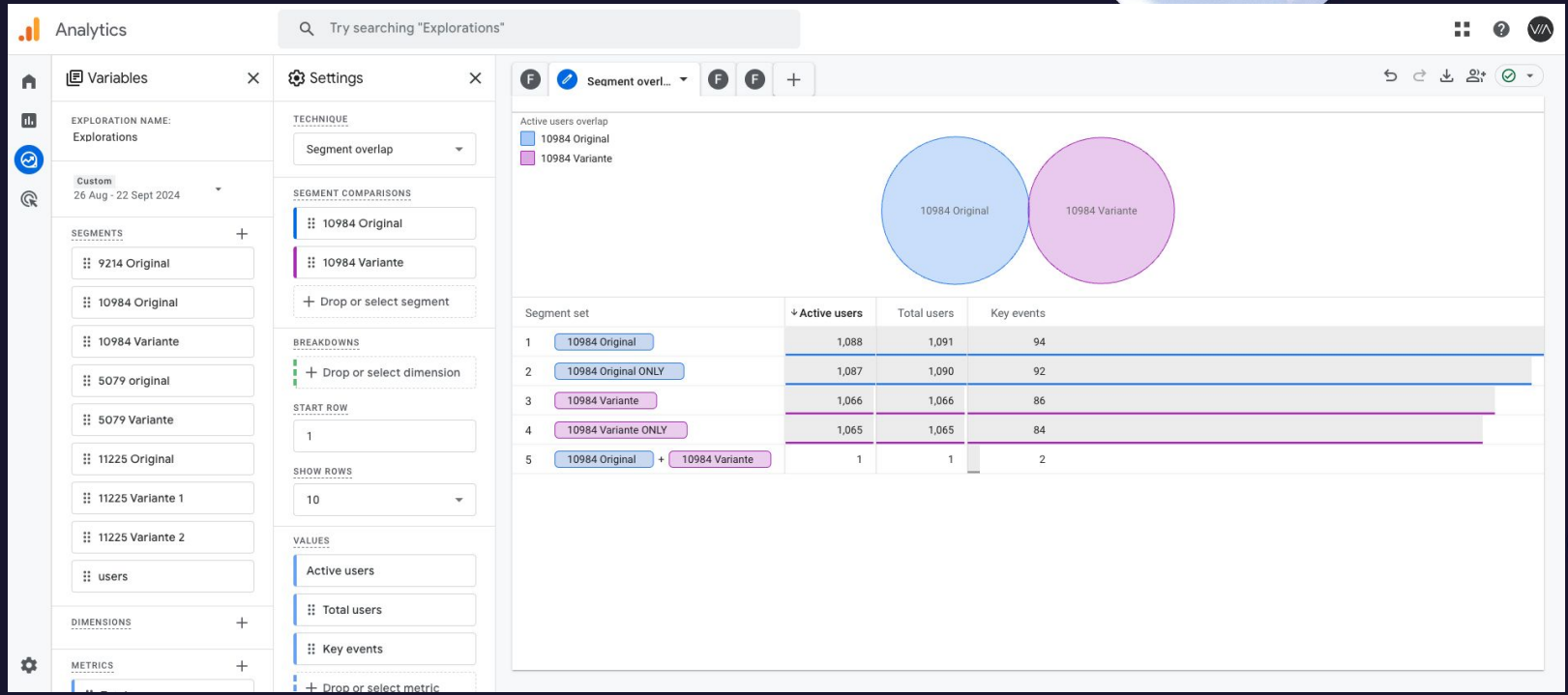
GA4 reporting

A route many experimenters take, because it’s relatively easy to implement.

GA4


Analytics
varify.io – GA4
Try searching
Variables
Settings
EXPLORATION NAME:
Segment overlap example
Last 28 days
16 Apr - 13 May 2025
SEGMENTS
Original ex Variation1
Variation ex Original
DIMENSIONS
METRICS
TECHNIQUE
Segment overlap
SEGMENT COMPARISONS
BREAKDOWNS
START ROW
1
SHOW ROWS
10
VALUES
Active users
FILTERS
Untitled segment
Help Centre
Save to property
Apply
Provide a short description
Include users when:
Varify AB Test
Condition*
contains
Or
And
(not set)
11194:14637
11194:Original
9920:12647
9920:12646
Add condition group to include
Add group to exclude
One of your filter definitions is not complete.
Summary
USERS IN THIS SEGMENT
FROM 16 APR TO 13 MAY
100% of all users
INCLUDE
EXCLUDE
TOTAL SESSIONS
100% of all sessions




Analytics
Try searching "Explorations"
Variables
EXPLORATION NAME:
Explorations
Custom
26 Aug - 22 Sept 2024
SEGMENTS
9214 Original
10984 Original
10984 Variante
5079 original
5079 Variante
11225 Original
11225 Variante 1
11225 Variante 2
users
DIMENSIONS
METRICS
Settings
TECHNIQUE
Segment overlap
SEGMENT COMPARISONS
10984 Original
10984 Variante
Drop or select segment
BREAKDOWNS
Drop or select dimension
START ROW
1
SHOW ROWS
10
VALUES
Active users
Total users
Key events
Drop or select metric
Segment overl...
Active users overlap
10984 Original
10984 Variante
10984 Original
10984 Variante
Segment set
Active users
Total users
Key events
1 10984 Original 1,088 1,091 94
2 10984 Original ONLY 1,087 1,090 92
3 10984 Variante 1,066 1,066 86
4 10984 Variante ONLY 1,065 1,065 84
5 10984 Original + 10984 Variante 1 1 2

VARIFY.IO

Functions Prices For agencies Company Login Try for free Get a demo EN

A/B Test - Significance Calculator

More information

Control Visitors: 1091

Control Conversions: 94

Variation Visitors: 1066

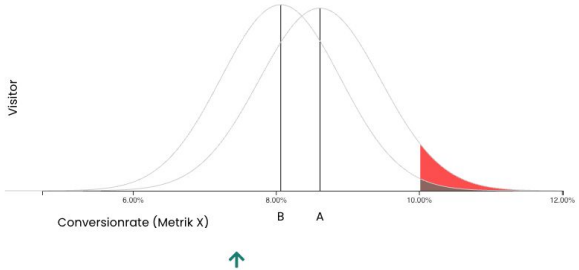
Variation Conversions: 86

Confidence Level: 95%

Calculate

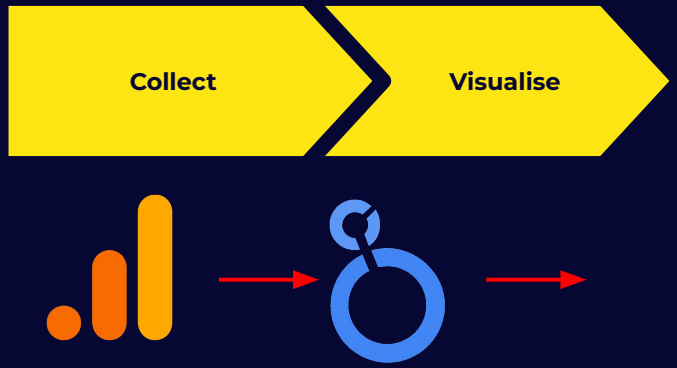
	Conversion Rate	Uplift (%)	Significant
Control	8.62%	--	--
Variation	8.07%	-6.37%	No

p-value: 0.3226
Confidence level: 67.74%



GA4 reporting

A route many experimenters take, because it’s relatively easy to implement.

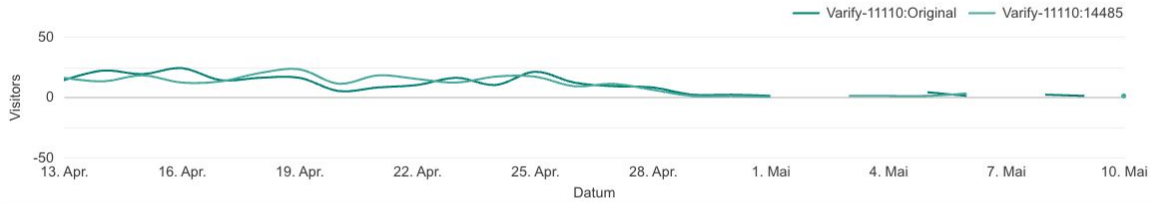


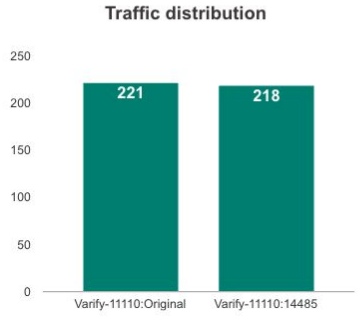
GA4 Looker Studio

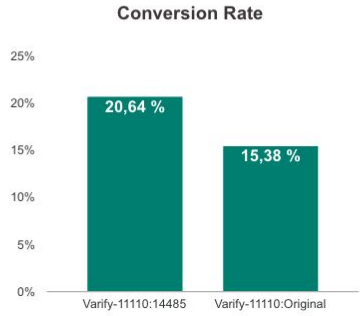
Experiment: varify-11110 (1) | 13.04.2025 - 10.05.2025 | Device | Reset

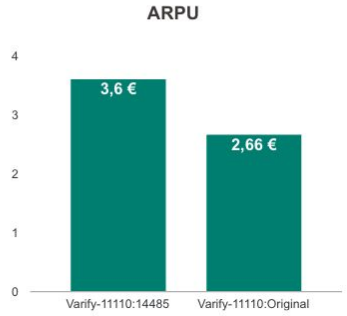
VARIFY.IO

Variation	Nutzer insgesamt	Purchases	Conversion Rate	Revenue	ARPU
Varify-11110:Original	221	34	15,53 %	582,16 €	2,66 €
Varify-11110:14485	218	45	20,93 %	774,12 €	3,6 €









Experiment ID: 11110

13.04.2025 - 10.05.2025

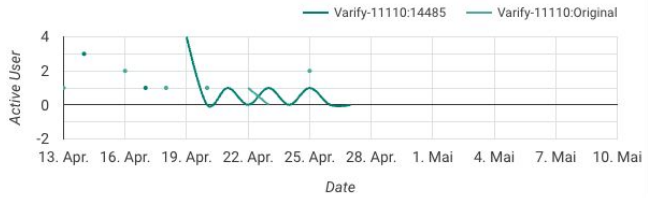
Device

Reset

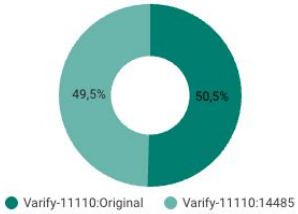
VARIFY.IO

Event: purchase (1)

Variation	Visitors	Event Count
Varify-11110:Original	219	34
Varify-11110:14485	215	45



Traffic distrbution



Signifikanzniveau:* 5%

Testfragestellung: einseitig

	Stichproben-größe	Erfolgreiche Versuche	Erfolgs-quote	Effekt	Konfidenz (P-Wert)	Signifikant* (ja / nein)
Kontrollgruppe						
Zielgruppe						

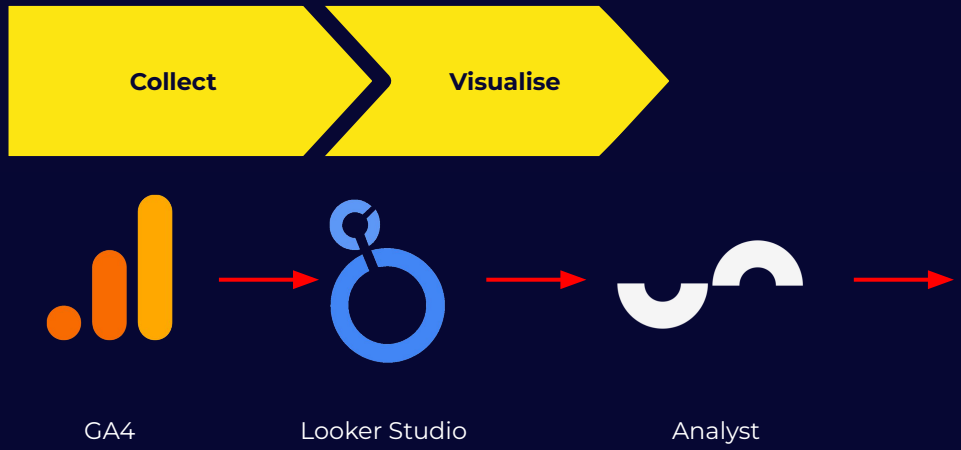
Signifikanz berechnen

CREATIVE DIGITAL AGENCY

happy horizon

GA4 reporting

A route many experimenters take, because it's relatively easy to implement.



The issue with GA4 data

Users | B | V | +

Segment	594 control	594 treatment
Naam gebeurtenis	Actieve gebruikers	Actieve gebruikers
Totalen	**51.022**	**50.162**

Visitors

How many visitors did variant A have?

51022 − +

How many visitors did variant B have?

50162 − +

Expected Percentage

What percentage of users should be in variant A?

50,00 − +

What percentage of users should be in variant B?

50,00 − +

Check for Sample Ratio Mismatch

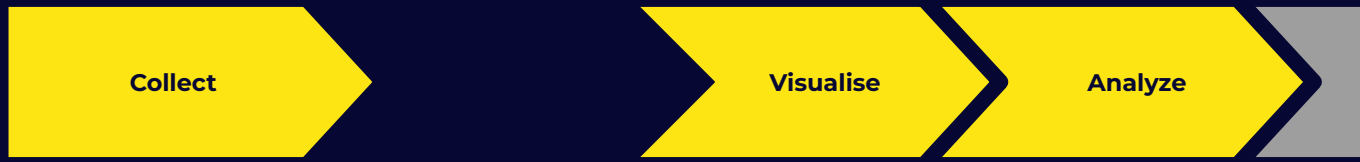
Conclusion

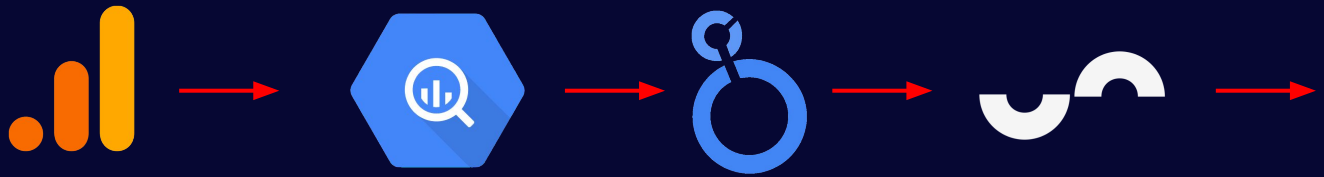
P-value: 0.0069. The expected amount of visitors per variant on average is 50592.

This suggests a **possible sample ratio mismatch**! The distribution of data between your variants significantly deviates from the expected proportions of [0.5, 0.5]. Check the distribution.

BigQuery reporting

The route everybody who's serious about experimentation should take, because it's the most accurate one.

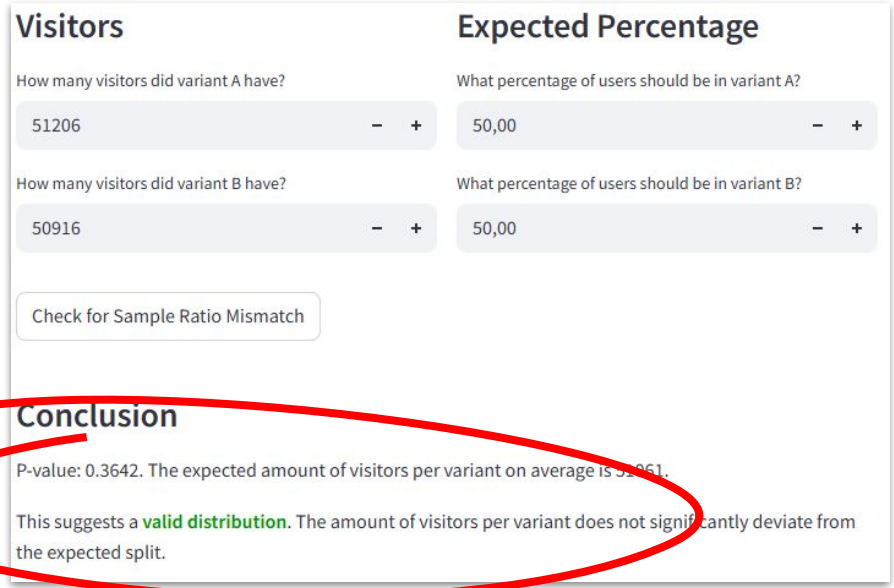




GA4 BigQuery Looker Studio Analyst

BigQuery data

```sql
WITH variant_data AS (
  SELECT
    user_pseudo_id AS variant_user_pseudo_id,
    CASE
      WHEN params.value.string_value = variant_a THEN 'A'
      WHEN params.value.string_value = variant_b THEN 'B'
      ELSE params.value.string_value
    END AS exp_variant_label
```



BigQuery for Experiment Analysis

Differences between GA4 and BQ

What is BigQuery?

"An enterprise system used for the **analysis and reporting** of structured and semi-structured data from multiple sources"

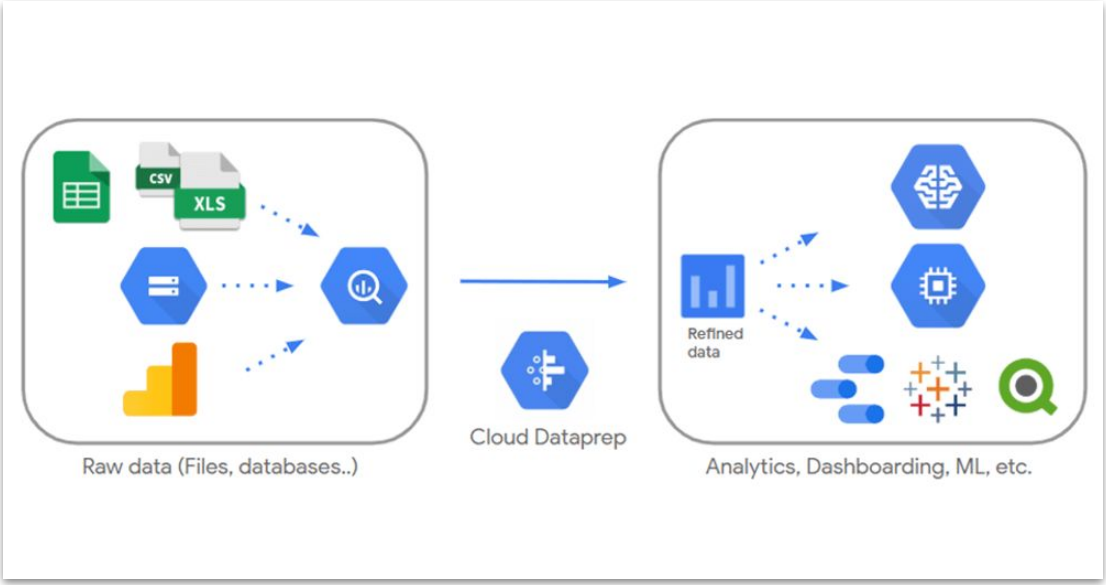


Image source: alteryx.com

BigQuery vs. Google Analytics 4

- Stores **raw event-level data**
- Data is always **unsampled**
- Data is organized in **daily tables**, with an event-based schema
- Requires writing **SQL** to access and analyze data
- Data is typically exported the **following day** (but can take up to 72 hours to finalize)
- It's **free** (up to the predefined limits)

- Presents **processed** and aggregated data
- **Can apply sampling** in reports
- Data is presented through **predefined reports** and customizable explorations
- Uses a **graphical user interface**
- Data can take **up to 48 hours** to fully process
- It's **free** (with GA4 360 as an upgrade)

BigQuery vs. Google Analytics 4

- Stores **raw event-level data**
- Data is always **unsampled**
- Data is organized in **daily tables**, with an event-based schema
- Requires writing **SQL** to access and analyze data
- Data is typically exported the **following day** (but can take up to 72 hours to finalize)
- It's **free** (up to the predefined limits)

- Presents **processed** and aggregated data
- **Can apply sampling** in reports
- Data is presented through **predefined reports** and customizable explorations
- Uses a **graphical user interface**
- Data can take **up to 48 hours** to fully process
- It's **free** (with GA4 360 as an upgrade)

In short

BigQuery 

Provides **maximum flexibility and depth**, requiring technicals skills (and potentially incurring cloud computing costs).

Google Analytics 4 

Provides **convenience and pre-processed insights** with limitations.

Incoming **STATISTICS**

😱📊

Unseen GA4 pitfall: HLL(++)

HyperLogLog is a **probabilistic algorithm** that estimates unique user counts to preserve memory, but there are issues for experiment analysis.

“The results from any A/B test with a decent sample size which relies on user-based metrics extracted from a Google Analytics property with the User Analysis toggle switched on are completely compromised.”

This starts being noticeable when a variant exceeds 12.000 users. Using BigQuery to extract raw user counts is, in most cases, the **most accurate way** to extract experiment results.

$$E^{\star} = -2^{32} \log\left(1 - \frac{E}{2^{32}}\right)$$

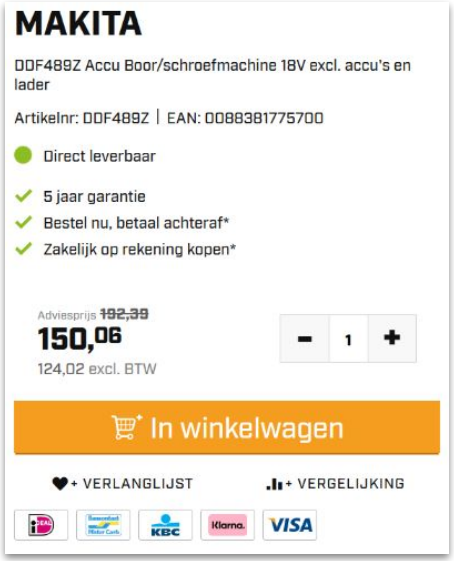
Text source: blog.analytics-toolkit.com
Image source: Wikipedia

BigQuery for Experiment Analysis

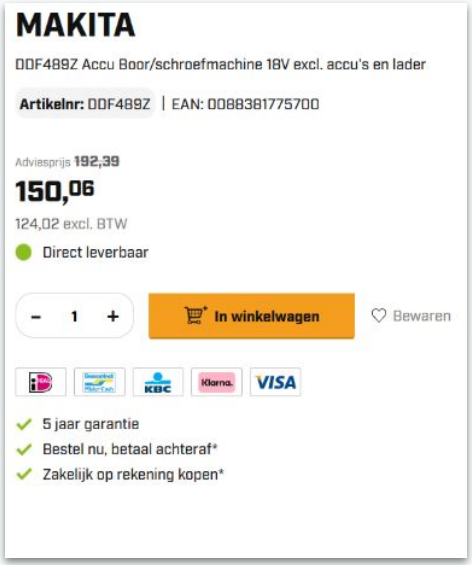
A single source of truth

And why BigQuery is the better choice (in most cases)

A/B test: Buybox redesign



VS



A

B

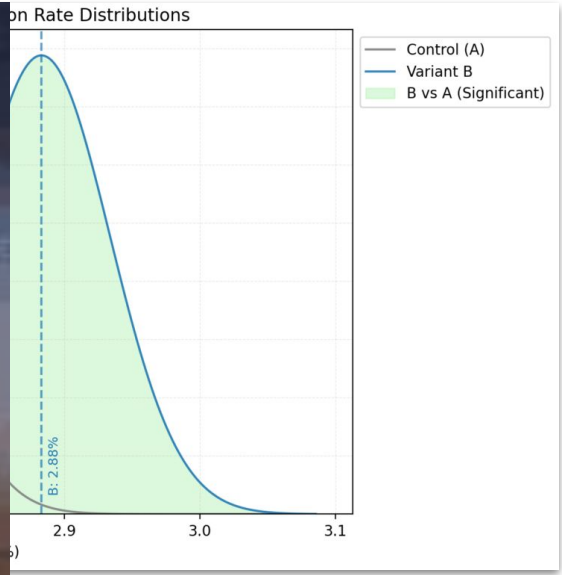
Results with GA4 Reporting

Experience variant	Users
A	107742
B	106554

Test results for B vs A

- **Statistically significant result** for B with
- **Observed power**: 62.56%
- **Conversion rate change** for B: 5.12%

Variant **B** is a **winner**, congratulations!



Chance to win: 97,5%

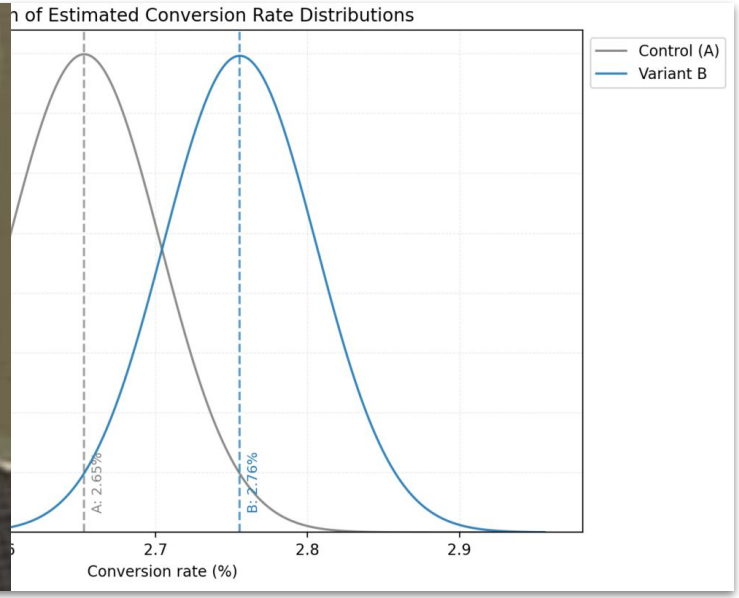
Results with BigQuery Reporting

Experience variant	Users	T
A	105495	
B	104809	

Test results for B vs A

- **Confidence interval for difference in conversion rates:** (-0.0004, 0.0024)
- **Observed power:** 42.11%
- **p-value:** 0.0741
- **Conversion rate change for B:** 3.86%
- **P-value (non-inferiority test):** 0.0000

Although the Z-test is not statistically significant (p = 0.0741), the non-inferi
is **not significantly worse** than A within the predefined margin. This sugges
perform similarly to the control.



Chance to win: 92,49%

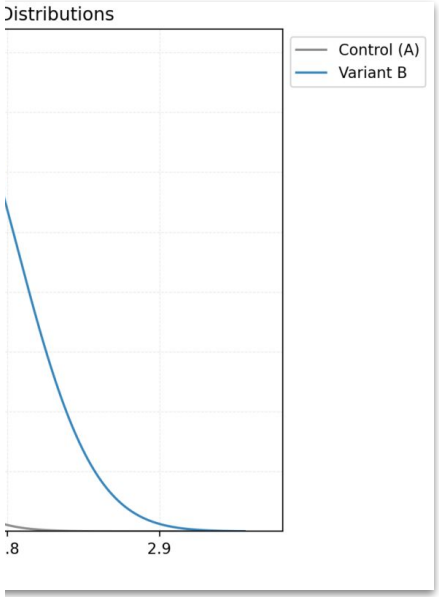
Results with BigQuery Reporting

Experience variant	Use
A	
B	

Test results for B vs A

- **Confidence interval for difference in conversion rat**
- **Observed power:** 42.11%
- **p-value:** 0.0741
- **Conversion rate change for B:** 3.86%
- **P-value (non-inferiority test):** 0.0000

Although the Z-test is not statistically significant (p = 0.
is **not significantly worse** than A within the predefinec
perform similarly to the control.



Small Differences - Big Impact

GA4 modeled part of the data. This estimation of cardinality inflated total visitors and adjacent conversions, **exaggerating the true effect**.

- Visitors +1,89%
- Conversions +3,14%

This is called a **Type-M error** (= an over-estimation of the effect size)

In other words: overestimating effects doesn't just mislead - it can lead to costly decisions and real business risks.

Source: Assessing type S and type M errors - Gelman, Carlin (2014)

BigQuery for Experiment Analysis

Connecting experiments and BigQuery

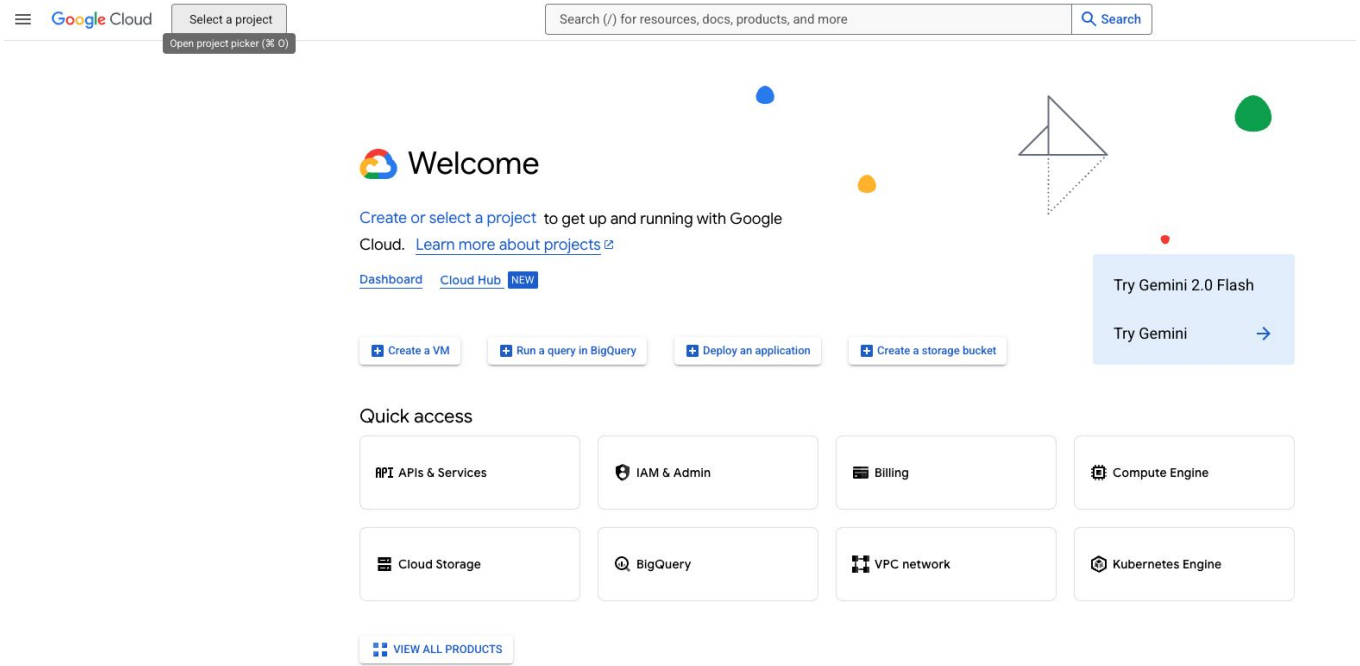
happy horizon

CREATIVE DIGITAL AGENCY

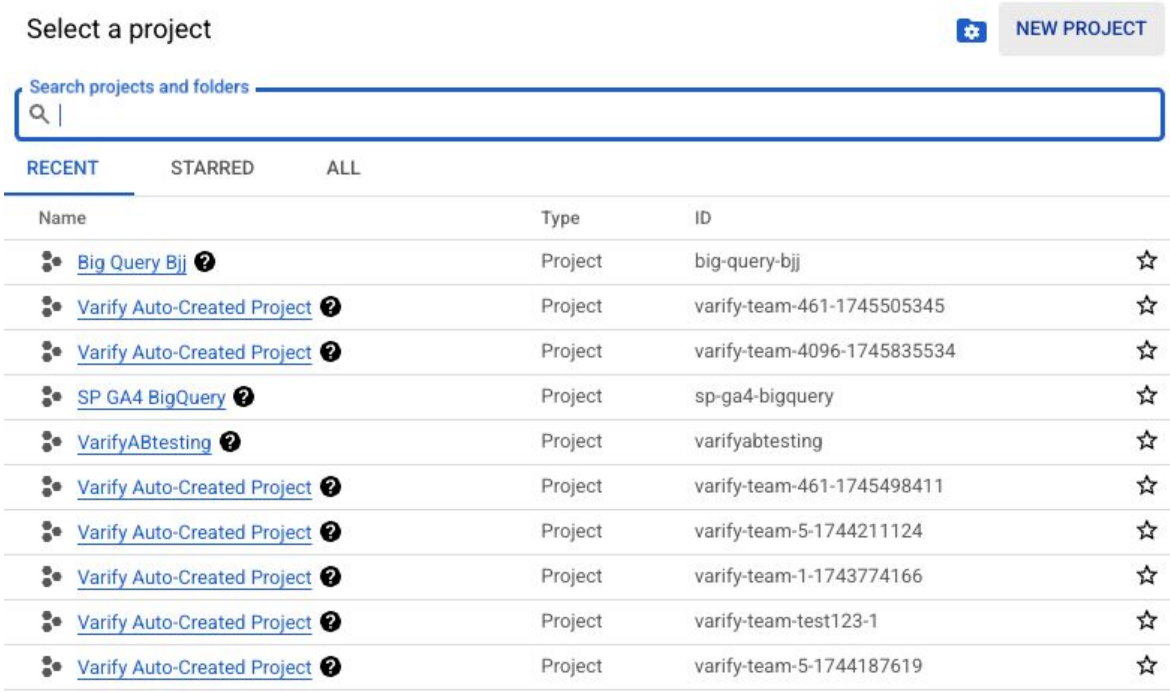
Connect GA4 to BigQuery

Because **reporting** is flawed, **not the collecting mechanism** (if implemented correctly).

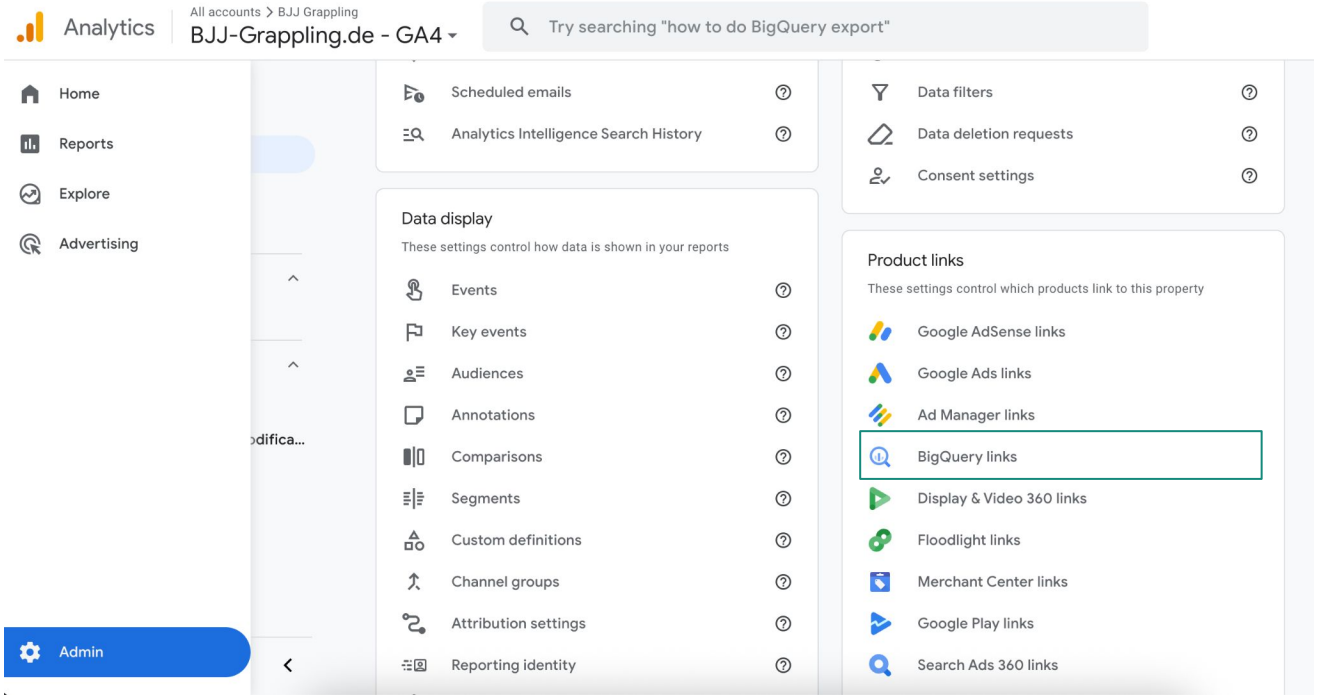
How-To:
Connect GA4 to BigQuery



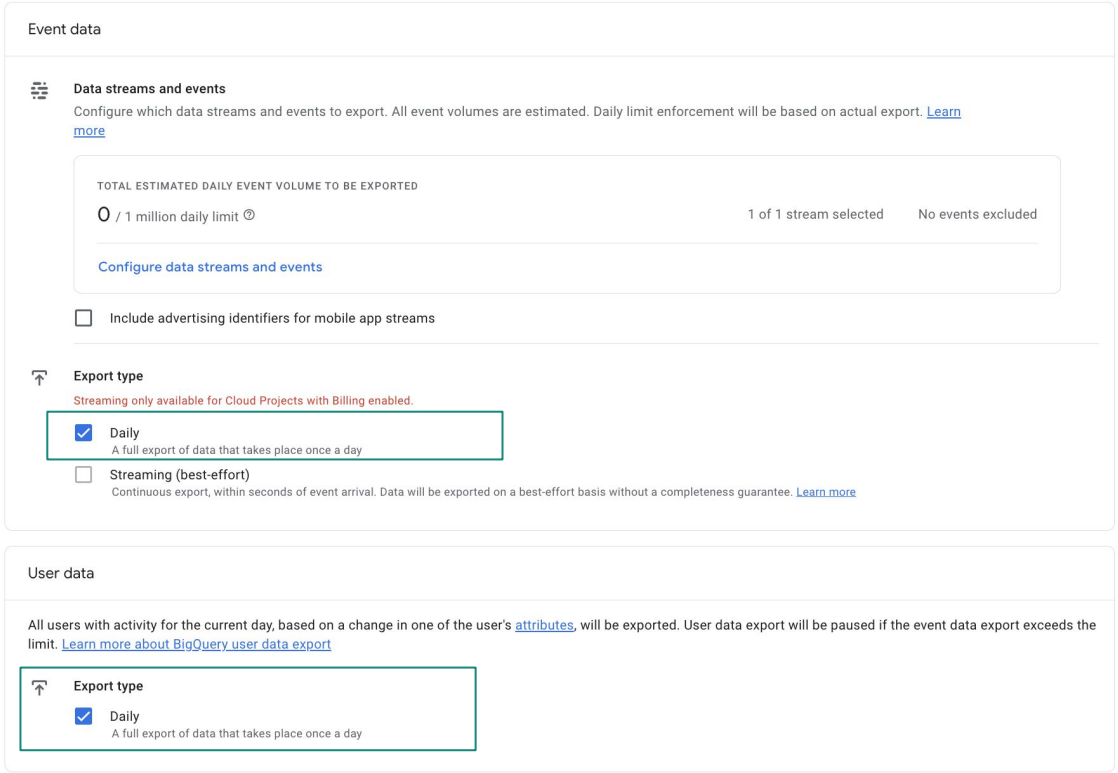
How-To:
Connect GA4 to BigQuery



How-To:
Connect GA4 to BigQuery



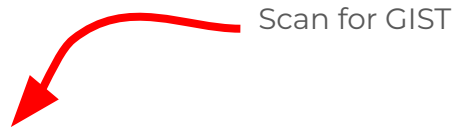
How-To:
Connect GA4 to BigQuery



How to Query Data of Experiments

1. Identify the **date range**
2. Identify **experiment variants**
3. Identify the **table** where the metrics of interest come from
4. Select **metrics to compare**
5. **Structure** the query logic with CTEs
6. **Combine** experiment variants with filtered metrics
7. **Aggregate** results per variant
8. **View** the aggregated table

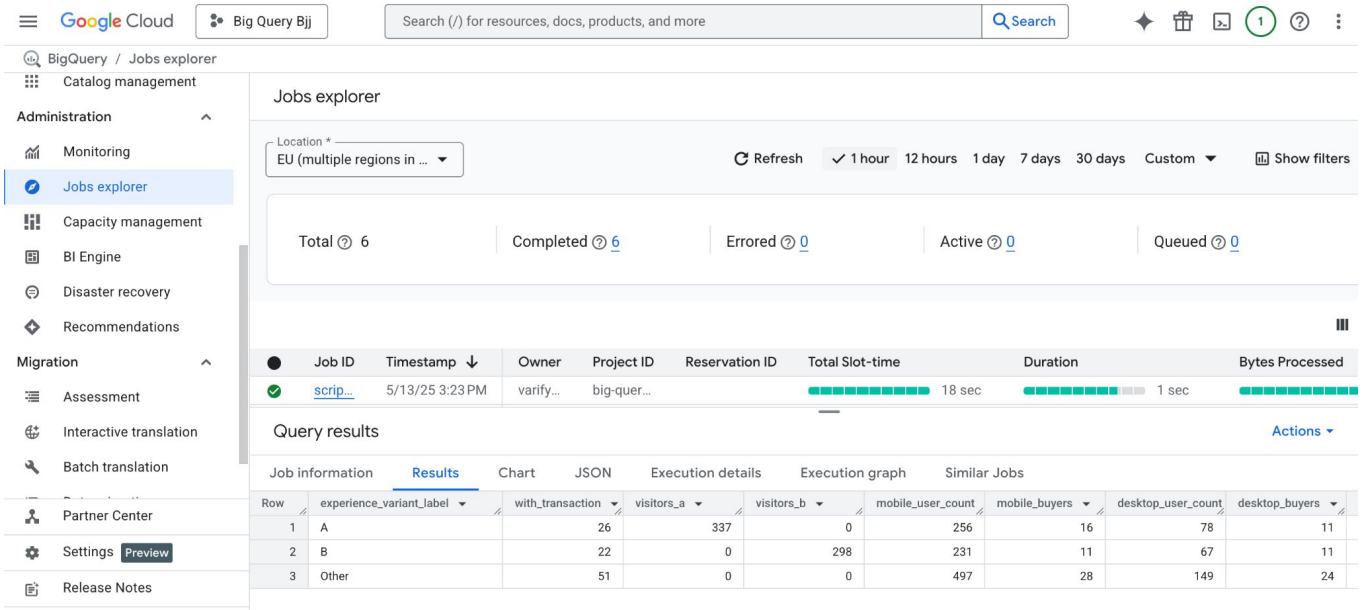
Query: Transactional Experiment Data




```
-- Binomial analysis
SELECT * FROM aggregated_data;
```

Query: Transactional Experiment Data

Adjusted for varify.io



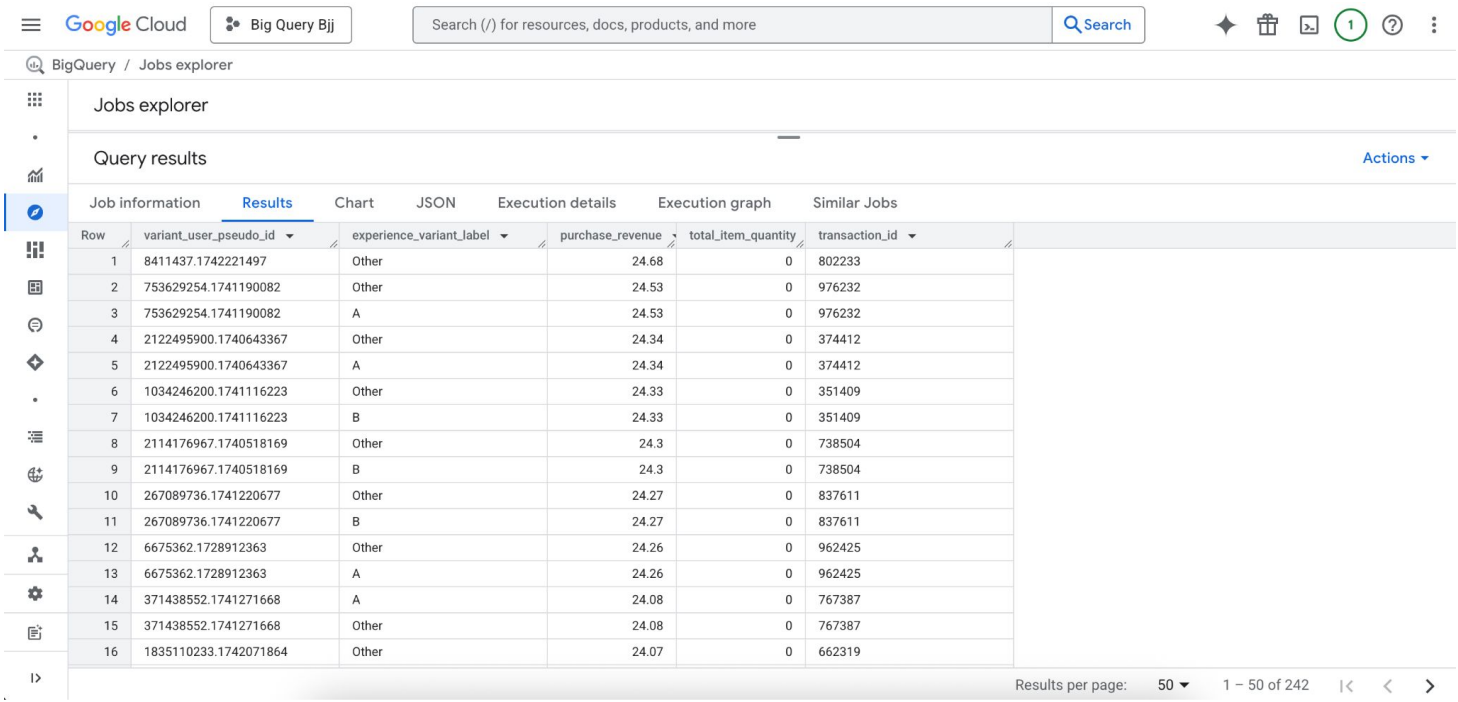
Query: Revenue Experiment Data

```sql
-- Continuous data analysis

SELECT
  variant_user_pseudo_id,
  experience_variant_label,
  purchase_revenue,
  total_item_quantity,
  transaction_id
FROM
  combined_data
WHERE
  purchase_revenue IS NOT NULL and purchase_revenue <> 0.0
GROUP BY
  variant_user_pseudo_id, experience_variant_label, purchase_revenue, total_item_quantity, transaction_id
ORDER BY
  purchase_revenue DESC
```

Scan for GIST

Jobs Explorer View



Row	variant_user_pseudo_id	experience_variant_label	purchase_revenue	total_item_quantity	transaction_id
1	8411437.1742221497	Other	24.68	0	802233
2	753629254.1741190082	Other	24.53	0	976232
3	753629254.1741190082	A	24.53	0	976232
4	2122495900.1740643367	Other	24.34	0	374412
5	2122495900.1740643367	A	24.34	0	374412
6	1034246200.1741116223	Other	24.33	0	351409
7	1034246200.1741116223	B	24.33	0	351409
8	2114176967.1740518169	Other	24.3	0	738504
9	2114176967.1740518169	B	24.3	0	738504
10	267089736.1741220677	Other	24.27	0	837611
11	267089736.1741220677	B	24.27	0	837611
12	6675362.1728912363	Other	24.26	0	962425
13	6675362.1728912363	A	24.26	0	962425
14	371438552.1741271668	A	24.08	0	767387
15	371438552.1741271668	Other	24.08	0	767387
16	1835110233.1742071864	Other	24.07	0	662319

BigQuery for Experiment Analysis

Tips for using BigQuery

happy horizon

CREATIVE DIGITAL AGENCY

Understand Your GA4 Export Schema & Experiment Setup

Familiarize yourself with the events_* tables generated by the GA4 BigQuery export. Key fields you'll likely use are:

- **event_date:** For partitioning and date filtering.
- **event_timestamp:** For precise timing.
- **event_name:** To filter specific events (conversions, exposure events).
- **user_pseudo_id:** The primary identifier for anonymous users (this connects Varify with GA4).
- **event_params:** This is a nested field (an ARRAY of STRUCTs) where most custom data, including experiment details, resides. Each element typically has a key (string), and associated value fields (value.string_value, value.int_value, etc.).
- **ga_session_id:** Found within event_params, useful for session-level analysis.

Understand Your GA4 Export Schema & Experiment Setup

Familiarize yourself with the events_* tables generated by the GA4 BigQuery export. Key fields you'll likely use are:

- **event_date:** For partitioning and date filtering.
- **event_timestamp:** For precise timing.
- **event_name:** To filter specific events (conversions, exposure events).
- **user_pseudo_id:** The primary identifier for anonymous users (this connects Varify with GA4).
- **event_params:** This is a nested field (an ARRAY of STRUCTs) where most custom data, including experiment details, resides. Each element typically has a key (string), and associated value fields (value.string_value, value.int_value, etc.).
- **ga_session_id:** Found within event_params, useful for session-level analysis.

Identify Experiment Parameters

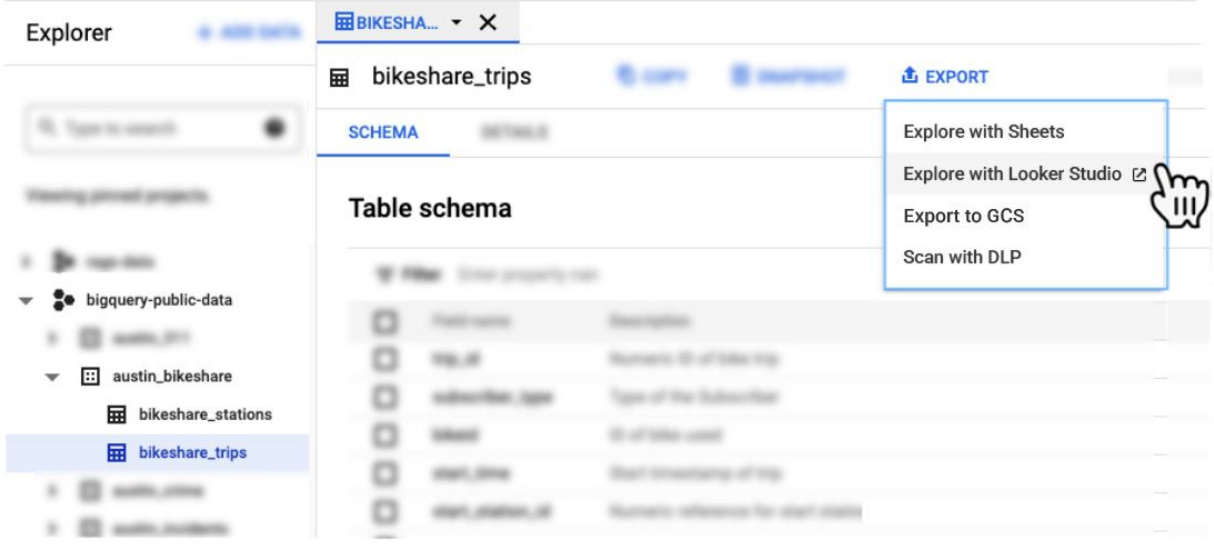
Crucially, you need to know the exact `event_params.key` names used to record experiment participation.

Some examples of values that you're looking for:

- Varify: `varify_abTestShort`
- Convert: `exp_variant_string`
- VWO: `vwo_variation_id`
- Webtrends Optimize: `wt_variation_id`

Use BigQuery Studio Schema Viewer

Before writing queries, use the schema viewer in BigQuery Studio (usually on the left panel when a table is selected) to explore the `events_*` table structure, especially the nested `event_params`. This helps confirm data types and exact field names without running costly `SELECT *` queries.



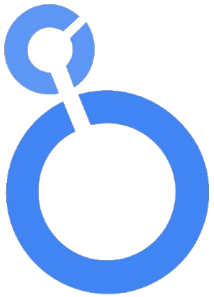
Master UNNEST for Experiment Params

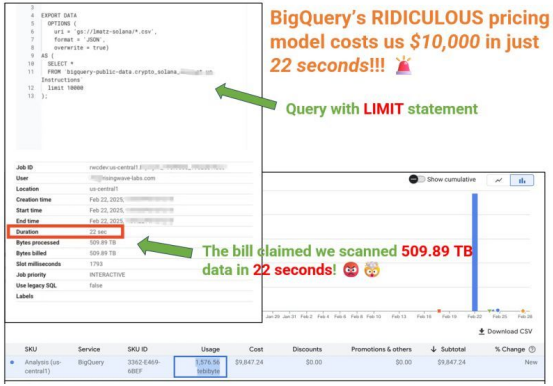
Since experiment details are usually within the `event_params` array, you must use `UNNEST` to flatten this structure so you can filter and select based on the experiment/variant keys.

```
SELECT
   user_pseudo_id AS variant_user_pseudo_id,
   CASE
     WHEN params.value.string_value = '15039:Original' THEN 'A'
     WHEN params.value.string_value = '15039:21017' THEN 'B'
   END AS experience_variant_label
FROM
  `table_id`,
  UNNEST(event_params) AS params
```

Optimize for Cost & Performance

Why?





https://www.linkedin.com/posts/yingjun-wu

Optimize for Cost & Performance

- **ALWAYS Filter by Date:** Use the `_TABLE_SUFFIX` pseudo-column in your `WHERE` clause. This is the most important optimization as it limits the amount of data scanned.
- **Select Only Necessary Columns:** Avoid `SELECT *` on unfiltered data. Explicitly list the columns you need, especially before joins or aggregations.
- **Filter Early:** Apply filters (like `event_name`, `_TABLE_SUFFIX`) as early as possible in your query or CTEs.
- **Use `SAFE_DIVIDE`:** Prevent division-by-zero errors when calculating rates.
- **Preview Data:** Use BigQuery Studio's "Run Preview" option (if available/enabled) or add a `LIMIT` clause during development to test logic on a small data sample quickly and cheaply.

Recommendations

- **Compare** your data sources
- **Understand** the differences – then define your single source of truth
- Ensure **clean** experiment tracking in BigQuery
- Use **tested SQL templates** (or “safe spaces”) – and adapt them
- **Optimize** your queries for performance and cost
- Get help **early** – not only when things break

	Visitors	Conversions	Conversion Rate	Improvement	Confidence	Significant
Original	84,895	1,730	2.04%	-	-	-
Variation 1	84,897	1,790	2.11%	4.46	83.8%	No

No statistically significant differences were found.
Please continue testing to ensure a sufficient sample size.

Goal 3: **Add to carts** GA4

	Visitors	Conversions	Conversion Rate	Improvement	Confidence	Significant
Original	84,698	6,314	7.45%	-	-	-
Variation 1	84,717	6,754	7.97%	6.94%	100.0%	Yes

Goal 4: **Add to carts (count)** Big Query

	Visitors	Conversions	Conversion Rate	Improvement	Confidence	Significant
Original	84,895	6,340	7.47%	-	-	-
Variation 1	84,897	6,790	7.99%	7.27%	100.0%	Yes

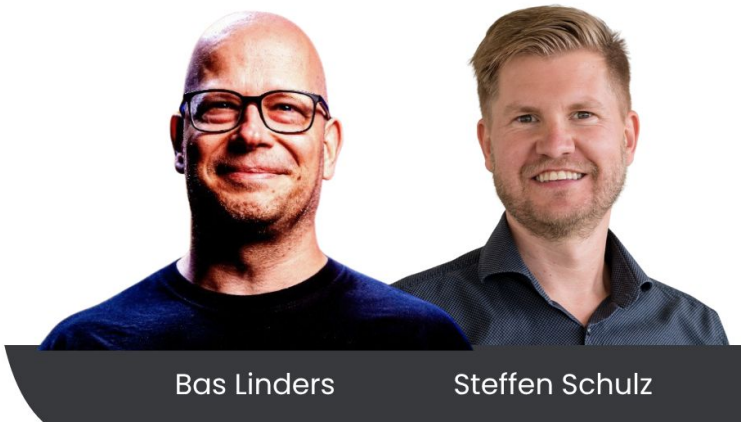
Goal 5: **Abtesting** GA4

	Visitors	Event Count	Conversion Rate	Improvement	Confidence	Significant
Original	84,698	84,662	99.96%	-	-	-
Variation 1	84,717	84,687	99.96%	0.01%	77.0%	No

No statistically significant differences were found.
Please continue testing to ensure a sufficient sample size.

Next steps

- **Compare your data**
- **Consult with experts**
- **Try BQ through varify.io**



Bas Linders

CRO Specialist

Steffen Schulz

CPO varify.io

Next steps

- Compare your data
- Consult with experts
- Try BQ through varify.io
- Join our (actually) exclusive Community

Bas Linders
CRO Specialist

Steffen Schulz
CEO varify.io

happy horizon

LookerStudio