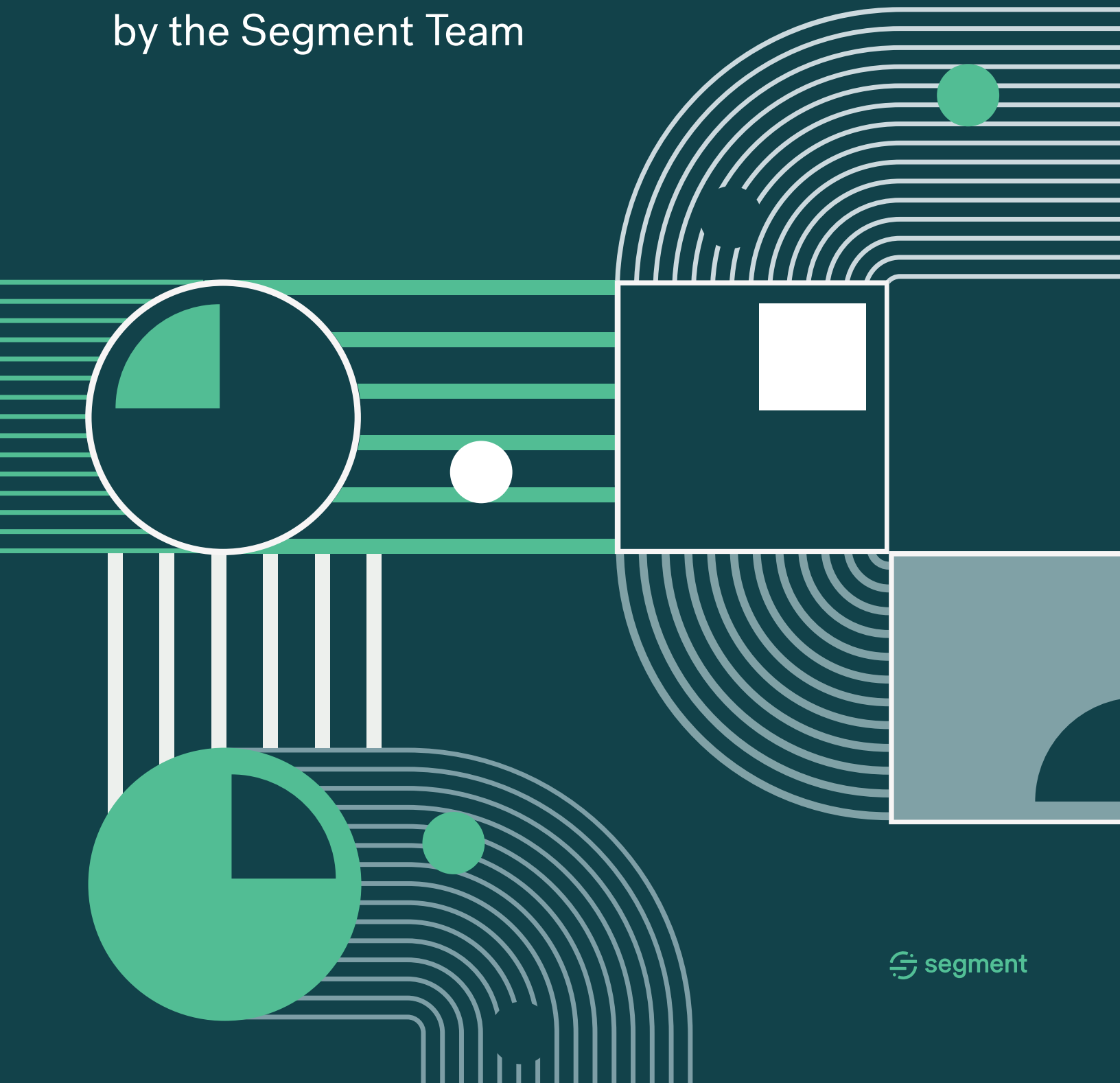
Segment on Customer Data

by the Segment Team



segment

Table of Contents

CHAPTER 1

Why Customer Data Matters

Contributor: Stephanie Evans
Customer Experience Product Lead @ Segment

Picture this: your company just launched a new product.

Countless hours went into this launch. There's the initial time spent researching your ideal customer and the problems this product will solve for them. There are weeks (perhaps months?) of engineering and design hours it takes to actually build your product and test that it works. And, let's not forget, all the messaging, training, and coordination efforts that go into getting the word out about this launch.

Just getting your product to market feels like a herculean effort in itself. And it's definitely cause for celebration when your product finally does launch. But, what really makes a product a success is what happens after launch.

After a big launch, it can be overwhelming to decide what to do next and what will yield the most traction — add the "killer" feature? Optimize your onboarding experience? Invest in SEO? Run ads on Facebook and LinkedIn?

Whether you're the head of your own one-person company, or you're a product manager in a 10,000+ person organization, the list of things you could do to grow product adoption or improve your user experience is infinite—which makes effective prioritization all the more challenging and critical.

So, what should you focus on to grow your business?

You need to know your customers to answer this question. And knowing your customers comes through data collection. Data gives you answers to questions like: *Who are your ideal customers? Where did they come from? How are they using your product? What does it help them do?*

Collecting, analyzing, and acting on data about how customers interact with your product (and combining that data with the traits that make them unique) will give you a clear direction for building a better product experience and growing product adoption for the long term.

A customer data story

Here is a quick story to illustrate why customer data is critical to the success of any business.

Imagine you just launched a sneakers app in iOS and Android app stores. In the first week, 500 people download your app, but none bought any sneakers. *What do you do now?*

You need to understand why app downloaders are not buying sneakers. To do that, you need to know more about them, such as who they are, how they're finding you, how they're using your app, and where they're dropping off. Without this data, you'll never know the cause of the problem at hand. And if you don't know what the problem is, there's no way to overcome it.

The solution? Data! Specifically, customer data that will help you make fast, informed decisions to grow your business.

But just having heaps of customer data in and of itself is not the goal.

Without a focused customer data strategy, you'll have a hard time attracting, activating, and retaining customers.

At Segment, we've helped tens of thousands of teams collect, analyze, and act on billions of pieces of customer data. They use this data to improve their product experience, develop lasting customer relationships, and ultimately grow their businesses.

We wrote this book to share with you what we've learned from our experiences.

CHAPTER 2

Choosing Metrics That Matter

Contributor: Sudheendra Chilappagari
Product Manager @ Segment

Whether you are a developer, product manager, or CEO, it's important that you identify, measure, and improve the right metrics. If you do, you'll have clear understanding of when you're adding value to customers and when you're not.

In the ideal scenario, as your metrics go up and to the right, your customer satisfaction moves in the same direction. When that's the case, you'll also benefit by knowing what direction to take your product and to build a roadmap that's aligned with improving your customer experience.

Defining the right metrics isn't rocket science, yet many teams get it wrong. While there are no hard and fast rules, here are some easy-to-follow guidelines for how you can go about choosing metrics that matter.

Guidelines for choosing metrics that matter

EASY TO UNDERSTAND

A good metric should be easy to understand, access, and lead to action.
When you share metrics amongst your team (and ideally your entire organization), they should be able to comprehend what each metric signifies and how it aligns with your vision/objectives.

EASY TO ACCESS

Quality metrics are also easy to access. This means they are not overly complicated to calculate and are available in tools used across your company, like Google Analytics, Amplitude, Looker, Slack, and others

LEAD TO INSIGHTS

Great metrics lead to insights (which in turn lead to action and results). Analyzing them informs decisions like reducing onboarding hurdles, building features that lead to value-added usage, or personalizing user experiences to grow engagement.

AVOID TOTALS

We often see teams make the mistake of tracking feel-good metrics like Total Active Users or Total API Requests. Measuring totals gives you only the half picture. To get the full picture, you need to measure things like conversions (%), growth (Δ change), customer engagement, and customer satisfaction. Totals = vanity metrics.

Note: There is one exception where a total makes sense — your revenue.

HERE ARE A FEW EXAMPLES OF BAD METRICS VS. THEIR BETTER COUNTERPARTS

BAD METRIC	GOOD METRIC
Total active users	% of active users = total active users / total sign-ups % of weekly growth of total active users Net promoter score
Total API requests	% of successful API requests API response time API uptime

Great outcomes require great inputs

Results are outcomes of inputs. And great outcomes require great inputs combined with a bit of patience.

Even for the most widely-used products, there's always a delay (lag time) to knowing if an outcome is moving in an intended direction. During this time of uncertainty, it's important to know you're heading in the right direction.

In such cases, you'll need to identify and define the differences between two types of metrics: leading indicators and lagging indicators.

- **Leading indicators** are like inputs. They measure the activities necessary to achieve your goals.
- **Lagging indicators** are like outputs. They measure the actual results.

Let's take an example of B2B SaaS company, Banjo Analytics, who wants to measure and improve Annual Recurring Revenue (ARR). Since it takes several months for Banjo to close a new deal, they define a handful of leading indicators around their sales pipeline to ensure that they're on track to achieve their lagging key metric — ARR.

LEADING INDICATOR	LAGGING INDICATOR
# of leads added to top of funnel/ month	Annual Recurring Revenue ($)
# of product demos with prospects/ month	
# of opportunities created/ month	
# of proposals sent/ month	

Choosing the right metrics also depends on many factors like type of product or offering, business model (B2C/B2B), industry (e-commerce/SaaS/marketplace), job function (product/marketing/sales/recruiting), and many more.

To give you a starting point for what metrics to use for your business, we've synthesized a handful of popular frameworks below.

Pirate metrics

Commonly used for: User-generated content or engagement apps like Instagram, Headspace, and Strava.

AARRR (Pirate Metrics) is a metrics framework developed by Dave McClure and is widely used by business to consumer software companies. This framework outlines fives stages of a user's journey, starting from acquisition and progression to a successful referral, which ultimately leads to a virtuous loop for growth.

E-commerce metrics

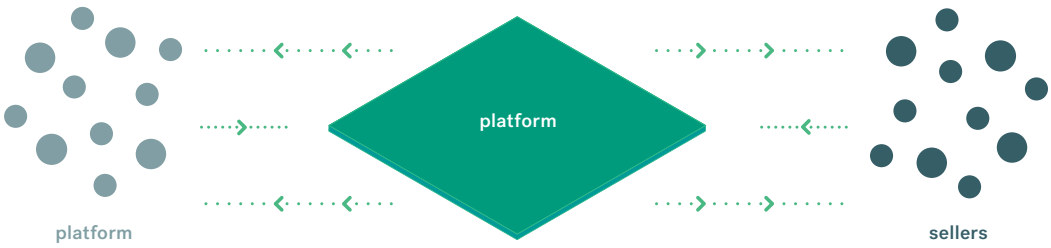
Commonly used for: E-commerce websites and apps including but not limited to Bonobos, Crate & Barrel, and Levi's.

If you're selling products online, you'll want to track metrics about how customers are finding your app, how they're engaging with it, how and when they're making purchases, and whether or not they're coming back for more.

STAGE/ TYPE	EXAMPLE METRICS
Marketing/ Traffic	No. of visitors — daily, weekly, monthly % of growth of no. of visitors % of conversion rate (visitor → purchased)
Revenue Metrics	Total sales — daily, weekly, monthly % of shopping cart abandonment rate Average order value
Customer Loyalty Metrics	% of returning customers monthly, quarterly, yearly Net Promoter Score
Unit Economics	Customer acquisition cost (CAC) by channels — such as organic, ads, referrals, and affiliates Customer lifetime value (CLTV) Gross profit margin

Two-sided marketplaces

Commonly used for: companies that have two-sided marketplaces, like DoorDash, Instacart, Uber, and Opendoor.



Marketplaces are incredibly powerful and create new consumer habits by re-imaging the way buyer ←→ seller transactions traditionally happen. Building and running a two-sided marketplace is quite the juggle, as you have to keep both buyers (demand) and sellers (supply) continually motivated.

At surface level, a two-sided marketplace may seem like twice the work. But when proven successful, two-sided marketplaces can have much more than twice the payoff.

Some of the useful metrics to measure when you are running a two-sided marketplace are:

STAGE/ TYPE	EXAMPLE METRICS
Supply-Side Metrics	# of sellers % of new seller growth rate % of activation rate (sign-up → active seller, 1st transaction) Cost of acquiring seller (seller CAC)
Demand Metrics	# of buyers % of new buyers growth rate % of activation rate (sign-up → active buyer, 1st transaction) Cost of acquiring buyer (buyer CAC)
Core Metrics	# of transactions on the platform % of growth of transactions Average transaction value Gross merchandise value (GMV) Net revenue % of net revenue per transaction Gross margin
Liquidity	% of listings that get booked within X days, optimize for higher % and lower time period
Retention	% of repeat orders

SaaS metrics

Commonly used for: B2B Software/SaaS companies, like New Relic, Zendesk, and Segment.

B2B Software or SaaS (Software-as-a-Service) is software that is built for business use cases rather than consumer. SaaS products are often purchased through a subscription model and typically have a long selling cycle (up to years in some cases). Since the time to purchase SaaS can be very long, it's imperative to measure leading indicators to know you're moving in the right direction before it's too late.

Common metrics at SaaS-led companies include:

STAGE	EXAMPLE METRICS
Marketing and Sales Metrics	Marketing qualified leads (MQLs) per month Sales qualified leads (SQLs) per month Sales pipeline per quarter Lead → opportunity conversion rate Lead → purchase velocity rate Annual ecurring Revenue (ARR) Average revenue per customer (ARPC
Product Metrics	Activation — % of users who sign up and reach a-ha moment Retention — N day retention, i.e., % of users using the product feature on nth day Feature adoption — % users who are using product / feature
Customer Success	Revenue churn Customer churn Customer health score
Unit Economics	Customer acquisition cost (CAC) Customer lifetime value (LTV) CAC Payback (months to recover CAC) Quick ratio — measures health of a SaaS business, an alternative to CAC:LTV. The higher the value the healthier the company. SaaS business with a Quick Ratio of four and above are generally considered to have "healthy" growth Gross margins

Analyzing metrics

While identifying and defining your key metrics is a great starting point, it's also helpful to use frameworks to add more context to insights. Think of a framework as a guideline that will help you breakdown your data into digestible parts, rendering it more actionable.

For example, let's say you're tracking time to pay back the cost of customer acquisition (CAC Payback) as a key metric. Just knowing your CAC Payback metric in and of itself is probably not all that insightful. You'll also want to look at more detailed metrics, like which types of customers yield a faster CAC Payback, or how a paid acquisition campaign compares to customers coming from organic or referral traffic sources.

Example frameworks

Ideal customer profile

Not all customers are created equal. They come in many shapes and sizes, and can become loyal customers at vastly different rates. That's why it's helpful to look into the traits and qualities that make up your ideal customer profile.

For simplicity's sake, let's define a segment as a group of customers that share a common characteristic. You could segment customers by range of information like demographics, industry, pricing plan, or behavior, and compare the metrics of one segment to another.

For example, if you are running an e-commerce startup, you could segment customers by region to analyze how people in large cities make purchases compared to suburban areas. If you're in SaaS, you might want to look into churn numbers by company size (small companies vs. mid-market vs. enterprise) to define a segment to focus on.

Cohort analytics

Another powerful tool to analyze your metrics is cohort analysis. Instead of looking at all users as one unit, cohort analysis breaks users into related groups and compares similar groups over time.

For example, let's say you're building a SaaS product and a key metric you care about is retention. After surveying a handful of churned customers, you find many were unaware of core features at the time of sign up. So you decide to run an experiment. You change your onboarding flow to increase feature awareness, which would hypothetically increase retention. You quickly get to work and launch the new onboarding flow in February of 2018. When you run a cohort analysis based on month of sign up, you can track how your key metric (% of retained customers) is improving for uses who went through the new onboarding flow.

In the example below, you can see that cohorts of customers who signed up after Feb'18 are sticking with your product at higher rates compared to those who signed up in the past. Hooray!

MONTH OF SIGN-UP	% OF RETAINED CUSTOMERS IN LIFETIME MONTH				
	0	1	2	3	4
Jan'18	90.50%	87.78%	80.99%	78.87%	75.00%
Feb'18	91.00%	88.45%	83.40%	77.54%	
Marr18	95.80%	92.50%	91.40%		
Apr'18	96.10%	93.80%			
May'18	95.45%				

Aligning your team towards a single metric

Metrics are a great way to set direction and alignment inside your team (and entire company for that matter). As we know from earlier, it's easy to lose focus and fall victim to "shiny object syndrome" — that is to say, an affinity toward the latest technologies and trends. This is where the concept of "One Metric that Matters" (OMTM) can be a useful tool.

OMTM is a powerful concept that can align your entire company around a single metric. Your OMTM measures the positive impact as a whole, and all other metrics tracked will essentially cascade from this one metric.

To help illustrate this, here are a handful of OMTM examples of few well-known companies:

COMPANY	ONE METRIC THAT MATTERS
Square (SaaS-enabled marketplace)	Gross processing volume (GVP)
Airbnb (marketplace)	Nights booked
Amplitude (SaaS)	Weekly querying users (WQUs)
Salesforce (SaaS)	Average records per account

Understanding why metrics matter and what model to start with is an important step. After that, you'll need to get more specific and define the right customer data to track that's particular to your use case.

About Segment

Segment provides the customer data infrastructure that businesses use to put their customers first. With Segment, companies can collect, unify, and connect their first-party data to over 200 marketing, analytics, and data warehousing tools. Today, over 15,000 companies across 71 countries use Segment, from fast-growing businesses such as Atlassian, Bonobos, and Instacart to some of the world's largest organizations like Levi's, Intuit, and Time. Segment enables these companies to achieve a common understanding of their users and activate their own data to make customer-centric decisions and create individualized experiences.

CHAPTER 3

Track the Right Things, Not All the Things.

Contributor: Calvin French-Owen
Co-Founder and CTO @ Segment

"What should I track?"

Ever since the early days of Segment, customers have asked us for advice on what sorts of data they should collect. Even after helping thousands of companies answer this question, it's still surprisingly hard to answer.

What makes tracking so challenging? It comes down to the fact that no business is exactly the same. **What you track depends on your business and only your business.**

There is no "shortcut" for thinking critically about your goals and measurement. A company is a machine that works to create value. You should measure your progress toward creating that value. And, more often than not, your progress and measurement will rely on And the key to measuring that progress is instrumenting an accurate representation of your customer data and actions.

That said, here are a few tips and techniques we've seen work again and again when it comes to understanding user behavior and building out tracking for your product.

Walk before you run

When first creating a tracking plan, we recommend starting with a limited number of metrics to track—even if it's just one. Having only one metric provides an unparalleled focus for teams working towards a common goal. It simplifies progress to a single yardstick measure success and failure with.

> **WHAT IS A TRACKING PLAN?**
> A tracking plan clarifies what events to track, where those events live in the code base, and why those events are necessary from a business perspective.

As a way of narrowing down your metrics, a "future self" test will give you confidence that you're tracking the right things. To do so, propose a timeline (typically three to six months out) and ask yourself questions like the following:

- What things would have to be true about my metrics to make my **future self** happy with the intended outcome?
- If I increase the number of active users by ten times in the next six months, will my **future self** be happy?
- If it's only doubled, will my **future self** still be happy?

If you can give an emphatic "YES" to these future self questions, then you're on the right track! If not, we suggest pressure testing different metrics until you find one that creates happiness for your future self.

Note: In our experience, it's easy to fall into the trap of a 'red herring' metric. These are metrics that at first glance seem like they should do what you want, but after closer inspection, don't actually deliver the results that you need. A red herring metric might be something like pageviews to your homepage, when what you actually want are paying, engaged users.

Measuring your metric(s)

Once you've chosen your key metrics, it's time to get started implementing them across your product, app, or website.

We've found the following key principles helpful for tracking metrics:

CLEARLY DEFINE WHAT YOU TRACK

It's possible (and even probable) that the definitions of your metrics will change over time. You may, for example, move from monthly revenue to annualized, add or remove pricing plans, or come up with new definitions for user engagement.

If you don't provide crisp definitions for what you want to track, it's still possible to optimize for metrics that don't have an impact on (or even hurt) business objectives. Just as important, it's worth spreading these definitions around your organization. We've done this in a number of places internally at Segment, so here's a quick example to help give you the idea of what makes for a good metric:

> *The "habit moment" occurs when a user is sending data and enables their second destination within seven days of sign-up.*

This metric is time-bound, clearly defined, and makes sense to anyone at Segment reading it.

What's more, notice that our habit moment is made up of a combination of various leading metrics as inputs—sending data, two destinations, and signed-up in a set time. So to improve this metric over time, we can run experiments that move our input metrics like reducing friction to sign-up, suggesting tools to start with, or surfacing tutorials for onboarding.

Start from a single source of truth

You want one reliable means of calculating the metrics you track. Whether that's a single line of code templated into all pages, a well-defined ETL (Extract Transform Load) path that pulls from Stripe, or a metric used across Salesforce reports.

If possible, you'll want the source of the metric to be as close as possible to wherever that data is generated. In our experience, the hierarchy of data fidelity looks something like the following:

Your database or system of record has the most correct view of the data (though is probably the most complex to get at). Sending data from your servers has the next best fidelity (though also requires more work to implement). And finally, client-side scripts have the lossiest data due to ad blockers or network disconnects from the browser.

No matter how you're tracking your metrics, it's critical that they're all reproducible and generated in one place. If you have twenty different lines to track a **sign up** event, you're going to see inconsistencies in many of the tools your teams use.

At Segment, we do this in a variety of ways. We have revenue and other critical business metrics pulled directly from the sources of truth: Salesforce and Stripe, respectively. Important metrics (like sign-ups or payments) are tracked directly from our servers. For more "indicative" metrics like page views or other user interactions, we track those directly from the page.

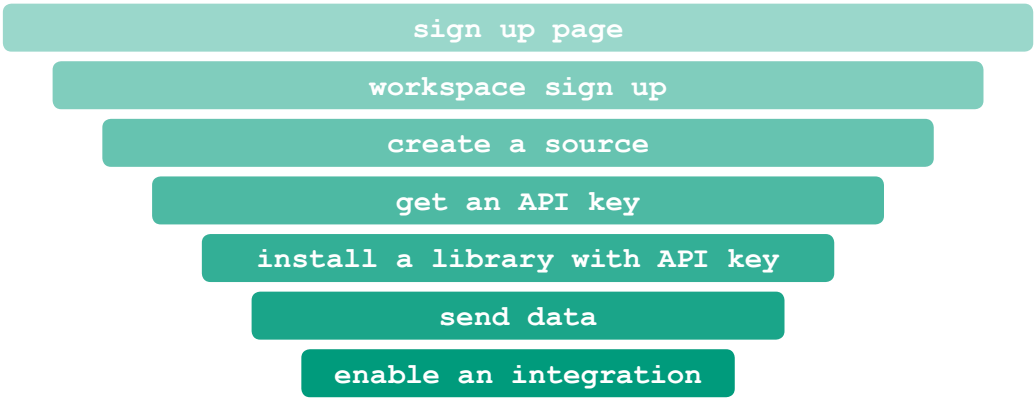
Back out the funnel

Of course, tracking only a handful of metrics won't suffice in the long run. While your goal metric could be your *One Metric that Matters (OMTM)* or a handful of metrics you've committed to using the "future self" test, you'll still need a set of waypoints to get there.

That's where your "funnel" comes in play. A funnel is really just a series of steps that lead up toyour goal metric. Measuring the core steps in your funnel is incredibly helpful as it allows you to easily pinpoint which area of your funnel has the biggest drop off in conversion.

For example, let's say a simplified version of your funnel goes from 1) a web visit to 2) a free trial sign-up to 3) an activated user and ultimately ends with a 4) paid customer. If your conversion rate from sign-up to active user is 5%, and your conversion from activated user to paid customer is 40%, it's pretty obvious where to focus your efforts—incentivizing those new users to activate!

Mapping out each part of your funnel into more granular subsets will give you more clarity on where to focus. At Segment, the activation piece of our funnel goes something like this:



Using the example above, we want users to successfully send data through Segment to an integration. But there's another set of steps that lead up to that point!

BEFORE...	USERS MUST...
sending data to an integration	enable that integration
enabling an integration	be sending data
data can be sent	install a library with their API Key
getting an API Key	create a source
creating a source	sign up for Segment
signing up for Segment	visit the sign-up page

Whew! Suddenly we have a more realistic picture of the steps that drive our metric. And we know exactly what steps we need to measure to evaluate progress.

Patching a leaky bucket

There's a lot of reasons why users won't make it to the end of your funnel. Maybe they are unclear on the benefits your product offers? Maybe the person who signed up for your product is not the right stakeholder? Maybe they just get bored and bounced?

Whatever the reasons for users dropping off, once you have the stages of your funnel defined, it should be much easier to dig in and triage the stages of drop-off.

One last thing to consider when thinking about your funnel is that it's also useful to distinguish *leading* metrics from *lagging* ones. Leading metrics (sign-ups, page views, leads) are typically much more moveable on a short-time horizon than lagging metrics (retention, churn, revenue, length of sales cycle). By focusing on leading metrics first, you'll be able to learn faster.

The what *and* the who

While measuring exactly what a user did is critical, it doesn't always paint the full picture of what's happening inside your funnel. Additional contextual user data points (we like to call them traits) like business size, role, industry, and location could have an impact on funnel metrics. If your product is built to serve mom-and-pop retail business, then your conversion metrics will have a noticeable difference depending on user industry and company size.

So how do you go about finding *who* is using your product?

First, you'll want to make sure that the metrics you track are associated with some sort of user ID—a unique identifier that lets you tie together events across user traits to give you a full picture of the customer journey.

> **Pro Tip:** Generally we recommend making this ID a unique identifier which never changes (the userId in your database). If you make the ID an email, you'll run into issues if your user ever changes their email.

Additionally, you'll want to assign a unique, temporary identifier for any user who isn't logged in. You may do this via localstorage, a first-party cookie, or a device identifier. But you need some way of tying together a user's path from landing on your homepage to when they actually sign up.

At Segment, we solve this problem via the **identify** call. An **identify** lets you tie a user to their actions and record traits about them. It includes a unique User ID and any optional traits available like email, name, location, etc.

So when should you use an **identify** call? We recommend doing so at key moments such as:

• After a user registers

• After a user logs in

• When a user updates their info (e.g. changes or adds a new address)

• Upon loading any pages that are accessible by a logged in user (optional)

The first three examples are pretty self-explanatory, but you might ask: why you would call identify on every page load if we're storing the userId in the cookie?

> **LET'S IMAGINE THIS SCENARIO:**
> A user logs into your app. Identify is called. For whatever reason, the user closes the browser and does not return until later. There's no way of knowing where that user will re-enter your app from. They could start my session from anywhere. And because there are many tools out there that require an initial identify call for certain features (e.g. Intercom chat widget) it's important to tell your end tools who the user is when they first start their session.

A few examples to tie it all together

Once you establish what actions to track and can identify *who* is taking those actions, you can combine the two to get a real sense of what is driving your core metrics.

To help you get started, we wanted to share a few tracking examples you can use. Below, you'll find top events that we see today at Segment. Some of these are simple funnel metrics (how many users viewed the homepage), while others are transactional (who completed an order).

Events worth tracking

SaaS	ECOMMERCE	MOBILE APP
Account Created	Product Clicked	Application Installed
Account Deleted	Product Viewed	Application Opened
Page Viewed	Product Added	Application Updated
Signed Up	Checkout Started	Application Backgrounded
Signed In	Checkout Step Completed	Application Crashed
Signed Out	Order Completed	Application Uninstalled
Invite Sent	Order Updated	Push Notification Received
Account Added User	Order Cancelled	Push Notification Tapped
Account Removed User	Order Refunded	Push Notification Bounced
Trial Started	Payment Info Entered	Install Attributed
Trial Ended	Cart Viewed	Deep Link Clicked
		Deep Link Opened

In addition to tracking events, you'll want to also capture properties (the what) associated with your tracked events. Here's an example of properties for a **Signed Up** event which is applicable across pretty much all industries.

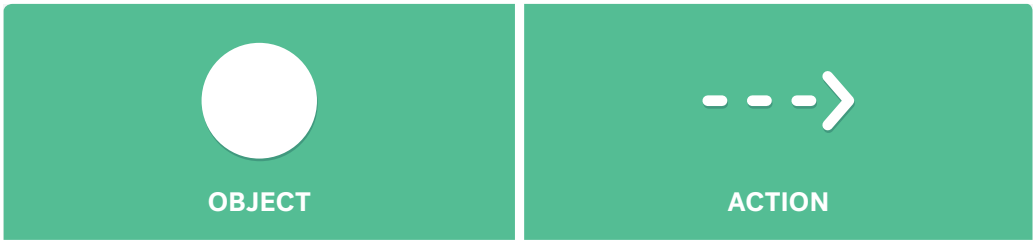
PROPERTY	TYPE	DESCRIPTION
type	String	The type of signup, e.g. invited, organic.
first _ name	String	The first name of the user.
last _ name	String	The last name of the user.
email	String	The email of the user.
phone	String	The phone number of the user.
username	String	The username of the user.
title	String	The title of the user.
context.groupId	String	The id of the account the user is joinning.

And finally, here's an example of the full track call for Javascript and iOS:

```
analytics.track('Signed Up', {
      type: 'organic',
      first _ name: 'Peter',
      last _ name: 'Gibbons',
      email: 'pgibbons@initech.com',
      phone: '410-555-9412',
      username: 'pgibbons',
      title: 'Mr'
}, {
      context: {
        groupId: 'acct _ 123'
      }
});
```

```
[[SEGAnalytics sharedAnalytics] track:@"Signed Up", properties: @{
      @type: @"organic",
      @first _ name: @"Peter",
      @last _ name: @"Gibbons",
      @email: @"pgibbons@initech.com",
      @phone: @"410-555-9412",
      @username: @"pgibbons",
      @title: @"Mr"
}];
```

Using the object-action framework



Song	—	Played
Account	—	Created
Product	—	Viewed

One last note to consider when it comes to data collection, is that the best way roll out a successful tracking plan is to establish consistent naming conventions from the start. This will not only make it easier to read your code, but it will also mean that everyone at your company can understand what your events mean.

As far as naming conventions go, consistency is key to scale. At Segment, we implement analytics using the object-action framework. We've developed this naming convention after working with thousands of customers on their analytics setup.

The idea is simple. First, choose your objects (e.g., Product, Application, etc.). Then define actions your customers can perform on those objects (e.g., Viewed, Installed, etc.). When you put it all together, your event reads Product Viewed or Application Installed.

We like the object-action framework because it makes it easy to do the following:

• Analyze the performance of a particular feature: "I want to build a funnel to see how many people who view products also add them to their cart. Righteous!" The events related to products are all next to each other in alphabetical order.

• Quickly scan a list of events to find what you're looking for: "Where are all of the product events? Nope, Nope. Got it."

• Impose a standard any marketer, analyst, or developer can understand: "I'm guessing this event called Product Viewed is about folks viewing products."

If you only take away one thing from this chapter, remember that the most important thing you can do when it comes to tracking is: **pick a single naming framework and stick with it.**

About Segment

Segment provides the customer data infrastructure that businesses use to put their customers first. With Segment, companies can collect, unify, and connect their first-party data to over 200 marketing, analytics, and data warehousing tools. Today, over 15,000 companies across 71 countries use Segment, from fast-growing businesses such as Atlassian, Bonobos, and Instacart to some of the world's largest organizations like Levi's, Intuit, and Time. Segment enables these companies to achieve a common understanding of their users and activate their own data to make customer-centric decisions and create individualized experiences.

CHAPTER 4

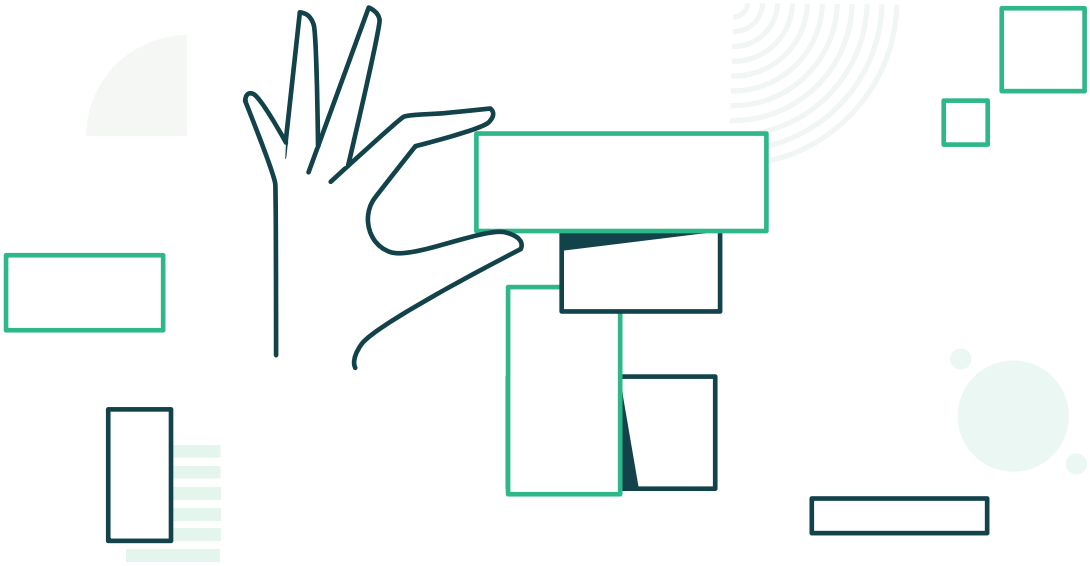
Choosing the Right Stack

Contributor: Eric Kim
Solutions Architect @ Segment

Now that we've established the what and how of customer data collection, we can move on to the where.

So where will you send your data? It's not an easy question to answer.

At the time of this writing, there are nearly 7,000 marketing technology tools listed in the **Marketing Technology Lumascape**. Most of these tools are built to help you make sense of and act on your data.



Needless to say, there's a lot of noise. Identifying **which companies are growing rapidly** and which categories are ripe for disruption is no simple task. And with more and more marketing and data tools coming on the market each year, finding the right one can be overwhelming.

In this chapter, we'll help you navigate this crazy world by breaking down how vendors compare. Let's get started.

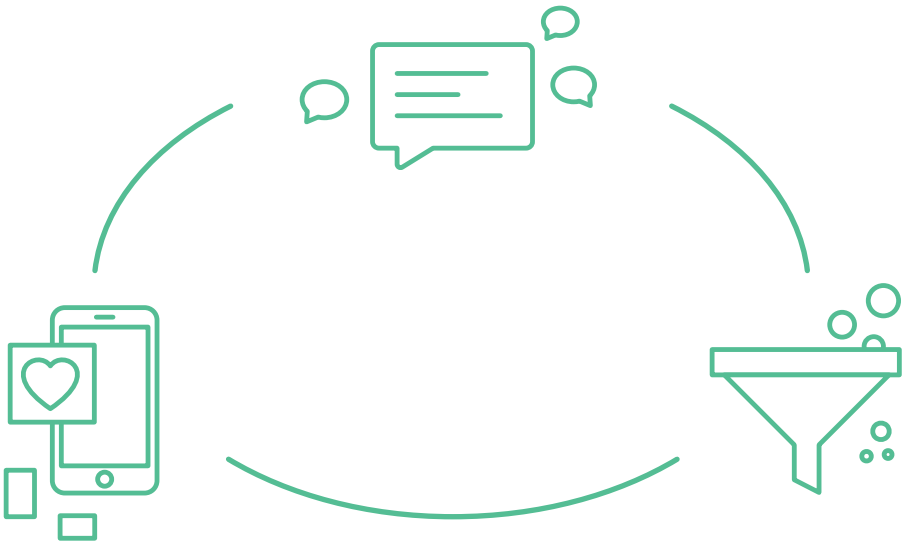
Start with business objectives

Finding the right for your use case starts with a bit of self-reflection. A helpful mental model to make tool selection more manageable is to group tools by use case or category. While Segment integrates with upwards of **300 tools**, many help accomplish similar outcomes. To help simplify things, we like to group tools into categories such as: Analytics, Email Marketing, Advertising, Customer Support, Attribution, and Push Notifications.

You can also prioritize tool selection based on your current initiatives. For example, if you're looking to convert more free trials into paid customers, perhaps you'll want to look into a messaging automation tool to surface reasons to go paid. Or maybe you're looking to get more insights about customers and Google Analytics is not cutting it? You could prioritize an advanced analytics tool.

We find it helpful to ask yourself a few preliminary questions like: *What are my business objectives? Where do the majority of customers interact with my brand? What industries do we serve best? What's the composition of my team (engineers, product, and marketing)?*

The general idea is to map your business objectives to a class of tools first, then determine which specific vendor will be best for your specific use case. Below we've mapped out some groupings to help you do just that.



Tools for understanding customers

Without question, the foundational class of tool to better understand your customers is **Analytics**.

Within this class alone, there are a wide variety of vendors that provide robust user analytics services—things like engagement tracking, funnel or cohort analysis, retention, and journey mapping. Some of the most widely used analytics tools include Google Analytics, Amplitude, Mixpanel, and Adobe Analytics.

A FEW TIPS FOR CHOOSING ANALYTICS TOOLS

Start with a list of must-have features
If you've used analytics tools before, then you will probably have some sense of what features are must-haves, nice-to-haves, and features you never use. When searching, it can be helpful to create a matrix of features that you care about.

REQUIRED FEATURES	SUPPORTED?
Digital Analytics	✓
iOS/ Android SDKs	✓
Standard Event Tracking	✓
Custom Event Tracking	✓
Funnel Analysis	✓
Conversion Tracking	✓
NICE-TO-HAVE FEATURES	**SUPPORTED?**
Cohort Visualization	✓
Predictive Analytics	✗
A/B Testing	✓
Automatic Notifications	✗

MOBILE VS. WEB

One thing to watch out for is analytics vendors who started with web analytics and developed mobile SDKs as an afterthought. If you're a mobile-first company, then you'll want to use an analytics tool that was built with mobile in mind. Tools like Mixpanel and Amplitude, for example, were built in the mobile era and support a multitude of mobile SDKs, in-app events, push messages, and more.

LOOK BACK TO YOUR TRACKING PLAN

In the last chapter, we outlined a path for creating a tracking plan. With this tracking plan in hand, you'll be able to identify tools that can easily support your event tracking requirements. If you don't have a tracking plan, it's helpful to first think through questions you'll need an analytics tool to answer—how are leads converting from one page (or screen) to the next? How do conversion rates compare week over week? Did this UX change impact revenue?

Tools for communicating with customers

In addition to leveraging an analytics tool to better understand your customers, you'll likely need a tool to efficiently communicate with them.

EMAIL MARKETING

The primary method for communicating with customers has traditionally been over email, using an **email marketing** tool—sometimes referred to as an Email Service Provider (ESP). Although some claim that email is dead, email communications is still used by the vast majority of companies no matter what industry or size.

Recent changes in the ESP landscape have made choosing an email service provider (ESP) less straightforward than it used to be. Some of those trends include the following:

LARGER VENDORS KEEP ACQUIRING EMAIL TOOLS.

ExactTarget and Pardot were acquired by Salesforce. Marketo was acquired by Adobe. Oracle acquired Eloqua. And the list goes on and on.

EMAIL TOOLS ARE OPTING FOR LESS SPECIALIZATION TO GENERATE BROADER APPEAL.

Email providers that were originally geared toward small and mid-sized businesses like Constant Contact and Campaign Monitor moved up-market in an attempt to meet the needs of enterprise companies. Similarly, ESPs that had been primarily transactional auto-response email providers, like Sendgrid, began adding more functionality as they matured.

UP-AND-COMERS HAVE ENTERED AND SHAKEN UP THE MARKET.

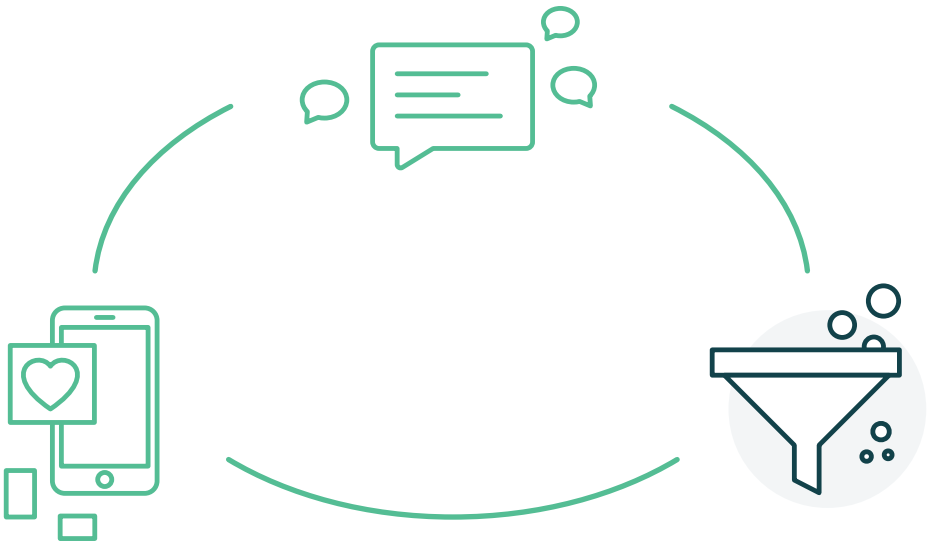
Some new vendors in the space—many of whom were originally sales tools or mobile marketing platforms—are now trying to expand their offerings and market themselves as ESPs or marketing suite alternatives. These email tools include companies like Customer.io, Drip, Autopilot, and Iterable.

MOBILE MESSAGING

If you're at a mobile-first company, there are many tools that provide you with ways to communicate with customers via **SMS** or **push notifications**. Braze, Iterable, and Kahuna are great examples of mobile analytics providers that also offer tools to communicate with your customers across mobile devices via SMS and Push Notification. It's also worth noting that many email and mobile messaging tools are merging feature sets. Many companies such as Braze, Iterable, and Customer.io offer support for both email and mobile communications with customers.

LIVE CHAT AND OTHER MESSAGING

Another way to communicate with users and customers is through a live chat or messaging app. Tools like Intercom and Drift allow go-to-market teams to start a conversation with their users or customers in real time through a live chat plugin that appears within an app or website.



Tools for acquiring more customers

If a key channel for growth is paid acquisition, then you'll most likely want to send user data to **advertising** platforms for use cases like retargeting non-converted visitors, enhanced audience targeting, and suppressing existing customers.

To make your paid advertising efforts more effective, you can pass first-party user data into ad platforms like Google Ads, Facebook, Twitter, LinkedIn, and other niche platforms. Doing so will allow you to target audiences based on their stage in the buying cycle, familiarity with your brand, and user traits (e.g. industry, job title, geolocation).

You can also pass your first-party customer data into a variety of other non-direct advertising platforms. In Ad Tech, these are referred to as Demand Side Platforms, some of which are Criteo, MediaMath, AppNexus, and AdRoll.

Tools for delivering a better customer experience

Another category of tools to consider is **experience optimization**, also commonly referred to as experimentation or A/B testing tools.

These tools offer powerful ways for product and marketing teams to test, learn, and deploy winning digital experiences to engage more users and drive more conversions. Most of these tools also take the custom development and measurement work out of setting up an experiment or personalizing the user experience by surfacing relevant content.

We also like to breakdown experience optimization tools into the following two categories–A/B testing and personalization.

A/B TESTING

Optimizely, VWO (Visual Website Optimizer), Google Optimize, Apptimize, and Leanplum are all examples of prominent players in the A/B testing space. Like the analytics tool category, some of these tools were built for mobile-first companies and others for web. When making a selection, we suggest building a table similar to the analytics features matrix above.

WEB AND IN-APP PERSONALIZATION

In addition to running experiments to improve conversion, you can also use tools to deliver a more personalized experience to users. Tools like Appcues, Optimizely, Webengage, and Leanplum allow you to dynamically modify the user experience on your website or app. Personalized experiences can be made via a number of dimensions such as user demographics, product usage, stage in customer lifecycle, etc.

Dig deeper to understand your customers

Data Warehousing is the last category of tools we'll highlight here. It's an important one, as a data warehouse is often the source of truth for data at many organizations. You can think of a data warehouse as a home for all of your data. Companies use a data warehouse to aggregate data from a number of different data sources so it's easy to analyze.

With a data warehouse, you have ultimate flexibility for how you store and later query your data. It helps you answer those tough analytical questions that your board may be asking about that aren't possible to do with your standard analytics tool.

DATA WAREHOUSE CONSIDERATIONS

You should consider a data warehouse if you want to do the following:

- Centrally store all of your business-critical data
- Analyze your web, mobile, CRM, and other applications together in a single place
- Dive deeper than traditional analytics tools by querying raw data with SQL
- Provide multiple people access to the same data set simultaneously

If you do decide a data warehouse is necessary for your team's needs, there are a number of important factors to consider when making a selection:

- **Data types:** what type of data you want your warehouse to store
- **Scale:** the amount of data you plan to store
- **Performance:** how quickly you need your data when you query it
- **Maintenance:** how much engineering effort you're willing and able to dedicate to your warehouse
- **Cost:** how much you are willing to spend on your data warehouse
- **Community:** how connected your warehouse is to other critical tools and services

Many of the factors listed will directly influence one another, and tradeoffs may be necessary. For example, opting for less scale may decrease performance but will typically be more cost-effective.

For more info data warehouse selection and considerations, check out our in-depth guide here: **How to choose the right data warehouse.**

Hopefully, this framework gets the wheels spinning on where to begin with selecting the appropriate tools for your stack.

If you want to go even deeper, we've written a seven lesson course that goes into detail on the categories above and other tools for attribution, performance monitoring, and business intelligence. Get started with that course here: **Choosing the Right Stack.**

Finally, if you want to get a full list of all classes of tools available on Segment, check out our **integration catalog here.**

About Segment

Segment provides the customer data infrastructure that businesses use to put their customers first. With Segment, companies can collect, unify, and connect their first-party data to over 200 marketing, analytics, and data warehousing tools. Today, over 15,000 companies across 71 countries use Segment, from fast-growing businesses such as Atlassian, Bonobos, and Instacart to some of the world's largest organizations like Levi's, Intuit, and Time. Segment enables these companies to achieve a common understanding of their users and activate their own data to make customer-centric decisions and create individualized experiences.

CHAPTER 5

Infrastructure for Your Customer Data

Contributor: Eric Kim
Solutions Architect @ Segment

Look into any category leading company—Nike in footwear, Netflix in streaming, Amazon in e-commerce and cloud computing—and you'll find there's a high correlation around customer obsession.

These companies dominate market share with one simple strategy: *They put customers first*. And they do so by applying a next-level focus to understanding their core customers, delivering hyper-personalized experiences, and offering high-quality products or services. The most effective customer-first companies today use data and technology solutions to deliver seamless experiences no matter how customers engage with their brand.

Because most customer interactions now occur over a digital mediums (mobile app, website, in-store purchases, etc.) you need a reliable data infrastructure to collect, unify, and act on your customer data. What this entails technically, is the ability to collect every first-party user interaction and integrate that data into the many platforms your teams use.

This is what we at Segment call **Customer Data Infrastructure**, or **CDI** for short.

Throughout the rest of this chapter, we'll outline a path for you to build a sound data infrastructure so that you can truly put customers first and use data in a way that's mutually beneficial for your customers and your business.

Three Pillars of Customer Data Infrastructure (CDI)

Customer Data Infrastructure is the technical foundation for any customer-first organization. By taking an infrastructure-based approach to the unification, standardization, and activation of data, you can ensure your data is consistent across all of your tools, and it's high quality. As a result, you'll have more opportunities to personalize user experiences, faster time to insight, a full picture of every customer. In other words, **you can trust in your data**.

CDI is made up of three critical components to make all of this happen:

1) **Data integration** to connect and unify your first-party data
2) **Data governance** to ensure your data is accurate and trusted across teams
3) **Audience management** to cater to customer preferences to deliver better experiences

Together these three pillars of CDI help you connect and unify your first-party data, ensure your data is accurate, and cater each customer interaction to that individual's preferences. In the section below, we'll detail how each pillar of CDI helps to unify, standardize, and activate your customer relationships.

DATA INTEGRATION

To get a complete picture of customers and their journeys, you're going to need to collect first-party data. CDI gives you the ability to connect every meaningful first-party customer interaction on every channel—from your mobile app(s) to website(s), in-store sales to backend systems, and from payment services to your CRM.

With the proper infrastructure in place, data will be available and consistent across all of the tools your teams prefer to use. These tools span from analytics tools (Google Analytics, Mixpanel, Amplitude) to marketing tools (Mailchimp, Optimizely, Braze, Intercom) to data warehouses (Redshift, BigQuery, Postgres, Snowflake) and many more.

> **Note:** The data we are referring to here is all first-party data, which is the data you collect first hand from direct user interactions across your owned platforms. It is **NOT** purchased or collected from an external entity like a data aggregator or data management platform.

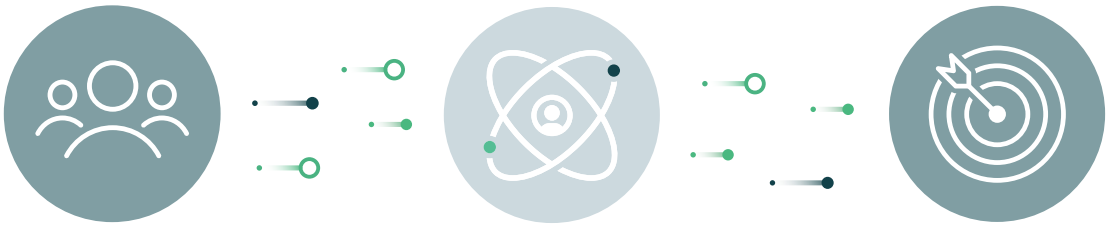
By decoupling data collection from vendor implementation, your engineers will save time writing individual integrations, your go-to-market teams can get up and running new tools faster than ever, and everyone can work from a unified history of the customer journey.

DATA GOVERNANCE

To trust that data in your downstream tools is accurate, consistent, and complies with internal privacy and security policies, you need a robust set of tooling for data governance. No matter how strict or laissez-faire your data implementation process is, you're still bound to eventually uncover data errors, missing fields, and duplicate information that slip through to production.

As data discrepancies surface, negative cascading effects will begin to infiltrate your organization. Marketing teams won't use data to deliver personalized customer experiences, or even worse, send inaccurate communications that cause customer confusion. Analytics teams will spend unnecessary hours piecing together an accurate representation of the customer journey, resulting in delayed insights and action. Product teams could be misinformed and introduce features that have negative impacts to activation and churn. The list goes on and on.

On the other hand, properly instrumented customer data infrastructure addresses such problem areas. CDI gives you confidence that your data is accurate by enforcing common data standards across your organization. It ensures consistency and security by giving you total control over what constitutes good, clean data vs. bad, unnecessary data at the source of generation.



AUDIENCE MANAGEMENT

Once you've collected the broadest set of raw data and trust that it's accurate, your next move is to make that data actionable. This is where audience management (our final component of CDI) comes in to play.

Audience Management resolves all of your user actions into profiles and allows you to identify what's most important to each user, like their favorite brands and average purchases per month. It allows you to build rich, relevant audience profiles that sync/update in real time as customers engage. This ensures a fast, consistent, personalized experience as your customer interacts with your brand.

What CDI Unlocks

The combination of the three pillars of CDI will equip your organization with reliable data infrastructure to understand and interact effectively with your customers across all channels. This unlocks organizational benefits in the following areas:

COMPANY-WIDE CUSTOMER DATA FOUNDATION

CDI democratizes high-quality data throughout your organization. Gone are the days of analytics teams being a bottleneck for accurate reporting and audience building. Because CDI federates customer data across all integrations and tools where data is used, all teams can be confident their working with the same, accurate data set. On top of that, turning on new tools for marketing and analytics can be done in an instant and can be easily backfilled with historical data. And finally, your engineering teams can focus on building your core product rather than integrations, data pulls, and cleanup.

NEVER MISS A DATA POINT

Because CDI is centers around first-party data collection at the source, you can feel confident your customer data gets to from where it originates to where it needs to be in the most efficient way possible. CDI unifies customer data across all data sources — web, mobile, servers, and more — and stores historical data so your customer profiles can become more complete over time.

GOVERN YOUR DATA

CDI safeguards the integrity of your data and ensures your team has access to high-quality data in each and every tool used. With CDI, you can apply a single data spec to multiple data sources—website, apps, servers, etc.—to ensure you have consistent and accurate data anywhere you need it. This unlocks the ability to diagnose data quality issues before they impact production and block unexpected data types and events from reaching your data warehouse or marketing tools. On top of that, it becomes much easier to comply with emerging privacy standards and customer preferences by carrying them through your whole marketing stack.

ACT ON YOUR DATA

The ability to act on your customer data with confidence unlocks enormous business impact. The list of ways to act on your data is practically infinite, but here are a handful of ways we've seen customers take action:

- **Build granular audiences** for marketing campaigns to deliver a personalized message and offer at scale
- **Identify target audiences** by identifying common characteristics of your best, most active customers and using that to inform look-a-like audiences
- **Capitalize on abandoned carts** by identifying and targeting abandoners across more channels than just email
- **Reduce churn** by looking into leading behavioral indicators like log-ins, purchases, email opens to determine who's at risk
- **Enrich customer profiles** with computations derived from first-party data like last product purchased, favorite topics, or average order value

CDI stops bad data in its tracks

Good customer experiences rely on accurate and complete customer data. And, that accurate and timely data requires the right infrastructure. Without the right infrastructure, your teams won't be aligned, your data won't be right, and your tools won't be operating off of full view of the customer.

Ultimately, all of this misalignment will be reflected in your customer experience. With CDI, this does not have to be your reality. Instead, you can use data for good, deliver awesome user experiences, and make data a true differentiator for your business.

About Segment

Segment provides the customer data infrastructure that businesses use to put their customers first. With Segment, companies can collect, unify, and connect their first-party data to over 200 marketing, analytics, and data warehousing tools. Today, over 15,000 companies across 71 countries use Segment, from fast-growing businesses such as Atlassian, Bonobos, and Instacart to some of the world's largest organizations like Levi's, Intuit, and Time. Segment enables these companies to achieve a common understanding of their users and activate their own data to make customer-centric decisions and create individualized experiences.

CHAPTER 6

Focus on Impact, Not Integrations

Contributor: Brennan Gamwell
Engineering Product Manager @ Segment

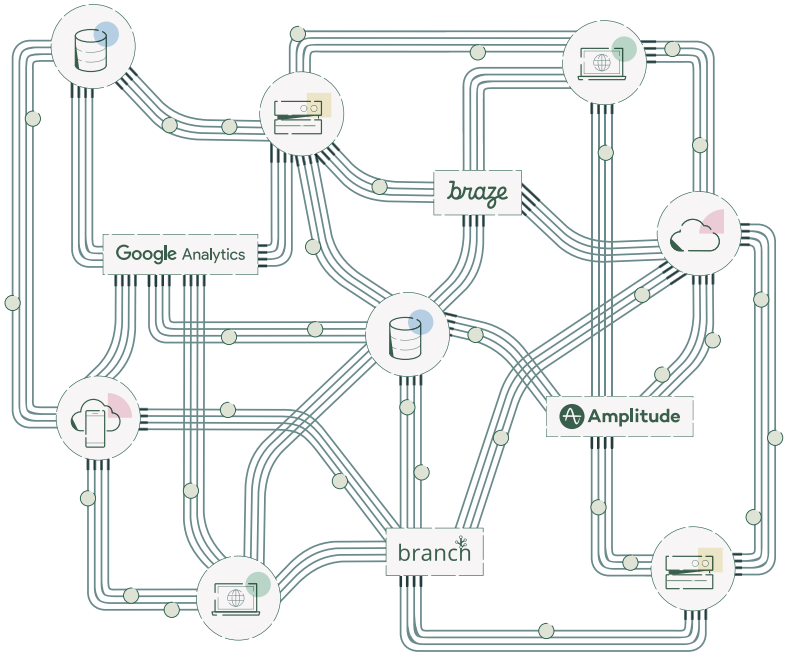
What Does It Take to Track Customer Data?

In a phrase, a lot.

Understanding user behavior via event tracking is a complex, choreographed dance among Product, Analytics, Marketing, and Engineering teams.

Product sees data as a means to innovate. For Analytics, the sky's the limit when they can derive data-driven insights. Marketing's benchmarks vary from optimizing the customer lifecycle to return on ad spend. And Engineering (perhaps being overworked) look to circumvent data integrations altogether to focus efforts on building the product.

When data collection is done right, each of these teams reap the benefits. *Doing so, however, is much easier said than done.*



Complexities with Customer Data Collection

After surveying Segment customers, we've found that gathering customer data is no small task. On average, a mid-sized company (200-1000 employees) dedicates 6-8 weeks per year to instrument a new data integration and 2-3 additional weeks to maintain it.

On top of that, there's rarely a use case where data is collected from a single source and routed to a single tool. So every new source of data collection (web, mobile, server-side, etc.) and every tool where that data needs to be delivered (analytics, email marketing, ad platforms, data warehouses, CRMs, etc.) compounds in time and resources needed for instrumentation.

As you can imagine, this is not ideal when your team is trying to innovate and grow as fast as possible.

We'll lay that a better path to collect, unify, and act on your data in this chapter. But first, let's take a closer into the obstacles that come up time and time again.

INTEGRATIONS WITH UNIQUE APIS

Implementing side-by-side data integrations with unique APIs introduces two unnecessary challenges:

> 1) Learning a new API for each new integration
> 2) Time required for quality assurance
>
> Although instrumenting any new tool that requires data is a relatively repeatable task—install the javascript snippet or SDK, setup event tracking, configure user identification, QA—each new tool comes with it's nuances.
>
> On the whole, this doesn't make much sense. Each new tool you implement relies on the same data, so why would you instrument tracking for each tool separately?

2. QA TIME

Imagine spending 6-8 weeks setting up a new tool, only to learn that the initial implementation included a crucial logic error.

Mistakes like this require not only more time to fix the issue, but also necessitate that code be re-deployed, and potentially, past data to be sanitized. Further, if your point of data collection is a mobile app, you may be cursed with bad data forever — there's no way to force a user to upgrade to the latest version of your app with the correct SDK.

OPPORTUNITY COSTS

Opportunity costs pile up. Instead of working on new features, engineers spend an increasing amount of time building and maintaining data integrations.

Let's think through the math. If an engineering team implemented 5 new data integrations, they would need to allocate between 30 and 40 weeks and between 10 and 30 additional weeks per year to maintain it. In the first year, that's between 40 and 70 weeks of employee time. Imagine hiring a team member and dedicating a substantial portion of another team member's time just to writing and maintaining data integrations.

The negative effects of opportunity costs inevitably cascade to other teams:

- Product innovation suffers because engineers are spending more time keeping up with a maintenance schedule rather than churning out new product features.
- Analytics insights are limited because the team doesn't have time to implement each data API in a way that gathers all the data needed to derive actionable insights.
- Marketing's ability to take action is at the mercy of other teams. Unless there's an engineer dedicated to the marketing team, it's unlikely that their team has the bandwidth to make small (but crucial) tweaks based on Marketing's analysis.

DATA SILOS LEAD TO DATA DISCREPANCIES

Each team in your company relies on a different source of truth for customer data.

Your sales team's go to is a customer relationship management system (CRM), your success team to a help desk, your marketing team to a data management platform (DMP) or customer data platform (CDP), and your analytics team to a data warehouse.

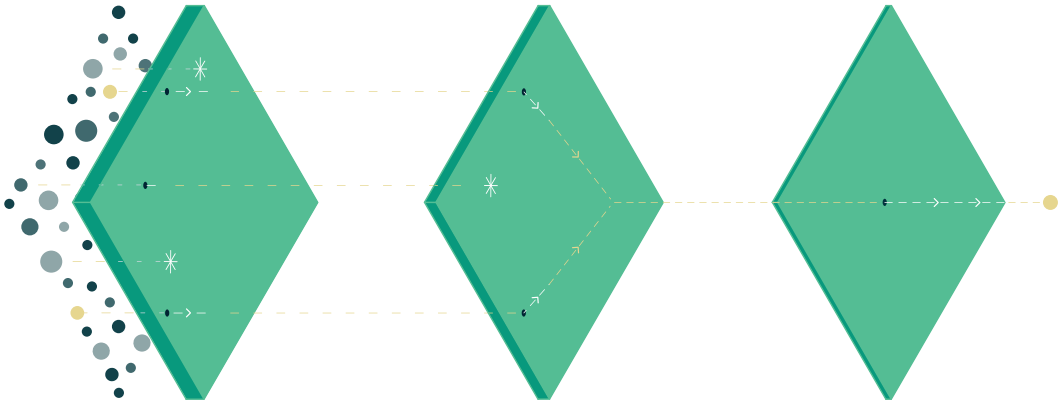
When data flowing to and from these tools is not consistent and updated it leads to an incomplete picture of your customer and variance in key metrics driving your business. On top of that, it often results in teams spending more time arguing about whose data is right than they do putting it to use.

A Better Way Forward

Clearly, there's a problem afoot — either allocate valuable human resources to maintaining data integrations or force teams that rely on that data to operate on gut feeling rather than being data informed.

So what's the alternative?

It just so happens that, many of the unique APIs needed to instrument tools for Analytics, A/B testing, Advertising, and other categories covered above, operate using much of the same data — clicks, page views, video views, purchases, etc. So it doesn't really make sense to instrument each and every tool you use individually on your website or app.



HERE ARE 3 THINGS YOU CAN DO:

1) Centralize your integations to a single platform
(hint: Segment can help with this)

2) Create a common data standard to enforce across all data sources
for trust and governance

3) Give your teams autonomy to add new tools either through the platform
or with helpful documentation about how to track data consistently

Out-Of-the-Box Solutions to Common Challenges

The benefits of leveraging a solution like Segment are many:

- Engineering teams are free to focus on building products rather than maintaining data integrations and pipelines
- Product teams get the insight needed to build sticky products and improve acquisition, activation, and retention metrics
- Marketing teams can analyze and optimize campaigns, targeting ideal customer profiles, and be more efficient with spend. Further, they can trust in data to send personalized communication to customers.
- Finally, Analytics teams have a clean dataset to work with which makes easier to write SQL queries and pull reports all the more faster

Let's dig in with even more specifics of how a customer data infrastructure (CDI) solution like Segment can provide out-of-the-box solutions to common intra- and cross-team challenges.

UNIFIED APIS

Instead of addressing data collection, connectivity, and access issues in a case-by-case fashion, a CDI addresses data collection at the source. A CDI unifies APIs across the tools your teams use, allowing Engineering teams to collect data once and route it to many downstream tools. Segment, for example, routes data to 200+ downstream connections via our standardized API. Further, gathering data using a single API guarantees data consistency when sent downstream to a source-of-truth repository such as an Amazon S3 bucket or a raw data warehouse. Immediately, Product and Analytics teams enjoy access to perfectly formatted data.

SINGLE POINT OF CONTROL

CDIs also offer a single point of control in the form of a user interface. Product, Analytics and Marketing teams, as a result, don't need to request Engineering time for small tweaks to settings, names, and other config issues that arise. Instead, they can leverage the interface to adjust settings on the fly, and watch their updates take effect right away.

Instead, they can leverage the interface to adjust settings on the fly, and watch their updates take effect right away.

TEAM AUTONOMY

CDIs empower users to create new audiences on the fly, and populate these audiences to one or many data integrations at a time. In particular, our Personas offering allows Marketing teams to build custom audiences that combine user traits and activity — all from a user friendly GUI that doesn't require knowledge of SQL.

Finally, CDIs collect data not only from their own API, but from other popular tools as well. Segment fetches data from 28 downstream tools, routing it to a raw data warehouse so your Product, Analytics and Marketing teams have access to not only Segment-collected data, but data originally stored or segmented elsewhere.

In short, leveraging a CDI will empower teams to innovate and optimize and focus on impact, not on the minutiae of implementing and maintaining data integrations. Moreover, the availability of CDIs should make the answer to at least one question crystal clear: Buy, don't build, when it comes to data integration solutions.

About Segment

Segment provides the customer data infrastructure that businesses use to put their customers first. With Segment, companies can collect, unify, and connect their first-party data to over 200 marketing, analytics, and data warehousing tools. Today, over 15,000 companies across 71 countries use Segment, from fast-growing businesses such as Atlassian, Bonobos, and Instacart to some of the world's largest organizations like Levi's, Intuit, and Time. Segment enables these companies to achieve a common understanding of their users and activate their own data to make customer-centric decisions and create individualized experiences.

CHAPTER 7

Building Trust in Your Data

Contributor: Kevin White
Head of Growth Marketing @ Segment

Data can be your best friend or your worst enemy.

When it's collected in a standardized way and consistent across the tools where it's used, data acts as your foundation for unlocking growth. As a result, teams who rely on data will trust that it's accurate and understand what it's telling them so they can use it effectively.

On the other hand, messy data causes less than ideal outcomes like misdirected communication with customers, confusing user experiences, ill-informed strategy decisions, and delays in time to business insight.

To gain widespread trust in data across your organization, it's critical to lay the right infrastructure and process for how it's collected, cleaned, governed, and acted on.

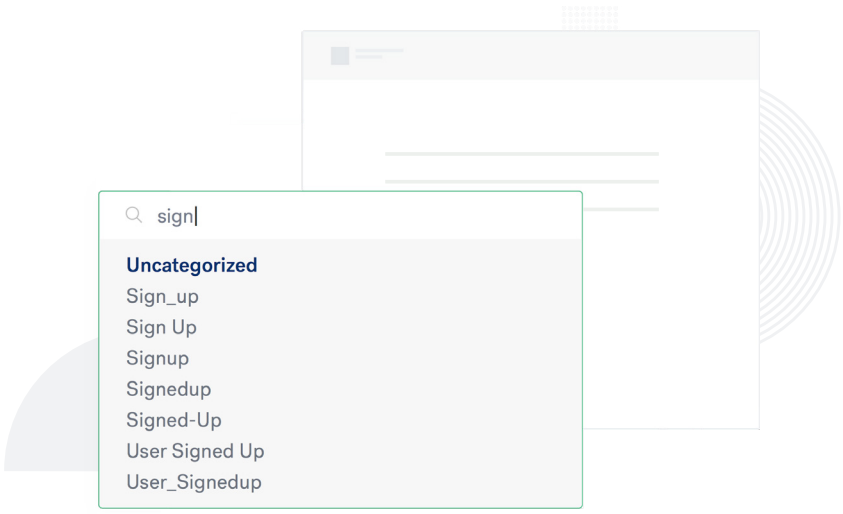
In this chapter, we'll outline a path to do just that. We'll start with a method for getting everyone on the same page, share proven frameworks that you can use data collection, and provide a few pointers for building trust in data across your organization.

Get everyone on the same page

Your first step to cleaning up your data? **Naming conventions.**

Naming conventions ensure your data is collected in a uniform fashion so that the teams who use it are all on the same page and speaking the same language.

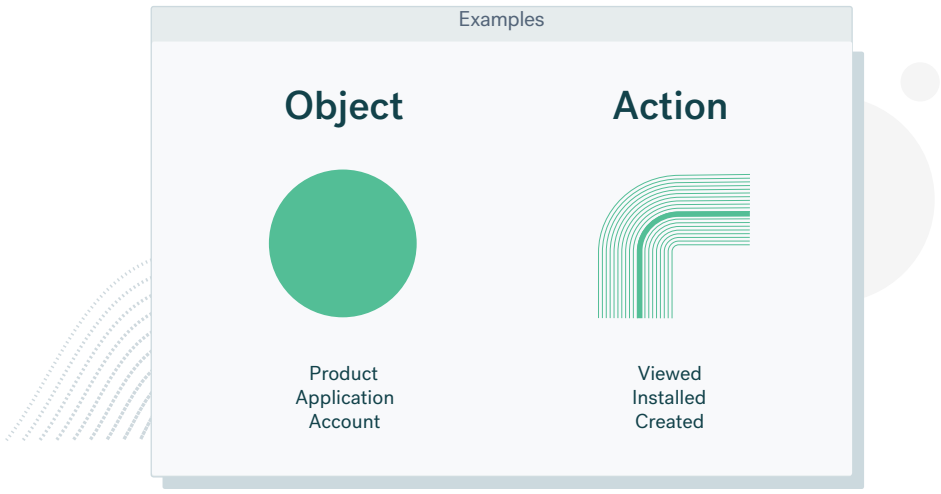
You may not realize it, but there are many ways to name the same user interaction. Take, for example, the simple action of a user signing up for a newsletter. You could implement the event as “Sign up,” “Signup,” or “User Signed Up.”



Without a consistent and agreed upon naming convention, you’d inevitably collect a mixed bag of data that varies by the preferences of whomever implemented it. In the example above, your teammates would be left guessing which event actually corresponds to a user signing up for your newsletter. This may not seem like a big deal, but imagine how confusing this will become as your traffic, customer base, and number of meaningful events grow over time.

To avoid getting into this situation and enable your company to actually put data to use, there are two simple things you can do today:

1. Align on a framework for naming your events and properties (down to the casting).
2. Put a process in place to enforce your company’s framework



When it comes to naming conventions, we highly recommend using the Object → Action framework as it's easy to understand and get everyone to adhere to.

For example, imagine you're building a music app that functions in a way that's similar to Spotify. To understand user activity and the context of that activity, you can apply an Object → Action naming convention. Here's how it to use the Object → Action framework when a song is played: First, choose your objects (ex: Song). Then define actions your users can perform on those objects (ex: Played). When you put it all together, your event reads `'Song Played'`.

The Object-Action Framework makes it easy to:

- Analyze a particular feature's performance
- Quickly scan a list of events to find what you're looking for
- Impose a standard that's easy for teams to understand

Of course, the object-action framework isn't the only way to do this.

You can use any order of actions and objects, and any type of casing. You can also use the present or past tense. What really matters is that you keep data collection consistent!

Developing your data dictionary

We've helped thousands of companies implement customer data collection and found that the most successful teams have one thing in common: *they use a data dictionary or data collection spec*.

A data collection spec clarifies what user actions to collect, where those events live in the code base, and why those events are necessary from a business perspective. These documents of record typically live in a spreadsheet. They serve as both as a project management tool and as collateral to align your team (and the entire organization for that matter) around what data to measure success by.

When first getting started, it's helpful to limit data collection to a handful of core user events. These events should also have a rich set of properties that can be used to give context on the action taking place. For example, if one of your core events was a user signing up for a free trial with your event being `'user signed up'` you'd probably also want to collect properties that give context as to who is taking that event, where they are coming from, etc.

Here's an example of what that event could look like in your code base:

```
analytics.track('1794362141', 'Order Updated', {
      'brand',:'Supreme',
      'browser': 'Chrome',
'category': 'Outerwear',
'currency': 'US Dollars',
'operating _ system': 'Web',
'position': 4,
'price': 79.99,
'producat _ id': 41232,
'quantity': 2,
'value': 150
});
```

When getting started, follow these rules when spec-ing out your data dictionary to keep it neat, tidy, and semantically useful:

- Don't create event names dynamically
- Don't create events to track properties
- Don't create property keys dynamically
- Make sure every event helps you answer a question about your business
- Start with your core customer lifecycle to construct your funnel
- Only add events as you feel they're missing

Data dictionary examples

Because data dictionaries can be a somewhat new concept for teams to become acquainted with, we've developed sample tracking plans for a variety of industries and use cases.

BASIC DATA DICTIONARY

Here is a simplified version of a data dictionary. We recommend starting with a plan like this before digging into more complicated tracking.

ADVANCED DATA DICTIONARY

Here are tracking plans we use to organize to help customers get started (and also for our own Segment tracking). Some of the event properties have been trimmed to keep things clean, but everything is here.

- SaaS Tracking Spec →
- Mobile Tracking Spec →
- E-commerce Tracking Spec →
- Video Tracking Spec →

Ensuring data quality and consistency

Any team who uses data benefits from pristine data quality. Product teams can iterate faster and build immersive user experiences with confidence. Analytics teams can build queries without heavy workarounds and inform-cross-functional decisions faster than ever. And marketing teams can inspire user action and improve advertising efforts by personalizing messaging according to user behaviors and traits.

Getting your organization to a state of high-quality data that all stakeholders trust takes a combination of alignment, validation, and enforcement.

Let's dig into each of those...

ALIGNMENT

All teams need to be aligned around the importance of data before it can be clean (or cleaned up) and trusted across your org. This means standardizingdata collection with an actionable data collection dictionary.

As mentioned above, we recommend using a "data dictionary" to document and standardize customer activity. We've found it helpful to assign a single owner of this document to oversee and enforce data standards throughout your organization. This owner should ensure:

1. Your naming convention and data schema is documented in a way that's easy for any team across your organization to understand
2. Collected customer data is collected in a uniform fashion that matches spec
3. Resources are available and secured to diagnose any areas where "dirty data" is introduced and apply a fix

VALIDATION

Once your data dictionary is agreed upon, set, and implemented, you'll want to make dirty data stays out of the picture. To do so, you'll need a method for validating how new user actions make their way to your codebase.

Even with rigorous naming conventions and instrumentation instructions, errors will be introduced if engineers don't receive automated feedback to help them identify and resolve issues during implementation. And when you're responsible for reviewing thousands of lines of code across dozens of events, it's inevitable that mistakes will happen.

A single tracking error on a business-critical event, like `'Lead Captured'`, can cost your business hundreds of thousands of dollars. The problem is that these bugs are typically detected weeks or months later, and by that time, the damage has been done.

Time is of the essence, so it's important to detect mistakes before they make their way to your production environment. Rather than manually trying to compare event payloads against your data dictionary, you'll want a way to automatically confirm when data matches your spec and alert you when it doesn't. There's a lot of ways to go about this, but (naturally) we prefer using our Protocols product to either 1) send a daily digest of current and new violations or 2) enable violation event forwarding to send violations as .track() calls to a Segment Source.

ENFORCEMENT

To really take data quality to the next level, you can implement a system and standards for data enforcement.

There's a wide range of variables to consider when it comes to enforcement. On the lighter side, you may want to only block PII data from reaching tools where it can be seen by anyone with access to said tool. Or (in more developed use cases) you may want to completely block all data which does not match your spec or schema from reaching any downstream tools. At first, it's probably best to start on the lighter spectrum of enforcement and slowly make your way to the other end of the spectrum.

If discarding data from blocked events sounds scary, there are precautions you can take to ensure no data is lost while still enforcing standards. For example, you could configure an isolated data warehouse to send data which doesn't meet your enforcement standards. Doing so will ensure that no data loss, and you could retroactively get discarded data into necessary tools with a bit of analytics and data engineering help.

ABOUT SEGMENT

Segment provides the customer data infrastructure that businesses use to put their customers first. With Segment, companies can collect, unify, and connect their first-party data to over 200 marketing, analytics, and data warehousing tools. Today, over 19,000 companies across 71 countries use Segment, from fast-growing businesses such as Atlassian, Bonobos, and Instacart to some of the world's largest organizations like Levi's, Intuit, and Time. Segment enables these companies to achieve a common understanding of their users and make customer-centric decisions.

CHAPTER 8

First-party Data

Contributor: Andy Schumeister
Product Marketing Manager @ Segment

With data protection regulations like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), it's no surprise that customers expect companies they engage with to respect and protect their personal data.

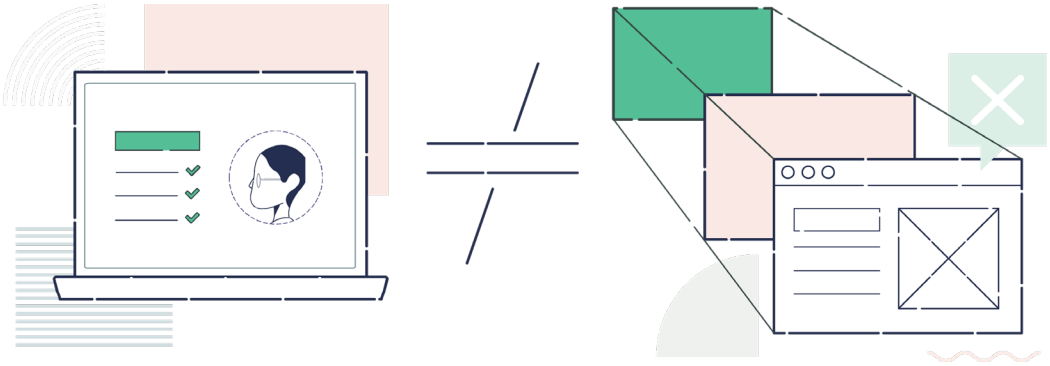
According to a study from Accenture, "87% of consumers believe it is important for companies to safeguard the privacy of their information." At the same time, "58% of consumers would switch half or more of their spending to a provider that excels at personalizing experiences without compromising trust."

*So what does it take to deliver respectful **and** personalized customer experiences?*

The short answer is first-party customer data. Activating this data requires the technical capability to collect every first-party interaction and integrate that data into the marketing and analytics tools your teams use to provide customer-first experiences.

In this chapter, we'll help you construct a data strategy to deliver private, respectful, and personalized experiences. We'll explain:

- The differences between first-party data and third-party data
- The benefits of using first-party data
- What it takes to activate first-party data



First-party data vs. third-party data

When it comes to customer data, *not all data is created equal*.

There are two distinct classes of data that marketers and businesses rely on to engage with their customers: first-party data and third-party data. These types of data are collected from different sources, are used for different purposes, and are even subject to different requirements under regulations like the General Data Protection Regulation (GDPR).

First-party data defined

First-party data is data you collect directly from your customers based on how *your* customers use *your* products or services. This includes information on which products a customer views or purchases from *you*, how often they visit *your* website or mobile app, and even *data that's stored in your* CRM. For the most part, your customers understand that you are collecting this data—for instance, providing it via a form completion—and expect that you use it to provide an intuitive user experience as they continue to engage with you going forward.

FIRST-PARTY DATA EXAMPLE

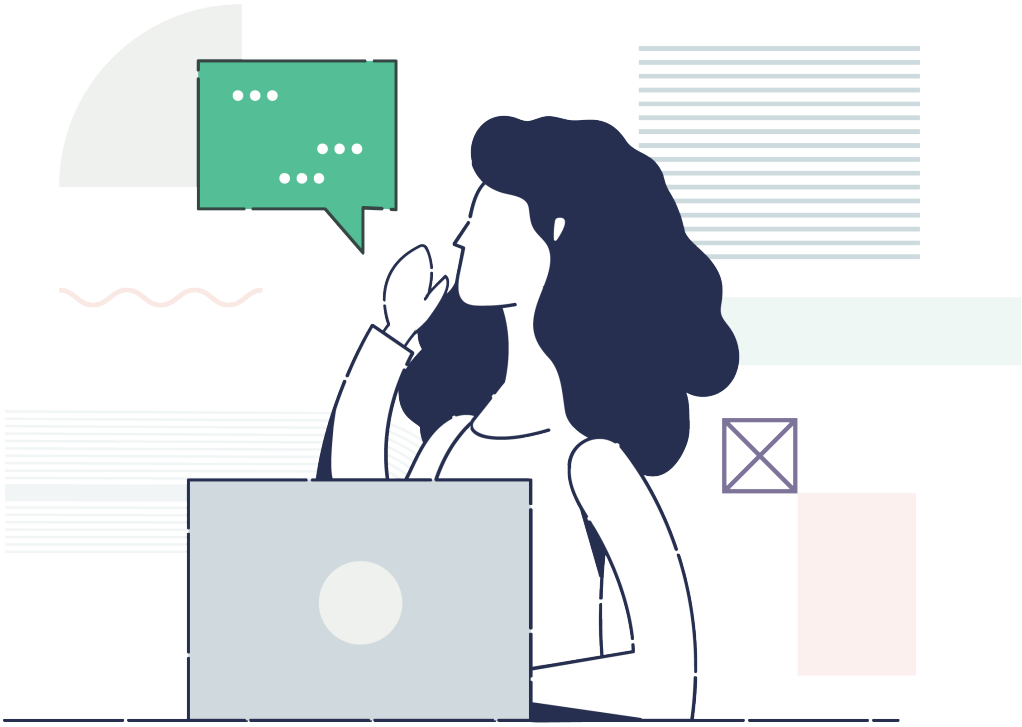
Here's an example: Let's say I go to Bonobos.com to buy a new pair of pants. Before I complete my purchase, Bonobos asks me if I want to create an account. I fill out a form and tell them my name, share a bit about my clothing style, and let them know that my preferred store location is in San Francisco. When I complete the purchase, the pair of pants also becomes part of my profile. This is all first-party data, or information, that I have knowingly shared with Bonobos.

Third-party data defined

Third-party data, on the other hand, is user or behavioral information that companies purchase or acquire from 3rd parties. It's often aggregated from multiple websites and segmented based on user interests, demographics, shopping behaviors, and more. This data is often collected with questionable consent and shared across companies without explicit consumer permission.

THIRD-PARTY DATA EXAMPLE

Here's an example: I apply for a credit card and provide details about my job, income, and address. If the credit card company were to sell that information (along with information from other applicants) to a real estate company, that company would be purchasing third-party data. While I directly provided the information to the credit card company, I did not choose to share my information with the real estate company.



At Segment, we often refer to the act of companies sharing third-party data with each other as "data gossip." If you've ever received an email promotion from a company you never shared your email address with, you've experienced data gossip. Your customers wouldn't tolerate their grocer telling their banker what they just purchased. And data gossip is no different. Moving away from third-party data will improve customer trust, which in turn will boost your brand's reputation.

The power of first-party data

First-party data is valuable for showing customers that you're attentive to their needs, showcasing products that fit their interests, or removing irrelevant content. It also has many advantages over third-party data. First-party data is not shared with other businesses, which is beneficial for both your customers and your business. It's typically more accurate than third-party data as well, because it reflects actual customer behavior from your own channels (web, mobile, in-store, etc.).

Many benefits of first-party data are due to the fact that the data is collected from customers you have a direct relationship with.

Here are a few reasons why having a direct relationship matters:

ACCURACY

Having a direct relationship with your customer means the data you collect from your customers is likely more accurate than third-party data. That's because the information is either provided directly from the customer or is based on their actual use of your website, app, or service. When third-party data is purchased, this data reflects a single point in time and degrades in quality over time.

RESPECT

Unlike third-party data, first-party data is collected with consent from your customers. This means that your customers are aware of the type of information you're collecting as well as how it's being used.

FIRST-PARTY DATA	THIRD-PARTY DATA
Direct customer relationship	Indirect customer relationship
Individual insights	Aggregated insights
Collected with consent	Not typically collected with explicit consent
Not shared with other businesses	Shared with other businesses
High accuracy	Low accuracy

What it takes to activate first-party data

To take action on your data, you first need to understand what it's telling you. And before that, the teams who use that data need to have confidence that it's accurate and reliable. To help you get to that state, we've outlined four core requirements for data activation below:

Requirements for data activation

ALIGNMENT

Before making rash business decisions or building a personalized user experience, it's important to get alignment between data stakeholders within your organization. This means coming to an agreement about what first-party data you will be collecting and gaining a general consensus for how it will be used.

STANDARDIZATION

After getting everyone on the same page about first-party data collection, you'll want to be sure all of the data you collect is standardized across the various touch points from where it's collected. This means establishing a source of truth that clearly defines what data you're collecting, provides a consistent naming convention and schema, and provides context as to how to interpret the data. Much of this was already covered in the previous chapter, in the section on developing your data dictionary.

VALIDATION

Next, you'll want to be sure your data is collected in the expected format defined in the previous steps. Even with rigorous naming conventions and instrumentation instructions, data that does not match your spec will inevitably make its way to your marketing and analytics tools. That's why it's important to have quality assurance (QA) checks in place to catch dirty data before it reaches the tools where you want to use it. This is a tedious problem to solve, and it's why we've incorporated a data validation feature into our Protocols product which automatically catches every incorrect property or data type found.

CONSISTENCY

The last thing you'll want to confirm before acting on your first-party data is that it's consistent across all of the tools where it will be used. This means that when a data value changes in one place that it's also reflected in other tools to prevent a disjointed customer experience. For example, let's say that a user on a free plan upgrades to a pro plan, becoming a paid customer. You'd want that new event `(Plan Upgraded)` to be consistent across any tool you use to reach customers so that your future communications will reflect their new status: a pro-plan customer.

Now that you've got a grasp on what first-party data is, the benefits of using it, and requirements for taking action on it, you can start to formulate a strategy to uncover insights and put it to use. Both of which will be covered in the last two chapters to come.