Correlation vs. Causation: Recognize your Analytics Reality
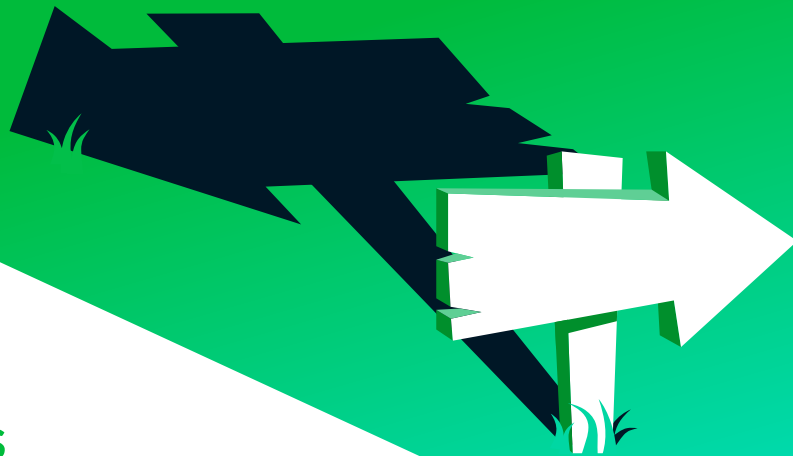
Make Meaning Count

Countly

Contents

Are your Website or Mobile App Analytics Leading you Astray?

The concepts of correlation and causation are often confused.

It's one of the biggest problems in product and customer experience analytics – in fact, in any web/mobile app analytics - leading to inefficiencies and business losses. This e-book clearly explains correlation versus causation, their role in website and mobile app analytics, and how they can impact your business. It also recommends solutions for gathering more relevant and robust customer and product data.

Who should read it?	Why?
Product Managers, Product Owners, and Journey Owners	Your customer and product data can be applied to the development of a better product and product map, and more intuitive, enhanced customer journeys
Product Line Managers	Your analytics data can help improve product and application performance
Business Leads	Your data can help improve product planning and feature development
Engineers	Your analytics data can improve automation, risk mitigation, error and crashes management, and product performance maintenance
Chief Information Officers (CIOs)	Your analytics can support the introduction or enhancement of existing digital products to reach a wider audience and enable greater customer satisfaction
Chief Executive Officers (CEOs)	To improve data-sharing practices and make teams data driven

Anyone who has influence over the development of web and mobile apps in your business, or is helping drive your organization's digital transformation will derive valuable insights from this e-guide.

Correlation and Causation

Correlation and Causation:
Not the Same Thing at All

It's crucial to minimize the mislabeling of correlation as causation. Before we unpack why, let's start by defining each of these terms:

Correlation

Correlation measures how two or more variables change simultaneously in relation to one another. In a positive or direct correlation, an increase in one variable happens simultaneously with an increase in the other. In a negative or inverse correlation, when there's an increase in one, there's a decrease in the other, or vice versa.

If two variables are correlated, one is not necessarily causing the change in the other. They are just co-existing. The change may be caused by a third or unknown variable.

Causation (or Causality)

Causation defines how two variables influence one another. One variable may cause the existence of the other, or it may cause a direct or inverse change in the second variable.

For example

What Happens when you Confuse Correlation and Causation?

Imagine you're launching a push notification campaign, sending messages at regular intervals. You track retention levels and notice that, while the campaign runs, churn has increased. You could assume that the increase in communication has led to users' dissatisfaction, that they aren't logging in anymore and may even be uninstalling your application.
In this case, you've identified the push messaging campaign and increased communication as the cause of churn.

When, in fact, the number of users engaging with the push notifications actually increased the amount of data processes running through the application, which led to a rise in crashes. Which is why users uninstalled the application.

So, the real cause of churn was the app's inability to handle increased activity – to cope with the success of the campaign - not the push notifications themselves.

When you see a correlation, and mistakenly identify it as causation, it can affect multiple parts of your business, impact your short-term goals and plans, and diminish your business's bottom line.

Why do these kinds of incorrect assumptions still happen so often?

The Power of Human Bias

The main culprit here is cognitive or 'human' bias.

Cognitive bias describes the way an individual interprets a situation – based on a preconceived, often strong opinion, resulting from a lack of information, actual or perceived.

We see a specific number of data points that indicate a pattern, and conclude that the pattern is the result of a cause-and-effect relationship.

What we often don't recognize – because our cognitive bias is blocking us – is that we're only seeing, or applying, some of the available information.

We also have a tendency to try and make sense of information as quickly as possible, influenced by our individual experiences, ways of thinking, observations, and general points of view.

!

Cognitive bias is unintentional, and generally instinctive. We don't choose to deliberately exclude data or information - we just don't think any information is missing.

Overcoming Cognitive Bias

To get real meaning from your data, you must keep seeking out information until you are absolutely confident you can see the full picture.

This can be a very tedious endeavor unless you have a defined process for maintaining sufficient data visibility, data sharing, and data depth.

Why do many businesses still lack such a process? There are 3 key roadblocks.

1. Siloed Data

Different teams often use different tools to collect the data that matters to them, leading to siloed data. Perhaps your product team uses a product analytics tool that focuses on customer and in-app behavior, while your engineering team has an application performance monitoring tool to keep an eye on any technical issues users are experiencing. Alongside this, your marketing team might use an attribution tool to track acquisition sources and a push notifications tool to boost engagement, and your customer success team has a feedback tool that collects information on user satisfaction. And even if teams do share web or mobile app analytics tools, they can still end up with siloed data if they're only tracking the specific metrics they need.

Consequently, each team will only be able to see one part of the whole picture. And they may not be able to cross-reference all the data points they're collecting – making it near-impossible to get a 360° view of their users. As a result, teams could easily miss key variables and their impact on others.

For example

Imagine your marketing team came up with a two-month campaign to boost sales from customers using a particular device. To do this, they relied on funnel analysis to map recent user behavior, and monitored overall revenue metrics to see if they improved.

However, they weren't aware that the design and support teams had already tracked a decrease in user satisfaction on the same device type, and were simultaneously running a month of A/B testing on the location of the 'Purchase' and 'Submit Feedback' buttons. When they finished the experiment, they implemented the most successful variation.

Ultimately, the marketing team's access to partial data revealed a sharp change a month into the campaign (when the A/B test ended), leading them to investigate further, without knowing that the cause was purely technical.
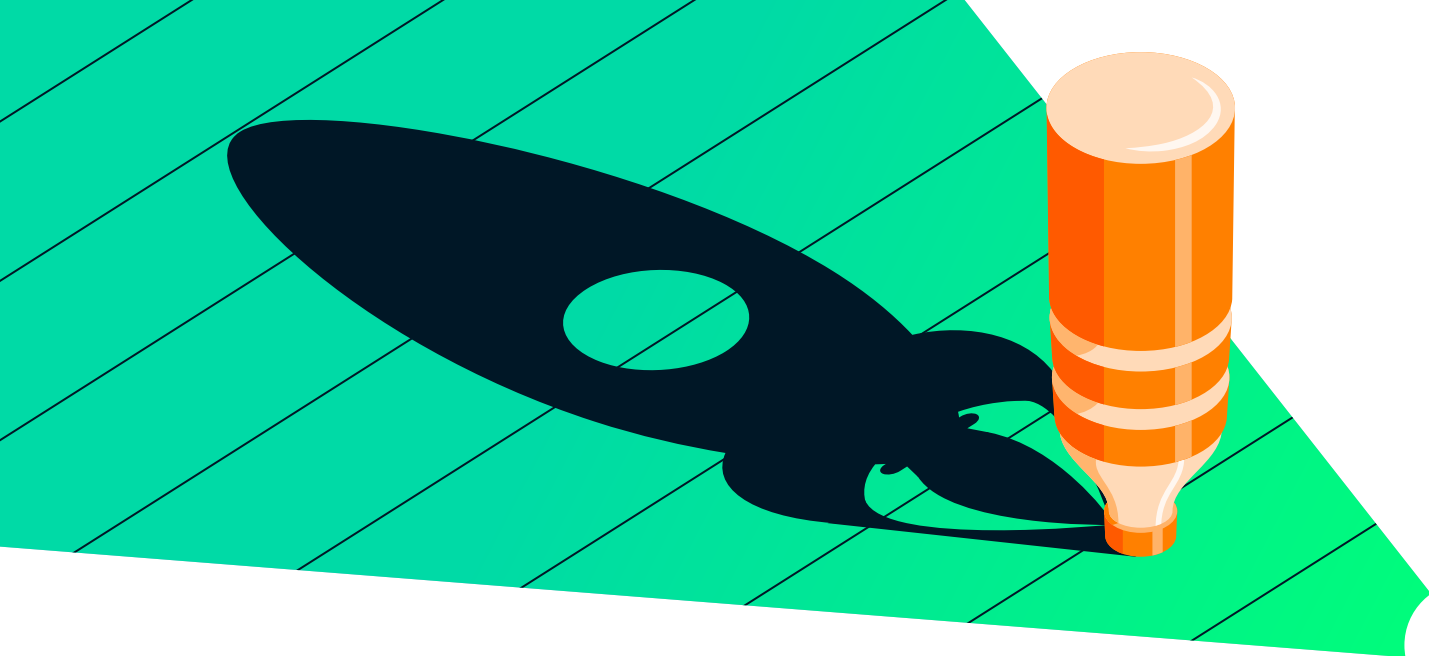
2. Incomplete Data

Tracking incorrect metrics, or fewer metrics than you need, results in incomplete data. This makes it more likely that you won't see all the factors causing a correlation, and jump to incorrect conclusions.

Much like siloed data, incomplete data collection is often unintentional. It can happen when team members don't know which metrics they need to calculate because they're not yet data-driven or their product analytics maturity is low. Or it could be because metrics are collected only as an afterthought instead of following a strategy that addresses the entire user experience. So anyone looking at incomplete data can find themselves incorrectly seeing relationships between certain variables as causal.

3. Superficial Data

Product and user data may also be misinterpreted if the depth of available information is insufficient.

Superficial data collection is often the result of policy related to privacy regulations and compliance. To remain privacy compliant, organizations forgo in-depth user and product data, typically when they rely on a third-party for data processing.

Despite multiple research findings that show users are willing to share personal information with their app/product providers in exchange for a more personalized customer experience, organizations still tend to address privacy regulations by collecting less data. This means they aren't just passing up an opportunity to provide an enhanced customer experience – they are also increasing the risk of misinterpreting the relationship between user behavior data points.

For example

You may rely on two or three data points, identifying a correlation between some of them.

However, you may be unaware of how other data points - specifically, more detailed information on a user could be causing the changes in the correlated variables.

The True Cost of Mistaking Correlation for Causation

The true cost of mistaking correlation for causation

When you see causation where there is none, it can have long-term effects for your business.

Imagine you run a mental health support app. You have raised the subscription cost slightly – and are now seeing a lower rate of acquisition. Attributing this to the price increase, you try to boost acquisition by launching a quick discount campaign. It seems like your quick fix has worked - you start seeing new subscribers taking advantage of the discount.

However, you've missed a critical data point. Both these variables – raising the subscription fee and launching the discount campaign – were preceded by the launch of a new intake form. Much longer and more detailed than the previous version, this form aims to gather additional information to help pair users with the right therapist or specialist. This is what led to user drop-off. People weren't comfortable sharing so much personal data at the first point of interaction, they weren't clear on how you handled their data, and they weren't sure if it was worth the effort because they didn't know if your app would help them.

However, when you launched the discount campaign, acquisitions increased because the greater effort required was offset by the price cut. People felt less at risk with the lower cost: if they found the app wasn't great, they wouldn't have lost money as well as time.

But the other issues still persist: the app still asks people to share extremely personal details. It asks them to share their vulnerabilities with a platform instead of a person, and they continue to remain worried about whether they would click with the specialist you connect them with, as well as how their data might be misused.

The ideal solution would have been twofold:
1) make your stance on user data privacy clear and transparent;

2) change the user journey so they have fewer questions to answer before they're connected with a therapist who can then work on a detailed intake with the user.

Changing the user journey would have addressed the causal issues. Instead, you've wasted time, money, and effort on a short-term solution that doesn't solve the underlying problem and that has limited results. You have had fewer acquisitions over more months than necessary and lower revenue for a longer period of time, as well as potentially negative reviews on app stores.

Don't let your bias come back to bite you.

Cost and Other Impacts Across Teams

Product

Product teams are the most affected when correlation is misinterpreted. Product Managers, Product Owners, And Journey Owners may misunderstand:

- How users are interacting with the app or product
- User preferences and behaviors
- Where friction points occur in users' interactions with the app/product
- The trends in app/product engagement, retention, and churn

Consequently, the Product team may be:

- Unable to correctly segment users
- Unable to identify the right feature updates or new features preferred by customers
- Unable to identify the markets where the app or product is or isn't performing well
- Unable to make effective corrections to feature flow

Overall, the team is wasting time, money, and effort on inaccurate engagement efforts, ineffective updates and product planning, and bad marketing investments.

Engineering/Development

Because engineering and development make product team visions a reality, they share many product issues, and some additional ones, including:

- Misunderstanding app or product performance in relation to technical specifications
- Misunderstanding device and region preferences
- Misunderstanding the reasons behind app/product failures

This can lead to cost implications, including:

- Ineffective performance management
- Ineffective planning of feature and improvements rollout
- Ineffective management of issue resolution

Overall, they are wasting time, money, and effort resolving issues that have been incorrectly mapped, working on ineffective updates and non-priority items, and investing in the wrong target groups.

The Rest of Your Business

When correlation is mistaken for causation, marketing teams can end up focusing on inaccurate engagement, retention, and even acquisition; data teams can struggle to interconnect various data points and understand their effect on multiple teams and stakeholders; and executive teams can see declining revenue and profits.

The effects of misinterpreting correlation may be further exacerbated in traditional organizations that are in the midst of digital transformation, as they struggle with new operational models and trying to drive a complete mindset shift in their teams. For a business in this scenario, the likelihood of mistaking correlation for causation is higher, at a time when they most need real causative data for decision making.

Misinterpretation can raise multiple issues in a business, including:
– Focusing on long-term product development without accurate information and missing the preferences of their target users, slowing down acquisition and growth;

– Focusing on short- and long-term product digitalization, without a clear understanding of their users' behaviors, performance issues in MVPs, and applicability of the digital product in the entire range of services

– Having no clear understanding of data requirements – especially for digitalization of a product – due to not being able to read the data or plan requirements accurately

– Sharing incorrect data, based on one person or team's subjective outlook – leading to a cascading effect at every subsequent point of further application or action

In any business, once data has been read inaccurately, it has a snowball effect – affecting any information shared, subsequent action taken, calculation made, or metric subsequently calculated using the incorrect finding. The consequences can be numerous, and costly.

Moving from Correlation to Causation

How to avoid joining dots that aren't there?

Use Testing

History and numerous experts have shown that testing is the best way to ensure the relationship between two variables is in fact causation and not correlation. The most common is A/B Testing, where you change one variable of the two or more that seem causal. If your outcome remains the same as in the unchanged example, i.e., showing the same trend, then you can confirm that the link is in fact causal.

However, there are drawbacks to testing. Not only is it done after a relationship has been identified, it also requires more time to be spent on confirming a relationship that could potentially have been clear from the outset.

Making Causation the Default

When you see correlation and causation clearly from the start, your data practices become faster, more insightful, and more accurate in the long run; and the changes you make have greater positive impacts for the business.

There are four practices that will give that clarity of vision:

1. Take a strategic, considered approach to analytics
2. Get the right feature set
3. Gain data-depth
4. Enhance data-sharing

Continued on the next page...
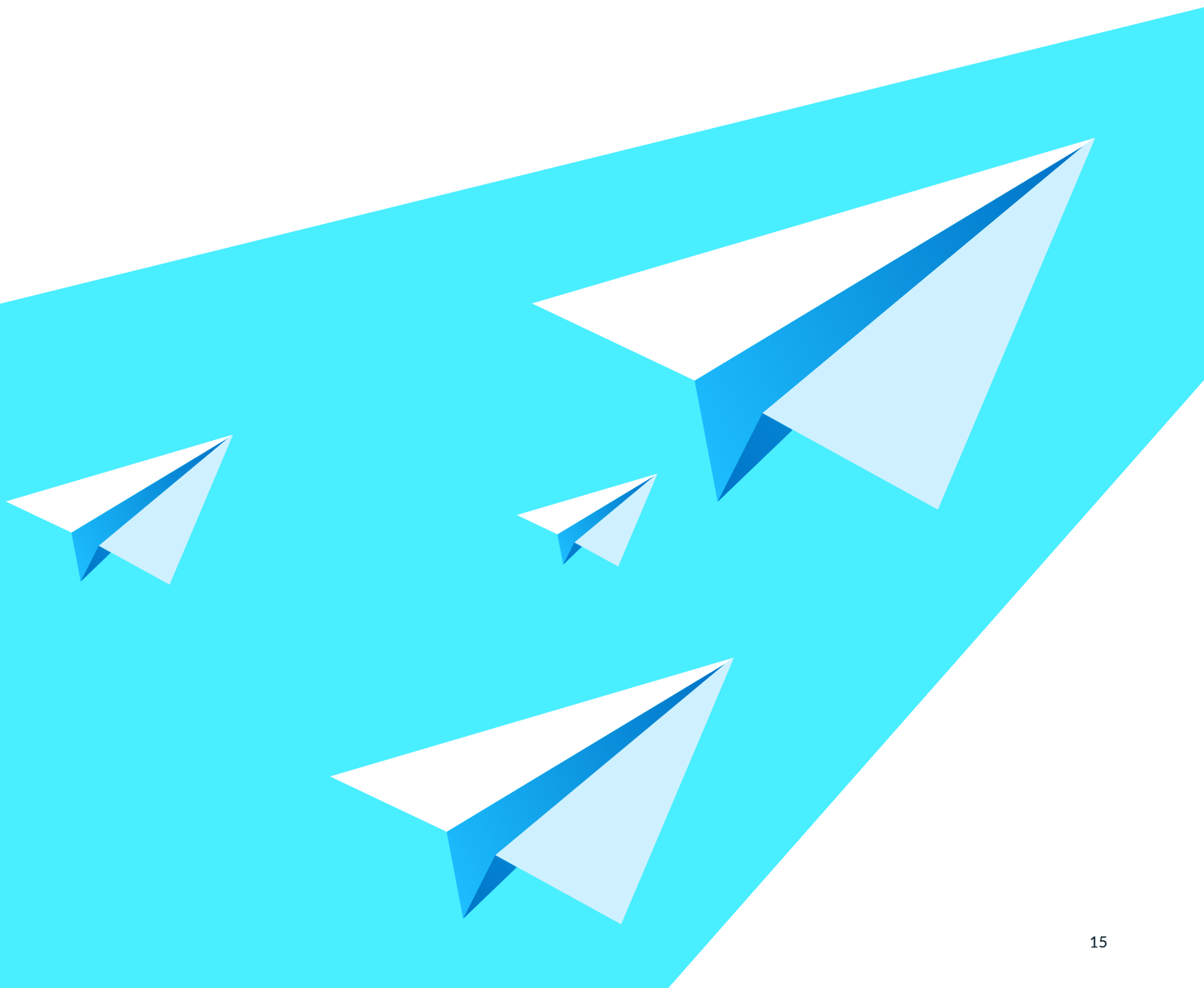
1. Take a Strategic, Considered Approach to Analytics

Many organizations fear they won't have enough web or mobile app data, or feel they have to amass data because that's what others do. They rush into data gathering without a strategy.

A better approach is to take a step back and get a clear idea of what you need to collect and why. Understand the data you could collect, and the relationship between data sets. Then it's easier to identify the relationship between variables.

Importantly, ask yourself these foundational questions:

- **Why do you want to collect data?** Is your team or organization simply following a trend or doing what others do? Or do you believe you're missing out on key insights that could help your app/product and business?
- **What is the end goal of the data collection?** What problems do you hope to resolve by getting more insight?
- **Who needs to use this data?** Which teams can use it and for what purposes?

- **Do you have someone who can put this information together?** Or do you need to work with someone who can help you gain a comprehensive view of your analytics requirements?
- **What could be positively/negatively affected by the collection and/or absence of data?** Which teams, which aspects of the product, and which parts of the business may be hampered by lack of data, and why?
- **Are you already collecting information that could be a part of this?** Do you have information in Excel or Google Sheets that you need to build on?

A strategic approach to data ensures you aren't collecting too much and then trying to make sense of it. The questions above are a good start, when combined with advice from experts who can guide you along your strategic path.

2. Get the Right Feature Set

If you are short on data, and you can't see all the relevant factors, you can easily misread trends as causal. That's why getting the right feature set is critical.

The solution is to make sure your web and mobile app analytics tools are giving you the full picture. Once you define your strategy and know what you are trying to understand, you can choose the best tool for the job.

3. Gain Data-depth

Gathering in-depth user data is a priority for most organizations.

And as mentioned above, app and product users are willing to share personal information with the provider if it means getting a more personalized, intuitive customer experience.

The challenge is balancing the collection of detailed data with user privacy. Compliance with leading privacy laws, such as HIPAA and GDPR, is often seen as an obstacle to business and product innovation.

However, it is completely possible to have both – user trust and innovation. More and more organizations are making user trust and data privacy central to their data collection practices, using tools that automatically ensure data privacy. They can help you delve much deeper into your users' data, giving the insights you need to create truly memorable customer experiences, boost retention, and build brand loyalty.

When you truly understand user behavior, the risk of misreading trends or causality is dramatically lower.

4. Enhance Data-sharing

Many teams still operate in silos, approaching and using analytics in different ways, unable – or unwilling – to share information.

For example

A Product Manager on the product team knows that drop-offs from Point X in an app are due to a poor call-to-action. However, he doesn't relay this information to the engineering team – who have concluded that drop-offs are due to users being on a specific smartphone.

Two different departments are tackling the exact same issue in completely different ways, due to lack of communication, collaboration and insight sharing. Company time and resources are being wasted on fixing issues that may not have been the problem to begin with.

Usually this happens when every department across an organisation uses different tools for data collection.

Using multiple tools also results in a mismatch in the basic data collected, because every tool has its own data logic, feature set, and limitations. So, even if an organization wanted to collaborate and share insight, it might not be able to – or may end up spending hours arguing about which data set is accurate.

With such ambiguity in the data, misinterpretation is inevitable.

We recommend a twofold solution for tackling these problems:
1. Use fewer – or even just one – data tool that meets the needs of all teams, ensuring everyone is looking at the same data.
2. Democratize access to data across teams, so more people can look into the same data and share insights.

17

The Upshot for your Business

In our 10 years of experience, we have seen firsthand the benefits of using minimal analytics tools across a whole business, from minimizing complications around integration and collaboration to improving efficiencies.

We have witnessed teams being able to access, for the first time, the data they need, and instantly seeing the bigger picture; not to mention, data-sharing encourages cross-team collaboration.

These tools open up opportunities for different ways of thinking - people are now equipped to consider every angle. They also help avoid jumping to bias-driven conclusions, signaling the end of the correlation versus causation dilemma.
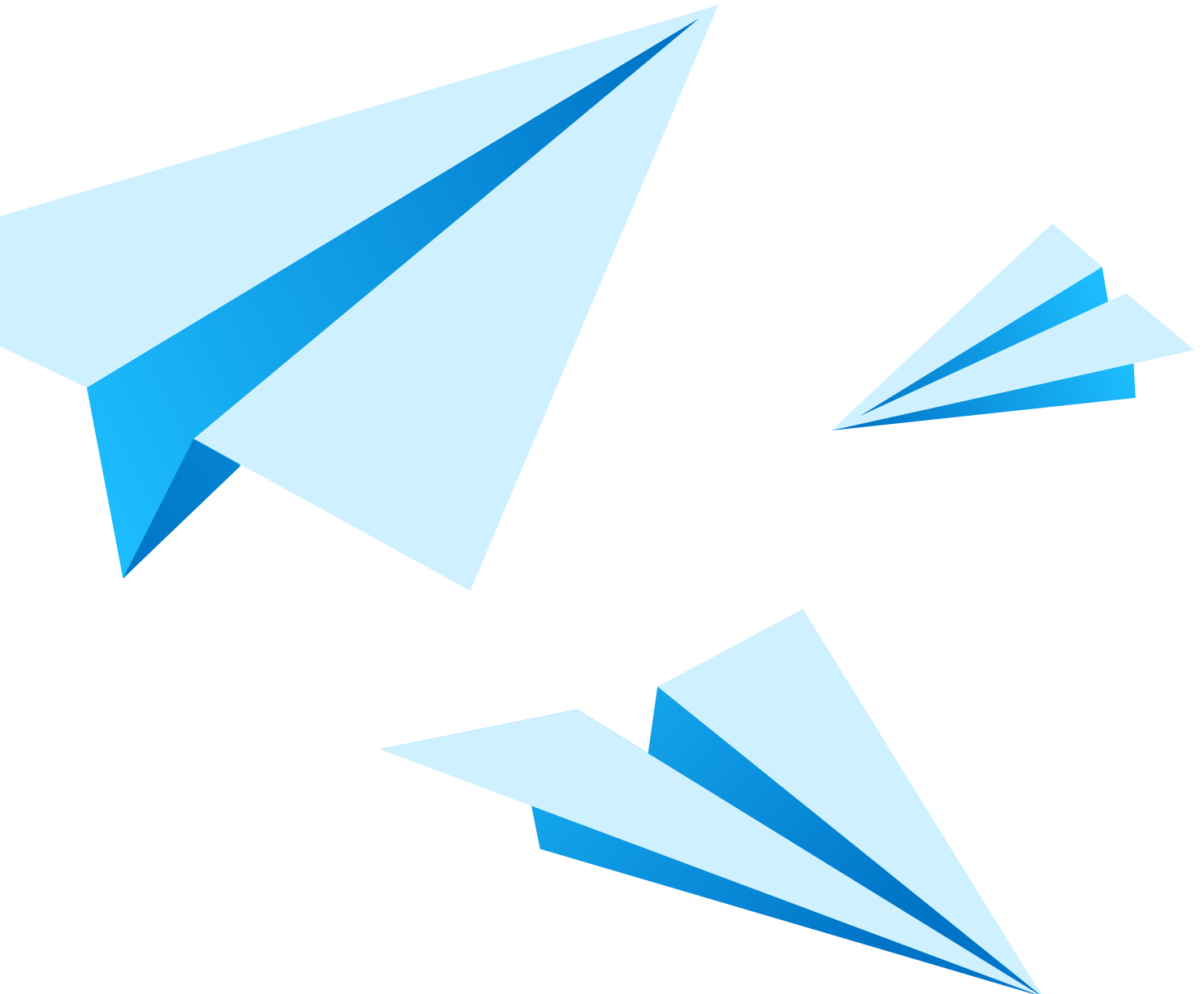
Use Cases and Examples

These examples show, in more detail, how limited data collection can increase the chances of joining dots that aren't there; and how seeing the bigger picture helps you avoid costly mistakes.

Use Case No.1

Lack of the Right Features

The Challenge

Imagine you want to find out how users' in-app purchases are contributing to the overall app revenue. You decide to track metrics using three features:

1. User activity: Your 'Visitors Overview' feature tracks the number of new and returning users.
2. Funnels: The 'Funnels' feature tracks how new users behave and to what extent they complete the funnel you've created. This works as follows:
Step 1: New user joins or downloads and installs the app.
Step 2: User reaches a specific number of sessions without making any in-app purchases.
Step 3: User responds to in-app purchase offer/prompt.
Step 4: User spends time on the purchase page.
Step 5: User makes in-app purchase.
Step 6: User meets a specific number of sessions, or user logs no more sessions, or uninstalls app.
3. Revenue: The 'Revenue' feature lets you track revenue per user session and per user.

You then make the following assessments:
– Your number of new users is healthy and shows consistent growth.
– However, your number of returning users has dropped – inconsistent with the new user growth.
– Your revenue per user and per session is dropping.
– You see that users are falling off at Step 5 -

they're not completing their purchases.
– You also see that users who do complete Step 5 drop off at Step 6. The user either doesn't meet a specific number of sessions - indicating reduced interest - or doesn't log in at all, indicating that the app was uninstalled or deleted.
– Your revenue per user is also dropping, corroborating your insight that customers are not seeing purchases through, or are deleting the app following a purchase and not contributing additional revenue.

With these insights, you may conclude the reason your revenue and returning users are falling is either that the paid features are dissatisfactory, or the in-app purchases are too expensive.

You then decide to roll out a discount for the in-app purchases, rewarding a larger number of sessions for a limited period. While this is relatively quick to implement, you do need to adjust the number of sessions you think should be rewarded by a lower price, and you need to set up more funnels to track these conversions.

Once implemented, you notice a positive effect at Step 4 - more users are clicking on the prompts. However, the rest are still unaffected.

The Challenge continued...

You move on to the next solution - updating the features themselves, because you believe those currently being offered aren't interesting enough for your users. You spend time, money, and effort on this, possibly altering your product roadmap as well.

You eventually roll out the new features, and once again see a rise in people completing Step 4. Yet, once again, the rest don't change.

You're now stuck, wondering, "What's the issue?" You've seen everything there is to see in terms of interactions between users and the app, and one of the solutions should have worked. Is the price or the perceived value of the feature itself not the cause of the drop in in-app purchases and revenue after all? And, if not, what exactly is wrong?

Did you misinterpret the root cause of the problem because you didn't have the entire picture? Did a lack of data lead you astray?

The Bigger Picture

The solution lies in tracking the right metrics through a more relevant combination of analytics features, adding two additional metrics to the equation:

1. Feedback forms/prompts to users who have logged in after a purchase but with diminishing frequency.
2. Crash analysis or performance monitoring, to find out if the app is crashing at some point.

With these two additional insights, you would now see a completely different picture. Some users who tried to make an in-app purchase were unable to complete it, because a bug was making the app crash, or those that logged in after a purchase also faced more crashes. This led to growing dissatisfaction and the user eventually:

– stopped logging in, or only logged in occasionally to see if the issue had been resolved; or
– simply stopped using the app, some perhaps feeling cheated.

Key Learnings

As it turned out, the cause of user dissatisfaction here wasn't the feature or price, but the technical challenges they faced following the purchase.

In this scenario, your team doesn't just lose users and waste time and effort, but you also run the risk of receiving negative reviews about your app, affecting new user acquisition in the long run.

If you had used a tool that provided a complete overview, using a range of features important to you, this outcome could have been avoided altogether. The problem would have been identified and resolved in time too.

This example shows how not having the right features can limit access to the right insights, leading people to wrongly assume there was causation where it didn't exist.

Use Case No.2
Lack of Data-sharing Between Teams

The Challenge

Imagine you're a Marketing Manager and you're launching a new campaign on social media, offering users a discount to sign up to your app.

On Day 1 of the campaign, you are pleased to see there has been really good engagement with your ad, and the number of sign-ups is increasing.

However, within two days, you see that although engagement with the ad is still promising, the number of sign-ups is falling. At the same time, users who signed up recently are also showing lower daily retention.
Based on the information you have, you conclude the cause of these two issues is that the campaign offering wasn't interesting enough for your audience – so you report that it wasn't a success and start working on something else.

The Bigger Picture

In the background, the tech team – who didn't know about the campaign – had failed to tell you about a rise in fatal app crashes on both the sign-up page and the app in general. This was possibly driven by the sudden increase in users.

By the time they were able to solve the issue, users had tapered off – which made finding a solution even harder.

Key Learnings

You wrongly assumed the campaign wasn't interesting enough for the audience – it actually failed because the app kept crashing. However, due to a lack of data-sharing – the silo effect – this information wasn't discovered.

If the tech team had been informed about the campaign, or if the marketing team had the complete picture, the issue could have been identified and resolved – the campaign would have been judged fairly and reported as a success.

Use Case No.3
Lack of Data-depth

Background

Most apps collect basic personal information from their users, e.g., username, full name, and e-mail address. Additional details may include phone number, address, location, call logs, message logs, and other similar information. While some apps only collect basic data, others may request more than they need, and put it towards specific data output in the future.

The amount of data collected depends on what the app does, what it needs to meet user expectations, and company policy. Plus, certain privacy laws, such as GDPR and HIPAA, can limit companies from collecting too much information unless the collection, processing, and use are strictly necessary, extremely secure, and not privy to third-party access.

GDPR compliance has often (incorrectly) been considered a limitation to business innovation. Many companies believe that limiting the depth of data collected to remain compliant with privacy policies holds them back from truly understanding their users and their customers' journeys. In their view, this superficiality of data prevents them from providing personalized customer experiences.

However, the truth is quite the opposite. Companies can remain privacy compliant and provide excellent customer experiences by:

1. Ensuring their data collection practices are governed by a data collection strategy, which clearly outlines what data they require and how they would use this to create a better customer experience and improved product performance.
2. Opting for a product analytics solution that is inherently privacy compliant, ensuring data collected from users remains secure.

Ultimately these practices allow companies to collect a depth of data which will avoid misunderstandings around correlation and causation.

The Challenge

Imagine you have an app that collects basic personal and device data, as well as data on two events you've set up to measure the number of engaged users:

1. User spends at least three minutes on at least one session.
2. User visits the 'Extras' page at least three times to see and participate in new, weekly offers and contests.

You then set up a cohort using these two events to track, daily, the number of users who have performed both of these events, at least once, in the past week. You notice the number of users in this cohort is increasing – but not at the rate you achieved last month.

On further exploration, you realize most of these users are from the latest app version. You assume something must be wrong, so ask your engineering team if there are an unusual number of crashes or any performance problems with it? But there aren't any issues.

So, you go back to the drawing board to figure out what could have changed to have had such a significant impact on user behavior. You spend hours, maybe even days, assuming your users don't like something about the new app version. You might even roll out different updates trying to address this – but still see no significant difference. The problem remains unresolved, and you've wasted time, effort, and money in the process.

The Bigger Picture

If you had tracked one additional custom data point for each user – account type – you would have quickly noticed that the level of engagement was only declining in users who hold a basic account type.

If you had checked this with the team, you would have discovered the flow that leads users to trigger the two events was slightly modified in the latest app version for the basic account type. It was taking them a few more clicks to get to the 'Extras' page where they would also receive a pop-up offering a customized solution to upgrade their plan.

Key Learnings

Instead of leading to more upsells as hoped, the change resulted in less engagement - possibly because users found it harder to access the 'Extras' promotions they enjoyed. Not only did the change end up not meeting its intended purpose, it adversely affected the user experience and risked losing the loyalty of an otherwise content group of users.

This is how lack of data-depth can lead to not recognizing the cause of a problem, wasting time, money, and effort on possible solutions.

How do you know what personal information you can collect, and still maintain user trust and privacy compliance? The solution lies in an analytics tool for your website and mobile apps that lets you truly understand your user and gives you a 360° view of your customers - while ensuring you remain completely privacy-compliant and cognizant of your users' security needs.

That's how you discover what you're actually dealing with, and recognize your analytics data reality.

Let's talk

Ready to get started? Talk to one of our product analytics experts to learn how Countly's analytics for web, mobile, and desktop apps can help your business become data-driven.

To learn more about how we can help, contact Countly today.

![Countly logo] Countly