

# Building a Predictive Model

**Step-by-step recommendations for  
incorporating Intent data into your  
predictive models**

<b>Using Bombora data in a predictive model</b>	<b>3</b>
<b>Overview</b>	<b>3</b>
<b>How to use this guide</b>	<b>3</b>
<b>Recommended steps</b>	<b>4 - 5</b>
<b>Dive deeper, step-by-step</b>	<b>6</b>
<b>Setting goals</b>	<b>6</b>
Intent data goal examples	6
Example of use cases	6 - 7
<b>Identifying and organizing relevant topics</b>	<b>7 - 8</b>
<b>ADAT file dictionary</b>	<b>9</b>
<b>File delivery</b>	<b>10</b>
<b>Data ingestion best practices</b>	<b>10</b>
Domain considerations	10 - 11
Geography considerations	11
Mapping Bombora's Intent data	11
<b>Building a model with Intent data</b>	<b>12</b>
Predictive binary classification model	12
Identifying funnel stage	13
Gain momentum. What's next	13
<b>About Bombora</b>	<b>14</b>

# Using Bombora data in a predictive model

## Overview

Leveraging Bombora's Intent data in a predictive model can help data science teams gain insights into the interests and buying intent of potential customers, allowing them to create more effective models to influence business strategies and impact revenue growth.

This guide is intended to help establish goals for Intent data usage to drive quantifiable business outcomes.

## How to use this guide

The recommended steps on the following pages contain a high-level overview of incorporating Intent data into predictive outcomes. Steps are listed so as to build on each other, and they will call out other sections of the guide for you to reference in more detail.

Let's dive into how you can begin integrating Bombora's Intent data into your predictive model!

# Recommended steps

## 1

**Define the problem and set goals:** For example, you may want to identify which accounts are most likely to convert to customers, or which products are most likely to be purchased by a particular segment of your target audience. This will help you begin to map the potential outcomes from leveraging Bombora's Intent data and how you will use it to improve your model's accuracy.

See [“Setting goals” on page 6.](#)

## 2

**Create an initial topic list:** Bombora tracks the content consumption of millions of B2B decision-makers across the web, and provides insights into the topics they are researching. Determining which topics are most relevant to your business will allow you to leverage Bombora's Intent data effectively.

See [“Identifying and organizing relevant topics” on page 7.](#)

## 3

**Available datasets and file format:** Bombora provides Intent data via an “ADAT” (All Domains All Topics) file, which is a custom formatted file designed for easy use and processing of Intent data. The ADAT format includes several key data fields, such as the Bombora company ID, the topic being researched, and the intensity of research activity.

See [“ADAT File Type” on page 9 for the full list of fields included.](#)

## 4

**Determine optimal file delivery methods:** Bombora provides multiple options for accessing and receiving Intent data files.

See [“File Delivery” on page 10 for the list of available delivery methods.](#)

## 5

**Connect Bombora data:** When ingesting Bombora Intent data into your internal tools and integrating it with your predictive model, there are specific nuances to the Bombora dataset that need to be considered.

See [“Data ingestion” on page 10 for additional Bombora-specific items.](#)

6

**Build and integrate your models:** To connect Bombora's Intent data to the use cases that you defined in step 1, you'll need to first narrow your lists of topics to only those that drive meaningful business outcomes. The narrowed list of topics will help you begin to build your model.

See [“Building your models” on page 12.](#)

7

**Train and validate the model:** Now that you have generated your Bombora specific models, you'll want to integrate them into your company's predictive model. This includes using a subset of your Bombora Intent data to train the model, and then using another portion to validate the model's accuracy. This will help you determine how well the model is predicting buying behavior and identify areas where the model can be improved.

8

**Deploy the model:** Once the model is trained, validated and you are satisfied with the results, the model can be deployed to your production environment.

9

**Monitor and refine the model:** As new Bombora Intent data becomes available, continue to monitor the model's performance and refine it as necessary to ensure that it continues to accurately predict buying behavior.

See [“Identifying and organizing relevant topics” on page 7 for ways to refine the initial topics selected.](#)

# Dive deeper, step-by-step

As you navigate through the details of this section, reach out to your Bombora contact to provide you with additional resources that will keep your goals on track.

## Setting goals

Connecting business problems and goals to your business objectives will help your team establish your Intent data north star. This can be done by defining your use cases for Intent data and identifying intended outcomes. Below are some common Intent data goals and example use cases to consider when establishing how you will leverage Intent data.

### Intent data goal examples

- 1. Funnel stage prediction:** Identify topics that are relevant to your products and services, group them into clusters and map clusters to the stage of the funnel they represent.
- 2. Account intelligence:** Analyze the research activity of a domain across these clusters to understand where they are in the funnel and when they are likely to move further down.
- 3. Inform sales and marketing:** Utilize Intent data to provide insights to a specific sales or marketing team, so they can connect with prospects and customers with the right message at the right time.

### Example use cases

#### Demand Gen:

A domain is conducting increased research on topics related to product benefits. This indicates they are at the top of the funnel and marketing should activate large scale campaigns to increase product awareness while monitoring for additional topics being researched, indicating they are likely to move down the funnel.

#### Sales Prioritization:

A domain is conducting increased research on topics related to your product or your product capabilities. This indicates they are in a product consideration stage and should be flagged to sales for direct outreach to increase the likelihood of generating an opportunity.

## Account Intelligence:

A domain is researching competitors + products. This indicates they are at the bottom of the funnel and are likely comparing solutions. If your sales team is not already engaging with them, they should immediately reach out using competitive talk tracks.

The above examples can help your team start to identify which use cases align to your business objectives and how your predictive model should be constructed to support each use case.

## Identifying and organizing relevant topics

Evaluating Company Surge® Scores across a set of topics that are grouped by product, service, or use case is more indicative of buying behavior than examining scores for individual topics. All topics have different qualities, so choosing a combination of topics which represent your product or service and its use cases will provide a stronger signal.



Factors that influence topic quality consist of:

- Frequency of occurrence across Bombora's B2B Data Co-op
- Level of competition between topics in our Natural Language Processing (NLP) algorithms
- Relevance to specific use cases

To begin identifying topics of interest for your business, you'll want to first specify a list of products and services sold. Then gather a list of keywords, website pages, and PDF documents which correspond to each product. Once you have a list indicating which keywords, site URLs, and pdf documents map to each product or service, you can provide it to Bombora's global support team who will return a list of potential topics based on Bombora's proprietary taxonomy of 13,000+ topics\*.

If you'd like to generate a list of potential topics on your own, please visit our [Customer Resource Center](#) for recommendations.

After you have a list of potential topics, you'll want to narrow them down to include only the best fit topics for your business. When seeking to identify if a topic is a fit and how it maps to the products and services offered, we recommend using a Historical Buyer Journey Analysis (HBJA).

HBJA reports allow your team to evaluate which topics correspond to meaningful outcomes for your business by comparing research activity that leads to a closed won deal vs. research activity among a group of control accounts and deriving meaningful behavioral patterns. The data from a HBJA report can be used to group, score, and assign weights to topics based on their ability to predict closed won deals.

**Please see the “Building a model with Intent data” for additional details on this process.**

**\*It's important to note that Bombora's topic taxonomy changes over time, so it's crucial to evaluate your topics on a regular basis to ensure topics are still relevant to your products, services and use cases.**

## ADAT file dictionary

The ADAT file provided by Bombora includes the following data fields:

**Company:** The legal name for the parent company or holding company. Note that a single company name can map to several domains when the company is a franchise or has subsidiary companies.

**Domain:** The domain of a company. Domain can represent either the email or web domain of an organization, but in most cases it represents an email domain. This is predominantly used across the industry to represent an individual business. Domains are the primary key used by Bombora when assigning topic activity.

**Firmographic information:** This includes information such as company size, industry, and other firmographic information for the domain. Note that sometimes when a domain is a subsidiary company, the firmographic information will map to the parent or holding company.

**Topic:** The topic in Bombora's B2B taxonomy. Note: the taxonomy continues to grow, so the number of unique topics will increase over time.

**Category:** This is the Bombora assigned category for a given topic.

**Composite\_Score:** This is the score given to each domain+topic pair for all localities indicating intent in the target period compared to the baseline period. All composite scores roughly follow a normal distribution with values  $\geq 60$  indicating increased intent. This score includes activity from unknown geographic areas.

**Bucket\_Code:** This is the confidence score that indicates the certainty of the validity of the composite score. Scores range from an A (high confidence) to an F (low confidence). The score is based on the number of events generated by a domain for all localities. The score is biased towards large companies.

**Metro\_Area:** This is the metro area used to calculate a more granular composite score. This is the metro area associated with the IP addresses for a given company. Note that there are geographic considerations to take into account.

**Metro\_Composite\_Score:** This is the score given to each domain topic pair in a given metro area, indicating intent in the target period compared to the baseline period. All composite scores roughly follow a normal distribution with values  $\geq 60$  indicating increased intent. This score does not include activity from unknown geographic areas.

**Metro\_Bucket\_Code:** This is the score that indicates the certainty of the validity of the metro composite score. Scores range from an A (high confidence) to an F (low confidence). The score is based on the number of events generated by a domain for a given metro area. The score is biased towards large companies.

**Domain\_Origin:** This is the country where Bombora sees the most activity for the given domain. Note that the domain origin may differ from the headquarters.

**Date\_Stamp:** This is the date when the file was generated.

## File delivery

Bombora offers a variety of files to our customers, however, most enterprise organizations leverage our ADAT file (“All Domains, All Topics”) which consists of all Company Surge® data available within a given week.

ADAT files are generated weekly and Enterprise organizations typically feed Bombora’s the files into a Data lake or Data warehouse for the purpose of connecting Bombora data to additional datasets to execute account scoring, propensity to buy, and other predictive models.

ADAT files are also available for delivery into SFTPS, warehouse locations and:



Please work with your Bombora contact to identify and set up the desired delivery method.

## Data ingestion best practices

### Domain considerations

When integrating Bombora’s Intent data, there are several considerations to keep in mind. One important aspect is organizing company/domain information. There are two main ways to do this: by compiling a single score that maps to a company name, or by using each domain score individually.

In a B2B context, domains are predominantly used to represent individual businesses, so it’s important to map each domain to its respective business and associate firmographic information with each domain. This mapping process takes time but can improve internal adoption of the data as it ensures that sales and marketing colleagues understand B2B behavior.

It’s important to note that when a domain is a subsidiary, the firmographic information provided by Bombora may map to the parent or holding organization. Additionally, it’s possible there may be multiple email domains, so mapping a visitor by email domain rather than business name makes more sense in the B2B case of determining where a website visitor works.

It’s also worth noting that not all domains have an associated website, as a business domain can also be an email domain. In the case of larger organizations, it’s common to have multiple email domains depending on the size and branch.

For example, Verizon has multiple email domains across its branches, such as [verizon.com](https://www.verizon.com) and [verizonenterprise.com](https://www.verizonenterprise.com). Facebook, on the other hand, has coverage for both its web and email domains, with [fb.com](https://www.fb.com) as the employee domain and [facebook.com](https://www.facebook.com) as the web domain.

If you choose to evaluate Company Surge® data at the business level instead of the domain level, it's important to consider using a weighted average to create a single score. This can provide a more accurate representation of overall intent.

## Geography considerations

Geo scores available in the ADAT file can be leveraged to improve predictive capabilities. The overall composite score is derived from all activity from a domain, regardless of the geographic location. However, for country, state, and metro composite scores, only activity from the resolved geographic area is included.

It's worth noting that domains can have geographic scores for regions outside of where their offices are located. This is because the location of the IP address which generated the activity is used to determine the geographic location. It's possible for activity to come from other global offices, traveling employees or remote workers, companies masking the location of their IP traffic, or an employee using a VPN. Therefore, it's important to consider these geo factors when building your predictive model.

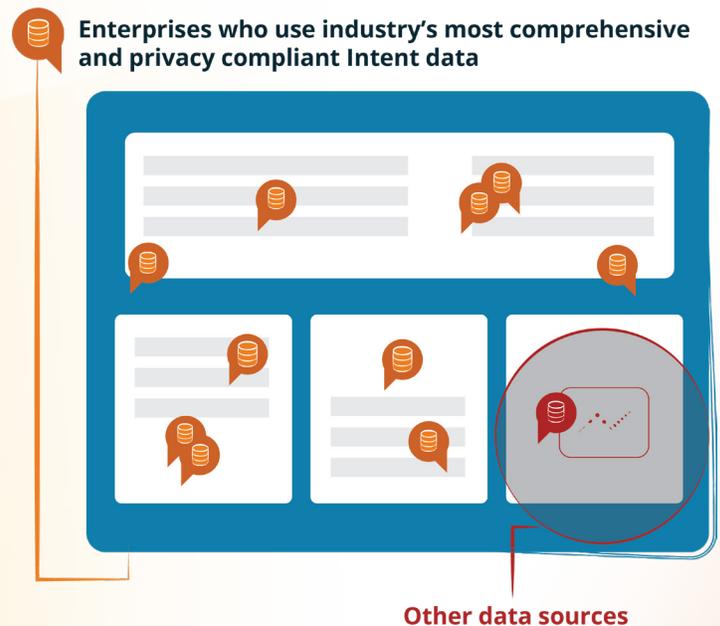
## Mapping Bombora's Intent data

Data privacy is of the utmost importance to our customers and Bombora takes this responsibility seriously. When it comes to data collection, all our data comes from our unique privacy-compliant Data Co-op of 5,000+ of the most highly trafficked the top publishers and websites in the B2B space. Co-op members are responsible for collecting data and passing along the applicable content.

A key component to remaining privacy compliant is the aggregation of Intent data to a domain instead of an individual. Proprietary identity resolution methodologies ensure we are able to resolve visitor data and remain privacy compliant.

As you begin mapping Bombora's domain based Intent data to your internal records, you'll want to consider if Intent data will reside at the business level, or if it should also map to buying units and individuals within a organization. If your organization would like to provide Intent data at the individual level, you can choose to either map the domain data to the individual, or map Bombora's geo scores to individuals. Either option may be best for your organization's selected uses cases.

For additional information regarding our proprietary approach to Intent data, visit [this article](#) from our VP of Data Science, Elena Drodz.



## Building a model with Intent data

Regression techniques can be used to understand the relationship between topics and the desired outcome, such as likelihood to close. Knowing that some topics are more indicative of a closed won deal and have a stronger predictive power, while others are less indicative of buying behavior, will ensure topics are grouped and weighted properly within your a predictive model.

The below steps for creating a predictive model should be completed for each product/service+ set of topics, and combined into one model to deliver a corresponding predictive indicator for each account.

### Predictive binary classification model

Using the consumer purchase use case as an example, we will walk through a potential process of building a predictive binary classification model.

1. Gather the set of topics that represent each product or service that you plan to build a predictive model for.
2. Aggregate a list of closed won accounts for each corresponding product and set of topics.
3. Submit a request to our Global Support team for a HBJA with a control group. The team will be able to generate a raw file containing a 1:1 match of control accounts to closed won accounts. Each control account will be assigned the close date of the matched target account. Each close date represents T0 and historical Company Surge<sup>®</sup> data for each account will be mapped backwards up to an 18 month lookback window.
4. Build a binary regression model to predict if an account will close or not. The dependent variable is if an account will close. The independent variables are historical intent research trends for each topic in your cluster. Some example variables are time series Company Surge<sup>®</sup> sums, aggregates for a given time period, or binary sums if a domain had a Company Surge<sup>®</sup> Score  $\geq 60$ .

As a reminder, a set of topics that represent a product/service sold are more indicative of intent, so using more than one topic as input features or combining all of your topics into a single feature can provide an increased predictive value.

## Identifying funnel stage

HBJA reports are also helpful for identifying which topics correspond to different stages of the funnel. Knowing your topic to funnel stage mapping can be leveraged in your predictive models and will help contribute to the defined use cases from the “Setting goals” section.

1. Aggregate all closed won accounts and generate a historical plot of Company Surge® Scores within a defined time frame prior to each close date. (If you have already run a HBJA report, the raw file will contain this information.) Keep in mind that the length of your sales cycle will determine how far back you will need to look. Ex: if your sales cycle is typically 12 months, we recommend looking back 18 months.
2. Next, standardize the Company Surge® Scores by subtracting 50 from each score to help determine peaks.
3. From the historical plot above, determine where the research peaks exist. (You should see multiple peaks across the plot.)
4. For each peak you’ll want to note when the peak starts and ends.
5. Once start and end dates for each peak are identified, refer to the raw HBJA file to identify which topics contain the highest volume of research spikes and the highest average scores and group them into the applicable funnel stage bucket.

These funnel stages can be mapped back to the use cases defined in the “Setting goals” section.

**\*This content is informational only and is subject to change without notice. Bombora makes no representations or warranties that this information is accurate, complete, and/or all-inclusive.**

## Gaining momentum. What’s next?

At this point, you’ve got the right foundation for incorporating Intent data into your predictive models. The customer journey is not a straight line, so don’t forget to monitor the performance of your models regularly so you can refine as needed, and accurately.

If you need help along the way, please reach out to your Bombora contact.

## About Bombora

Bombora tells businesses which accounts are researching their products and services. With this understanding, sales and marketing teams can be more relevant and consistent, and improve performance across all activities. This Intent-driven approach revolutionizes the way businesses market and sell to other businesses through transparent data built on an ecosystem of quality, collaboration, and innovation. With direct integrations with dozens of leading data and media-buying platforms, Bombora is building a world in which business buyers value sales and marketing for its relevance, timeliness, and accuracy.

To learn more, visit [bombora.com](https://bombora.com)